

UNIVERSIDADE FEDERAL DO RIO GRANDE - FURG
CENTRO DE CIÊNCIAS COMPUTACIONAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO
CURSO DE MESTRADO EM ENGENHARIA DE COMPUTAÇÃO

Dissertação de Mestrado

Moluscos bivalves como biossensores de poluição aquática

Bruna de Vargas Guterres

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Computação da Universidade Federal do Rio Grande - FURG, como requisito parcial para a obtenção do grau de Mestra em Engenharia de Computação

Orientador: Prof^ª. Dr^ª. Silvia Silva da Costa Botelho
Co-orientador: Prof^ª. Dr^ª. Juliana Zomer Sandrini

Rio Grande, 2021

Ficha Catalográfica

G983m Guterres, Bruna de Vargas.
Moluscos bivalves como biossensores de poluição aquática /
Bruna de Vargas Guterres. – 2021.
129 f.

Dissertação (mestrado) – Universidade Federal do Rio Grande –
FURG, Programa de Pós-Graduação em Computação, Rio
Grande/RS, 2021.

Orientadora: Dra. Sílvia Silva da Costa Botelho.

Coorientadora: Dra. Juliana Zomer Sandrini.

1. Biossensores 2. Análise Comportamental 3. Valvometria
4. Poluição Aquática I. Botelho, Sílvia Silva da Costa II. Sandrini,
Juliana Zomer III. Título.

CDU 351.777.6

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



MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DO RIO GRANDE
CENTRO DE CIÊNCIAS COMPUTACIONAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO
CURSO DE MESTRADO EM ENGENHARIA DE COMPUTAÇÃO

DISSERTAÇÃO DE MESTRADO

**MOLUSCOS BIVALVES COMO BIOSSENSORES DE
POLUIÇÃO AQUÁTICA**

Bruna de Vargas Guterres

Banca examinadora:

Prof. Dr. Rubens Mendes Lopes

Prof. Dr. Ricardo Nagel Rodrigues

Prof. Dr. Paulo Lilles Jorge Drews Junior

Profa. Dra. Silvia Silva da Costa Botelho
Orientadora

Profa. Dra. Juliana Zomer Sandrini
Coorientadora

Dedico a presente dissertação de mestrado aos meus pais, Maria e Soni, que são os pilares de tudo me tornei e conquistei. Em especial, dedico este trabalho ao meu irmão, Denilson Guterres, que sempre foi meu exemplo e motivo de orgulho.

AGRADECIMENTOS

Gostaria de agradecer principalmente a minha família por todo apoio ao longo do mestrado e da vida, por me ensinarem os valores que carrego e a importância da educação como ferramenta de transformação. Agradeço aos meus amigos e colegas que se fizeram presentes ao longo desta caminhada e aos diversos professores que auxiliaram na construção dos conhecimentos adquiridos.

Agradeço ainda à minha orientadora Prof^a. Dr^a. Silvia Botelho pela confiança em mim depositada, pelo apoio e por todo auxílio ao longo da execução da presente dissertação de mestrado. Gostaria de agradecer a minha co-orientadora Prof^a. Dr^a Juliana Sandrini por aceitar auxiliar na construção do biossensor proposto e por toda ajuda fornecida.

Da mesma forma, agradeço a Amanda Guerreiro pelo auxílio nos testes laboratoriais executados e na escrita de artigos e a Jhordan Silveira, Je Junior e Eduardo Leivas pela parceria e disponibilidade a qualquer tempo. Por fim, agradeço a CAPES pelo fomento financeiro.

*Tenho a impressão de ter sido uma criança brincando à beira-mar,
divertindo-me em descobrir uma pedrinha mais lisa
ou uma concha mais bonita que as outras, enquanto o imenso oceano
da verdade continua misterioso diante de meus olhos.*

— ISAAC NEWTON

RESUMO

GUTERRES, Bruna de Vargas. **Moluscos bivalves como biossensores de poluição aquática**. 2021. 128 f. Dissertação (Mestrado) – Programa de Pós-Graduação em Computação. Universidade Federal do Rio Grande - FURG, Rio Grande.

Além dos poluentes aquáticos conhecidos, poluentes emergentes são gerados através de fontes antropogênicas diversas. Como se proteção de um ecossistema complexo, pode-se fazer uso de organismos presentes nesse meio na biomonitoração de ambientes aquáticos. Trata-se da associação de componentes biológicos e detectores físico-químicos como dispositivos analíticos conhecidos como biossensores que possibilitam o monitoramento em tempo real de todo o espectro de substâncias tóxicas e efeitos sinérgicos. Respostas comportamentais de moluscos bivalves podem ser utilizadas na biomonitoração automática provendo uma maior sensibilidade e relevância ecológica que testes de toxicidade padrões. O presente trabalho propõe e valida em laboratório modelos computacionais de visão e inteligência artificial na construção de biossensores de poluição aquática baseados na análise comportamental de moluscos bivalves tendo em vista uma arquitetura modular. Um modelo baseado em visão computacional é apresentado como técnica de valvometria não invasiva capaz de eliminar a necessidade de fixação de componentes às conchas de bivalves. Modelos de classificação baseados em grupo de bivalves e na combinação de classificadores são apresentados como ferramentas capazes de inferir a presença de contaminantes a partir de parâmetros comportamentais atingindo acurácias superiores à 90%. O emprego de redes neurais artificiais auto-regressivas não lineares na modelagem comportamental de moluscos individuais é apresentado com o intuito de considerar sua natureza. Os resíduos entre os comportamentos real e modelado são empregados em um gráfico de controle capaz de inferir a presença de contaminantes. Os modelos propostos demonstraram vantagens em relação ao emprego de métodos estatísticos tradicionais na detecção de poluentes.

Palavras-chave: Biossensores, análise comportamental, valvometria, poluição aquática.

LISTA DE FIGURAS

Figura 1	Diagrama em blocos acerca da estrutura de um biossensor de monitoramento aquático baseado na análise comportamental de moluscos bivalves.	24
Figura 2	Modelo de um modelo artificial (HAYKIN, 2007).	25
Figura 3	Estrutura de uma NARNET (BENMOUIZA; CHEKNANE, 2013, Adaptado) para estimar o valor atual (\hat{y}_t) de uma série temporal com base em seus últimos p valores. Cada entrada da rede encontra-se atrasada em relação à outra em uma unidade de tempo (z^{-1})	26
Figura 4	Espaços de características linearmente separáveis (A) e linearmente inseparáveis (B) tendo em vista as classes vermelha e azul. A linha escura esboça a fronteira ideal de decisão no espaço original (SANTOS, 2002, Adaptado).	27
Figura 5	Mapeamento de um espaço de entrada X não linearmente separável e um espaço de características linearmente separável Z através de uma função <i>kernel</i> ϵ . (SANTOS, 2002).	28
Figura 6	Estrutura exemplo de uma árvore de decisão binária em que nós terminais são representados através de quadrados e nós não terminais são representados por meio de círculos (OSHIRO, 2013).	29
Figura 7	Diagrama em blocos sobre o funcionamento dos métodos <i>Bagging</i> (OSHIRO, 2013).	30
Figura 8	Matriz de Confusão gerada a partir de um conjunto de instâncias (SILVA, 2006, Adaptado).	31
Figura 9	Matriz de confusão expandida na qual métricas de avaliação como acurácia e precisão são facilmente visualizadas.	32
Figura 10	Exemplo de curva com ROC com área sob a curva (AUC) igual à 0,78. Linha em vermelho ilustra a curva de um classificador com saídas totalmente aleatórias (CORTES; MOHRI, 2004, Adaptado).	32
Figura 11	Exemplo de curva de Precisão-Revocação através de uma linha azul. A região cinza destaca a área sob a curva PR (AUC-PR) (CHICCO, 2017, Adaptado).	33
Figura 12	Ponto $Q=(X,Y,Z)$ projetado na imagem através de um raio que passa pelo centro de projeção resultando em um ponto $q = (z, y, f)$ projetado na imagem através de uma câmera com distância focal f . (BRADSKI; KAEHLER, 2008)	34

Figura 13	Exemplo de operação morfológica de erosão na qual a imagem original (A) passa por um processo de convolução com um elemento estruturante retangular (B) resultando na imagem a direita (C) (BARELLI, 2018).	37
Figura 14	Exemplo de operação morfológica de dilatação na qual a imagem original (A) passa por um processo de convolução com um elemento estruturante em forma de cruz (B) resultando na imagem a direita (C) (BARELLI, 2018).	38
Figura 15	Exemplo da aplicação do operador morfológico de abertura à uma imagem em escala de cinza (A). O resultado (B) quando subtraído da imagem original fornece uma imagem com iluminação mais uniforme (C) que pode ser melhorada através do ajuste de seu contraste (D) (BARELLI, 2018).	38
Figura 16	O espaço RGB é representado por um cubo (A) através do qual é possível visualizar todas as cores passíveis de obtenção a partir da combinação das componentes vermelha, verde e azul variando de 0 à 1 deste espectro. É comum visualizar este cubo ao longo de seu eixo neutro (eixo da origem ao ponto 1,1,1). A seção transversal do formato resultante corresponde ao cone que representa o espaço HSV (B). Fonte: Corrêa ¹ et al. (2014) (adaptado).	39
Figura 17	Modelo de regressão linear baseado em pontos amostrais de duas variáveis dependente (y) e independente (x). A reta modelada não estima corretamente todos valores de y tendo como base os valores de x havendo, dessa forma, erros definidos como a diferença entre os valores real e estimado da variável dependente (BLACK, 2009)(adaptado)	41
Figura 18	Estimativa normal de densidade por <i>kernel</i> (linha sólida) e contribuições do <i>kernel</i> (linhas pontilhadas) centrado em cada X_i da amostra x_1, x_2, \dots, x_4 . A EDK em qualquer x é a soma das contribuições do <i>kernel</i> centrado em cada X_i (GIVENS; HOETING, 2012)(adaptado).	43
Figura 19	Exemplo de gráfico de controle EWMA para uma série temporal. As linhas de limite central (CL), superior (UCL) e inferior (LCL) encontram-se ilustradas. Os círculos indicam os valores resultantes da média ponderada e as cruces indicam os valores da série analisada a cada intervalo de tempo (SIREGAR et al., 2013, Adaptado).	44
Figura 20	Contribuições propostas	72
Figura 21	Modelo de uma técnica de valvometria baseada em segmentação de cores. Moluscos bivalves com suas conchas superior e inferior pintadas em cores distintas, como por exemplo laranja e verde, são fixados de forma a fornecer vista frontal à uma câmera. Um sistema embarcado é responsável por capturar as imagens, processá-las e inferir a amplitude de abertura de cada concha.	73
Figura 22	Diagrama em blocos ilustrando o pré-processamento e extração de ROIs. As conchas dos moluscos bivalves são pintados em cores distintas como, por exemplo, verde e laranja para posterior análise.	74

Figura 23	Diagrama de processamento de cada ROI. Ele pode ser dividido em segmentação de cores, redução de ruído através de operadores morfológicos de abertura e determinação do pixel mais próximo da extremidade da concha para cada cor de interesse.	74
Figura 24	Modelo computacional proposto para a construção de um módulo ferramenta de análise comportamental baseado na utilização de um classificador para, a partir dos parâmetros comportamentais horários de um grupo de moluscos bivalves, inferir o <i>status</i> de contaminação ambiente (contaminado ou não).	76
Figura 25	Modelo computacional proposto para a construção de um módulo ferramenta de análise comportamental baseado na combinação de classificadores. Quatro classificadores (<i>C1</i> à <i>C4</i>) são empregados para inferir a presença de contaminantes (classificação binária) a partir de parâmetros comportamentais horários. A saída destes classificadores (votos intermediários) são processadas através de um classificador de votação responsável por determinar o <i>status</i> final de contaminação, isto é, se há ou não contaminantes no meio aquático.	77
Figura 26	Modelo para a construção de um módulo ferramenta de análise comportamental baseado no treinamento de NARNETs para prever o comportamento individual mexilhões não expostos a contaminantes. Os erros entre os valores previstos e reais de diversos indivíduos são empregados em um gráfico de controle EWMA a fim de identificar o <i>status</i> de contaminação ambiente.	78
Figura 27	<i>Layout</i> do experimento com mexilhões <i>P. perna</i> . Uma câmera é posicionada de forma externa ao aquário de forma a obter uma vista frontal dos mexilhões instrumentados com ímãs e sensores pintados em verde e laranja. A câmera é conectada a um Odroid XU4 com o intuito de capturar e armazenar imagens de acordo com um <i>trigger</i> externo. Um Arduino Mega em conjunto com uma placa de aquisição de dados é responsável por armazenar as saídas dos sensores e enviar um <i>trigger</i> à câmera.	80
Figura 28	<i>Layout</i> do experimento realizado em triplicatas no qual mexilhões <i>P. perna</i> foram instrumentados com sensores de efeito Hall e ímãs. As saídas dos sensores foram conectadas a uma placa de aquisição de dados que, em conjunto com um Arduino Mega, salvou as respostas comportamentais em um arquivo CSV. Após um período não toxicológico de três dias, os mexilhões foram expostos a 0%, 5% e 20% de óleo diesel WAF	82
Figura 29	Divisão do experimento em período de exposição não tóxica e período de exposição toxicológica. Ambos períodos tem duração de 24 horas A partir das primeira 24 horas, os animais foram expostos à 5% e 20% de WAF de diesel S-500	84

Figura 30	Esquema do classificador de votação no qual quatro classificadores foram treinados com base em um conjunto de dados de treinamento contendo a amplitude média de abertura, percentual de tempo aberto, frequência de transição e quantidade de micromovimentos. As classificações (P_1, P_2, P_3 e P_4) adquiridas por meio de um novo conjunto de dados foram usadas para obter uma previsão final (P_f) através de um classificador de votação.	87
Figura 31	Arquitetura da NARNET para prever $Y(t + 1)$ com base nos dados comportamentais das últimas seis horas (últimas sessenta AOA) de $Y(t)$. A rede possui sessenta atrasos de entrada, uma camada oculta com quinze neurônios cuja função de ativação é uma tangente hiperbólica. A camada de saída abrange um neurônio com uma função de ativação linear.	87
Figura 32	Imagem antes (a) e após (b) melhoria de contraste e a respectiva enumeração dos animais.	90
Figura 33	Região de interesse (a), regiões de coloração laranja (b) e verde (c) detectadas a partir dos valores HSV médios. Os pontos verde e vermelhos ilustrados em (c) correspondem aos pontos máximo e mínimo das figuras (b) e (c), respectivamente. A distância d destacada em amarelo corresponde ao valor de amplitude detectado para o caso apresentado.	90
Figura 34	Comportamento de abertura e fechamento de mexilhões <i>Perna perna</i> tendo em vista o emprego dos sensores de efeito Hall e algoritmo de visão computacional	91
Figura 35	Erro entre sensores de efeito Hall e algoritmo de visão computacional para cada animal ao longo de duas horas de experimento.	92
Figura 36	Gráficos de dispersão do comportamento de mexilhões <i>Perna perna</i> de acordo com sensores de efeito Hall e algoritmo de visão computacional. A linha de regressão linear ($y = ax + b$) minimiza a soma residual dos quadrados entre os dados obtidos pelos sensores de efeito hall (y) e o algoritmo de visão computacional (x) de acordo com os coeficientes a e b . Os coeficientes de correlação de Pearson encontram-se ilustrados para cada bivalve.	93
Figura 38	Estimativa de densidade por kernel (KDE) de histogramas de erro entre sensores de efeito hall e algoritmo de visão computacional. As funções densidade de probabilidade e os histogramas de erro são demonstradas como linhas e barras azuis, respectivamente.	94
Figura 37	Estimativa de densidade por kernel (KDE) de histogramas de erro entre sensores de efeito hall e algoritmo de visão computacional para cada mexilhão <i>Perna perna</i> . As funções densidade de probabilidade e os histogramas de erro são demonstradas como linhas e barras azuis, respectivamente.	96
Figura 39	Exemplos de respostas comportamentais de mexilhões <i>Perna perna</i> expostos a 5% e 2 % de WAF de diesel. Sensores de efeito Hall foram empregados para adquirir dados comportamentais em um experimento que compreendeu 24 horas de exposição não toxicológica e toxicológica, como demonstrado pela linha tracejada preto.	97

Figura 40	Matriz de confusão, curva ROC e AUC referentes ao processo de treinamento de um classificador <i>ensemble bagged</i> para inferir o <i>status</i> de contaminação ambiente a partir de parâmetros comportamentais horários de um grupo de cinco moluscos bivalves.	98
Figura 41	Matrizes de confusão referentes ao processo de teste do classificador <i>bagged trees</i> (A) e ao emprego de métodos estatísticos na identificação de diferenças significativas entre os grupos de controle e de teste ($p < 0,05$) (B)	99
Figura 42	Matrizes de confusão, curvas ROC e AUC dos classificadores <i>bagged trees</i> (A), SVM cúbica (B), SVM quadrática (C) e SVM gaussiana (D) obtidas durante a primeira etapa de treinamento do modelo baseado na combinação de classificadores para detecção de compostos petroquímicos.	100
Figura 43	Matriz de Confusão, curva ROC e valor AUC obtidos durante o treinamento do classificador de votação para detecção de compostos petroquímicos.	100
Figura 44	Matrizes de Confusão, curvas ROC e valores AUC obtidos durante o teste do classificador de votação tendo em vista 5% (A) e 20% (B) de WAF de diesel.	101
Figura 45	Curvas Precisão-Revocação (PR) para 5% (A) e 20% (B) de WAF de diesel.	101
Figura 46	Exemplo de resposta obtida através uma NARNET durante o treinamento, validação e teste (A) e o respectivo erro ao longo do tempo (B).	103
Figura 47	Exemplo de desempenho de uma NARNET ao longo das épocas de treinamento (A) e respectivo gráfico de dispersão entre sua saída e a amplitude média esperada (B). Neste exemplo é possível verificar uma correlação muito forte em um intervalo de confiança de 95% de acordo com o coeficiente de correlação de Pearson (r).	103
Figura 48	Exemplo de funções de autocorrelação (ACF) para a amplitude média de abertura (A) e para os resíduos decorrentes do uso da NARNET (B) a um nível de significância de 5%.	104
Figura 49	Gráfico de controle EWMA para mexilhões não expostos à WAF de diesel S-500. Dados de calibração foram empregados para estabelecer os limites do gráfico de controle. As cruces representam as amplitude média de abertura e pontos vermelhos e pretos indicam os valores da média móvel exponencialmente ponderada que ultrapassam ou não os limites de controle superior (UCL) e inferior (LDL) a cada intervalo de 0,1 hora, respectivamente.	106
Figura 51	Gráfico de controle EWMA para mexilhões expostos a 20% de WAF de diesel. Dados de calibração foram empregados para estabelecer seus limites. A linha tracejada azul indica o início da exposição toxicológica e as cruces representam as amplitudes médias de abertura. Pontos vermelhos e pretos indicam os valores EWMA que ultrapassaram ou não os limites de controle superior (UCL) e inferior (LDL), respectivamente.	106

Figura 50	Gráfico de controle EWMA para mexilhões expostos a 5% de WAF de diesel. Dados de calibração foram empregados para estabelecer seus limites. A linha tracejada azul indica o início da exposição toxicológica e as cruzes representam as amplitudes médias de abertura. Pontos vermelhos e pretos indicam os valores EWMA que ultrapassaram ou não os limites de controle superior (UCL) e inferior (LDL), respectivamente.	107
Figura 52	Matrizes de confusão referentes às saídas do modelo baseado na associação de NARNETs e gráfico de controle EWMA para os grupos expostos à 5% (A) e 20% de contaminante.	108
Figura 53	Resposta da NARNET, valores almejados de amplitude de abertura e resíduos ao longo do tempo para os grupos de animais expostos a 5% (A) e 20% (B) de WAF de diesel. Linhas pontilhadas em cinza indicam o início do período de contaminação.	110
Figura 54	Matrizes de confusão referentes ao emprego dos testes estatísticos na comparação das amplitudes de abertura de mexilhões <i>Perna perna</i> expostos à 5% (A) e 20% (B) de WAF de diesel S-500 em relação ao grupo de controle a cada 0,1 hora.	110
Figura 55	Comportamento médio de moluscos bivalves expostos à 5% de WAF de diesel S-500 a cada intervalo de 0,1 hora. A linha tracejada em azul indica o início da exposição toxicológica na qual a WAF de diesel foi diluída nos aquários. Pontos em vermelho e preto identificam diferenças significativas erroneamente e corretamente identificadas a um nível de significância de 5%, respectivamente.	111
Figura 56	Comportamento médio de moluscos bivalves expostos à 20% de WAF de diesel S-500 a cada intervalo de 0,1 hora. A linha tracejada em azul indica o início da exposição toxicológica na qual a WAF de diesel foi diluída nos aquários. Pontos em vermelho e preto identificam diferenças significativas erroneamente e corretamente identificadas a um nível de significância de 5%, respectivamente.	111

LISTA DE TABELAS

Tabela 1	Tabela de Análise de Variância	49
Tabela 2	Unidade de análise, período de aclimatação, período de monitoramento (PM) e frequência de aquisição empregados em trabalhos relacionados.	65
Tabela 3	Técnicas de valvometria e algoritmos empregados em trabalhos relacionados.	66
Tabela 4	Técnicas de valvometria e algoritmos empregados em trabalhos relacionados.	67
Tabela 5	Parâmetros comportamentais e variáveis de influência em trabalhos relacionados. Dentre os parâmetros comportamentais encontram-se a amplitude média de abertura (AA), período de tempo aberto (PA), período de tempo fechado (PF), frequência de transição (FT) e <i>avoidance behavior</i>	68
Tabela 6	Parâmetros comportamentais e variáveis de influência em trabalhos relacionados. Dentre os parâmetros comportamentais encontram-se a amplitude média de abertura (AA), período de tempo aberto (PA), período de tempo fechado (PF), frequência de transição (FT) e <i>avoidance behavior</i>	69
Tabela 7	Coeficiente (r) e valor p resultante da análise de correlação de Pearson para cada animal inspecionado. A reta de regressão linear ($y = ax + b$) que minimiza a soma residual dos quadrados entre os dados obtidos pelos sensores de efeito hall (y) e algoritmo de visão computacional (x) de acordo com os coeficientes a e b . Também são apresentados o coeficiente de determinação (r^2) e o erro quadrático médio (MSE) entre o sensor de efeito Hall e o algoritmo de visão computacional.	92
Tabela 8	Performance dos classificadores durante o teste tendo em vista a exposição à 5% e 20% de WAF de diesel.	102
Tabela 9	Resumo da performance resultante do emprego de NARNETs na previsão da amplitude média de abertura de mexilhões <i>P. perna</i> a cada intervalo de 0,1 hora. A performance foi avaliada com base nos valores de acurácia, coeficiente de correlação de Pearson (r) e erro quadrático médio (MSE) referentes aos períodos de exposição não toxicológica e toxicológica.	105

LISTA DE ABREVIATURAS E SIGLAS

ACF	Função de Autocorrelação
ANOVA	Análise de Variância
ANP	Agência Nacional de Petróleo, Gás Natural e Biocombustíveis
AOA	Amplitude Media de Abertura
AUC	Área sob a Curva
AUC-PR	Área sob a Curva PR
BEWS	Sistema Biológico de Aviso Prévio
CL	Limite Central
CONECO	Laboratório de Microcontaminantes Orgânicos e Ecotoxicologia Aquática
CROSERF	<i>Chemical Response to Oil Spills Ecological Effects Research Forum</i>
CSV	Valores Separados por Vírgula
EDK	Estimativa de Densidade por Kernel
EWMA	Média Móvel Exponencialmente Ponderada
FAO	<i>Food and Agriculture Organization of the United Nations</i>
FDP	Função Densidade de Probabilidade
FN	Falsos Negativos
FP	Falsos Positivos
FURG	Universidade Federal do Rio Grande
HSV	Matiz (<i>Hue</i>), Saturação (<i>Saturation</i>) e Brilho (<i>Value</i>)
ICB	Instituto de Ciências Biológicas
LCL	Limite de Controle Inferior
LED	Diodo Emissor de Luz
MANOVA	Análise de Variância Multivariada
MSE	Erro Quadrático Médio
NARNET	Rede Neural Artificial Auto-regressiva Não-linear

OpenCV	Biblioteca Aberta de Visão Computacional
PERMANOVA	Análise de Variância Multivariada com Permutações
PR	Precisão-Recuperação
RNA	Rede Neural Artificial
RGB	Vermelho (<i>Red</i>), Verde (<i>Green</i>) e Azul(<i>Blue</i>)
ROC	Características Operacionais do Receptor
ROI	Região de Interesse
SDK	<i>Kit</i> de Desenvolvimento de <i>Software</i>
SVM	Máquinas de Vetores de Suporte
TN	Verdadeiros Negativos
TP	Verdadeiros Positivos
UCL	Limite de Controle Superior
UNEP	Programa de Meio Ambiente das Nações Unidas
VBM	Medição Baseada em Visão
WAF	Fração Acomodada em Água

SUMÁRIO

1	INTRODUÇÃO	18
1.1	Objetivo geral	22
1.2	Objetivos específicos	22
2	REVISÃO BIBLIOGRÁFICA	23
2.1	Redes Neurais Artificiais	24
2.1.1	Rede Neural Artificial Auto-regressiva Não-linear (NARNET)	25
2.2	Modelos de Classificação	26
2.2.1	Máquinas de Vetores de Suporte (SVM)	27
2.2.2	Árvores de Decisão	28
2.2.3	Métricas de Avaliação	30
2.3	Ferramentas de Visão Computacional	34
2.3.1	Calibração de Câmeras	34
2.3.2	Operadores Morfológicos	37
2.3.3	Representação de cores em imagens	38
2.4	Ferramentas de análise estatística	40
2.4.1	Regressão linear	40
2.4.2	Correlação	41
2.4.3	Função densidade de probabilidade	42
2.4.4	Gráficos de Controle de Qualidade	43
2.4.5	Métodos ANOVA e alternativas não-paramétricas	45
2.5	Considerações Finais	51
3	TRABALHOS RELACIONADOS	53
4	MODELOS COMPUTACIONAIS	72
4.1	Módulo Técnica de Valvometria	72
4.1.1	Modelo com Visão Computacional	73
4.2	Módulo Ferramentas de Análise Comportamental	74
4.2.1	Modelo com Classificador Baseado em Grupo	75
4.2.2	Modelo com Combinação de Classificadores	75
4.2.3	Modelo com NARNET e Gráfico de Controle EWMA	77
5	VALIDAÇÃO	79
5.1	Setup dos Experimentos	79
5.1.1	Módulo Técnica de Valvometria	79
5.1.2	Módulo Ferramentas de Análise Comportamental	81
5.2	Ferramentas de Avaliação	83

5.2.1	Módulo Técnica de Valvometria	83
5.2.2	Módulo Ferramentas de Análise Comportamental	84
6	RESULTADOS E DISCUSSÃO	89
6.1	Módulo Técnica de Valvometria	89
6.2	Módulo Ferramentas de Análise Comportamental	97
6.2.1	Modelo com Classificador Baseado em Grupo	98
6.2.2	Modelo com Combinação de Classificadores	99
6.2.3	Modelo com NARNETs e Gráfico de Controle EWMA	102
7	CONCLUSÃO	113
8	PUBLICAÇÕES	115
	REFERÊNCIAS	116
	APÊNDICES	124
A	Moluscos Bivalves	125
A.1	Mexilhões <i>Perna perna</i>	125
B	Sensores de efeito Hall	127
C	Fração Acomodada em Água (WAF)	128

1 INTRODUÇÃO

A indústria petrolífera fornece bens de difícil substituição na matriz produtiva de qualquer país sendo base para o modo de operação, consumo e mesmo cultura da sociedade moderna (CANELAS, 2007). A importância da indústria petrolífera, dentro da cadeia produtiva do petróleo, não se resume apenas à produção final de derivados, ela também se dá na proteção da qualidade do meio ambiente. Isto porque suas atividades de exploração, transporte, distribuição e armazenamento podem gerar danos ambientais. Dentre os ambientes suscetíveis a estes riscos, o meio aquático é um dos mais preocupantes em virtude da possibilidade de ocorrência de vazamentos acidentais de óleo e efluentes de refinarias que podem comprometer recursos hídricos e conferir sérios riscos aos organismos expostos.

Além dos poluentes aquáticos decorrentes da indústria petrolífera, poluentes emergentes são gerados através de fontes antropogênicas diversas e se distribuem através das matrizes ambientais (GAVRILESCU et al., 2015). Embora tenham sido feitos grandes avanços na detecção e análise de traços de poluentes durante as últimas décadas, há grande variedade de contaminantes não detectados de preocupação ambiental emergente que necessitam identificação e quantificação em vários componentes ambientais e tecidos biológicos (GAVRILESCU et al., 2015). Além disso, as atividades humanas resultaram na contaminação dos recursos hídricos com micropoluentes biológicos como vírus e bactérias, que apresentam potencial patogenicidade.

Justifica-se, portanto, a necessidade de pesquisa e inovação em tecnologias para detecção dos mais variados contaminantes ambientais mesmo que emergentes. Nesse sentido torna-se necessário adotar medidas e procedimentos para monitorar e proteger ambientes aquáticos a fim de que atitudes corretivas aplicáveis sejam empregadas o mais rápido possível. Compostos químicos hidrofóbicos tendem a se acumular nos tecidos de organismos aquáticos por diferentes mecanismos como absorção direta da água, absorção de partículas em suspensão e consumo de alimentos contaminados (OOST; BEYER; VERMEULEN, 2003) dando origem à bioacumulação. Dessa forma, métodos de bioacumulação baseiam-se na análise de tecidos animais a fim de inferir a contaminação ambiente já que a concentração de compostos químicos nestes tecidos pode refletir a condição de seu meio (NEFF, 2002).

Nesse sentido, o monitoramento ambiental de poluentes baseado em análises toxicológicas de bioacumulação e mortalidade apresentam como desvantagens o fato do equilíbrio de concentração normalmente ser obtido após semanas de exposição. Estes métodos são, portanto, inapropriados para aplicações que almejam uma rápida detecção da presença de poluentes (KRAMER; JENNER; ZWART, 1989). Embora existam técnicas analíticas tradicionais altamente sensíveis e seletivas como, por exemplo, cromatografia e espectrometria de massa, estas são demoradas, caras, demandam pessoal altamente treinado e não permitem o monitoramento frequente e em tempo real de contaminantes (DORST et al., 2010). Ademais, estes métodos apresentam certas desvantagens e limitações pois baseiam-se apenas em princípios físico-químicos e não fornecem uma avaliação eficiente e econômica acerca da situação ecossistêmica geral (STARODUB et al., 2005).

Os sistemas de controle para ambientes aquáticos baseiam-se principalmente em métodos físico-químicos que são demorados, caros e capazes de detectar uma estreita gama de poluentes sem prover dados acerca de sua influência sobre a biota (GREKOV; KUZMIN; MISHUROV, 2019). Dessa forma, tais métodos não são apropriados para detectar contaminantes de forma imediata nem avaliar seu efeito ecossistêmico rapidamente. A incorporação de componentes biológicos aos sistemas de monitoramento, em contraste com métodos físico-químicos tradicionais, permite avaliar a característica toxicológica integral do meio aquático independente da origem e composição dos poluentes (GREKOV; KUZMIN; MISHUROV, 2019).

Tratam-se da associação de componentes biológicos e detectores físico-químicos como dispositivos analíticos conhecidos como biossensores. A utilização de biossensores permite a análise de alterações na resposta biológica de organismos (resposta molecular, celular, bioquímica, fisiológica ou comportamental) que podem estar relacionadas com a exposição e efeitos de elementos tóxicos (NEWTON; COPE, 2007). Neste contexto, a escolha do componente biológico utilizado como biossensor é importante uma vez que alguns animais apresentam características mais propícias que outros.

Moluscos bivalves são animais invertebrados de corpo não articulado protegido por duas conchas (valvas) unidas por um ligamento e apresentam características convenientes à utilização como biossensores uma vez que alterações comportamentais de abertura e fechamento de suas conchas representam uma resposta biológica à presença de contaminantes sendo uma forma relativamente fácil e barata de monitorar ambientes aquáticos (NEWTON; COPE, 2007). Respostas comportamentais podem ser utilizadas na biomonitoração automática de ambientes aquáticos provendo uma maior sensibilidade e relevância ecológica que testes de toxicidade padrões (LIAO et al., 2009). O emprego destes organismos aquáticos como biossensores fornece monitoramento em tempo real de todo o espectro de substâncias tóxicas e considera seus efeitos sinérgicos (LYUBIMTSEV; KHOLODKEVICH; DRUZHININ, 2019). Além disso, o uso da análise do com-

portamento de bivalves como biomarcadores¹ sensíveis e subletais a poluentes pode não se restringir à detecção de compostos químicos sendo adequado para monitorar os efeitos da poluição difusa, como sedimentos finos e seus efeitos nos serviços ecossistêmicos (LUMMER; AUERSWALD; GEIST, 2016).

A análise comportamental de moluscos bivalves requer o emprego de técnicas de valvometria que permitem adquirir dados referentes à amplitude de abertura de suas conchas. Dentre estas técnicas, sensores de efeito Hall apresentam uma série de vantagens como durabilidade, leveza, facilidade de fixação e oferecem um menor estresse ao animal facilitando a medição de seus movimentos (NAGAI et al., 2006). Entretanto, dados resultantes do monitoramento contínuo de organismos bivalves são bastante complexos devido ao comportamento não linear de seu movimento tornando assim sua análise estatisticamente desafiadora (HARTMANN et al., 2016; BAE; PARK, 2014).

Ademais, a resposta individual dos animais, principalmente sob exposição toxicológica, varia e reflete a natureza adaptativa individual que leva organismos a diferirem em sua capacidade de resposta às variações ambientais (DINGEMANSE et al., 2010). Tem-se, portanto, uma variabilidade inerente dos padrões comportamentais entre indivíduos. Além disso, suas respostas comportamentais encontram-se fortemente relacionadas não só à atividades vitais como respiração, alimentação e excreção, mas também à condições ambientais como a presença de contaminantes e predadores. Biossensores baseados no monitoramento comportamental ou fisiológico de animais sentinela² fornecem informações sobre o bem-estar animal e suas respostas às mudanças ambientais e às ações de manejo (ANDREWARTHA et al., 2015). Logo, o apropriado entendimento do movimento desses animais sésseis pode auxiliar no gerenciamento da qualidade da água em habitats naturais (HARTMANN et al., 2016).

Dentre os moluscos bivalves, o mexilhão *Perna perna* é um molusco da família *Mytilidae* de maior tamanho e mais abundante na costa brasileira apresentando grande importância sócio-econômica em virtude de sua utilização na mitilicultura (JR; BRASIL; SALOMÃO, 2007). A análise comportamental de moluscos bivalves através de sensores de efeito Hall em conjunto com ímãs foi originalmente proposta em Wilson, Reuter e Wahl (2005) com base em sua aplicação na análise comportamental de pinguins. A partir de então, sensores de efeito Hall em conjunto com ímãs consolidaram-se como técnica de valvometria sendo empregados para monitoramento contínuo em tempo real de águas superficiais, água do mar e efluentes através de soluções comerciais³ e na realização de experimentos em ambientes naturais (COMEAU et al., 2018; GREKOV; KUZMIN;

¹Biomarcadores ou marcadores biológicos são entidades que podem ser medidas experimentalmente e indicam a ocorrência de uma determinada função normal ou patológica de um organismo ou uma resposta a um agente farmacológico (LDA, 2020)

²animais localizados em um ambiente comum capaz de fornecer um aviso prévio de condições abaixo do ideal que podem afetar o bem-estar ou sobrevivência individual ou da população

³Aquadect (2020) e MolluSCAN (2020)

MISHUROV, 2019).

Algumas técnicas de valvometria propostas na literatura são categorizadas como não invasivas embora necessitem fixar dispositivos às conchas dos moluscos bivalves (CO-MEAU et al., 2018; BASTI et al., 2009; NAGAI et al., 2006). Isso pode comprometer sua mobilidade e exigir uma adequação dos animais em virtude dessa nova configuração de peso em suas valvas. Além disso, O emprego de técnicas de exploração funcional não invasiva, ao contrário das técnicas letais ou invasivas, permitem que os organismos individuais sejam monitorados pelo tempo que for necessário (MOULLAC et al., 2020). Logo, surge a possibilidade do estudo acerca de outras formas de monitoramento comportamental que não exijam a anexação de quaisquer materiais a esses animais como, por exemplo, através de sinais visuais.

Em virtude dos avanços rápidos e significativos das tecnologias em sistemas de câmeras e computadores, o acesso a câmeras e unidades de computação mais baratas, mais rápidas e de maior qualidade foi facilitado. Como resultado, há uma tendência crescente no uso de medição baseada em visão (*Visual Based Measure* - VBM) em uma ampla variedade de aplicações e cenários automatizados (SHIRMOHAMMADI; FERRERO, 2014).

Os rápidos avanços no campo da inteligência artificial têm profundas implicações para a economia e também para a sociedade em geral através de contribuições importantes em áreas como a agricultura, engenharia e medicina (COCKBURN; HENDERSON; STERN, 2018). A inteligência artificial melhorou muito a coleta, análise e previsão de informações biológicas e médicas. A exemplo disso, algoritmos de inteligência artificial (IA) foram capazes de solucionar um dos maiores desafios biológicos prevendo como proteínas de uma cadeia linear de aminoácidos se enrolam em formas 3D (SERVICE, 2020).

A integração de inteligência artificial (IA) e biossensores pode preencher a lacuna entre a aquisição de dados e análise com melhor precisão diagnóstica e terapêutica (JIN et al., 2020) em aplicações biomédicas. Biossensores vestíveis tem ganhado atenção devido ao seu tamanho mínimo, característica não invasiva e capacidade de monitoramento contínuo. A versatilidade de algoritmos de inteligência artificial tem revolucionado a construção de biossensores vestíveis através de novas possibilidades de desenvolvimento como, por exemplo, biossensores para monitoramento de saliva e suor baseados em biomarcadores como *proxy* para diagnóstico de doenças (JIN et al., 2020).

Além disso, fazendas de aquicultura inteligentes equipadas com biossensores, sensores para monitorar variáveis ambientais e sistemas de gerenciamento tem potencial para revolucionar todos os sensores da indústria (ANDREWARTHA et al., 2015). Dados fisiológicos e comportamentais em tempo real que fornecem *insights* sobre o bem-estar animal têm a capacidade de auxiliar práticas de manejo e otimizar os processos de produção. As informações sobre o ambiente são frequentemente necessárias para interpretar as respostas fisio-comportamentais e o animal sentinela torna-se integrado em uma série de

sensores ambientais (ANDREWARTHA et al., 2015).

Inteligência artificial na construção de biossensores de poluição aquática baseados no comportamento de moluscos bivalves pode oferecer vantagens em relação ao uso de métodos estatísticos tradicionais. Ademais, uma técnica de valvometria baseada em visão computacional pode simplificar o sistema com uma menor quantidade de cabos e ligações, eliminar a necessidade de fixação de componentes às conchas, facilitar o processo de instrumentação e reduzir erros devido à falhas de impermeabilização e fixação.

De forma multidisciplinar, este trabalho busca, portanto, averiguar o uso de técnicas de visão computacional na construção de uma técnica de valvometria não invasiva capaz de eliminar a necessidade de fixação de componentes às conchas de moluscos bivalves. Ademais, almeja-se investigar o emprego de métodos computacionais de inteligência artificial na biomonitoração aquática de poluentes baseada na análise comportamental de mexilhões *Perna perna*. Nesse sentido, as subseções 1.1 e 1.2 apresentam os objetivos geral e específicos do presente trabalho.

1.1 Objetivo geral

A presente dissertação de mestrado tem como objetivo geral propor modelos computacionais para construção de biossensores de poluição aquática baseados na análise comportamental de moluscos bivalves. De forma mais precisa, pretende-se desenvolver uma arquitetura modular com contribuições baseadas em visão computacional e inteligência artificial para a construção destes biossensores.

1.2 Objetivos específicos

Pode-se elencar os seguintes objetivos específicos:

- Realizar o estudo bibliográfico acerca do emprego de moluscos bivalves como biossensores de poluição aquática;
- Identificar os requisitos e módulos principais associados a um biossensor baseado na análise comportamental de mexilhões *Perna perna*;
- Propor métodos computacionais baseados em visão e inteligência artificial para os módulos e seus componentes;
- Validar a metodologia por meio de comparação com técnica de valvometria estado da arte.

2 REVISÃO BIBLIOGRÁFICA

O emprego de moluscos bivalves como biossensores aquáticos baseados em sua análise comportamental requer o emprego de técnicas de valvometria para adquirir dados acerca da amplitude de abertura de suas conchas. Dentre estas técnicas, destacam-se pares de indutores, eletrodos, células de carga e sensores de efeito Hall (HARTMANN et al., 2016; TRAN et al., 2010; LIAO et al., 2009). Além disso, empregam-se parâmetros de análise comportamental capazes de descrever esse comportamento no tempo como, por exemplo, amplitude média de abertura, frequência de transição e percentual de tempo em que as conchas encontram-se abertas (COMEAU et al., 2018; HARTMANN et al., 2016).

Por fim, torna-se necessário o emprego de ferramentas de análise comportamental capazes de inferir alterações nestes parâmetros decorrentes de alterações ambientais de interesse. Dentre estas ferramentas pode-se citar verificação de critérios de comportamento considerados atípicos a fim de emitir sinal de alerta por meio de um BEWS (AQUA-DECT, 2020) e métodos estatísticos como, por exemplo, ANOVA (*Analysis Of Variance*) (HARTMANN et al., 2016).

Dessa forma, biossensores baseados na análise comportamental de bivalves podem ser entendidos como ferramentas capazes de verificar a condição ambiental e emitir sinais de alerta associados às condições ambientais inóspitas sejam elas decorrentes da presença de contaminantes, algas nocivas ou condições físico químicas inadequadas. Nesse sentido, pode-se definir uma sequência de módulos necessários à construção destes biossensores tendo em vista as funções desempenhadas por cada qual na determinação de sinais de alerta ambiental conforme ilustrado pelo diagrama em blocos da Figura 1.

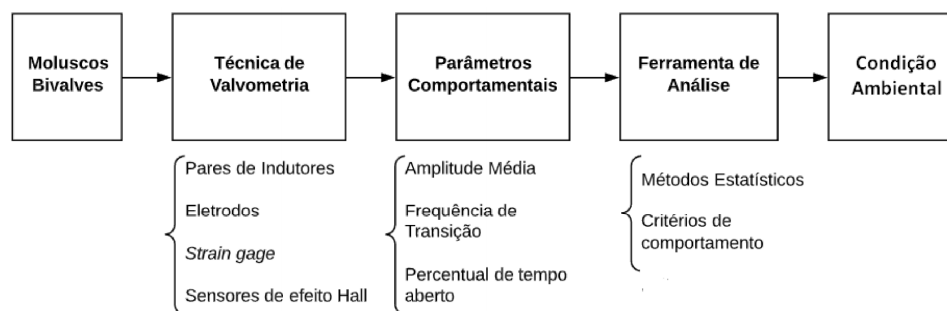


Figura 1: Diagrama em blocos acerca da estrutura de um biossensor de monitoramento aquático baseado na análise comportamental de moluscos bivalves.

A construção de biossensores de poluição aquática baseados na resposta comportamental de moluscos bivalves requer uma fundamentação teórica multidisciplinar abrangente sobre as principais características dos animais empregados, bem como a importância e propriedades dos meios de exposição selecionados para testes laboratoriais. Além disso, torna-se necessário apresentar os principais dispositivos e ferramentas empregados na aquisição de dados comportamentais acerca da amplitude de abertura de suas conchas. Nesse contexto, os apêndices A, B e C versam sobre os conceitos básicos como moluscos bivalves, sensores de efeito Hall e Fração Acomodada em Água (WAF) como meio de exposição padronizado, respectivamente.

Tendo os objetivos da presente dissertação de mestrado, propõe-se uma técnica de valvometria baseada em visão computacional capaz de eliminar a necessidade de fixar componentes às conchas de moluscos bivalves. Ademais, tendo em vista a necessidade de inferir a presença de contaminação ambiente a partir de parâmetros comportamentais, propõe-se o emprego de métodos inteligentes de classificação como ferramentas de análise. Por fim, almejando considerar a natureza adaptativa de bivalves individuais, propõe-se a modelagem de seu comportamento através de redes neurais auto-regressivas não lineares e posterior inferência do *status* de contaminação ambiente a partir do emprego de gráficos de controle de qualidade.

Dessa forma, as subseções 2.1, 2.2 e 2.3 versam sobre Redes Neurais Artificiais (RNAs), modelos de classificação e ferramentas de visão computacional empregados ao longo da presente dissertação. A validação dos modelos propostos é efetuada através de comparações com métodos estatísticos tradicionalmente empregados na construção de biossensores baseados em moluscos bivalves e apresentados através da subseção 2.4.

2.1 Redes Neurais Artificiais

Redes Neurais Artificiais (RNAs) são modelos matemáticos inspirados na estrutura neural de organismos inteligentes que aprendem por meio da experiência (HERCULANO-HOUZEL, 2009). Ela se assemelha ao cérebro humano por ser cons-

tituída por unidades básicas que possibilitam adquirir conhecimento através de um processo de aprendizado e mantê-lo com base em forças de conexão entre neurônios (pesos sinápticos) (HAYKIN, 2007).

Conjuntos de neurônios são organizados de modo a formar uma rede neural artificial. Estes neurônios são matematicamente modelados por meio de pesos sinápticos ($w_{k1}, w_{k2}, \dots, w_{km}$) associados à cada uma das entradas (X_1, X_2, \dots, X_m) da rede neural (Figura 2). O somatório dos sinais de entrada ponderados pelos respectivos pesos sinápticos e um *biases* (b_k) são entradas de uma função de ativação ($\varphi(\cdot)$) responsável por modelar influência sobre a saída (y_k) conforme equação 1.

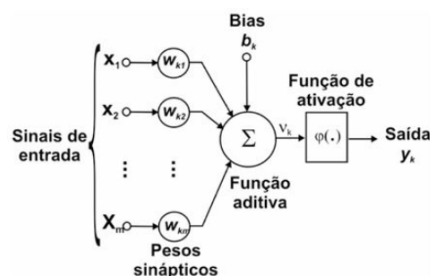


Figura 2: Modelo de um modelo artificial (HAYKIN, 2007).

$$y_k = \varphi\left(\sum_{i=1}^n w_{kn} \cdot x_n + b_k\right) \quad (1)$$

Em geral pode-se identificar duas classes de arquiteturas fundamentalmente diferentes: redes alimentadas em adiante (*feedforward*) e redes recorrentes (HAYKIN, 2007). Redes recorrentes são caracterizadas pelo fato de existir pelo menos um laço de realimentação da saída de neurônios como entradas de neurônios de camadas anteriores (HAYKIN, 2007). Isto impacta fortemente a capacidade de aprendizagem da rede e envolve o emprego de elementos de atraso unitário (z^{-1}) que resulta em um comportamento dinâmico não linear (HAYKIN, 2007).

Dentre as redes neurais recorrentes, encontram-se as Redse Neural Artificial Auto-regressiva Não-linear (NARNET) que possibilitam a modelagem de séries temporais não lineares conforme subseção a seguir.

2.1.1 Rede Neural Artificial Auto-regressiva Não-linear (NARNET)

Redes neurais artificiais auto-regressivas não-lineares (*Non-linear Auto regressive Neural networks- NARNETs*) podem ser utilizadas para modelar séries temporais uma vez que são capazes de reconhecer seus padrões e características não lineares (IBRAHIM et al., 2016). Por meio destes modelos é possível determinar a série temporal no instante t tendo em vista os valores da série nos últimos p instantes de tempo ($t - 1, t - 2, \dots$,

$t - p$) (BENMOUIZA; CHEKNANE, 2013) através de um sistema discreto não linear auto-regressivo conforme equação 2.

$$\hat{y}(t) = f(y(t - 1), y(t - 2), \dots, y(t - p)) \quad (2)$$

Em que \hat{y} representa os valores estimados da série temporal e f uma função não linear de aproximação na qual os valores estimados dependem unicamente dos últimos p valores da série temporal. A Figura 3 ilustra a estrutura de uma NARNET com uma única camada oculta.

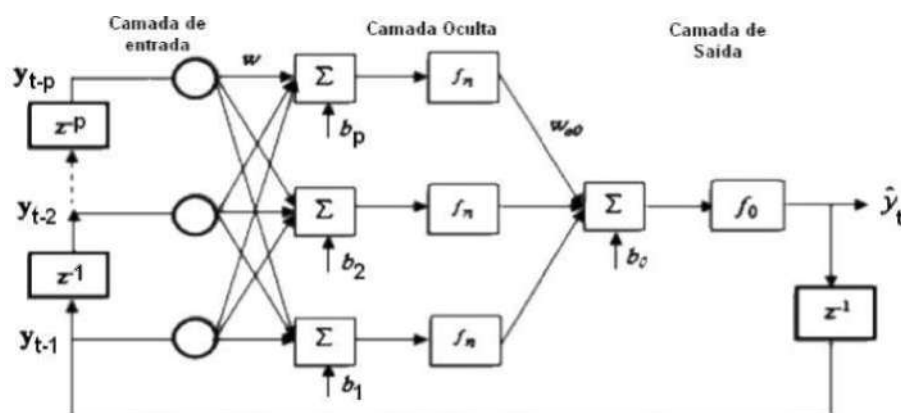


Figura 3: Estrutura de uma NARNET (BENMOUIZA; CHEKNANE, 2013, Adaptado) para estimar o valor atual (\hat{y}_t) de uma série temporal com base em seus últimos p valores. Cada entrada da rede encontra-se atrasada em relação à outra em uma unidade de tempo (z^{-1})

Conforme ilustrado através da Figura 3, NARNETs são redes neurais recorrentes baseadas em um modelo linear regressivo nas quais as p últimas saídas são realimentadas como entradas (BENMOUIZA; CHEKNANE, 2013).

2.2 Modelos de Classificação

A classificação consiste em, dado um conjunto de dados de treinamento, encontrar regras ou maneiras de classificar um novo exemplar tendo em vista suas características e atributos bem como um conjunto de classes previamente definido (PAULA, 2002). Embora existam metodologias não supervisionadas para a construção de classificadores, a subseções a seguir versam sobre as abordagens supervisionadas associadas à Máquinas de Vetores de Suporte (SVM) e árvores de decisão. Além disso, métricas de avaliação da *performance* de classificadores são apresentadas.

2.2.1 Máquinas de Vetores de Suporte (SVM)

Máquinas de Vetores de Suporte (*Support Vector Machines* - SVMs) são uma categoria de redes *feedforward* que podem ser usadas para classificação de padrões e regressão linear. A ideia base de SVMs é definir um hiperplano como superfície de decisão capaz de maximizar a separação entre duas classes de dados (HAYKIN, 2007). Nesse sentido, SVMs podem fornecer boa generalização em problemas de classificação de padrões.

O modelo mais simples de SVMs consiste em um Classificador de Margem Máxima que permite a separação entre dados linearmente separáveis. Embora esta característica limite a empregabilidade deste modelo, ele é fundamental ao entendimento de SVMs mais modernas. A Figura 4 ilustra espaços de características linearmente separáveis e inseparáveis nos quais uma linha escura representa o hiperplano de decisão (separação) ideal entre as classes.

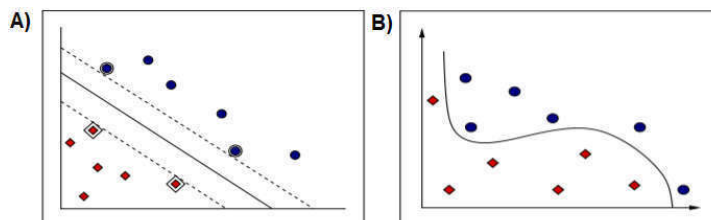


Figura 4: Espaços de características linearmente separáveis (A) e linearmente inseparáveis (B) tendo em vista as classes vermelha e azul. A linha escura esboça a fronteira ideal de decisão no espaço original (SANTOS, 2002, Adaptado).

Um hiperplano de Margem Máxima (ou de Separação Ótima) é aquele que separa um conjunto de vetores sem erro maximizando a distância entre os vetores (classes opostas) mais próximos ao hiperplano (SANTOS, 2002). Com o intuito de tolerar ruídos e *outliers*, as seguintes equações definem o modelo SVM.

$$x_i \cdot w + b \geq +1 - \xi_i \quad \text{para} \quad y_i = +1 \quad (3)$$

$$x_i \cdot w + b \leq -1 + \xi_i \quad \text{para} \quad y_i = -1 \quad (4)$$

$$\xi_i \geq 0 \quad \forall i \quad (5)$$

onde x_i , y_i e ξ_i definem as i -ésimas entrada, saída/classe alvo e diferença entre a saída alvo e o valor estipulado pela SVM. Da mesma forma, w e b representam os pesos e *biases* da rede. Com o intuito de separar conjuntos de dados não linearmente separáveis, SVMs realizam, antes da definição do hiperplano, um mapeamento não-linear dos vetores de entrada em um espaço de características com alta dimensão (SANTOS, 2002).

Trata-se de uma estratégia de pré-processamento dos dados através de funções *kernel* ($\epsilon(x)$) responsáveis por mapear o espaço de entrada X em um novo espaço $Z =$

$\{\phi(x)|x \in X\}$ denominado espaço de características. A Figura 5 ilustra um mapeamento de um espaço de entrada linearmente inseparável para um espaço de características de maior dimensão no qual os dados podem ser linearmente separados.

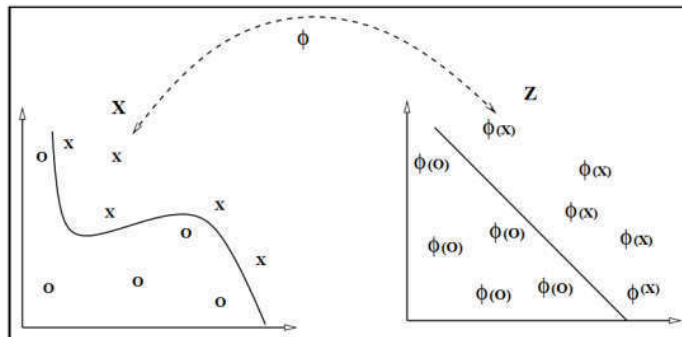


Figura 5: Mapeamento de um espaço de entrada X não linearmente separável e um espaço de características linearmente separável Z através de uma função *kernel* ϵ . (SANTOS, 2002).

Logo, a escolha da função *kernel* empregada em SVMs é importante (SANTOS, 2002) e define diferentes modelos SVM como, por exemplo, SVMs quadrático, cúbico e gaussiano.

2.2.2 Árvores de Decisão

Árvores de Decisão são algoritmos de aprendizagem de máquina que utilizam a estratégia de dividir para conquistar dividindo, portanto, um problema de classificação em subproblemas recursivamente (OSHIRO, 2013). Sua construção baseia-se na escolha de atributos para particionar as entradas em subconjuntos correspondentes. Dessa forma, uma árvore de decisão consiste em nodos (nós) e ramos. O nodo inicial é denominado nó raiz. A cada nó um atributo é verificado e, de acordo com o resultado, a árvore se ramifica até atingir os nodos terminais onde o resultado da classificação é designado (PAULA, 2002).

A Figura 6 ilustra a estrutura de uma árvore de decisão binária capaz de classificar dados em duas classes (C_1 e C_2). Independente de sua arquitetura, árvores de decisão apresentam caminhos determinados por uma sequência de verificações (efetuadas a cada nodo) que conduzem dados de entrada a uma classificação (PAULA, 2002).

O processo de aprendizagem da estrutura de uma árvore de decisão (ou de regras equivalentes) é conhecido como indução (PAULA, 2002). Uma árvore pode ser induzida pela seleção de algum atributo ou teste que servirá para particionar, a cada nodo, os exemplos em subconjuntos de acordo com o resultado do teste (OSHIRO, 2013). Dessa maneira, nodos tornam-se terminais quando todos exemplares restantes em um subconjunto pertencem a uma única classe ou quando o número de casos restantes for menor que um dado limiar (PAULA, 2002). Nesse caso, o nodo é associado à classe que apresenta maior

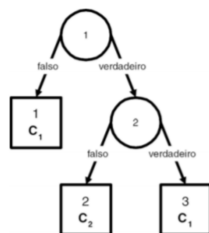


Figura 6: Estrutura exemplo de uma árvore de decisão binária em que nós terminais são representados através de quadrados e nós não terminais são representados por meio de círculos (OSHIRO, 2013).

frequência no grupo.

A seleção randômica de atributos resulta em árvores extensas e específicas para os dados de treinamento (*overfitting*) (OSHIRO, 2013). Nesse caso, o modo de seleção dos atributos (regras) consiste em avaliar sua qualidade ou representatividade de modo a aumentar a performance da árvore. A função de avaliação é, portanto, uma heurística que objetiva selecionar bons atributos para cada nodo tendo em vista informações incompletas. Após inferida, uma árvore de decisão pode ser empregada para classificar novas entradas (OSHIRO, 2013).

As classificações efetuadas por árvores de decisão podem ser combinadas na classificação de novas entradas (OSHIRO, 2013). Nesse sentido, métodos *ensemble* são algoritmos de aprendizado que constroem um conjunto de classificadores que combinam suas saídas (votos) na classificação de novos dados. Dentre os métodos *ensembles* populares, destacam-se os métodos *boosting*, *bagging* e *Random Forests*. Na subseção a seguir o método *ensemble bagging* é apresentado.

2.2.2.1 Bagging

O diagrama em blocos referente ao algoritmo *ensemble bagging* encontra-se ilustrado através da Figura 7.

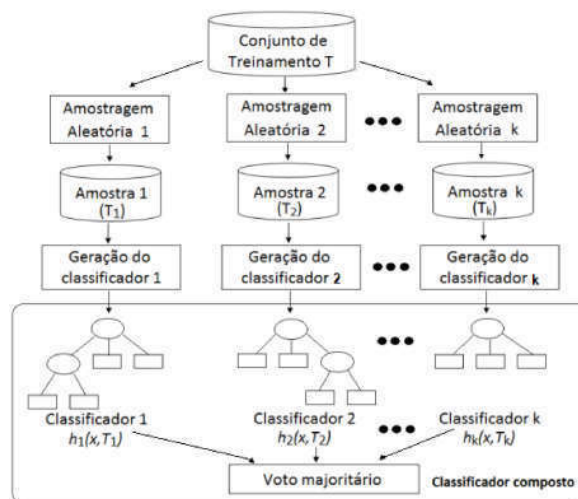


Figura 7: Diagrama em blocos sobre o funcionamento dos métodos *Bagging* (OSHIRO, 2013).

O método *bagging* é um algoritmo de aprendizado *ensemble* que emprega amostras *bootstrap* do conjunto de treinamento para treinar diferentes classificadores (OSHIRO, 2013). Na amostragem *bootstrap* E_0 , por exemplo, um conjunto de treinamento T com n exemplos é dividido em amostras (subconjuntos) T_k com reposição contendo n exemplos.

A seguir, cada T_k é empregado para treinar um classificador diferente (h_k). A estratégia de combinação dos classificadores consiste no voto majoritário (OSHIRO, 2013) conforme ilustrado através da Figura 7).

2.2.3 Métricas de Avaliação

Em aprendizado de máquina utilizam-se algumas ferramentas e métricas para avaliar o desempenho de classificadores. Dentre estas, pode-se citar matrizes de confusão que fornecem informações relevantes ao cálculo de diferentes parâmetros necessários à utilização de curvas de Características Operacionais do Receptor (ROC) e curvas de Precisão-Revocação (*Precision-Recall* - PR) (SILVA, 2006). A exemplo disso, pode-se citar a taxa de positivos verdadeiros ou revocação/sensibilidade, taxa de falsos positivos, acurácia e a métrica *F-score*. Dessa forma, as subseções a seguir abordam matrizes de confusão, curvas ROC e curvas PR bem como métricas associadas.

2.2.3.1 Matrizes de Confusão

Matrizes de confusão auxiliam na visualização do desempenho de classificadores. Tratam-se de matrizes quadradas cuja dimensão corresponde à quantidade de classes de um problema de classificação. Elas relaciona as saídas do classificados com as saídas almeçadas de forma que em sua diagonal principal encontram-se a quantidade de acertos tendo em vista cada classe enquanto que os valores fora da diagonal principal representam

a quantidade de dados erroneamente classificados (OSHIRO, 2013).

Em um classificador binário é possível definir valores positivos verdadeiros (*True Positives* - TP) como a quantidade de exemplos rotulados corretamente como positivos, falsos positivos (*False Positives* - FP) como a quantidade de entradas negativas incorretamente classificadas como positivas, verdadeiros negativos (*True Negatives* - TN) correspondem a negativos rotulados corretamente como negativos e os falsos negativos (*False Negatives* - FN) compreendem exemplos positivos rotulados incorretamente como negativos (SILVA, 2006). Figura 8 exemplifica a matriz de confusão de um classificador binário.

As informações disponibilizadas através de uma matriz de confusão são utilizadas no cálculo de uma série de métricas como a taxa de positivos verdadeiros ou revocação/sensibilidade (equação 6) e a taxa de falsos positivos (*False Positive Rate* - equação 7) que servem como base para a construção das curvas ROC (SILVA, 2006). De forma similar, pode-se calcular a precisão como a fração de exemplos classificados como positivos que são verdadeiramente positivos (DAVIS; GOADRICH, 2006)(equação 8). Por fim, define-se a acurácia como a taxa de acerto geral do classificador, isto é, a razão entre a soma dos acertos e o número total de instâncias (equação 9).

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

Figura 8: Matriz de Confusão gerada a partir de um conjunto de instâncias (SILVA, 2006, Adaptado).

$$Recall = \frac{TP}{TP + FN} \quad (6)$$

$$FPR = \frac{FP}{FP + TN} \quad (7)$$

$$Precision = \frac{TP}{TP + FP} \quad (8)$$

$$Accuracy = \frac{TP + TN}{P + N} \quad (9)$$

onde $P = TP + FN$ e $N = FP + TN$ são as quantidades totais de dados pertencentes às classes positiva e negativa. Da mesma forma, *Recall*, *FPR*, *Precision* e *Accuracy* representam as métricas de revocação, taxa de falsos positivos, precisão e acurácia, respectivamente.

Podem ser utilizadas algumas variantes deste modelo de matriz de confusão capazes demonstrar mais facilmente métricas importantes como acurácia e precisão. Dessa forma,

na presente dissertação de mestrado uma matriz de confusão expandida será empregada conforme ilustrado pela Figura 9.

Classe Predita	0	TN	FN	B 1-B
	1	FP	TP	Precision 1-Precision
		A 1-A	Recall 1-Recall	Accuracy 1-Accuracy
		0	1	Classe Verdadeira

Figura 9: Matriz de confusão expandida na qual métricas de avaliação como acurácia e precisão são facilmente visualizadas.

Nota-se por meio da Figura 9 que a matriz de confusão expandida demonstra claramente as métricas acurácia (*accuracy*), precisão (*Precision*) e revocação ou sensibilidade (*Recall*) bem como quantidade de verdadeiros negativos (TN), verdadeiros positivos (TP), falsos negativos (FN) e falsos positivos (FP). Nota-se ainda o percentual de instâncias corretamente classificadas como negativas em relação à quantidade de classificados como negativos (B). Por fim, é possível visualizar o percentual de instâncias corretamente classificadas como negativas em relação ao total de instâncias que deveriam ter sido classificadas como negativas (A).

2.2.3.2 Curvas ROC

A curva ROC de um classificador binário discreto plota a sensibilidade (taxa de positivos verdadeiros) em função da taxa de falsos positivos conforme ilustrado na Figura 10.

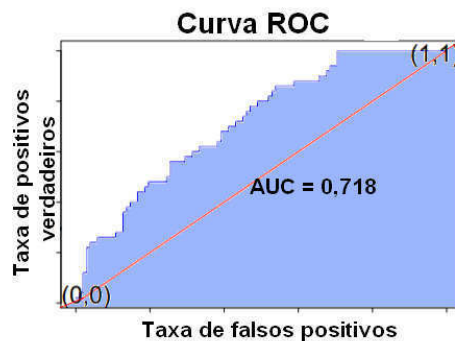


Figura 10: Exemplo de curva com ROC com área sob a curva (AUC) igual à 0,78. Linha em vermelho ilustra a curva de um classificador com saídas totalmente aleatórias (COR-TES; MOHRI, 2004, Adaptado).

Um classificador com saídas totalmente aleatórias apresenta uma linha reta desde a origem até (1,1) em sua curva ROC. Curvas pelo menos um pouco acima dessa linha apresentam aprimoramento em relação à realização de classificações aleatórias (CORTES; MOHRI, 2004). Ademais, como curvas ROC correspondem à uma demonstração bidimensional da *performance* de um classificador, torna-se necessário reduzir essa curva à um valor escalar para comparar classificadores. Dessa forma, a área sob a curva ROC (*Area Under the Curve* - AUC) varia de 0 à 1 e resume estatisticamente a qualidade do ranking de uma solução de classificação (CORTES; MOHRI, 2004).

2.2.3.3 Curvas PR

O emprego de curvas ROC pode proporcionar resultados demasiadamente otimistas quanto à *performance* de um classificador quando há desbalanceamento considerável na distribuição entre as classes. Dessa forma, as curvas de Precisão-Revocação apresentam-se como uma alternativa à curvas ROC para tarefas que envolvam conjuntos de dados desbalanceados (DAVIS; GOADRICH, 2006). Nessas curvas, a precisão (equação 8) é plotada em função da revocação (equação 6). Em outras palavras, a fração de exemplos classificados como positivos que são verdadeiramente positivos é esboçada em função da fração de exemplos positivos que foram corretamente classificados conforme exemplificado na Figura 11.

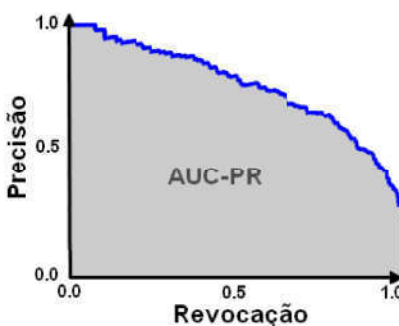


Figura 11: Exemplo de curva de Precisão-Revocação através de uma linha azul. A região cinza destaca a área sob a curva PR (AUC-PR) (CHICCO, 2017, Adaptado).

Assim como nas curvas ROC, a área sob as curvas PR (*Area Under the Curve Precision-Recall* - AUC-PR) é empregada como métrica na comparação de classificadores e quanto maior o seu valor, melhor o desempenho dos classificadores (DAVIS; GOADRICH, 2006). Além disso, como a precisão e a revocação descrevem aspectos complementares da correção de um sistema de classificação, pode-se combinar estes valores através da métrica *F-score* (SILVA, 2006) conforme equação 10.

$$F - score = \frac{2.Precision.Recall}{Precision + Recall} \quad (10)$$

2.3 Ferramentas de Visão Computacional

Com o intuito de desenvolver uma técnica de valvometria para a construção de biossensores baseados na análise comportamental de bivalves, propõe-se o emprego de câmeras em conjunto com algoritmos de visão computacional. Nesse sentido, as subseções que seguem versam sobre as ferramentas de visão computacional referentes à calibração de câmeras, operadores morfológicos e a representação de cores em imagens importantes ao desenvolvimento proposto.

2.3.1 Calibração de Câmeras

Calibração de câmera consiste no processo de determinar dados geométricos e ópticos da câmera, isto é, seus parâmetros intrínsecos e extrínsecos. Os parâmetros extrínsecos possibilitam relacionar o sistema de coordenadas da câmera com o sistema de coordenadas global (do “mundo”) enquanto que os parâmetros intrínsecos permitem modelar sua geometria e a distorção causada por suas lentes (BARELLI, 2018) conforme abordado nas subseções a seguir.

2.3.1.1 Parâmetros Intrínsecos

Estimar os parâmetros intrínsecos da câmera é essencial pois possibilita estimar a estrutura da cena no espaço Euclidiano e remover distorções causadas pelas lentes que comprometem a acurácia de algoritmos de visão computacional (BRADSKI; KAEHLER, 2008). Um modelo simples, mas útil da aquisição de imagens é o uso do modelo de câmera *pinhole* ilustrado na Figura 12. Neste modelo tanto o objeto quanto sua projeção apresentam a mesma orientação. Nesta imagem o local do orifício que permite passagem de luz corresponde ao centro de projeção e o ponto de intersecção entre o plano da imagem e o eixo óptico é referido como ponto principal.

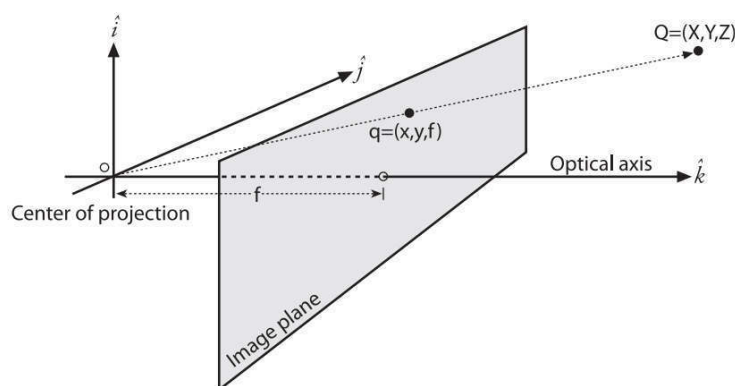


Figura 12: Ponto $Q=(X,Y,Z)$ projetado na imagem através de um raio que passa pelo centro de projeção resultando em um ponto $q = (z, y, f)$ projetado na imagem através de uma câmera com distância focal f . (BRADSKI; KAEHLER, 2008)

Nota-se através da Figura 12 que um ponto Q situado nas coordenadas X, Y, Z no mundo é representado no plano da imagem através de um ponto q de coordenadas x, y , f sendo f a distância focal da câmera. Dois parâmetros (c_x e c_y) são adicionados ao modelo com o intuito de considerar a diferença entre a posição do eixo óptico e o centro das coordenadas no plano de projeção. Consideram-se ainda duas distâncias focais f_x e f_y já que pixels individuais em câmeras de baixo custo são retangulares e não quadrados (BRADSKI; KAEHLER, 2008). As equações 11 e 12 descrevem o sistema em que um ponto $Q(X, Y, Z)$ no mundo é projetado no plano de projeção nas coordenadas x_{proj} e y_{proj} .

$$x_{proj} = f_x \cdot \frac{X}{Z} + c_x \quad (11)$$

$$y_{proj} = f_y \cdot \frac{Y}{Z} + c_y \quad (12)$$

A relação que descreve um ponto Q_i de coordenadas (X_i, Y_i, Z_i) no mundo em um ponto projetado com coordenadas (x_i, y_i) é chamada de transformação projetiva (BRADSKI; KAEHLER, 2008). É possível rearranjar os parâmetros que definem a câmera em uma única matriz 3×3 denominada matriz intrínseca da câmera (M) que relaciona a projeção de pontos do mundo (Q) no plano de imagem (q) conforme abaixo expresso:

$$q = MQ \quad (13)$$

Onde:

$$q = \begin{bmatrix} x \\ y \\ z \end{bmatrix}, \quad M = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}, \quad Q = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \quad (14)$$

A matriz intrínseca da câmera não considera distorções porque um modelo pinhole não utiliza lentes (BRADSKI; KAEHLER, 2008). Existem dois tipos básicos de distorção causadas por lentes: distorções radiais e distorções tangenciais. A distorção radial ocorre quando os raios de luz dobram mais perto das bordas de uma lente do que no centro óptico causando o efeito "olho de peixe" (MATHWORKS, 2019). Nesse fenômeno, a distorção é nula no centro óptico e aumenta à medida que se aproxima das bordas sendo modelada com base na expansão em série de Taylor em torno de um raio nulo ($r = 0$) centrado no centro óptico (BRADSKI; KAEHLER, 2008) conforme segue.

$$x_{\text{corrigido}} = x (1 + k_1 \cdot r^2 + k_2 \cdot r^4 + k_3 \cdot r^6) \quad (15)$$

$$y_{\text{corrigido}} = y (1 + k_1 \cdot r^2 + k_2 \cdot r^4 + k_3 \cdot r^6) \quad (16)$$

onde (x, y) é a coordenada original distorcida e $(x_{\text{corrigido}}, y_{\text{corrigido}})$ corresponde à coordenada corrigida. A distorção tangencial decorre de defeitos de fabricação que fazem com que a lente não esteja posicionada perfeitamente paralela ao plano da imagem (BRADSKI; KAEHLER, 2008). Os coeficientes de distorção tangencial modelam esse tipo de distorção de acordo com a equação abaixo:

$$x_{\text{distorcido}} = x + [2 \cdot p_1 \cdot x \cdot y + p_2 \cdot (r^2 + 2 \cdot x^2)] \quad (17)$$

$$y_{\text{distorcido}} = y + [p_1 \cdot (r^2 + 2 \cdot y^2) + 2 \cdot p_2 \cdot x \cdot y] \quad (18)$$

onde (x, y) são a localização do pixel sem distorção, p_1 e p_2 são os coeficientes de distorção tangencial da lente e $(x_{\text{distorcido}}, y_{\text{distorcido}})$ são a localização dos pixel distorcidos.

2.3.1.2 Parâmetros extrínsecos

Para cada imagem capturada pela câmera de um objeto específico, pode-se descrever a pose do objeto em relação ao sistema de coordenadas da câmera em termos de rotação e translação (BRADSKI; KAEHLER, 2008). O vetor de translação (T) descreve uma mudança de um sistema de coordenadas para outro sistema cuja origem é deslocada (BRADSKI; KAEHLER, 2008).

Uma matriz de rotação em três dimensões pode ser decomposta em uma rotação bidimensional em torno de cada eixo, no qual as medições do eixo pivô permanecem constantes (BRADSKI; KAEHLER, 2008). Se girarmos em torno dos eixos x , y e z em sequência com os respectivos ângulos de rotação ψ , φ e θ , o resultado é uma matriz de rotação total R que é dada pelo produto das três matrizes $R_x(\psi)$, $R_y(\varphi)$ e $R_z(\theta)$ (BRADSKI; KAEHLER, 2008), em que:

$$R_x(\psi) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \psi & \sin \psi \\ 0 & -\sin \psi & \cos \psi \end{bmatrix}, R_y(\varphi) = \begin{bmatrix} \cos \varphi & 0 & -\sin \varphi \\ 0 & 1 & 0 \\ \sin \varphi & 0 & \cos \varphi \end{bmatrix} \quad (19)$$

$$R_z(\theta) = \begin{bmatrix} \cos \theta & \sin \theta & 0 \\ -\sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (20)$$

2.3.2 Operadores Morfológicos

Operações morfológicas são operações que modificam a estrutura dos objetos representados em uma imagem. Embora possam ser utilizadas em qualquer tipo de imagem, elas são normalmente empregadas em imagens binárias para amenizar ruídos ou realçar objetos de interesse (BARELLI, 2018). Estas operações são realizadas pixel a pixel com base na convolução de um elemento estruturante (normalmente representado por uma matriz quadrada) com a imagem a ser modificada.

As operações morfológicas de erosão e dilatação são as mais fundamentais uma vez que todas as demais podem ser descritas basicamente como uma determinada sequência de erosões e dilatações (BARELLI, 2018). A operação de erosão permite a corrosão de arestas do objeto de interesse e pode levar à eliminação de pequenos objetos presentes na imagem sendo empregada na eliminação de pequenos ruídos. Sua realização consiste em sobrepor cada pixel da imagem com o centro do elemento estruturante. O pixel analisado somente será representado na imagem resultante se todos os pixels demarcados no elemento estruturante coincidirem com pontos do objeto de interesse conforme exemplificado na Figura 13.

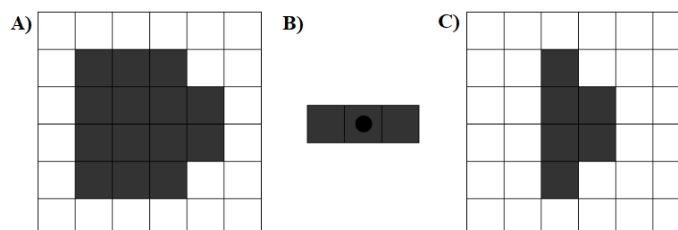


Figura 13: Exemplo de operação morfológica de erosão na qual a imagem original (A) passa por um processo de convolução com um elemento estruturante retangular (B) resultando na imagem a direita (C) (BARELLI, 2018).

A operação morfológica de dilatação possibilita alargar a área de objetos representados. Dessa forma, os artefatos presentes na imagem tendem a se tornar maiores (BARELLI, 2018) conforme exemplificado na Figura 14. No processo de convolução se pelo menos um pixel do elemento coincidir com um pixel representado na imagem, a posição do pixel central fará parte da imagem resultante.

Com o intuito de eliminar ruídos e ao mesmo tempo preservar as regiões de interesse pode-se fazer uso do operador morfológico de abertura que consiste em uma operação de erosão seguida de uma operação de dilatação (BARELLI, 2018). De forma análoga, o operador morfológico de fechamento é caracterizado pela utilização destes operadores em ordem inversa e permite preencher pontos de objetos que foram danificados durante o processo de binarização.

As operações de abertura e fechamento são utilizadas com o objetivo de uniformizar a iluminação dos objetos uma vez que tendem a suprimir pequenas regiões brilhantes e

escuras, respectivamente (BARELLI, 2018). A Figura 15 ilustra o operação de abertura a uma imagem em escala de cinza e sua subtração com a imagem original na correção da iluminação da imagem.

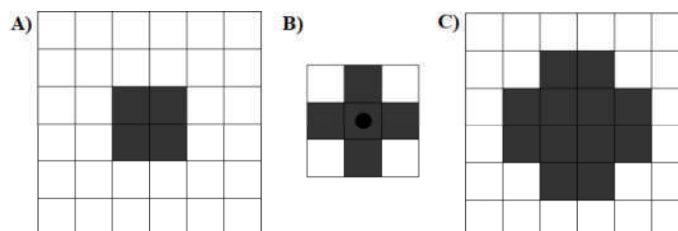


Figura 14: Exemplo de operação morfológica de dilatação na qual a imagem original (A) passa por um processo de convolução com um elemento estruturante em forma de cruz (B) resultando na imagem a direita (C) (BARELLI, 2018).

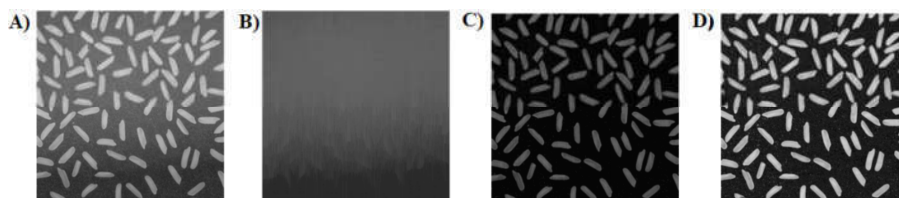


Figura 15: Exemplo da aplicação do operador morfológico de abertura à uma imagem em escala de cinza (A). O resultado (B) quando subtraído da imagem original fornece uma imagem com iluminação mais uniforme (C) que pode ser melhorada através do ajuste de seu contraste (D) (BARELLI, 2018).

2.3.3 Representação de cores em imagens

Cor é uma experiência rica e complexa geralmente causada pelo fato do sistema visual humano responder de forma alterada a diferentes comprimentos de onda da luz (FORSYTH; PONCE, 2003). A representação de imagens coloridas requer o emprego de múltiplos canais enquanto que imagens em escala de cinza (monocromáticas) possuem apenas um canal que representa a luminância (Y) em todos os pontos de uma dada cena (DAWSON-HOWE, 2014). As informações de cor podem ser representadas de diversas formas como, por exemplo, através dos espaços de cor RGB (*Red-Green-Blue*) e HSV (*Hue, Saturation and Value*) abordados nas subseções a seguir e ilustrados por meio da Figura 16.

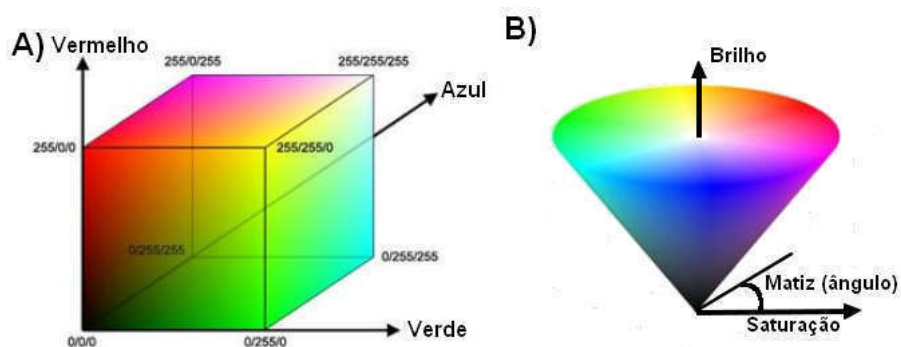


Figura 16: O espaço RGB é representado por um cubo (A) através do qual é possível visualizar todas as cores passíveis de obtenção a partir da combinação das componentes vermelha, verde e azul variando de 0 à 1 deste espectro. É comum visualizar este cubo ao longo de seu eixo neutro (eixo da origem ao ponto 1,1,1). A seção transversal do formato resultante corresponde ao cone que representa o espaço HSV (B). Fonte: Corrêa¹ et al. (2014) (adaptado).

2.3.3.1 Espaço RGB

De acordo com Dawson-Howe (2014), a representação de imagens coloridas mais comum consiste na utilização de três canais que representam, de forma resumida, os comprimentos de onda vermelho (*Red* - 700 nm), verde (*Green* - 546 nm) e azul (*Blue* - 435 nm). Trata-se do espectro de cores linear RGB no qual cada pixel é representado pela mistura de componentes vermelha, verde e azul cujos valores variam de 0 à 255 conforme ilustrado na Figura 16A.

Embora a utilização deste espaço possibilite representar aproximadamente 16,8 milhões ($256 \times 256 \times 256$) cores, segundo Dawson-Howe (2014) algumas cores não podem ser representadas através desse modelo.

2.3.3.2 Espaço HSV

Embora o espaço de cor RGB seja amplamente utilizado, coordenadas de cor em um espaço linear não capturam intuições humanas acerca da topologia de cores (FORSYTH; PONCE, 2003). Dessa forma, pode-se construir espaços de cor que refletem essas relações através da transformação não linear do espaço RGB. Dentre estes encontra-se o espaço HSV (*Hue, Saturation and Value*) cujas componentes encontram-se abaixo descritas conforme Forsyth e Ponce (2003). Ademais, este espaço é representado através de um cone (Figura 16B).

- Matiz (*hue*)- propriedade de uma cor que varia na passagem de vermelho para verde;
- Saturação (*saturation*) - propriedade de uma cor que varia na passagem de vermelho

para rosa;

- Brilho - também chamado de valor (*value*), corresponde a propriedade que varia na passagem de preto para branco;

2.4 Ferramentas de análise estatística

A presente dissertação de mestrado propõe uma técnica de valvometria baseada em visão computacional e sua validação através de um comparativo com o emprego de sensores Hall por meio de ferramentas estatísticas. Nesse sentido, analisam-se as curvas de regressão linear e a correlação entre os dados obtidos por estas abordagens bem como a Função Densidade de Probabilidade (FDP) entre os erros. Empregam-se ainda gráficos de controle de qualidade como ferramenta de análise a fim de identificar alterações decorrentes de exposição toxicológica.

Por fim, diferenças significativas entre grupos de animais expostos e não expostos a contaminantes são investigadas a fim de comparar a performance dos modelos propostos com o emprego de métodos estatísticos tradicionais. Dentre estes métodos, pode-se citar Análise de Variância (ANOVA - *Analysis of variance*) de um ou múltiplos fatores e o teste não paramétrico de Kruskal-Wallis. Dessa forma, as subseções a seguir versam sobre regressão linear, correlação, FDP, gráficos de controle de qualidade, métodos ANOVA e Kruskal-Wallis.

2.4.1 Regressão linear

A análise de regressão linear bivariada permite a obtenção de um modelo matemático capaz de estimar o valor de uma variável (y - variável dependente) a partir de outra (x - variável independente) (BLACK, 2009). Nesta análise tem-se, portanto, a modelagem da relação entre as variáveis dependente e independente através de uma linha reta. Em estatística, o modelo probabilístico que descreve a linha de regressão através de pontos amostrais pode ser definida conforme equação 21. Figura 17 ilustra o modelo regressão linear para uma dada amostra de variáveis dependentes e independentes.

$$\hat{y} = b_0 + b_1 \cdot x \quad (21)$$

onde \hat{y} corresponde ao valor estimado para a variável dependente, x representa a variável independente, b_0 o valor onde a reta intercepta o eixo y e b_1 é a inclinação da curva.

Torna-se necessário, portanto, determinar os coeficientes b_0 e b_1 . Nesse sentido, pode-se empregar o método dos mínimos quadrados uma vez que ele possibilita determinar os coeficientes em questão de forma a produzir a mínima soma dos valores de erro ao quadrado (BLACK, 2009). Deve-se, portanto, definir a soma do quadrado dos erros (SQ)

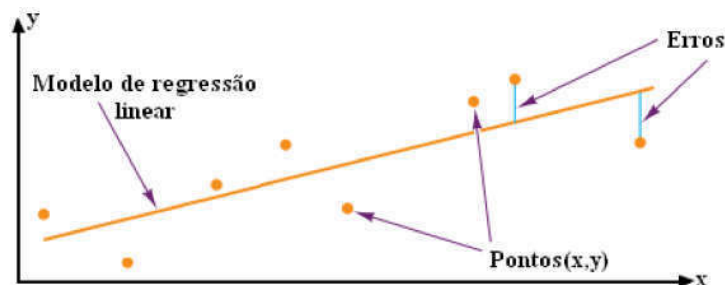


Figura 17: Modelo de regressão linear baseado em pontos amostrais de duas variáveis dependente (y) e independente (x). A reta modelada não estima corretamente todos valores de y tendo como base os valores de x havendo, dessa forma, erros definidos como a diferença entre os valores real e estimado da variável dependente (BLACK, 2009)(adaptado)

sendo estes definidos como as diferenças entre os valores estimados e reais da amostra da variável dependente ($e = \hat{y} - y$) de tamanho n conforme equação 22.

$$SQ = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - (b_0 + b_1 \cdot x_i))^2 \quad (22)$$

Com o intuito de encontrar os coeficiente que minimizam SQ , pode-se derivar a equação 22 em função de b_0 e b_1 e igualar a zero conforme as equações 23 e 24. Solucionando o sistema de equações resultante pode-se obter os coeficientes do modelo de regressão linear baseado no método de mínimos quadrados conforme equações 25 e 26 (BLACK, 2009).

$$n \cdot b_0 + b_1 \sum_{i=1}^n x_i = \sum_{i=1}^n y_i \quad (23)$$

$$b_0 \sum_{i=1}^n x_i + b_1 \cdot \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i \cdot y_i \quad (24)$$

$$b_0 = \bar{y} - b_1 \cdot \bar{x} \quad (25)$$

$$b_1 = \frac{\sum_{i=1}^n x_i \cdot y_i - n \cdot \bar{x} \cdot \bar{y}}{\sum_{i=1}^n x_i^2 - n \cdot \bar{x}^2} \quad (26)$$

2.4.2 Correlação

Em muitas aplicações torna-se interessante analisar o relacionamento entre duas ou mais variáveis. Nesse sentido, os coeficientes de associação ou correlação permitem avaliar quantitativamente o grau de associação ou dependência linear entre duas variáveis (PUBLISHERS, 2010). A exemplo disso, pode-se citar o coeficiente de Pearson (r) que fornece valores entre -1 e +1 (equação 27). Valores iguais a -1, +1 e 0 indicam associações perfeitas negativas, perfeitas positivas e a não existência de associação li-

near entre as variáveis, respectivamente (PUBLISHERS, 2010). Além disso, a análise estatística r pode indicar diferentes níveis de correlação como correlações lineares muito fracas ($0 < |r| < 0.2$), fracas ($0.2 \leq |r| < 0.4$), moderadas ($0.4 \leq |r| < 0.6$), fortes ($0.6 \leq |r| < 0.8$) ou muito fortes ($0.8 \leq |r| < 1$) (CHRISTMANN; BADGETT, 2009).

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}} \quad (27)$$

onde r é o coeficiente de correlação de Pearson entre as variáveis x e y cujos valores médios são \bar{x} e \bar{y} , respectivamente. Por fim, o quadrado deste coeficiente é conhecido como coeficiente de determinação (r^2) e possibilita avaliar a capacidade de um dado modelo de regressão linear de explicar um resultado (BLACK, 2009). Ele indica, portanto, a proporção da variância na variável dependente que é explicada pelo modelo de regressão linear e a variável independente.

2.4.3 Função densidade de probabilidade

A FDP é um conceito fundamental em estatística (SILVERMAN, 1986). Considerando a função densidade de probabilidade f de uma variável aleatória X , pode-se afirmar que f possibilita encontrar a probabilidade da variável X estar situada entre os valores a e b conforme descrito em 28.

$$P(a \leq X < b) = \int_a^b f(x)dx \quad \forall a < b \quad (28)$$

Nesse sentido, dado um conjunto de pontos amostrais de um dado processo ou fenômeno com uma função de densidade de probabilidade desconhecida torna-se interessante estimar $f(x)$ a partir dos dados observados. Uma abordagem paramétrica pode ser empregada com esta finalidade. Entretanto, nesta abordagem supõem-se que os dados extraídos pertencem a uma família paramétrica conhecida de distribuições (SILVERMAN, 1986). Muitas vezes, a utilização de abordagens não paramétricas torna-se mais adequada aos dados amostrais, visto que suposições menos rígidas são feitas acerca de sua distribuição.

Em estatística, a Estimativa de Densidade por *Kernel* (EDK) é uma forma não-paramétrica para estimar a FDP de uma variável aleatória (SILVERMAN, 1986). Sendo x_1, x_2, \dots, x_n uma amostra de uma variável aleatória independente e identicamente distribuída com uma densidade desconhecida, pode-se estimar sua FDP (\hat{f}) através de uma EDK conforme 29.

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \quad (29)$$

onde \hat{f} é a FDP estimada, K é a função *kernel* empregada, h é um parâmetro de

distância denominado de largura de banda que controla a influência de valores vizinhos em um *kernel*. Uma função *kernel* centrada em X_i atribui pesos às contribuições dadas por cada vizinho x à FDP estimada \hat{f} de acordo com a proximidade de X_i a x . Normalmente, funções *kernel* são positivas em todos os lugares e simétricas em relação a zero. A Figura 18 exemplifica uma EDK utilizando uma função normal.

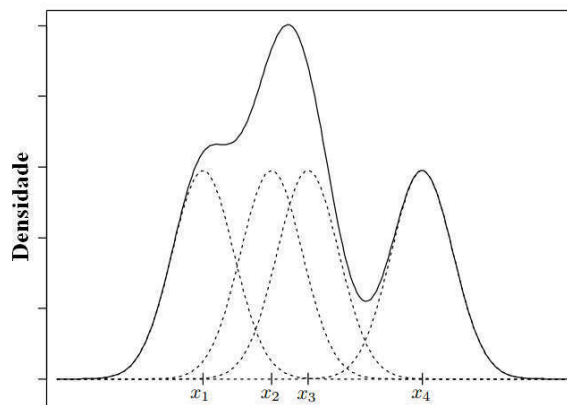


Figura 18: Estimativa normal de densidade por *kernel* (linha sólida) e contribuições do *kernel* (linhas pontilhadas) centrado em cada X_i da amostra x_1, x_2, \dots, x_4 . A EDK em qualquer x é a soma das contribuições do *kernel* centrado em cada X_i (GIVENS; HOE-TING, 2012)(adaptado).

2.4.4 Gráficos de Controle de Qualidade

De acordo com Black (2008), qualidade é habilidade de um produto em fornecer características condizentes com suas especificações. Controle de qualidade pode ser entendido como um conjunto de estratégias, técnicas e ações utilizadas por uma organização com o intuito de assegurar a produção de produtos de qualidade.

Além disso, processo pode ser definido como uma série de ações, alterações ou funções que levam a um resultado. Nesse sentido, muitas ferramentas tem sido desenvolvidas com o intuito de auxiliar no controle de qualidade de processos. Dentre estas ferramentas, gráficos de controle são ferramentas estatísticas que englobam métodos gráficos para avaliar se um processo está ou não em um estado de controle estatístico (BLACK, 2008).

Cada gráfico de controle apresenta uma linha central ou limite de controle médio (*Central Line* - CL), e linhas de controle superior (*Upper Control Limit* - UCL) e inferior (*Lower Control Limit* - LCL). Estas ferramentas podem ser utilizados para verificar a previsibilidade de séries temporais fazendo uso de dados anteriores para determinar CL, UCL e LCL. A Figura 19 ilustra um exemplo de gráfico de controle EWMA.

Os limites estabelecidos no gráfico de controle (UCL, CL e LCL) podem ser empregados para concluir se a variação do processo é consistente (no controle) ou imprevisível (fora de controle) e, portanto, afetada por causas especiais de variação (TAGUE, 2005).

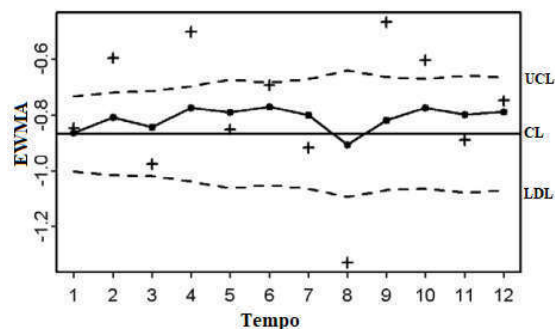


Figura 19: Exemplo de gráfico de controle EWMA para uma série temporal. As linhas de limite central (CL), superior (UCL) e inferior (LCL) encontram-se ilustradas. Os círculos indicam os valores resultantes da média ponderada e as cruzes indicam os valores da série analisada a cada intervalo de tempo (SIREGAR et al., 2013, Adaptado).

Dentre outros, o gráfico de controle baseado na média móvel exponencialmente ponderada (*Exponentially Weighted Moving Average* - EWMA) é um procedimento quase perfeitamente não paramétrico, muito eficaz contra pequenas variações no processo e sensível à autocorrelação dos dados (MONTGOMERY, 2009).

A EWMA é uma média ponderada das observações atuais e passadas. Os pesos são baseados no tempo de forma que as observações atuais apresentam peso exponencialmente maior que as mais antigas. Dessa forma, seu emprego baseia-se na configurações de dois parâmetros sendo eles o fator peso λ dado à observação mais recente e o fator k que determina o número de desvios padrões empregados para estabelecer os limites superior e inferior. Dessa forma, a estatística empregada nestes gráficos de controle pode ser calculada conforme equação 30.

$$z_i = \lambda \sum_{j=0}^{i-1} (1 - \lambda)^j x_{i-j} + (1 - \lambda)^i z_0 \quad (30)$$

onde $z_0 = \bar{x} = \mu_0$, z_i é o valor da média ponderada no i -ésimo intervalo e x representa os valores inspecionados através do mapa de controle EWMA. Nota-se através desta equação que os valores dos pesos decaem geometricamente. Além disso, UCL e LCL são estabelecidos com base na linha central (z_0) e na variância (σ) conforme as equações (31) e (32), respectivamente.

$$UCL = \mu_0 + k\sigma \sqrt{\frac{\lambda}{2 - \lambda}} \quad (31)$$

$$LCL = \mu_0 - k\sigma \sqrt{\frac{\lambda}{2 - \lambda}} \quad (32)$$

2.4.5 Métodos ANOVA e alternativas não-paramétricas

Segundo Publishers (2010), a análise da variância (*Analysis Of Variance* - ANOVA) pode ser pensada como um método para testar a hipótese de igualdade entre as médias de diferentes grupos por meio da análise das variâncias de suas amostras. Dessa forma, o método ANOVA pode ser empregado para determinar a existência de diferenças significativas entre grupos de acordo com um ou mais fatores.

O método ANOVA de um fator possibilita a comparação de médias univariadas de um ou mais grupos. Por sua vez, o método ANOVA multivariado (*Multivariate Analysis of Variance* - MANOVA) corresponde a uma extensão da análise de variância de um fator capaz de identificar diferenças significativas entre grupos com base em diversas variáveis respostas (fatores) simultaneamente (REIS, 2001).

As informações fornecidas pela modelagem univariada são contempladas pela extensão multivariada sem, no entanto, levar em consideração um nível de significância conjunto dos testes e um aproveitamento das correlações existentes entre variáveis (DEMÉTRIO, 1995). Ademais, existem diversas razões para adotar testes multivariados no lugar de testes univariados repetidos. Primeiramente, a utilização de testes estatísticos separados para cada variável provoca incremento demasiado de erros tipo I, ou seja, aumenta a probabilidade de rejeitar a hipótese nula quando ela é verdadeira (COUTO, 2006).

Além disso, embora não sejam verificadas diferenças significativas entre grupos na análise de cada variável individual, quando estas são inspecionadas em conjunto por meio de um método multivariado diferenças significativas poderão ser identificadas devido à combinação de pequenas diferenças entre variáveis individuais (COUTO, 2006). Nesses casos, testes multivariados tornam-se mais poderosos com maior probabilidade de rejeitar a hipótese nula quando ela é falsa (COUTO, 2006). Por fim, testes univariados não consideram informações importantes contidas nos dados como, por exemplo, a correlação entre variáveis que é incorporada na análise multivariada através de matrizes de variâncias e covariâncias.

O emprego do método MANOVA requer a verificação de alguns pressupostos derivados do método ANOVA de um fator sendo eles a normalidade multivariada e a igualdade de matrizes de variância e covariância (COUTO, 2006). Quando as suposições necessárias ao emprego dos métodos paramétricos não são verificados, alternativas não paramétricas devem ser empregadas. Dentre estas alternativas podem-se citar os métodos de Kruskal-Wallis para o caso univariado e a análise de variância multivariada permutacional (*Permutational Multivariate Analysis of Variance* - PERMANOVA) (FACHADA et al., 2016).

O método PERMANOVA foi proposto por McArdle e Anderson (2001) e, de acordo com Sun (2020), é um dos testes estatísticos não-paramétricos multivariados mais amplamente usados em ecologia microbiana. Ele é, portanto, um teste não paramétrico multivariado ANOVA com hipótese nula de não haver diferenças significativas entre os grupos

e baseia-se na análise e particionamento da soma dos quadrados das distâncias e processo de permutação.

Com o intuito de verificar os conceitos básicos referentes ao método ANOVA de um fator a partir do qual os métodos MANOVA e PERMANOVA são derivados bem como acerca de sua opção não paramétrica, as subseções a seguir apresentam os métodos ANOVA de um fator e Kruskal-Wallis.

2.4.5.1 ANOVA

Métodos ANOVA podem ser empregados para investigar a influência de um determinado fator sob um fenômeno observado. Cada valor observado pode, portanto, ser modelado conforme a equação 33 sendo composto por uma parte previsível (ou controlada) afetada pelo fator investigado e outra aleatória (não previsível).

$$y_{i,j} = \theta_i + e_{i,j} \quad (33)$$

onde $y_{i,j}$, θ_i e $e_{i,j}$ correspondem à j -ésima observação do grupo (nível) i , ao efeito comum a todas as observações do grupo (nível) i e ao efeito aleatório (não controlado) da j -ésima observação deste mesmo nível. O erro pode ser considerado como o efeito resultante de características não explicadas pelo modelo.

Considerando-se um fator de n níveis, deve-se estimar $\theta_1, \theta_2, \dots, \theta_n$ capazes de minimizar o somatório do quadrado dos ruídos e verificar se o conhecimento do fator escolhido influencia o fenômeno estudado. No entanto, esta abordagem requer o emprego de algumas restrições conforme expresso pelas equações 34 e 35 nas quais e_i corresponde ao erro não controlado das observações do nível (grupo) i .

$$E(e_i) = 0, \quad \forall i \in [1, n] \quad (34)$$

$$Var(e_i) = \sigma_i^2, \quad \forall i \in [1, n] \quad (35)$$

Considerando-se um fator de n níveis e que cada nível i possui n_i observações, o somatório do quadrado dos resíduos pode ser expresso pelas equações 36 e 37.

$$SQ(\theta_1, \dots, \theta_n) = \sum_{i=1}^n \sum_{j=1}^{n_i} e_{i,j}^2 \quad (36)$$

$$SQ(\theta_1, \dots, \theta_n) = \sum_{j=1}^{n_1} e_{1,j}^2 + \sum_{j=2}^{n_2} e_{2,j}^2 + \dots + \sum_{j=n}^{n_n} e_{n,j}^2 \quad (37)$$

Observa-se através da equação 37 que se as variâncias residuais $\theta_1^2, \theta_2^2, \dots, \theta_n^2$ não forem iguais, a soma dos quadrados dos resíduos será mais afetada pelo grupo com menor

variância. Para fins de simplificação considera-se variância constante dos erros, propriedade também conhecida como homoscedasticidade (equações 38 e 39). Admite-se, dessa forma, a mesma variabilidade residual entre os grupos.

$$\text{Var}(e_1) = \text{Var}(e_2) = \dots = \text{Var}(e_n) \quad (38)$$

$$\sigma_1^2 = \sigma_2^2 = \dots = \sigma_e^2 \quad (39)$$

Com esse intuito de encontrar $\theta_1, \theta_2, \dots, \theta_n$ capazes de minimizar a soma dos quadrados dos resíduos dos diferentes níveis, efetua-se sua derivada função da parte previsível de cada grupo e iguala-se a zero conforme segue:

$$\frac{\partial SQ(\theta_1, \dots, \theta_n)}{\partial \theta_i} = 0 \quad (40)$$

$$-2 \sum_j^{n_i} (y_{i,j} - \hat{\theta}_i) = 0 \therefore \hat{\theta}_i = \frac{\sum_{j=1}^{n_i} y_{i,j}}{n_i} \quad (41)$$

Em que $i \in [1, n]$, n é a quantidade de níveis do fator considerado e n_i é a quantidade de observações de cada nível i . Analisando-se a expressão 41 nota-se que $\hat{\theta}_i = \bar{y}_i$. Pode-se afirmar, portanto, que a quantidade total de informação perdida pela adoção do modelo é dada por 42.

$$SQ(\theta_1, \dots, \theta_n) = \sum_{i=1}^n \sum_{j=1}^{n_i} (y_{i,j} - \bar{y}_i)^2 \quad (42)$$

Analisando-se a variância residual de cada nível e a minimização proposta, pode-se afirmar que $s_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (y_{i,j} - \bar{y}_i)$. Contudo, considerando a suposição de homoscedasticidade, pode-se afirmar que:

$$s_e^2 = s_i^2 \quad \forall i \in [1, n] \quad (43)$$

$$s_e^2 = \frac{\sum_{i=1}^n (n_i - 1) \cdot s_i^2}{\sum_{i=1}^n n_i - n} \quad (44)$$

$$s_e^2 = \frac{SQ(\theta_1, \dots, \theta_n)}{m - n} \quad (45)$$

Em que $m = \sum_{i=1}^n n_i$ e s_e^2 corresponde a quantidade média de informação perdida. A homoscedasticidade entre os níveis deve ser observada e torna-se necessário estabelecer suposições acerca do efeito residual e independência das observações tanto dentro quanto entre os níveis do fator considerado conforme segue.

- $e_{i,j} : N(0, \sigma_e^2) \quad \forall i \in [1, n] \text{ e } j \in [1, n_i]$

- $E(e_i.e_j) = 0 \quad \forall i \neq j$, indicando independência entre as observações dentro de cada população;

As suposições necessárias a adoção do modelo em questão requer, portanto, que os resíduos sigam uma distribuição normal com média 0 e variância σ_e^2 e que resíduos associados a pares de observações sejam independentes. Com base nessas suposições, pode-se afirmar que o conjunto de observações y_1, y_2, \dots, y_n é uma amostra aleatória simples da população $N(\theta, \sigma_e^2)$.

As informações utilizadas para empregar a ANOVA podem ser resumidas através de uma tabela tendo em vista o modelo adotado. Conforme anteriormente exposto, a equação 42 corresponde a quantidade total de informações perdidas dentro dos grupos sendo, conforme a equação 46, referenciada pela tabela de análise de variância como SQ_{Den} . De forma análoga, a quantidade total de informações perdidas pela adoção do modelo em questão é dada pela equação 47.

$$SQ_{Den} = SQ(\theta_1, \dots, \theta_n) = \sum_{i=1}^n \sum_{j=1}^{n_i} (y_{i,j} - \bar{y}_i)^2 \quad (46)$$

$$SQ_{Total} = \sum_{i=1}^n \sum_{j=1}^{n_i} (y_{i,j} - \bar{y})^2 \quad (47)$$

em que \bar{y} corresponde à média de todas observações conforme descrito pela equação 48 e m é a quantidade total de observações dada por $m = \sum_{i=1}^n n_i$.

$$\bar{y} = \frac{\sum_{i=1}^n \sum_{j=1}^{n_i} y_{i,j}}{m} \quad (48)$$

Subtraindo-se a quantidade total de informação perdida (SQ_{Total}) da soma dos quadrados dos resíduos dentro dos grupos (SQ_{Den}) tem-se a economia obtida ao se passar de um modelo para o outro (SQ_{Ent}) conforme descrito pela equação 49.

$$SQ_{Ent} = SQ_{Total} - SQ_{Den} \quad (49)$$

$$SQ_{Ent} = \sum_{i=1}^n (n_i(\bar{y}_i - \bar{y}))^2 \quad (50)$$

Dessa forma, a expressão 50 mede uma certa distância entre a média de cada grupo e a média global dos dados analisados. De acordo com Bussab (1988), quanto maior a diferença entre as médias \bar{y}_i mais eficaz é a consideração do fator na descrição do fenômeno analisado. Além disso, pode-se afirmar que a quantidade total média de informação perdida QM_{Total} é dada pela equação 51. De modo análogo, é possível afir-

mar que a quantidade média de informação perdida entre os níveis (QM_{Den}) pode ser expressa pela equação 52.

$$QM_{Total} = \frac{1}{n-1} \sum_i^n (y_i - \bar{y}) = \frac{1}{n-1} SQ_{Total} \quad (51)$$

$$QM_{Den} = \frac{1}{m-n} \sum_{i=1}^n \sum_{j=1}^{n_i} (y_{i,j} - \bar{y}_i)^2 \quad (52)$$

$$QM_{Den} = \frac{1}{m-n} SQ_{Den} \quad (53)$$

em que $m = \sum_{i=1}^n n_i$ corresponde a quantidade total de observações. Estas informações são agrupadas através de uma tabela de análise de variância tal qual a Tabela 1.

Tabela 1: Tabela de Análise de Variância

Fonte de variação	Graus de liberdade (gl)	Soma dos quadrados (SQ)	Quantidade média (QM)	F
Entre	$n - 1$	SQ_{Ent}	QM_{Ent}	$\frac{QM_{Ent}}{QM_{Den}}$
Dentro	$m - n$	SQ_{Den}	QM_{Den}	
Total	$m - 1$	SQ_{Total}		

Pode-se reescrever a equação 33 com base nas observações de todos os níveis do fator considerado tal qual a equação 54. Nota-se, todavia, que a influência desse fator segue compondo a parte aleatória do modelo.

$$y_{i,j} = \theta + e_{i,j} \quad (54)$$

Tem-se, por conseguinte, a adoção da hipótese nula $H_0 : \theta_1 = \theta_2 = \dots = \theta_n$ contra a hipótese alternativa de que pelo menos uma das igualdades não se verifica. Segundo Bussab (1988), rejeitar a hipótese nula significa afirmar que o fator é importante para explicar a variável de interesse ou fenômeno observado.

Considerando-se, para fins de simplificação, que todos os níveis contam com a mesma quantidade de observações, pode-se estabelecer a variância comum σ_e^2 através da estimativa ponderada s_e^2 abaixo reescrita.

$$s_e^2 = \frac{\sum_{i=1}^n (n_i - 1) s_i^2}{m - n} = \frac{SQ_{Den}}{m - n} = QM_{Den} \quad (55)$$

Tendo em vista, entretanto, o Teorema do Limite Central, pode-se afirmar que:

$$\bar{y}_i : N(\theta_i : \frac{\sigma_e^2}{n_i}) \quad \forall i \in [1, n] \quad (56)$$

Quando, por outro lado, não são encontradas evidências para rejeitar a hipótese nula, tem-se:

$$\bar{y}_i : N\left(\theta : \frac{\sigma_e^2}{n_n}\right) \quad \forall i \in [1, n] \quad (57)$$

Dessa maneira, $(\bar{y}_1, \bar{y}_2, \dots, \bar{y}_n)$ pode ser considerada uma amostra aleatória simples da população $N\left(\theta : \frac{\sigma_e^2}{n_n}\right)$ e, nesse cenário, as equações 58 e 59 descrevem estimadores não viesados de θ e $\frac{\sigma_e^2}{n_n}$, respectivamente.

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n \bar{y}_i \quad (58)$$

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (\bar{y}_i - \bar{y})^2 \quad (59)$$

Com base na equação 59, pode-se afirmar que $n_n s^2$ se trata de um estimador não viesado de σ_e^2 . Tendo em vista tais condições, pode-se afirmar que se a hipótese nula for verdadeira, a estatística F descrita pela equação 60 será aproximadamente uma unidade enquanto que, se forem encontradas evidências para rejeitar H_0 , ela adquirirá valores maiores que 1. Dessa forma, a estatística F permite testar $H_0 = \theta_1 = \theta_2 = \dots = \theta_n$. Ademais, observando-se as equações 58 e 59, pode-se aproximar a estatística F pela razão de dois qui-quadrados conforme expresso pela equação 60.

$$F = \frac{ms^2}{s_e^2} = \frac{QM_{Ent}}{QM_{Den}} : \frac{\chi^2(k-1)}{k-1} : \frac{\chi^2(n-k)}{n-k} \quad (60)$$

A estatística F corresponde, portanto, ao quociente de dois qui-quadrados seguindo, por conseguinte, uma distribuição F de Snedecor com graus de liberdade $n - 1$ e $m - n$ (BUSSAB, 1988). Dessa forma pode-se afirmar por meio da tabela 1 as informações acerca tanto dos graus de liberdade quanto da estatística F são fornecidos.

2.4.5.2 *Kruskal-Wallis*

O teste não paramétrico de Wilcoxon permite verificar se duas populações podem ser consideradas semelhantes a partir de amostras independentes e é classificado como um teste não paramétrico. Esse teste baseia-se na classificação (posto) dos valores obtidos através da combinação das amostras de ambas populações para lidar com valores extremos que inviabilizam uma distribuição normal (OTT; LONGNECKER, 2015).

Considerando-se k o número de populações a serem comparadas e n_i a quantidade de observações da amostra da população $i \in [1, k]$, o método da soma dos postos das observações encontra-se descrito a seguir.

1. Ordena de forma crescente os dados resultantes da combinação das amostras populacionais;

2. Atribui valores de 1 à $n_T = \sum_{i=1}^k n_i$ a todas as observações sendo 1 associado ao menor valor e n_T ao maior;
3. Calcula T_i como sendo a soma das classes (postos) de todas as observações da população $i \in [1, k]$;

Com base na soma dos postos da observações, teste de Kruskal-Wallis adota a hipótese nula (H_0) de que todas as distribuições são idênticas. Conforme Ott e Longnecker (2015), a estatística utilizada nesse teste é descrita pela equação 61.

$$H = \frac{12}{n_T(n_T + 1)} \cdot \sum_{i=1}^k \frac{T_i^2}{n_i} - 3(n_T + 1) \quad (61)$$

2.5 Considerações Finais

Tendo em vista os objetivos da presente dissertação de mestrado, a seção de revisão bibliográfica apresentou os principais conceitos necessários ao entendimento dos modelos computacionais propostos bem como sua validação. A subseção 2.1 apresentou redes neurais artificiais de forma resumida bem como NARNETs como ferramentas propícias à modelagem de séries temporais que serão empregadas para modelar o comportamento de moluscos individuais com o intuito de considerar sua natureza adaptativa.

Ferramentas capazes de inferir a condição ambiente, como por exemplo a presença ou ausência de contaminantes, a partir da análise comportamental de bivalves são essenciais à construção dos modelos computacionais propostos. Dessa forma, a subseção 2.2 versou sobre os modelos de classificação e as métricas de avaliação que podem ser utilizadas na análise de sua *performance*.

A subseção 2.3 discorreu sobre ferramentas de visão computacional importantes à construção de uma técnica de valvometria totalmente não invasiva baseada em imagens e segmentação de cores. Por fim, a subseção 2.4 apresentou ferramentas de análise estatística empregadas não só na validação, mas também construção dos modelos computacionais propostos.

As ferramentas de regressão linear, correlação e função densidade de probabilidade são importantes na validação da técnica de valvometria proposta por meio de comparações com outra técnica de valvometria bem estabelecida. Gráficos de controle de qualidade permitem analisar a previsibilidade de séries temporais sendo utilizados, portanto, em conjunto com NARNETs na construção de biossensores de poluição aquática.

Métodos estatístico, como por exemplo ANOVA, são comumente empregados na construção de biossensores partir da identificação de diferenças significativas entre dados comportamentais de bivalves expostos e não expostos à condições ambientais de interesse. Dessa forma, estes métodos são empregados para determinar se os modelos

computacionais propostos apresentam vantagens para detecção de contraminantes apresentam vantagem em relação à estes métodos tradicionais, eles são empregados para fins de comparação na detecção de poluentes.

3 TRABALHOS RELACIONADOS

Sabendo-se que biossensores correspondem à associação de componentes biológicos e detectores físico-químicos, pode-se afirmar que sua construção parte da escolha dos animais e sensores empregados em sua confecção. Nesse contexto, animais bivalves cumprem os requisitos necessários sendo não só sésseis, abundantes e disponíveis ao longo do ano (KRAMER; JENNER; ZWART, 1989), como também manuseáveis e resistentes ao manuseio laboratorial (PHILLIS, 1980).

Tratando-se, portanto, do emprego desses animais pode-se afirmar que dentre as formas de análise toxicológica encontram-se métodos baseados em bioacumulação e mortalidade bem como métodos de análise comportamental. Métodos baseados em bioacumulação baseiam-se na análise laboratorial de seus tecidos a fim de inferir a contaminação ambiente. Estes métodos apresentam como desvantagem o fato do equilíbrio de concentração ser normalmente obtido após semanas de exposição sendo, portanto, inapropriados para aplicações que almejam uma rápida detecção de poluentes (KRAMER; JENNER; ZWART, 1989).

As respostas comportamentais dos bivalves são características adaptáveis e mensuráveis que contribuem para sua aptidão evolutiva e, portanto, podem servir como meios sensíveis, quantificáveis e ecologicamente relevantes para a avaliação do estresse ambiental (MARKICH, 1995). Parâmetros comportamentais refletem a toxicidade comportamental podendo indicar toxicidade de forma mais rápida e sensível em comparação aos métodos baseados em mortalidade. Da mesma forma, eles podem contribuir para uma maior sensibilidade e relevância ecológica aos testes de toxicidade padrão (FOURNIER et al., 2004). Reações fisiológicas e comportamentais são normalmente rápidas e potencialmente adequadas a um monitoramento biológico contínuo (KRAMER; JENNER; ZWART, 1989).

De acordo com Newton e Cope (2007), as alterações no comportamento de abertura e fechamento de conchas de moluscos bivalves representam uma resposta biológica à presença de contaminantes sendo uma forma relativamente fácil e barata de monitorar ambientes aquáticos. Entretanto, sua análise requer o emprego de técnicas de valvometria que permitem inferir a amplitude de abertura de suas conchas.

Técnicas de valvometria tem sido estudadas em ecotoxicologia há mais de 20 anos

sendo utilizadas na construção em biossensores de poluição aquática (TRAN et al., 2003). Dentre estas técnicas pode-se citar a utilização de quimógrafo, células de carga, pares de indutores, transdutores de deslocamento linear, eletromiografia e sensores de efeito Hall em conjunto com ímãs.

Métodos convencionais baseados em quimógrafo e células de carga não só requerem uma variedade de abordagens para conectar o sensor à concha (KUWATANI, 1963; HIGGINS, 1980; NAGAI et al., 2006; FUJII, 1979, 1977), como também mostram-se pesados e complexos dificultando a medição de movimentos em um ambiente próximo ao natural (NAGAI et al., 2006). Pares de eletrodos são fixados em valvas oposta de forma a averiguar a amplitude de abertura de acordo com a impedância entre estes terminais (TRAN et al., 2010; HABERKORN et al., 2011). A utilização de transdutores de deslocamento linear requer uma série de abordagens para inferir as amplitudes de abertura das conchas de moluscos bivalves bem como para realizar a calibração do sistema.

A exemplo disto, Markich (1995) fixa uma das extremidades de um fio de nylon na concha do bivalve enquanto a outra é anexada a um braço de alavanca de um transdutor de deslocamento linear. O sistema é montado de forma que o braço se encontre totalmente na horizontal quando a concha se encontra totalmente fechada mantendo, assim, o fio de nylon perfeitamente perpendicular nessa posição. O movimento da concha altera a tração exercida sobre o fio e permite o movimento do transdutor gerando, dessa forma, uma variação no sinal de tensão disponibilizada pelo dispositivo.

Esta técnica de valvometria foi empregada para verificar a resposta comportamental de bivalve *Velesunio angasi* a diferentes concentrações de urânio. A amplitude e duração média de abertura das conchas bem como a frequência de contrações musculares foram utilizadas como parâmetros de análise comportamental. Por fim, um modelo logístico foi empregado para determinar a relação sigmoideal entre o comportamento e a concentração de urânio apresentando resultados satisfatórios na identificação deste composto.

A utilização de pares de indutores como técnica de valvometria requer a fixação de duas bobinas em conchas opostas sendo uma alimentada através de um oscilador de alta frequência de forma a induzir uma tensão na bobina oposta de acordo com a amplitude de abertura da concha dos moluscos bivalves. A exemplo disso, pode-se citar os trabalhos de Kramer, Jenner e Zwart (1989), Liao et al. (2009) e Sow et al. (2011).

Moullac et al. (2020) revisou técnicas de exploração funcional não invasiva para bivalves com base no estado da arte e apresentou futuras ferramentas potenciais. Há uma ampla variedade de métodos instrumentais para monitorar bivalves fisiologicamente mantendo-os vivos e intactos. A gama atual de técnicas não invasivas para bivalves consiste em sistemas de medição para fluxos metabólicos e atividade valvar (MOULLAC et al., 2020). O autor apresentou as principais ferramentas empregadas na medição de fluxos metabólicos e propôs um sistema para registrar as características de rotação de pérolas situadas no interior de ostras *Pinctada margaritifera*.

Do ponto de vista de análise comportamental, Moullac et al. (2020) destacou a medição da atividade valvar (abertura e fechamento de conchas) como ferramenta potencialmente importante para o monitoramento biológico da qualidade da água e para a compreensão das estratégias de otimização metabólica. As técnicas de valvometria *strain gauges* e sistemas similares, sistemas de alta frequência baseado em pares de indutores e sensores de efeito Hall em conjunto com ímãs foram citadas pelo presente trabalho. Além destas, o trabalho cita sensores magnetorresistivos formados por quatro palhetas encapsuladas ativadas por um ímã com suportes fixados à cada concha.

Se o sinal monitorado é parte de uma série temporal, análises espectrais podem ser utilizadas para analisar os ritmos comportamentais. No estudo apresentado por Moullac et al. (2020) as atividades valvares apresentaram periodicidades de 23,25 h e 2,15 h seguidas por uma série de micromovimentos de baixa amplitude que podem estar associados às atividades de filtração, ciliar das guelras e de rotação de pérolas. As técnicas de valvometria apresentam como saída sinais elétricos (mV) que podem ser utilizados de forma binária com indicativos dos estados aberto/fechado ou transformados em amplitudes de abertura (mm) (MOULLAC et al., 2020). Além disso, algumas aplicações transformam este sinal em ângulo de abertura das conchas com base no tamanho conhecido de bivalves.

O autor, por fim, cita o monitoramento de atividade cardíaca como ferramenta complementar para inferir o *stress* animal por meio de princípios de impedância (método invasivo com perfuração de conchas e inserção de eletrodos) e infravermelho. Além disso, sistemas de imagem por ressonância magnética podem possibilitar uma série de informações atualmente adquiridas através da dissecação de animais bem como monitorar o processo de formação de pérolas no interior de ostras.

Kramer, Jenner e Zwart (1989) investigou o comportamento dos moluscos bivalves *Dreissena polymorpha* e *Mytilus edulis* na detecção de cloreto e sulfato cúprico. Como resultado, o autor destacou uma forte diminuição na frequência e na duração de abertura das conchas dos animais com o aumento da concentração desses poluentes. Esta relação foi observada tanto para a contaminação contínua com cloreto quanto para picos de concentração de sulfato cúprico.

Liao et al. (2009) monitorou a atividade dos moluscos de água doce *Corbicula fluminea* em resposta ao arsênio. Neste estudo, três respostas comportamentais foram consideradas apropriadas na representação de alterações comportamentais em virtude deste contaminante sendo elas o percentual de animais com conchas fechadas e abertas, a amplitude média de abertura e a duração média diária de abertura das conchas. O autor observou ainda a ocorrência de maiores amplitudes de abertura no período da manhã. Por fim, baixas concentrações de arsênio causaram alterações significativas na amplitude de abertura das conchas indicando que este parâmetro comportamental é apropriado ao monitoramento de ambientes aquáticos.

Sow et al. (2011) utilizou uma abordagem não paramétrica de EDK para resumir os

dados comportamentais complexos em um perfil de densidade simples para cada animal e período de 24 horas. Os dados comportamentais de ostras *Crassostrea gigas* permanentemente submersas em profundidades de 3 a 7 m foram coletados a cada 1,6 segundos durante 6 meses enviados à uma central acessível via conexão intranet ou internet. Os perfis resultantes foram utilizados para averiguar a relação entre o comportamento médio e o ritmo biológico das marés. O trabalho concluiu que o fechamento das conchas relacionou-se fortemente com a amplitude das marés e ocorreu de forma sincronizada com os períodos de maré baixa.

Porter e Breitburg (2016) monitorou o comportamento de ostras durante as fases normóxica, hipóxica e supersaturada do ciclo de oxigênio dissolvido (DO - *Dissolved Oxygen*) ao longo de vários períodos de 2 dias. Os animais foram equipados com *strain gauges* cujos sinais de saída foram adquiridos a cada 2,5 segundos com posterior extração de média a cada 15 segundos. Análises de variâncias (ANOVA) foram empregadas para verificar o efeito dos tratamentos de DO no percentual de tempo que as ostras mantiveram-se fechadas. A hipóxia severa induziu o fechamento de conchas por tempos significativamente mais longos em relação às condições normóxicas e ao fechamento de mais de 1/3 dos animais simultaneamente após o início desta condição. Animais expostos à hipóxia leve levaram um tempo significativamente mais longo para fecharem suas conchas em relação à condição de hipoxia severa e mantiveram suas valvas fechadas por um período de tempo semelhante à normóxia.

Tran et al. (2003) analisou o comportamento do molusco *Corbicula fluminea* em diversos experimentos. Por meio de um índice visual pessoalmente verificado e baseado no estado de abertura das conchas e posição do sifão, este trabalho demonstrou o *stress* provocado pela fixação vertical dos animais mesmo após 10 dias de experimento e que sua instrumentação por meio de pares de eletrodos não afeta significativamente seu comportamento. Este trabalho verificou ainda que entre 10h e 16h houve uma menor probabilidade de fechamentos espontâneos (microfechamentos) otimizando a sensibilidade do sistema de detecção de poluentes. Neste trabalho o limiar de sensibilidade para cádmio foi verificado através de três diferentes concentrações. A concentração efetiva média (CE50) foi considerada como aquela capaz de provocar o fechamento das conchas de 50% dos animais. ANOVA foi empregada para determinar diferenças significativas entre grupos de animais expostos não expostos ao contaminante. Concentrações de cádmio acima de 50 mg/L, entre 50 mg/L e 16 mg/L e abaixo deste valor foram detectadas em até 1 hora, em até 5 horas e não foram detectadas, respectivamente.

Fournier et al. (2004) analisou o fechamento das conchas dos bivalves de água doce *Corbicula fluminea* expostos à urânio em dois ambientes com valores distintos de pH durante um período de 5 horas. Os limiares de sensibilidade a este composto foram estabelecidos com base no EC50 e dependeram da concentração total de urânio, do tempo de

integração e do pH ambiente. Os bivalves mostraram-se mais sensíveis à urânio quando mantidos em um pH 5,5 em comparação com a utilização de um pH 6,5.

Tran et al. (2010) analisou o comportamento de ostras *Crassostrea gigas* durante uma exposição laboratorial simulada de floração de algas nocivas *Alexandrium minutum* (≈ 3500 células/*mL*) ao longo de 7 dias. O comportamento de ostras expostas a algas tóxicas e não tóxicas (*T-Isochrysis* ou *Heterocapsa triquetra*) foi significativamente diferente tendo em vista todos parâmetros comportamentais avaliados por meio do método ANOVA e teste não paramétrico de Kruskal–Wallis. Surpreendentemente, a duração diária de abertura das conchas bem como a quantidade de microfechamentos aumentou enquanto que a amplitude de abertura das conchas diminuiu na presença de algas nocivas (de predominantemente 50-80% para 30-40%). A presença de *A. minutum* foi detectada dentro de 1 hora de exposição e os animais recuperaram os padrões comportamentais em até 5 dias após exposição.

A acidificação dos oceanos está associada à uma redução de pH na superfície oceânica e à alterações no sistema carbônico da água do mar devido à absorção do excesso de CO_2 atmosférico (CLEMENTS et al., 2020). Dessa forma, Clements et al. (2020) investigou se a exposição de curto prazo (3 semanas) a pCO_2 elevada poderia impactar as respostas comportamentais à ameaça de predação em mexilhões *Mytilus galloprovincialis* adultos. O monitoramento comportamental foi realizado por meio de técnica de valvometria não invasiva de alta frequência. Sinais de predadores resultaram em uma redução prolongada (1 h) no grau de abertura da válvula (-20%), mas não teve efeito claro na atividade geral da válvula enquanto o nível elevado de pCO_2 não afetou nenhum parâmetro comportamental investigado. Os resultados corroboraram com a crescente de evidência de que os efeitos dos níveis de pCO_2 no comportamento de animais marinhos são provavelmente fracos.

Lassoued et al. (2019) avaliou em laboratório respostas fisiológicas de mexilhões juvenis (*Mytilus galloprovincialis*) expostos à diferentes condições de pCO_2 sob dois regimes de alimentação diferentes (ótimo e abaixo do ideal). As propriedades da casca, como resistência à compressão e composição foram afetadas negativamente pelo alto pCO_2 na água do mar, independentemente da disponibilidade de alimentos. Sob o regime alimentar ideal, as taxas de alimentação dos mexilhões aumentaram em resposta ao pCO_2 elevado, possivelmente como uma estratégia para manter uma forte fixação. Por outro lado, mexilhões com dieta abaixo do ideal demonstraram fixação fraca e abertura de válvula estreita sob alto pCO_2 . Em conclusão, sob disponibilidade alimentar ideal, os mexilhões foram resilientes à acidificação da água ao considerar métricas como atividade alimentar, abertura da válvula e força de fixação. Sob uma dieta abaixo do ideal, no entanto, a capacidade dos mexilhões de responder à acidificação foi comprometida. Esses resultados destacam interações ecofisiológicas complexas para organismos calcificadores sujeitos às mudanças climáticas.

A resposta a *A. minutum* pareceu ser diferente das respostas comportamentais apre-

sentadas em relação outros contaminantes. A exemplo disso, pode-se citar traços de metais frente aos quais bivalves rapidamente fecham suas conchas (KRAMER; JENNER; ZWART, 1989; TRAN et al., 2003). Espécies de bivalves consideradas altamente sensíveis como o molusco *Mercenaria mercenaria*, exibiram mecanismos fisiológicos e comportamentais para evitar a toxicidade reduzindo a amplitude de abertura de suas conchas e abrindo e fechando rapidamente suas valvas para limpar as guelras de algas tóxicas (TRAN et al., 2010). A resposta apresentada também difere do comportamento associado à falta de oxigênio. Diferentes níveis de hipóxia são caracterizados pelo aumento da duração da abertura da válvula (TRAN et al., 2005) que está associada ao aumento da ventilação para manter constantes os níveis de consumo de O₂ e oxigenação do sangue (TRAN; BOUDOU; MASSABUAU, 2000).

Os sensores de efeito Hall, diferente das técnicas de valvometria baseadas em indutores, são mais leves e não requerem a utilização de alta voltagem nem a entrega de energia à ambas valvas do animal, simplificando a fixação e melhorando consideravelmente a estabilidade a longo prazo (WILSON; REUTER; WAHL, 2005). Em adição a isso, pode-se afirmar que a fixação desses sensores é mais simples uma vez que ela não é tão crítica à qualidade de medição se comparada ao uso de células de carga, por exemplo (WILSON; REUTER; WAHL, 2005).

De acordo com Nagai et al. (2006), dentre as técnicas de valvometria, sensores de efeito Hall em conjunto com ímãs apresentam uma série de vantagens. Dentre estas pode-se citar durabilidade, leveza, facilidade de fixação e o oferecimento de um menor *stress* ao animal facilitando, dessa forma, a medição de seus movimentos. Dentre os trabalhos que investigam o comportamento de moluscos bivalves por meio de sensores de efeito Hall em conjunto com ímãs pode-se citar Wilson, Reuter e Wahl (2005), Nagai et al. (2006), Basti et al. (2009), Hartmann et al. (2016), Comeau et al. (2018) e Comeau et al. (2019).

Wilson, Reuter e Wahl (2005) demonstrou o emprego destes sensores em conjunto com ímãs para quantificar o comportamento de bivalves *Mytilus edulis* e *Astarte borealis* tanto em laboratório quanto em ambiente natural. Uma pequena haste de calibração foi disposta de forma adjacente ao ponto de abertura máxima da concha de cada bivalve. Imagens dos animais foram adquiridas e as saídas dos sensores Hall para cada bivalve foram regredidas em função do ângulo observado de abertura a fim de obter uma curva de calibração. As saídas dos sensores não foram linearmente relacionadas com a distância entre o ímã e o sensor.

Durante os experimentos (WILSON; REUTER; WAHL, 2005), condições ambientais de profundidade, intensidade da luz e temperatura foram adquiridas. O comportamento de mexilhões expostos *in situ* foi quantificado como a frequência de fechamento/abertura e ângulos máximos médios de abertura. Diferenças nos níveis de atividade foram comparadas a cada hora entre os indivíduos através de ANOVA de dois fatores. Animais colocados de volta na natureza após longos períodos no laboratório mostraram um ritmo diurno apa-

rente no ângulo de abertura com ângulos de abertura máximos durante os períodos com menor intensidade de luz.

Nagai et al. (2006) desenvolveu um sistema biológico de aviso prévio para detectar floração das algas nocivas *H. circularisquama* baseado na análise comportamental de ostras *P. fucata* em laboratório. Em água filtrada, os animais demonstraram fechamentos por curtos períodos de tempo (microfechamentos) esporádicos. A frequência de microfechamentos aumentou à medida que o número de células de *H. circularisquama* no ambiente aquático aumentou. Esta maior frequência de microfechamentos diminuiu quase imediatamente quando os animais novamente foram mantidos em ambiente sem algas nocivas.

Jr, Brasil e Salomão (2007) investigou respostas fisiológicas de mexilhões *Perna perna* com base em taxas de respiração, depuração, excreção e eficiência de absorção em diferentes temperaturas e salinidades. Os resultados mostraram variações nas taxas fisiológicas e na capacidade de aclimação que, em conjunto, permitiram compreender o seu comportamento no ambiente, bem como estimar a seu crescimento. A depuração e as taxas de respiração aumentaram duas vezes à medida que o mexilhão foi submetido a choques térmicos. A aclimação à salinidade foi claramente a capacidade mais bem desenvolvida enquanto que a eficiência de crescimento líquido reduziu com o aumento da temperatura. Embora respostas fisiológicas tenham sido investigadas, este trabalho mostrou que os níveis de temperatura e salinidade também influenciam na resposta ao meio com possível impacto comportamental.

Basti et al. (2009) averiguou os efeitos do dinoflagelado tóxico *H. circularisquama* sobre os movimentos valvares de *Ruditapes philippinarum* adultos mantidos sob três diferentes regimes alimentares por mais de 96 horas. Para cada regime alimentar os parâmetros de análise comportamental duração de abertura, amplitude da abertura e frequência de aduções foram analisados por meio do método ANOVA repetido ou teste de pares combinados de Wilcoxon. Em caso de diferenças significativas, o método da diferença mínima significativa (LSD) de Fisher foi empregado. O monitoramento em tempo real dos mariscos demonstrou que o comportamento de *R. philippinarum* é altamente sensível ao tóxico *H. circularisquama* detectando eficientemente sua presença em concentrações extremamente baixas na mistura com alimentos e de uma maneira dependente do tempo e da concentração.

Os resultados de (BASTI et al., 2009) demonstraram, portanto, que esta análise comportamental poderia ser empregada como um sistema biológico de aviso prévio para monitorar surtos dessa alga tóxica ao longo das áreas costeiras marinhas. A adição de algas nocivas induziu uma rápida e incompleta reação de fechamento das conchas e uma forte correlação entre o tempo destes fechamentos e a concentração de *H. circularisquama* foi encontrada. A reabertura das conchas foi caracterizada por um aumento significativo na frequência de aduções a partir de 5 células/*ml* e uma diminuição significativa na amplitude média de abertura a 500 células/*ml*. Após a exposição, os moluscos transferidos para

água do mar limpa retomaram a atividade normal, mas mostraram um aumento significativo na duração e amplitudes de abertura que podem estar associados ao comportamento de recuperação.

Haberkorn et al. (2011) explorou as relações entre a concentração de *Alexandrium minutum* no ambiente aquático, o acúmulo de Toxinas Paralíticas (PSPs) nas glândulas digestivas e o comportamento de ostras *Crassostrea gigas* por meio do método ANOVA seguido de um procedimento de diferença mínima significativa de Fisher (LSD). Uma regressão estatística foi usada para correlacionar o conteúdo de toxinas e variáveis comportamentais. As ostras, aclimatadas por 10 dias, foram expostas a quatro misturas de microalgas *A. minutum* em concentrações de 10×10^{-3} células/ml e 5×10^{-3} células/ml.

Durante os primeiros 2 dias de exposição à alga nociva, as ostras exibiram significativamente mais microfechamentos. Os moluscos que exibiram maiores aumentos na duração de abertura também exibiram um maior $[PSP]_{dg}$ (HABERKORN et al., 2011). Uma relação dose-resposta foi observada entre a frequência de microfechamentos e a concentração de *A. minutum* no ambiente sem, entretanto, relação com a concentração de PSP na glândula digestiva. A duração de abertura mostrou uma relação dose-resposta com a concentração de PSP na glândula digestiva, mas não com a concentração de *A. minutum* no ambiente aquático.

Da mesma forma, Comeau et al. (2019) investigou a resposta comportamental de *M. galloprovincialis* a diferentes concentrações do dinoflagelado nocivo *A. minutum* em condições laboratoriais controladas com o intuito de verificar seu emprego em sistemas de aviso prévio. Os mexilhões foram instrumentados através de sensores de efeito Hall e ímãs, fixados em estruturas de vidro para evitar agrupamento e aclimatados por um período de quatro dias antes da exposição toxicológica. Um grupo de animais não foi exposto à alga nociva (*H. triquetra*) enquanto a concentração de *A. minutum* nos demais animais foi incrementada a 3 horas.

Após o experimento, cada mexilhão foi empregado para calibrar o sistema de aquisição e adquirir sua amplitude de abertura em milímetros através da abertura das conchas em posições conhecidas repetidamente (50 vezes). Os parâmetros comportamentais percentual de tempo aberto, amplitude média de abertura e quantidade de microfechamentos foram analisados através de um modelo misto de análise de variância após verificar suposições de normalidade (Shapiro-Wilks), homocedasticidade (Levene) e, se necessário, transformação logarítmica dos dados (COMEAU et al., 2019).

A concentração e espécie de fitoplâncton não afetou significativamente os parâmetros comportamentais amplitude média de abertura e percentual de tempo aberto. A quantidade de microfechamentos aumentou significativamente em virtude de presença de algas nocivas mesmo em baixas concentrações sugerindo que este parâmetro comportamental é particularmente sensível à algas nocivas e associado à um comportamento de evitação.

O emprego de bivalves na detecção de concentrações relativamente baixas destas al-

gas (500 – 1200 células/L) demonstrou aumento na quantidade de microfocamentos em diferentes espécies dentre as quais pode-se citar *Mytilus galloprovincialis* (COMEAU et al., 2019), *Ruditapes philippinarum* (BASTI et al., 2009) e *Crassostrea gigas* (HABERKORN et al., 2011). Resultados similares frente a maiores concentrações foram reportados na análise comportamental de ostras *Crassostrea gigas* (TRAN et al., 2010, 2015) e *Pinctada fucata* (NAGAI et al., 2006) bem como da vieira *Pecten maximus* (COQUE-REAU et al., 2016).

Hartmann et al. (2016) investigou o efeito do sal de degelo no comportamento dos moluscos bivalves de água doce *Anodonta anatina* e demonstrou que a resposta comportamental destes mexilhões é um biomarcador adequado para o monitoramento de ambientes aquáticos. O experimento envolveu a simulação de três eventos de influxo subsequentes. Três *endpoints* subletais foram investigados em conjunto com os parâmetros comportamentais atividade de filtração, frequência de transição (número de mudanças de aberto para fechado, ou vice-versa) e comportamento de evitação. A atividade de filtração foi medida como a fração de tempo que as conchas de um mexilhão estavam abertas e consideradas filtrando. Para reduzir o ruído, uma média móvel de 10 minutos foi empregada. O método ANOVA foi empregado após verificação das suposições de normalidade (Shapiro-Wilks) e homoscedasticidade (Teste de Levene). Para determinar se a exposição toxicológica teve um efeito significativo sobre a atividade de filtração e frequência de transição estes parâmetros foram analisados através modelos lineares de efeitos mistos.

Em cenários de exposição ao sal de degelo, os padrões de comportamento de *A. anatina* foram significativamente afetados. A atividade de filtração dos mexilhões diminuiu significativamente durante os períodos de exposição muito alta e longa ao sal de degelo e aumentou durante a exposição curta. A frequência de transição dos mexilhões expostos aumentou durante os períodos de exposição alta e prolongada. O comportamento de evitação foi caracterizado por fechamentos de concha mais intensos com movimentos de concha repetidos, rápidos e irregulares e demonstrou ser potencialmente adequado para verificar o efeito do contaminante no comportamento de bivalves. Por fim, o tempo em que as conchas encontraram-se abertas variou significativamente tanto entre indivíduos quanto no comportamento padrão de cada qual e nenhum sincronismo nos padrões de filtragem dos mexilhões foi observado.

Utilizando de sensores de efeito Hall, Comeau et al. (2018) estudou o comportamento das valvas de mexilhões *Mytilus galloprovincialis* cultivados para consumo em cordas de até doze metros de comprimento. Os bivalves instrumentados foram igualmente divididos nas alturas 2 e 7 metros e envoltos por redes de algodão com o intuito de mantê-los nas posições desejadas. Utilizando análise de periodogramas, o autor avaliou cinco parâmetros comportamentais sendo eles o percentual de tempo em que as conchas se mantêm fechadas, máxima duração de fechamento, o percentual de tempo de abertura e amplitudes média, mediana e máxima de abertura registrada. Os animais apresentaram

comportamentos assíncronos de fechamento e não foram encontradas diferenças significativas entre os comportamentos adquiridos nas diferentes profundidades. Ademais, os mexilhões tenderam a exibir amplitudes máximas de abertura durante a noite e aberturas mínimas ao longo dos dias.

Andrewartha et al. (2015) apresentou biossensores baseados em animais sentinela junto ao monitoramento de variáveis ambientais como futura principal tecnologia para gestão de fazendas de aquicultura, bem-estar animal, consciência social e produção sustentável. Este trabalho propôs um estudo de caso acerca da aquisição em tempo real de variáveis ambientais e comportamentais de ostras em estações de aquicultura. Dados acerca da amplitude de abertura, temperatura e atividade cardíaca de ostras foram coletadas em conjunto com informações acerca da profundidade e intensidade luminosa por períodos de tempo que variaram de semanas à meses. As amplitudes de abertura foram adquiridas por meio de sensores de efeito Hall e ímãs enquanto que a atividade cardíaca e temperatura foram coletados através de pletismógrafos e termistores inseridos na concha dos animais através de pequenos furos.

Antes de realizar a interpretação dos dados coletados em campo, os biossensores foram calibrados em laboratório através do estabelecimento de relações importantes entre variáveis ambientais e comportamentais/biológicas através de faixas de condições ambientais relevantes à fazendas de aquicultura e seus possíveis eventos estressores (ANDREWARTHA et al., 2015). Dessa forma, ostras foram empregadas como biossensores junto a uma série apropriada de sensores ambientais que detectam variáveis relevantes para a saúde da ostra como, por exemplo, oxigênio dissolvido, salinidade, pH e concentração de clorofila *a*. Os dados foram coletados em tempo real por meio de tecnologia 4G e acessados remotamente.

Tran et al. (2020) Investigou o comportamento de ostras *C. gigas* equipadas com eletrodos por 18 dias em ambiente natural (submergidas em uma profundidade de 4 metros através de uma bolsa presa à uma estrutura fixa) e 7 dias em ambiente escuro laboratorial. A amplitude média de abertura horária foi empregada para estabelecer o ritmo biológico dos animais por meio de análises crono-biológicas através do método estatístico de Cosinor. Este método que possibilita o ajuste de uma função senoidal aos dados experimentais através do método dos mínimos quadrados (MARTINO; NETO, 2012). Os dados comportamentais indicaram ritmos comportamentais diários (24,4 horas) e ritmos influenciados pelas marés (12,4 horas devido ao nível da água e 6,2 horas devido à corrente) com amplitudes máximas e mínimas em momentos de maré alta e baixa, respectivamente. Além disso, trabalho em questão investigou as expressões gênicas associadas ao ritmos das marés e relógio circadiano.

Jou et al. (2013) apresentou o emprego de pares de indutores como técnica de valvometria não invasiva para monitorar o comportamento de mexilhões *Corbicula fluminea*. Com o intuito de verificar as amplitudes máximas de abertura das conchas dos bival-

ves, o sinal e tensão proveniente dos indutores foi convertido em amplitude de abertura através de um processo baseado em imagem. Os animais foram mantidos em aquários com areia no fundo de forma a ser possível se moverem e enterrarem. A cada hora os animais foram visualmente inspecionados a fim de determinar seu estado de abertura (aberto ou fechado) e extensão do sifão (estendido ou não). Estes dados foram estatisticamente comparados com as amplitudes de abertura provenientes da técnica de valvometria. Os resultados demonstraram que as magnitudes de abertura das conchas de 20% e 50% podem ser adotadas como os limiares determinantes dos estados abertura e extensão do sifão, respectivamente. Além disso, não foram identificadas diferenças significativas na extensão de sifões entre animais equipados e não equipados com indutores.

Bamber e Westerlund (2016) realizou uma série de exposições laboratoriais e bioensaios comportamentais para estabelecer a sensibilidade do molusco *Arctica islandica* a vazamentos simulados de CO_2 . Os movimentos de suas conchas foram coletados através de sensores de efeito Hall e ímãs tendo em vista ambientes com reduzido pH. Os bivalves foram fixados em uma estrutura plástica auxiliar contendo sensor e ímã com os sifões voltados para cima. O sistema de aquisição foi calibrado para fornecer a amplitudes de aberturas em milímetros por meio de conchas vazias de tamanho similar e geração de uma curva de calibração. Os parâmetros comportamentais amplitude média de abertura horária e distância percorrida (diferenças entre amplitudes de abertura subsequentes) foram investigados.

Dois experimentos laboratoriais foram realizados com o intuito de expor controladamente bivalves à reduções agudas de pH ao longo do tempo e exposição contínua à um pH de 6.2 (BAMBER; WESTERLUND, 2016). A influência desta exposição em sua capacidade de se enterrar foi analisada. Os indivíduos foram, portanto, colocados em aquários com sedimento e a quantidade de animais totalmente enterrados foi analisada após 24 e 48 horas. Da mesma forma, a atividade de estender o pé após pH reduzido foi analisada medindo-se a extensão pé mantido fora da concha por meio de um feixe de luz infravermelha e fototransistor.

Os resultados apresentados por Bamber e Westerlund (2016) sugeriram que a exposição a um pH reduzido não apresenta impacto evidente na capacidade de enterrar. Além disso, extensão do pé para fora da concha não correspondeu com as atividades valvares sugerindo que estes movimentos não estão diretamente relacionados à uma tentativa de se enterrar em sedimento. A redução gradativa do pH da água mostrou um aumento inicial na taxa e na extensão dos movimentos das conchas na maioria dos indivíduos quando o pH caiu para 6,2. A exposição contínua de bivalves à este pH resultou em um claro aumento nos movimentos da válvula durante as primeiras 40 h de exposição seguido por uma redução gradual na intensidade da atividade. Esta redução ao longo do tempo sugere aclimação dos animais ao ambiente (BAMBER; WESTERLUND, 2016).

Massabuau et al. (2015) monitorou o comportamento de vieiras *Chlamys islandica*

e mexilhões azuis (*Mytilus edulis*) situados à uma profundidade de 15 à 18 metros no Oceano Ártico por meio do emprego de eletrodos como técnica de valvometria não invasiva e comunicação via GPRS e internet. Os eletrodos foram conectados à um sistema de aquisição de dados que conecta-se, através de um cabo umbilical, à um sistema responsável pela transmissão dos dados comportamentais. Aquisição diária de dados durante um ano demonstrou a capacidade da tecnologia de reconstruir automaticamente e facilmente ritmos biológicos compreensíveis, taxas de crescimento e atividades de desova dos bivalves. Os autores destacaram o emprego de biossensores como potencial ferramenta de monitoramento inteligente de ambientes aquáticos permitindo avaliação sensível e contínua da nocividade dos impactos industriais, em particular para a indústria de óleo e gás. As gravações comportamentais foram disponibilizadas através do projeto MolluSCAN eye (MOLLUSCAN, 2020) onde era possível acessar o comportamento de ostras e outros animais ao longo do último dia.

Os resultados indicaram que em ambiente propício, tanto vieiras *Chlamys islandica* quanto mexilhões *Mytilus edulis*, apresentam um comportamento com maior predominância de valvas abertas e poucos fechamentos parciais transientes (microfechamentos). Quando estressados, ambas espécies apresentaram uma maior quantidade de microfechamentos. Dessa forma, Massabuau et al. (2015) apresentou um sistema remoto de biomonitoramento online capaz de fornecer informações apropriadas para detectar rapidamente substâncias e condições nocivas no meio ambiente.

Com o intuito, portanto, de construir biossensores de poluição aquática baseados na análise comportamental de moluscos bivalves, dados acerca da amplitude de abertura das conchas desses animais necessitam ser coletados para posterior análise. De acordo com Markich (1995) a frequência mínima necessária para medir de forma precisa os padrões de movimentos sutis dos bivalves é de 2 Hz e, a fim de caracterizar adequadamente as respostas de cada indivíduo, deve-se observá-los por um período mínimo que varia de 24 à 36 horas uma vez que cada animal apresenta padrões de movimento específicos, rítmicos e reprodutíveis.

A Tabela 2 relaciona unidades de medida, frequências de aquisição e períodos de aclimação e monitoramento com o intuito de melhor comparar os diferentes trabalhos relacionados. Da mesma forma, as Tabelas 3 e 4 relaciona as técnicas de valvometria e algoritmos empregados na análise comportamental. Por fim, as Tabelas 5 e 6 demonstram os parâmetros comportamentais investigados pelos diferentes trabalhos relacionados frente à diferentes variáveis de interesse.

Tabela 2: Unidade de análise, período de aclimação, período de monitoramento (PM) e frequência de aquisição empregados em trabalhos relacionados.

Trabalho	Unidade	Período de Aclimação	PM	Frequência [Hz]
Kramer, Jenner e Zwart (1989)	%	3 d	24 h, 20 h	0,2
Markich (1995)	-	10 d	96 h	0,2
Tran et al. (2003)	%	7 d	29 h	66,7 m
Fournier et al. (2004)	-	7d	29 h	277,8 μ
Wilson, Reuter e Wahl (2005)	°	48 h	4 d	5
Nagai et al. (2006)	mm	72 h	4 h	50
Basti et al. (2009)	mm	7-10 d	4 d	0,2
Liao et al. (2009)	%	3 sem	14 d	3,33m
Tran et al. (2010)	%	10 d	7 d	10
Haberkorn et al. (2011)	-	10 d	5 d	3,33
Sow et al. (2011)	mm	-	5 me	0,625
Jou et al. (2013)	%	49 d	14 d	3,33 m
Tran et al. (2015)	-	10 d	48 h	0,2
Massabuau et al. (2015)	mm		1 ano	10
Bamber e Westerlund (2016)	mm	16 d	7 h	0,2
Hartmann et al. (2016)	mV	3.5 d	48 d	16,7m
Porter e Breitbart (2016)	Binário	-	12 d	0,4
Comeau et al. (2018)	%	2 d	10 d	16,7m
Lassoued et al. (2019)	%	10 d	3 sem	16,7m
Comeau et al. (2019)	mm	4 d	9 d	10
Tran et al. (2020)	-	3,5 d	25 d	-
Clements et al. (2020)	%	24 h	2 d	66,67 m

Visando facilitar as análises estatísticas, os dados referentes a cada animal são muitas vezes normalizados a fim de obter valores de amplitudes de abertura de 0 à 100% (COMEAU et al., 2018) e diversos parâmetros comportamentais são analisados por período de tempo. A exemplo disso, pode-se citar a duração de abertura das conchas (TRAN et al., 2010), a quantidade de microfechamentos (TRAN et al., 2010), a amplitude média de abertura (HARTMANN et al., 2016; COMEAU et al., 2018), a frequência de transição (BASTI et al., 2009) e o percentual de tempo aberto (COMEAU et al., 2018).

Além disso, a análise comportamental de bivalves tem sido empregada em BEWS para detectar condições ambientais perigosas. A exemplo disso, pode-se citar as soluções comerciais *Musselmonitor* (AQUADECT, 2020) e *MolluSCAN*. *Musselmonitor* realiza o monitoramento comportamental de bivalves através de um BEWS contendo sensores de

efeito Hall e ímãs. O período de aquisição dos dados comportamentais pode ser ajustado entre 10s e 600s e os dados comportamentais de cada bivalve são normalizados entre 0 e 100% a fim de obter a amplitude de abertura percentual.

Tabela 3: Técnicas de valvometria e algoritmos empregados em trabalhos relacionados.

Trabalho	Técnica Valvometria	Algoritmos
Kramer, Jenner e Zwart (1989)	Indutores	MANOVA
Markich (1995)	Transdutor de deslocamento linear	Regressão logística
Tran et al. (2003)	Eletrodos	ANOVA, LSD, Kruskall -Wallis e transformação Box-cox
Fournier et al. (2004)	Eletrodos	Regressão logística
Wilson, Reuter e Wahl (2005)	Sensor Hall e ímã	MANOVA
Nagai et al. (2006)	Sensor Hall e ímã	Regressão logística
Basti et al. (2009)	Sensor Hall e ímã	ANOVA, Levene, Wilcaxon, LSD e Kolmogorov-Smirnov
Liao et al. (2009)	Indutores	Modelos Hill e Monte Carlo
Tran et al. (2010)	Eletrodos	ANOVA, Tukey Kramer, Kruskall-Wallis e Teste T
Haberkorn et al. (2011)	Eletrodos	ANOVA, teste T e regressão estatística
Sow et al. (2011)	Indutores	EDK
Jou et al. (2013)	Indutores	Regressão não linear, Kolmogorov-Smirnof, estimadores de máximo <i>likelihood</i> e método Monte Carlo

Ali Amin Ali (2019) explorou aspectos de hardware e *software* empregados no desenvolvimento de sistemas capazes de monitorar o comportamento de abertura e fechamento das conchas de moluscos bivalves. Ambos sistemas analisados empregavam sensores de efeito Hall e ímã para medir e relatar com precisão a abertura de conchas bivalves para um microcontrolador que registra e transmite os dados para uma estação terrestre. Este sistema foi projetado para operar em campo e também em ambiente de laboratório.

No campo devido à falta de disponibilidade de energia, foram estabelecidas duas configurações possíveis sendo a primeira caracterizada por uma unidade totalmente submersível com bateria (duração de aproximadamente 29 dias), microcontrolador e cartão SD Ali Amin Ali (2019). A segunda configura de campo emprega rede celular para comunicar seus dados a uma estação terrestre. Esta é uma solução viável, uma vez que a maioria dos recifes existem perto o suficiente da costa Ali Amin Ali (2019). Além disso, o sistema inclui um painel coletor solar para carregar a bateria durante o dia e manter constantemente a transmissão do sistema à noite.

Tabela 4: Técnicas de valvometria e algoritmos empregados em trabalhos relacionados.

Trabalho	Técnica Valvometria	Algoritmos
Tran et al. (2015)	Eletrodos	ANOVA e Kruskal-Wallis
Massabuau et al. (2015)	Eletrodos ¹	Regressão não paramétrica baseada em estimadores <i>kernel</i> para estabelecer ritmos diários, mensais e anuais <i>likelihood</i>
Bamber e Westerlund (2016)	Sensor Hall e ímã	-
Hartmann et al. (2016)	Sensor Hall e ímã	MANOVA, ANOVA, Levene, Tukey Kramer e Shapiro-Wilks
Porter e Breitburg (2016)	<i>strain gauges</i>	ANOVA
Comeau et al. (2018)	Sensor Hall e ímã	Teste Mann-Whitney
Lassoued et al. (2019)	Sensor Hall e ímã	ANOVA, MANOVA, Shapiro-Wilk, W-test, Tukey-Kramer e Levene
Comeau et al. (2019)	Sensor Hall e ímã	ANOVA, Levene, Shapiro-Wilks e transformação logarítmica
Tran et al. (2020)	Eletrodos	Método Cosinor
Clements et al. (2020)	Sensor Hall e ímã	ANOVA e modelo linear de efeitos mistos

Humos et al. (2019) propõe uma plataforma capaz de transmitir dados de diferentes sensores relacionados ao ambiente subaquático para a nuvem em tempo real através de uma antena por um período superior a um mês com visibilidade mínima a partir da superfície para evitar casos de vandalismo. Um sistema de medição de atividade valvar (ALI AMIN ALI, 2019) para fornecer dados comportamentais em conjunto com variáveis ambientais é empregado de forma a monitorar a saúde ambiental em tempo real. Humos et al. (2019) propõe ainda a geração de gráficos instantaneamente.

Dessa forma, as tendências diárias, semanais e mensais podem ser facilmente identificadas. Isso poderá permitir análises comportamentais frente à quaisquer mudanças ambientais ao longo do tempo por meio da integração de vários tipos de informações e, em seguida, derivar gráficos para descrever os vários aspectos da etologia, cronobiologia ou fisiologia de ostras. Por fim, Humos et al. (2019) propõe futuramente empregar a plataforma proposta junto à ferramentas de inteligência artificial para emitir alertas ambientais de acordo com os dados adquiridos sugerindo, portanto, que ferramentas computacionais de inteligência artificial podem fornecer uma medida de estado ambiental altamente sensível com potencial empregabilidade na mitigação de condições adversas (HUMOS et al., 2019).

Tabela 5: Parâmetros comportamentais e variáveis de influência em trabalhos relacionados. Dentre os parâmetros comportamentais encontram-se a amplitude média de abertura (AA), período de tempo aberto (PA), período de tempo fechado (PF), frequência de transição (FT) e *avoidance behavior*.

Trabalho	Comportamento frente à:	Parâmetros Comportamentais
Kramer, Jenner e Zwart (1989)	Cobre, hipoclorito e óxido tributilestanho	AA, PA, PF e FT
Markich (1995)	Urânio	PA, AA e FT por hora e período aberto
Tran et al. (2003)	Cádmio	PA, FT e microfechamentos
Fournier et al. (2004)	Urânio em diferentes níveis de pH	Período até fechamento e percentual de animais fechados
Wilson, Reuter e Wahl (2005)	Períodos do dia	FT e ângulo máximo de abertura
Nagai et al. (2006)	Alga nociva <i>H. circularisquama</i>	Microfechamentos
Basti et al. (2009)	Regimes alimentares, alga nociva <i>H. circularisquama</i> e após exposição	AA, PA e microfechamentos
Liao et al. (2009)	Arsênio e períodos do dia	Percentual de mexilhões abertos, AA, PA e PF diários
Tran et al. (2010)	Alga nociva <i>A. minutum</i>	PA, AA e microfechamentos diários
Haberkorn et al. (2011)	Alga nociva <i>A. minutum</i>	Microfechamentos e PA diários
Sow et al. (2011)	Maré e períodos do dia	AA
Jou et al. (2013)	Presença e ausência de indutores	AA e inspeção visual do estado de abertura e extensão do sifão (estendido ou não)

Em um ambiente não contaminado, as conchas dos animais bivalves encontram-se abertas em aproximadamente 75% do tempo e os fechamentos são esporádicos com reabertura iminente (AQUADECT, 2020). De acordo com Aquadect (2020), várias são

as alterações comportamentais indicadoras de contaminação. O animais podem manter as conchas fechadas por um período mais longo de tempo, aumentar a frequência de transição, isto é, de abertura e fechamento e reduzir a amplitude média de abertura.

Tabela 6: Parâmetros comportamentais e variáveis de influência em trabalhos relacionados. Dentre os parâmetros comportamentais encontram-se a amplitude média de abertura (AA), período de tempo aberto (PA), período de tempo fechado (PF), frequência de transição (FT) e *avoidance behavior*.

Trabalho	Comportamento frente à:	Parâmetros Comportamentais
Tran et al. (2015)	Alga nociva <i>A. minutum</i>	PA e microfechamentos
Massabuau et al. (2015)	Ambiente natural (oceano ártico)	AA e microfechamentos
Bamber e Westerlund (2016)	Redução de pH por bombeamento de CO_2	AA, distância de movimento, tamanho do pé e enterramento
Hartmann et al. (2016)	Sal de degelo	Atividade de filtração, FT, PA e Avoidance behavior
Porter e Breitburg (2016)	Hipoxia leve, severa e normoxia (DO)	PF
Comeau et al. (2018)	Profundidade e períodos do dia	PF e máximo PF, PA e amplitudes mediana e máxima.
Lassoued et al. (2019)	Diferentes níveis de pCO_2 e regimes de alimentares	Taxas de depuração, força de fixação e AA
Comeau et al. (2019)	Profundidade e períodos do dia	PA, AA e microfechamentos
Tran et al. (2020)	Ambiente natural sujeito a variações da maré e ambiente escuro laboratorial	AA horária
Clements et al. (2020)	Resposta à predadores frente à níveis de pCO_2	Percentual de abertura em relação à amplitude máxima e amplitude relativa entre duas medições

Neste sistema, o comportamento de cada mexilhão é medido e avaliado individualmente. Há portanto, uma redução na sensibilidade do sistema, causada por variações

naturais entre organismos (AQUADECT, 2020). Esta redução é amenizada analisando-se os valores médios de medição em um grupo de animais. O mecanismo de configuração de alarme considera a sincronia comportamental entre um grupo de bivalves conforme sugerido em Liao et al. (2009).

Ademais, pode-se citar o trabalho de Grekov, Kuzmin e Mishurov (2019) no qual um sistema similar ao *Musselmonitor* é distribuído ao longo da região a ser monitorada através de boias e comunicação *bluetooth* para a construção de um sistema automático e online de biomonitoração capaz de rapidamente identificar níveis perigosos de poluição aquática. Neste sistema, as reações individuais de 16 mexilhões presos à uma estrutura física contendo sensores de efeito Hall e ímãs são registradas simultaneamente e a reação sincronizada de um grupo de mexilhões (pelo menos 70% do número total) representa um sinal de alarme. O sistema precisa ser calibrado para cada um dos animais e registra temperatura e intensidade luminosa. A utilização a longo prazo deste sistema apresentou alta eficiência na análise da qualidade da água e nível de contaminação no mar, reservatórios de água doce e rios (GREKOV; KUZMIN; MISHUROV, 2019).

O projeto de olho MolluSCAN liderado pelo CNRS e pela Universidade de Bordeaux é baseado na atividade de um grupo multidisciplinar de pesquisadores com o intuito de obter informações de grupos de bivalves deixados sem intervenção humana local por longos períodos de tempo (1-2 anos) (MASSABUAU et al., 2015). O MolluSCAN consiste em um sistema de valvometria altamente especializado que transfere dados comportamentais de bivalves em tempo real através de rede móvel (ANDRADE et al., 2016) permitindo a servidores terrestres realizar o monitoramento comportamental de bivalves e verificar a ocorrência de comportamentos anormais no ambiente como, por exemplo, a ocorrência de algas nocivas (COMEAU et al., 2019), contaminantes. A versão PRO do sistema possibilita integrar diversos tipos de informação relacionadas a aspectos da etologia, cronobiologia ou fisiologia do bivalve (MOLLUSCAN, 2020).

Em geral, uma redução na amplitude média de abertura e um aumento na frequência de transição são algumas das mudanças comportamentais associadas à exposição dos bivalves a algas nocivas (BASTI et al., 2009; NAGAI et al., 2006; TRAN et al., 2010) e outros poluentes. Dentre estes, pode-se citar urânio (MARKICH, 1995), cádmio (TRAN et al., 2003) e arsênio (LIAO et al., 2009). Em adição a isso, diversos métodos estatísticos são empregados com o intuito de detectar diferenças significativas nos comportamentos de animais expostos e não expostos a uma variedade de contaminantes. Dentre estes métodos pode-se citar ANOVA (*Analysis Of Variance*) (HARTMANN et al., 2016), teste não paramétrico Kruskal – Wallis (TRAN et al., 2010) e teste de Fisher usando o método da menor diferença significativa (HABERKORN et al., 2011).

Embora a análise dos parâmetros comportamentais gerais tenha proporcionado mudanças significativas entre grupos de bivalves expostos e não expostos a contaminantes, ela investiga uma tendência geral. A resposta individual dos animais, principalmente

sob exposição toxicológica, varia e reflete a plasticidade fenotípica para uma tentativa individual de reduzir seu contato com um ambiente degradado (TIERNEY; SHAMCHUK, 2012). Trata-se da natureza adaptativa individual que leva organismos a diferirem em sua capacidade de resposta às variações ambientais (DINGEMANSE et al., 2010).

Tem-se, portanto, uma variabilidade inerente dos padrões comportamentais entre indivíduos causando, por exemplo, fechamento assíncrono das conchas dos animais (COMEAU et al., 2018), variação significativa no tempo em que as valvas encontraram-se abertas e assincronia nos padrões de filtragem (HARTMANN et al., 2016). Nesse sentido, considerar a natureza adaptativa individual de moluscos bivalves, isto é, modelar seus padrões comportamentais peculiares e individuais na construção de biossensores de poluição aquática pode fornecer vantagens na construção de biossensores de poluição aquática baseados em seu comportamento.

Embora métodos estatísticos sejam amplamente utilizados na determinação de alterações comportamentais significativas associadas à presença de contaminantes e variáveis ambientais inóspitas (Tabelas 3, 4 5 e 6), o emprego de ferramentas de inteligência artificial pode contribuir positivamente para o desenvolvimento de biossensores inteligentes, de maior acurácia e dedicados à determinação de poluentes específicos em ambientes aquáticos com características próprias.

Analisando-se os parâmetros comportamentais mais frequentemente investigados na construção de biossensores baseados na atividade de abertura e fechamento de bivalves ao longo dos trabalhos relacionados (Tabelas 5 e 6), destacam-se o emprego da amplitude média de abertura, percentual de tempo em que os animais permanecem com suas conchas abertas ou fechadas, frequência de transição e quantidade de microfchamentos. A frequência de transição é caracterizada pela quantidade de vezes em que os animais alteram seu estado de aberto para fechado e vice-versa (HARTMANN et al., 2016) enquanto que microfchamentos são caracterizados por fechamentos parciais e instantâneos das conchas ao longo do tempo (COMEAU et al., 2019).

Nesse sentido, o presente trabalho visa investigar o uso de visão computacional como técnica de valvometria não invasiva. Especialmente, busca-se investigar o uso de ferramentas computacionais e de inteligência artificial no desenvolvimento de biossensores de poluição aquática capazes de considerar a natureza adaptativa de mexilhões *Perna perna*.

4 MODELOS COMPUTACIONAIS

O presente trabalho objetiva investigar a empregabilidade de ferramentas computacionais de visão e de inteligência artificial na construção de biossensores de poluição aquática baseados na análise comportamental de moluscos bivalves. Dessa forma, propõe-se o desenvolvimento de uma técnica de valvometria não invasiva baseada em visão computacional bem como três modelos de ferramentas de análise comportamental através da combinação de classificadores, classificador baseado em grupo e da associação de NARNETS e gráfico de controle EWMA.

De forma mais precisa, propõe-se modelos computacionais como ferramentas de análise comportamental capazes de inferir a presença ou ausência de contaminantes (*status* de contaminação) no ambiente aquático a partir de parâmetros comportamentais associados à atividade de abertura e fechamento de conchas bivalves. A Figura 20 relaciona os módulos constituintes de um biossensor com os modelos propostos pela presente dissertação.

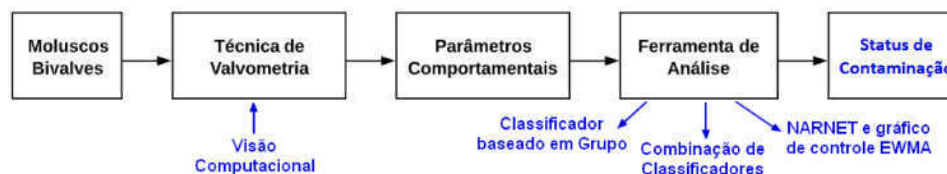


Figura 20: Contribuições propostas

As subseções 4.1 e 4.2 versam, portanto, sobre os modelos propostos para a construção dos módulos referentes à técnica de valvometria e ferramentas de análise comportamental, respectivamente.

4.1 Módulo Técnica de Valvometria

Com o intuito de investigar a empregabilidade de métodos de visão computacional como técnica de valvometria não invasiva na construção de biossensores, propõe-se um modelo computacional baseado em segmentação de cores conforme subseção a seguir.

4.1.1 Modelo com Visão Computacional

O modelo proposto para a construção de uma técnica de valvometria baseada em visão computacional fundamenta-se no emprego de um sistema embarcado responsável por adquirir e processar imagens de bivalves bem como inferir suas amplitudes de abertura ao longo do tempo. A Figura 21 ilustra o modelo proposto.

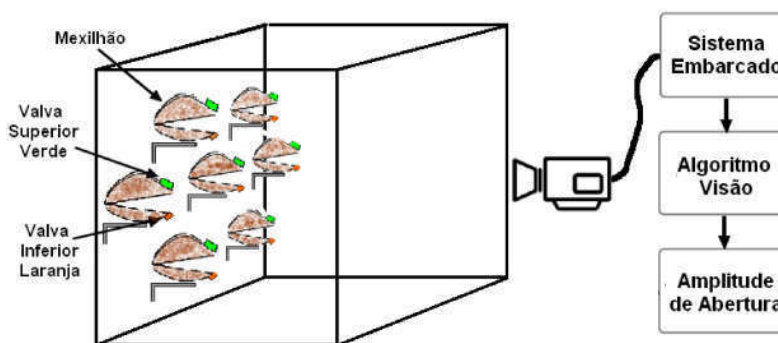


Figura 21: Modelo de uma técnica de valvometria baseada em segmentação de cores. Moluscos bivalves com suas conchas superior e inferior pintadas em cores distintas, como por exemplo laranja e verde, são fixados de forma a fornecer vista frontal à uma câmera. Um sistema embarcado é responsável por capturar as imagens, processá-las e inferir a amplitude de abertura de cada concha.

Primeiramente, propõe-se a fixação dos bivalves de forma a obter uma vista frontal através de uma câmera posicionada externamente ao ambiente aquático conforme ilustrado no diagrama da Figura 21. Tendo em vista uma abordagem baseada em segmentação de cores, propõe-se colorir através de esmalte as extremidades das conchas superior e inferior em cores distintas a serem determinadas conforme aplicação.

Um sistema embarcado é responsável por coletar imagens dos animais e processá-las a fim de inferir suas amplitudes de abertura. Com esse intuito, propõe-se que as imagens sejam pré-processadas e divididas em Regiões de Interesse (*Region of Interest - ROI*) contendo cada animal instrumentado (Figura 22). Durante a fase de pré-processamento, as imagens são distorcidas com base nos parâmetros extrínsecos e intrínsecos da câmera e emprega-se um algoritmo de melhoria de contraste.

As ROIs são analisadas para determinar a amplitude de abertura das conchas de cada bivalve com base na distância entre os pontos de cores distintas de interesse mais próximos à extremidade da concha. A Figura 23 resume o processamento de cada ROI e a estimativa da amplitude de abertura.

Visando avaliar a distância entre os pixels com as cores de interesse mais próximos da extremidade das conchas, cada ROI é processada no espaço de cores HSV (*Hue, Saturation and Value*). Para cada cor de interesse (por exemplo verde e laranja) um conjunto de valores de pixels é coletado e sua média utilizada durante o processo de segmentação considerando tolerâncias para as componentes de matiz, saturação e brilho.

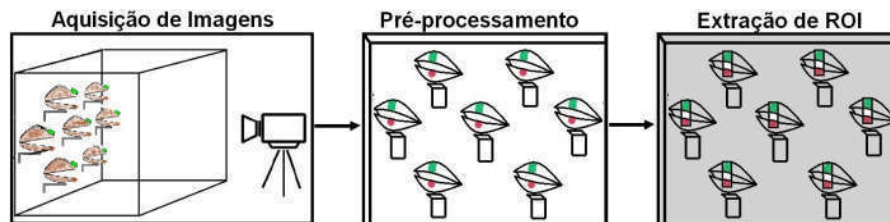


Figura 22: Diagrama em blocos ilustrando o pré-processamento e extração de ROIs. As conchas dos moluscos bivalves são pintadas em cores distintas como, por exemplo, verde e laranja para posterior análise.

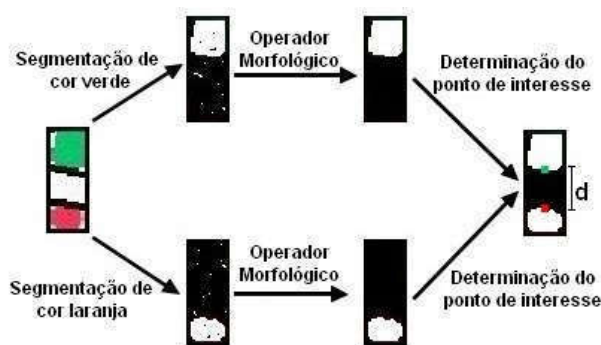


Figura 23: Diagrama de processamento de cada ROI. Ele pode ser dividido em segmentação de cores, redução de ruído através de operadores morfológicos de abertura e determinação do pixel mais próximo da extremidade da concha para cada cor de interesse.

A representação de cores no espaço HSV é selecionada pois possibilita isolar a cromaticidade em um único canal (matiz). Faixas de tolerância para as componentes matiz, saturação e brilho são importantes para selecionar adequadamente as cores de interesse e evitar a seleção de pixels demasiadamente escuros e claros, respectivamente.

As imagens binárias resultantes são filtradas através de operadores morfológicos de abertura (erosão seguida de dilatação) para eliminar ruídos e pequenas regiões detectadas erroneamente sem, entretanto, comprometer arestas de interesse. A distância em pixels entre os pontos mais próximos da extremidade da concha é extraída conforme ilustrado na Figura 23.

4.2 Módulo Ferramentas de Análise Comportamental

O módulo ferramentas de análise comportamental abrange métodos capazes de identificar diferenças no comportamento de bivalves devido à presença de compostos tóxicos no meio aquático. Dessa forma, três modelos computacionais são propostos como ferramentas de análise comportamental com o intuito de inferir o *status* de contaminação ambiente, isto é, a presença ou ausência de contaminantes.

Nesse sentido, as subseções 4.2.1, 4.2.2 e 4.2.3 abordam os modelos propostos com

base em um classificador baseado em grupo de moluscos bivalves, na combinação de classificadores e na associação de NARNETs e gráfico de controle EWMA, respectivamente.

4.2.1 Modelo com Classificador Baseado em Grupo

O modelo com classificador baseado em grupo objetiva inferir o *status* de contaminação ambiente (classificação binária acerca da presença ou ausência de contaminantes) a partir da análise comportamental de um grupo de moluscos bivalves. A análise comportamental proposta baseia-se em parâmetros comumente empregados na construção de biossensores sendo eles amplitude média de abertura, percentual de tempo aberto e frequência de transição (Tabelas 5 e 6).

O parâmetro comportamental quantidade de microfchamentos é comumente empregada na construção de biossensores tendo em vista trabalhos relacionados (tabelas 5 e 6). Este parâmetro é definido como a quantidade de fechamentos parciais instantâneos e com reabertura iminente.

Analisando-se gráficos comportamentais associados à exposição de bivalves a contaminantes em trabalhos relacionados, nota-se a presença de não só a ocorrências de microfchamentos, mas também microaberturas (quantidade de aberturas parciais instantâneas e com refechamento iminente). Dessa forma, propõe-se ainda o parâmetro de análise comportamental quantidade total de micromovimentos (soma de microfchamentos e microaberturas ao longo do tempo) na construção do modelo em questão.

O parâmetro frequência de transição corresponde ao número de vezes que o mexilhão muda seu estado aberto para fechado e vice-versa (HARTMANN et al., 2016). Por sua vez, a quantidade de micromovimentos corresponde à número total de microaberturas e microfchamentos, isto é, à quantidade de aberturas ou fechamentos instantâneos parciais ao longo do tempo.

Os parâmetros comportamentais associados à um grupo de moluscos bivalves são, portanto, utilizados como entradas de um classificador responsável por inferir a presença de contaminantes. A Figura 24 esboça o modelo computacional proposto como ferramenta de análise comportamental.

4.2.2 Modelo com Combinação de Classificadores

O modelo baseado na combinação de classificadores tem como objetivo inferir o *status* de contaminação ambiente (contaminado ou não) a partir da análise de classificações (votos) fornecidas por um conjunto de classificadores previamente treinados para detectar compostos de interesse. O principal objetivo do modelo questão é combinar as características de classificadores distintos a fim de inferir a presença de contaminantes com maior acerto.

Assim como para o modelo de classificação baseado em grupo, os parâmetros comportamentais horários amplitude média de abertura, percentual de tempo aberto, frequência

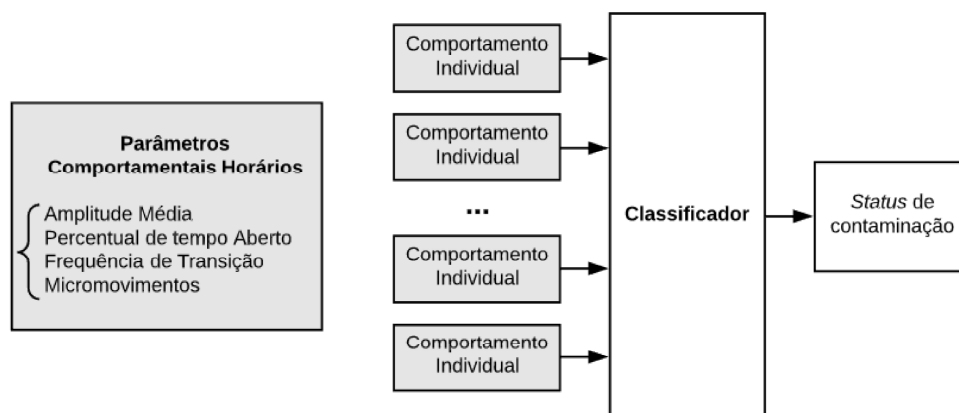


Figura 24: Modelo computacional proposto para a construção de um módulo ferramenta de análise comportamental baseado na utilização de um classificador para, a partir dos parâmetros comportamentais horários de um grupo de moluscos bivalves, inferir o *status* de contaminação ambiente (contaminado ou não).

de transição e quantidade de micromovimentos são empregados como entradas de cada classificador presente no modelo em questão.

Estes parâmetros comportamentais horários são inspecionados através da combinação de quatro classificadores para determinar se há ou não poluentes no ambiente. Com esse intuito, os votos destes classificadores são processados por meio de um classificador de votação responsável por inferir o *status* final de contaminação ambiente. A Figura 25 esboça o modelo computacional proposto para a construção deste módulo ferramenta de análise comportamental.

Quatro classificadores (C_1 à C_4) são treinados com base nos parâmetros comportamentais horários amplitude média de abertura, percentual de tempo aberto, frequência de transição e quantidade de micromovimentos de mexilhões *Perna perna*. A seguir, as classificações binárias por eles efetuadas (votos) acerca da presença de contaminantes são combinadas por meio de um classificador de votação responsável por inferir o *status* final de contaminação ambiente (classificação binária sobre a presença ou ausência de contaminantes). O treinamento deste classificador é realizado com base em votos adquiridos com base em exposições controladas de bivalves tanto em um ambiente não contaminado, quanto com a presença de contaminantes de interesse.

Estes classificadores podem ser treinados de forma especializada para detectar a presença de compostos distintos e a seguir ser combinados para inferir um *status* de contaminação geral ou representar diferentes abordagens na detecção de um único poluente e serem combinados com o intuito de gerar uma classificação binária final de melhor qualidade.

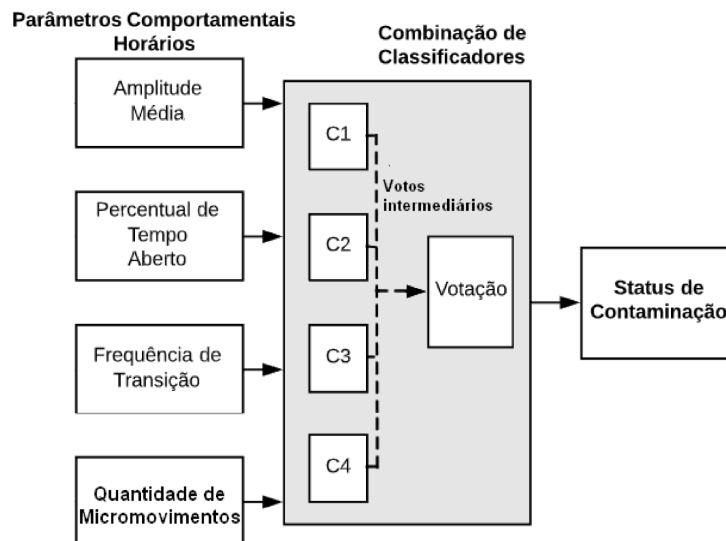


Figura 25: Modelo computacional proposto para a construção de um módulo ferramenta de análise comportamental baseado na combinação de classificadores. Quatro classificadores ($C1$ à $C4$) são empregados para inferir a presença de contaminantes (classificação binária) a partir de parâmetros comportamentais horários. A saída destes classificadores (votos intermediários) são processadas através de um classificador de votação responsável por determinar o *status* final de contaminação, isto é, se há ou não contaminantes no meio aquático.

4.2.3 Modelo com NARNET e Gráfico de Controle EWMA

Tendo em vista que há uma variabilidade inerente aos dados comportamentais entre indivíduos, considerar a natureza adaptativa de cada animal pode contribuir positivamente ao desenvolvimento de biossensores de poluição aquática. Nesse sentido, uma ferramenta de análise comportamental capaz de modelar e considerar os padrões de movimento específicos de cada bivalve na construção de biossensores é apresentada.

Com o intuito de considerar a natureza adaptativa individual, propõe-se o treinamento de uma NARNET para prever o comportamento de cada mexilhão *Perna perna* em um ambiente não contaminado. Como as NARNETs são treinadas através de dados comportamentais adquiridos em uma exposição não tóxica, supõe-se que o comportamento dos animais ao longo de uma exposição toxicológica tenderia a se desviar do sinal previsto por essas redes neurais. Nesse sentido, propõe-se inferir o *status* de contaminação ambiente (presença ou ausência de contaminantes) por meio da análise dos resíduos entre o comportamento previsto e o comportamento real através de um gráfico de controle EWMA. A Figura 26 ilustra, portanto, o modelo baseado associação de NARNETs e gráfico de controle EWMA.

Os dados comportamentais normalizados são pré-processados a fim de obter a amplitude média de abertura (*Average Opening Amplitude* - AOA) a cada intervalo de 0,1 hora conforme equação 62.

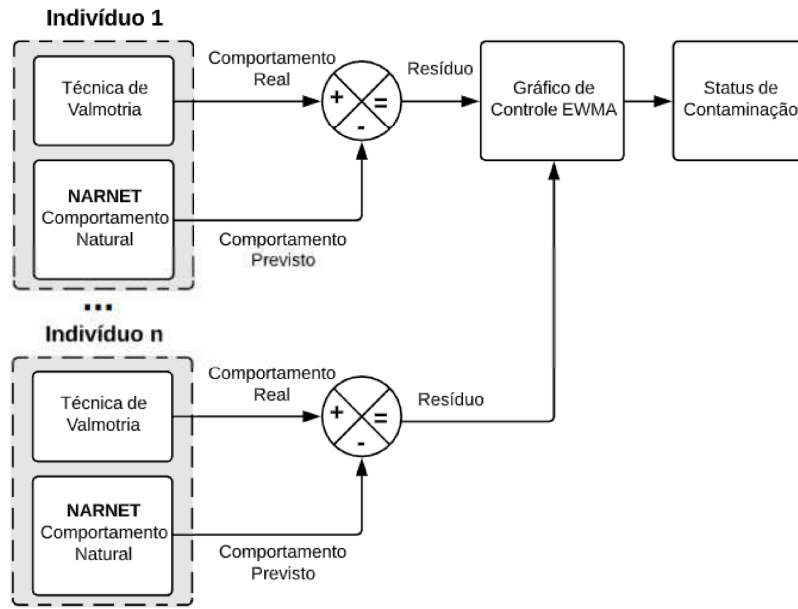


Figura 26: Modelo para a construção de um módulo ferramenta de análise comportamental baseado no treinamento de NARNETs para prever o comportamento individual mexilhões não expostos a contaminantes. Os erros entre os valores previstos e reais de diversos indivíduos são empregados em um gráfico de controle EWMA a fim de identificar o *status* de contaminação ambiente.

$$Y(i) = \frac{\sum_{j=(i-1).f.T_s+1}^{i.f.T_s} X_{norm}(j)}{T_s - 1} \quad \forall i \in [1, T/T_s] \quad (62)$$

onde X_{norm} , f e $Y(i)$ correspondem às amplitudes de abertura normalizadas de 0 à 100%, frequência de aquisição e amplitudes médias de abertura (AOA) a cada intervalo de 0,1 hora (i) ao longo do período de tempo T_s , respectivamente.

Para modelar o comportamento de cada bivalve no instante de tempo $t + 1$ ($Y(t + 1)$) com base em seu comportamento ao longo das últimas n horas (últimas $10 \times n$ observações de $Y(t)$), propõe-se o treinamento e utilização de NARNETs com $10 \times n$ atrasos de entrada. A determinação da quantidade de atrasos de entrada é realizada com base na função de autocorrelação comportamental de moluscos bivalves.

Após o treinamento das NARNETs, os resíduos entre o comportamento previsto e o comportamento real de cada animal são empregados em um gráfico de controle EWMA capaz de inferir a presença de contaminação ambiente com base em limites de controle estabelecidos por meio de resíduos associados à um ambiente não contaminado. Por fim, propõe-se o emprego de um fator de ponderação de 0,6 e 3,6 sigmas ($k = 3,6$) para estabelecer os limites de controle desta ferramenta.

5 VALIDAÇÃO

Com o intuito de validar os modelos que compõem os módulos técnica de valvometria e ferramentas de análise comportamental, dois experimentos laboratoriais foram executados. Dessa forma, as seções 5.1 e 5.2 apresentam os estes experimentos e as ferramentas empregadas na análise dos dados comportamentais resultantes, respectivamente.

5.1 *Setup* dos Experimentos

Os modelos computacionais propostos para a construção de biossensores baseados na análise comportamental de moluscos bivalves foram validados através de dois experimentos laboratoriais envolvendo mexilhões *Perna perna* (marrons). Estes moluscos foram obtidos em uma fazenda de mitilicultura no sudoeste de Santa Catarina, Brasil (-27,729769, -48,562973) e transportados imediatamente para a Universidade Federal do Rio Grande.

Antes dos experimentos, os mexilhões foram mantidos em tanques areados com água do mar (salinidade 30, temperatura 20°C e fotoperíodo 12L: 12D) por 15 dias. Além disso, a troca de água e alimentação utilizando algas *Isochrysis galbana*, *Chaetoceros muelleri*, *Nannochloropsis sp* e *Conticribra weissfloggi* foram realizadas a cada dois dias.

As subseções 5.1.1 e 5.1.2 apresentam o *Setup* dos experimentos destinados à validação dos módulos técnica de valvometria e ferramentas de análise comportamental, respectivamente.

5.1.1 Módulo Técnica de Valvometria

Com o intuito de validar o modelo baseado em visão computacional proposto como técnica de valvometria não invasiva, propõe-se a realização de um experimento que possibilite comparar as amplitudes de abertura obtidas com seu emprego e através de uma técnica de valvometria bem estabelecida. Nesse sentido, sensores de efeito Hall UGN3503 em conjunto com ímãs de neodímio ($8 \times 2mm^2$) foram utilizados para obter dados comportamentais através de sua fixação nas extremidades opostas das bordas ventrais médias das conchas de mexilhões *Perna perna*. Esses componentes foram fixados com cola cianoacrilato (COMEAU et al., 2018; BASTI et al., 2009; NAGAI et al., 2006)

e os sensores foram impermeabilizados com resina epóxi (BASTI et al., 2009; NAGAI et al., 2006).

A metodologia de aquisição de dados consistiu na leitura das tensões de saída dos sensores através de um dispositivo de prototipagem Arduino Mega cujo conversor analógico-digital que possuía 10 bits de resolução. Os dados adquiridos foram armazenados em um arquivo CSV (*Comma-Separated Values*) através de uma placa de aquisição de dados composta por um multiplexador de 16 canais e um módulo de cartão SD.

Cinco animais instrumentados foram presos em uma placa de vidro usando estruturas metálicas em forma de L e cola cianoacrilato. A quantidade de animais foi escolhida com base no tamanho dos aquários disponíveis e no espaço para realização de experimento no Instituto de Ciências Biológicas da FURG.

Como o algoritmo de visão computacional proposto baseia-se na detecção de cores, os sensores e ímãs foram pintados em cores distintas por meio de esmalte antes da instrumentação dos animais. Tendo em vista que o modelo proposto prevê o emprego de cores distintas para as valvas superior e inferior e que a validação em questão é realizada em ambiente controlado com água transparente, as cores verde e laranja aproximadamente complementares foram arbitrariamente selecionadas.

A placa de vidro foi completamente submersa em um aquário mantido a uma temperatura de 20 °C e salinidade de 30. Fitas LED (*Light Emitting Diode*) dispostas ao redor do aquário foram empregadas como um sistema de iluminação passiva. A Figura 27 esboça o *layout* do experimento.

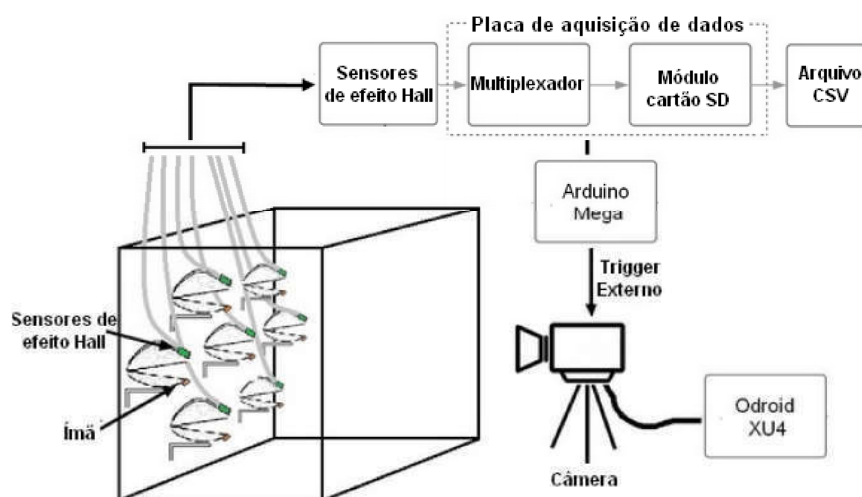


Figura 27: *Layout* do experimento com mexilhões *P. perna*. Uma câmera é posicionada de forma externa ao aquário de forma a obter uma vista frontal dos mexilhões instrumentados com ímãs e sensores pintados em verde e laranja. A câmera é conectada a um Odroid XU4 com o intuito de capturar e armazenar imagens de acordo com um *trigger* externo. Um Arduino Mega em conjunto com uma placa de aquisição de dados é responsável por armazenar as saídas dos sensores e enviar um *trigger* à câmera.

Uma câmera Basler Acer acA4600-10c foi posicionada externamente ao aquário para obter imagens dos diferentes animais para posterior análise. A câmera permitiu a aquisição de imagens coloridas, a sincronização através de gatilhos de *software* e *hardware* e forneceu uma resolução de 4608x3288 pixels. A fixação dos bivalves e a posição da câmera proporcionaram uma vista frontal dos animais. Dessa maneira, a distância entre os sensores e ímãs coloridos foi proporcional à amplitude de abertura das conchas.

O experimento teve como objetivo comparar os dados gerados pelos sensores de efeito Hall e processamento de imagem considerando 10%, 20% e 20% de tolerância para os componentes Matiz, Saturação e Brilho, respectivamente.

Com esse intuito, a aquisição das imagens e da tensão de saída dos sensores de efeito Hall foi sincronizada. O Arduino Mega possibilitou esta sincronização através do envio de um gatilho externo à câmera ao ler a tensão de saída dos sensores de efeito Hall através de uma placa de aquisição de dados. Para manter a câmera esperando estes gatilhos para capturar e armazenar imagens para posterior avaliação, ela foi conectada a um computador de placa única Odroid XU4 que executa um algoritmo desenvolvido através do *software* QT Creator, da biblioteca OpenCV (*Open Source Computer Vision Library*) e do *Software Development Kit* (SDK) da câmera. A Figura 27 ilustra o *layout* da experimento.

Como sistemas embarcados apresentam restrições de memória, as imagens e os sinais dos sensores de efeito Hall foram adquiridos a cada 1 minuto por 2 horas. Além disso, na literatura foram encontrados tempos de aquisição de 1 minuto ou mais para investigar o comportamento de bivalves como biomarcador em ecotoxicologia e como ele é influenciado por variáveis ambientais como por exemplo em Liao et al. (2009), Hartmann et al. (2016) e Comeau et al. (2018).

As imagens resultantes foram pré-processadas através de um algoritmo de melhoria de contraste e distorção das imagens com base nos parâmetros intrínsecos e extrínsecos da câmera. Estes parâmetros foram estimados através da *toolbox Camera Calibrator* do *Software* MATLAB. Essa ferramenta empregou o método Zhang de calibração que baseia-se em uma transformação projetiva entre o plano do objeto de calibração e o plano da imagem.

Após o pré-processamento, as ROIs contendo cada molusco bivalve foram manualmente determinadas e processadas através da segmentação proposta. Por fim, um operador morfológico de abertura foi empregado com o intuito de remover da imagem objetos contendo menos de 200 pixels.

5.1.2 Módulo Ferramentas de Análise Comportamental

A validação dos modelos propostos como ferramentas de análise comportamental requer dados comportamentais de moluscos bivalves imersos em ambiente controlado com e sem a influência de contaminantes. Nesse sentido, realizou-se o experimento toxicológico cujo *layout* encontra-se ilustrado através da Figura 28.

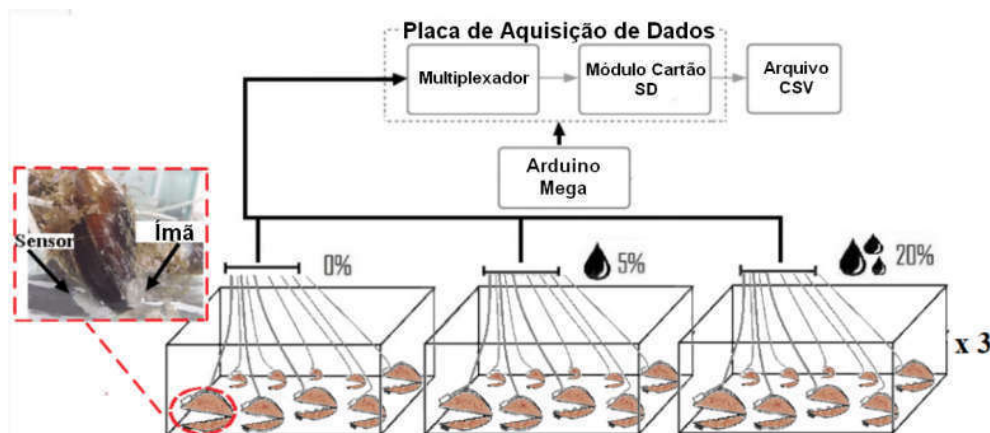


Figura 28: *Layout* do experimento realizado em triplicatas no qual mexilhões *P. perna* foram instrumentados com sensores de efeito Hall e ímãs. As saídas dos sensores foram conectadas a uma placa de aquisição de dados que, em conjunto com um Arduino Mega, salvou as respostas comportamentais em um arquivo CSV. Após um período não toxicológico de três dias, os mexilhões foram expostos a 0%, 5% e 20% de óleo diesel WAF

Tendo em vista a dificuldade de adquirir petróleo bruto, o experimento toxicológico foi realizado através de uma WAF de óleo diesel S-500 (Apêndice C). O preparo deste meio de exposição ocorreu de acordo com o protocolo proposto por Singer et al. (2000) no laboratório de toxicologia do Instituto de Ciências Biológicas (ICB) da Universidade Federal do Rio Grande (FURG).

Trinta e nove mexilhões *Perna perna* (comprimento > 30mm) foram distribuídos em tanques aerados de 5L e mantidos em condições constantes de salinidade (30), temperatura (20°C) e fotoperíodo (12 horas claro e 12 horas escuro). O experimento foi realizado em triplicatas (três aquários para cada tratamento) e três grupos experimentais foram analisados sendo eles controle (0% de WAF), 5% de WAF e 20% de WAF. As duas concentrações de diesel WAF foram preparadas através de diluições do WAF nos aquários e escolhidas com base nos experimentos acerca de biomarcadores bioquímicos e moluscos bivalves descritos por Lüchmann et al. (2011).

Os bivalves foram instrumentados com sensores de efeito Hall UGN3503 e ímãs de neodímio (10 mm de diâmetro por 4 mm de altura). O processo de instrumentação consistiu na fixação dos sensores e ímãs em extremidades opostas das bordas ventrais médias das conchas desses mexilhões (Figura 28). Esses componentes foram fixados por meio de cola cianoacrilato e os sensores impermeabilizados com resina epóxi (BASTI et al., 2009; NAGAI et al., 2006).

Os sensores de efeito Hall foram conectados a uma placa de aquisição de dados para gerar um arquivo CSV (*Comma-Separated Values*) com as respostas comportamentais adquiridas ao longo do tempo. A placa de aquisição de dados era composta por multiplexadores de 16 canais e módulo cartão SD. Ela operava em conjunto com um dispositi-

tivo de prototipagem Arduino Mega cujo conversor analógico-digital possuía 10 bits de resolução.

Os moluscos foram aclimatados por três dias e a aquisição de dados começou posteriormente a uma frequência de 2 Hz ($f = 2Hz$). Os mexilhões foram expostos às respectivas concentrações de WAF de diesel após o terceiro dia de gravação que permaneceu por mais 44 horas. As saídas dos sensores de efeito Hall (Z) referente a cada um dos moluscos foram normalizadas de forma que os valores mínimo (Z_{min}) e máximo (Z_{max}) correspondam à 0 e 100%, respectivamente (COMEAU et al., 2018; TRAN et al., 2010). A equação 63 demonstra o processo de normalização.

$$X_{norm}(i) = \frac{100 \cdot (Z(i) - Z_{min})}{Z_{max} - Z_{min}} \quad \forall i \in [0, f \cdot T[\quad (63)$$

onde X_{norm} corresponde aos dados comportamentais normalizados a cada instante de tempo i tendo em vista uma frequência de aquisição f durante T segundos.

5.2 Ferramentas de Avaliação

Os modelos computacionais propostos para a construção de biossensores de poluição aquática foram validados com base nos dados comportamentais adquiridos ao longo de experimentos laboratoriais. As subseções 5.2.1 e 5.2.2 versam, portanto, sobre as ferramentas de avaliação empregadas para, a partir dos dados comportamentais resultantes, validar os modelos referentes aos módulos técnica de valvometria e ferramentas de análise comportamental, respectivamente.

5.2.1 Módulo Técnica de Valvometria

O experimento para validar o modelo de técnica de valvometria proposto tinha como objetivo possibilitar a comparação entre os dados comportamentais adquiridos por meio de uma técnica de valvometria bem estabelecida e do modelo em questão. Os dados obtidos através do algoritmo de visão computacional foram comparados com a saída dos sensores de efeito Hall a fim de validar o método proposto.

Os sinais resultantes foram normalizados de modo que, para cada mexilhão, as amplitudes de abertura mínima e máxima correspondam à 0 e 1, respectivamente. A relação entre os sinais obtidos para cada animal foi inspecionada através do estabelecimento de um modelo de regressão linear e dos coeficientes de correlação de Pearson (r) e de determinação (r^2). Os erros entre o algoritmo de visão computacional e sensores de efeito Hall foram inspecionados conforme a equação 64.

$$E_j(i) = y_j(i) - \hat{y}_j(i) \quad \forall i \in [0, 120[, j \in [1, 7] \quad (64)$$

onde $E_j(i)$, $y_j(i)$ e $\hat{y}_j(i)$ correspondem ao erro e aos dados normalizados referentes

à saída do sensor de efeito Hall e ao resultado do algoritmo de visão computacional para o j -ésimo animal a cada intervalo de tempo i . Além disso, os histograma dos erros foi analisado com o intuito de estimar sua FDP através da EDK baseada em uma função gaussiana. Por fim, o erro quadrático médio (MSE) foi calculado.

5.2.2 Módulo Ferramentas de Análise Comportamental

A validação de ferramentas capazes inferir o *status* de contaminação ambiente a partir do comportamento de moluscos bivalves requer a utilização de dados comportamentais adquiridos com e sem poluentes. Dessa forma, os dados resultantes do experimento toxicológico foram analisados através das ferramentas apresentadas nas subseções a seguir para validar os diferentes modelos computacionais propostos.

5.2.2.1 Modelo com Classificador Baseado em Grupo

Os dados adquiridos 24 horas antes e após o início da exposição toxicológica foram empregados para validar o modelo classificador baseado em grupo (Figura 29). Os dados comportamentais adquiridos foram, portanto, pré-processados de acordo com os parâmetros comportamentais, agrupados em intervalos de uma hora e rotulados como '0' se não houve exposição toxicológica e '1' caso contrário.

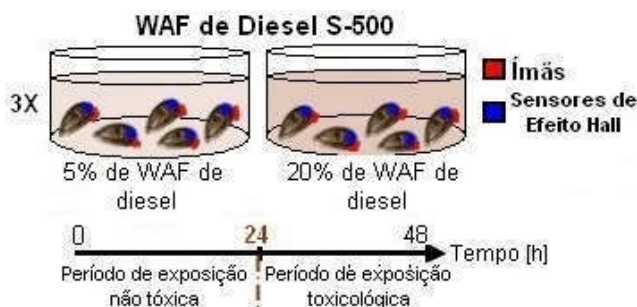


Figura 29: Divisão do experimento em período de exposição não tóxica e período de exposição toxicológica. Ambos períodos tem duração de 24 horas. A partir das primeiras 24 horas, os animais foram expostos à 5% e 20% de WAF de diesel S-500.

Para validar o modelo proposto, foram considerados o comportamento de 8, 8 e 10 mexilhões *Perna perna* pertencentes aos grupos de controle e expostos à 5% e 20% de contaminante, respectivamente. Com o intuito de realizar o treinamento e teste do modelo proposto através de um conjunto de dados balanceado, apenas o comportamento dos animais pertencentes aos grupos expostos à WAF de diesel foram utilizados.

Tendo em vista os dados comportamentais disponíveis e a posterior comparação com métodos estatísticos para determinação de diferenças significativas entre grupos, o modelo proposto foi validado com base nos parâmetros comportamentais amplitude média de abertura, percentual de tempo aberto, frequência de transição e quantidade de micromovimentos de cinco animais. Dessa forma, os dados comportamentais dos animais

foram agrupados de cinco em cinco através de uma análise combinatória simples para gerar dados compatíveis com a quantidade de entradas do empregadas no modelo.

Imediatamente antes do processo de agrupamento, dez e oito animais distintos foram aleatoriamente selecionados para compor os conjuntos de treinamento e teste. Tendo em vista que para cada animal há 48 conjuntos de parâmetros comportamentais horários, foram gerados conjuntos distintos de treinamento e teste contendo 12096 (81,82%) e 2688 dados (18,18%), respectivamente. Além disso, metade dos animais que compuseram o conjunto foram expostos à 5% WAF de diesel e a outra metade à 20% de WAF.

O aplicativo *Classification Learner* do *software* Matlab foi empregado para treinar de forma supervisionada diversos modelos clássicos de classificação como, por exemplo, árvores de decisão, SVMs (*Support Vector Machines*) e classificadores *ensemble*, baseados em regressão logística e KNN (K-nearest neighbors) através de uma validação cruzada de 10-fold. O classificador de maior acurácia de AUC resultante deste processo de treinamento foi selecionado para validar o modelo proposto.

A seguir, o classificador selecionado foi utilizado para inferir os *status* de contaminação de acordo com os dados de teste. Sua performance foi avaliada com base na acurácia, sensibilidade, especificidade e taxa de falsos positivos considerando como classe positiva a presença de contaminantes no ambiente.

Com o intuito de comparar a performance do modelo proposto em relação a métodos estatísticos, analisou-se a ocorrência de diferenças significativas entre o conjunto de dados de controle e de teste. Supôs-se que durante o período de exposição toxicológica deveriam ser identificadas diferenças comportamentais significativas entre os grupos de animais.

O método estatístico ANOVA de múltiplos fatores (*Multivariate Analysis Of Variance* - MANOVA) foi empregado após verificação das suposições de normalidade e homocedasticidade da variância através dos métodos estatísticos Shapiro-Wilk multivariado e Bartlett, respectivamente. Caso alguma das suposições não tenha sido verificada, o teste não paramétrico de Análise de Variância Multivariado com Permutação (PERMANOVA) foi empregado. Todos os testes estatísticos foram realizados a um nível de significância de 5%.

5.2.2.2 Modelo com Combinação de Classificadores

Assim como o para o modelo classificador baseado em grupo (Figura 29), os dados adquiridos 24 horas antes e após o início da exposição toxicológica foram empregados para validar o modelo baseado na combinação de classificadores. Os dados comportamentais adquiridos foram, portanto, pré-processados de acordo com os parâmetros comportamentais amplitude média de abertura, percentual de tempo aberto, frequência de transição e quantidade de micromovimentos. A seguir, eles foram agrupados em intervalos de uma hora e rotulados como '0' se não houve exposição toxicológica e '1' caso contrário. Dessa forma, o comportamento de mexilhões pertencentes ao grupo de con-

trole foram rotulados como '0' uma vez que não houve exposição à WAF de diesel. Além disso, foram considerados um total de 8 e 10 mexilhões expostos à 5% e 20% de WAF de diesel, respectivamente.

A abordagem empregada na validação deste modelo teve como objetivo verificar se a combinação de diferentes modelos de classificação, como por exemplo SVM, KNN e árvores de decisão, apresenta vantagens em relação ao emprego de classificadores individuais na determinação da presença de contaminantes a partir do comportamento de um molusco bivalve. Pretendeu-se, portanto, verificar se o emprego de animais individuais no monitoramento aquático era factível e se a combinação de diferentes classificadores contribuiria para uma determinação mais apropriada do *status* de combinação ambiente.

Os dados agrupados e rotulados foram divididos em dois conjuntos treinamento (primeiro e segundo) e um de teste. O conjunto de teste foi utilizado para testar a combinação de classificadores. O primeiro e o segundo conjuntos de treinamento foram empregados para treinar os modelos de classificação e o classificador de votação, respectivamente. Todos os conjuntos de dados eram disjuntos.

Dessa forma, 16,67% dos dados foram utilizados para teste e englobam o comportamento de um e dois mexilhões selecionados aleatoriamente dos grupos expostos a 5% e 20% de WAF, respectivamente. Os dados restantes foram divididos em primeiro (65%) e segundo (35%) conjuntos treinamento.

O aplicativo *Classification Learner* do *software* Matlab foi empregado para treinar os modelos de classificação. Ele permitiu a realização de um treinamento supervisionado diversos modelos clássicos como, por exemplo, árvores de decisão, SVMs (*Support Vector Machines*), KNN (*K-nearest neighbors*), classificadores *ensemble* e baseados em regressão logística. Além disso, este aplicativo possibilitou comparações entre os classificadores resultantes com base em métricas de acurácia, matrizes de confusão e curvas ROC. Nesse sentido, diferentes modelos de classificação foram treinados e aqueles com maior acurácia e valor AUC selecionados para compor e validar o modelo proposto.

Conforme demonstrado na Figura 30, o processo de treinamento foi dividido em dois estágios.

O primeiro conjunto de dados de treinamento foi utilizado para treinar diferentes modelos de classificação disponíveis no *software* em questão e determinar se há ou não poluentes no ambiente tendo em vista os parâmetros comportamentais horários de cada mexilhão. Os quatro classificadores com maior acurácia foram selecionados para compor o modelo proposto e inferir o *status* de contaminação de acordo com os dados comportamentais do segundo conjunto de treinamento.

Estas classificações foram empregadas para treinar um classificador de votação e, portanto, inferir um *status* final de contaminação (classificação binária sobre a presença ou ausência de contaminantes). Este classificador foi selecionado com base no modelo de classificação de maior precisão e AUC durante o segundo estágio de treinamento através

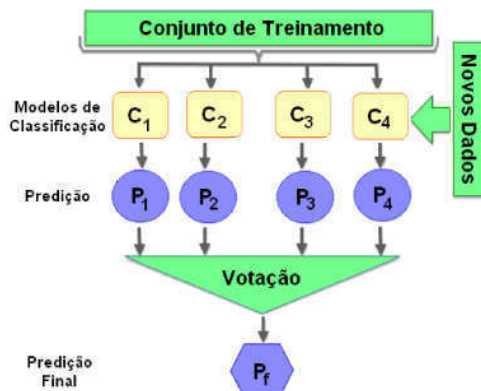


Figura 30: Esquema do classificador de votação no qual quatro classificadores foram treinados com base em um conjunto de dados de treinamento contendo a amplitude média de abertura, percentual de tempo aberto, frequência de transição e quantidade de micromovimentos. As classificações (P_1 , P_2 , P_3 e P_4) adquiridas por meio de um novo conjunto de dados foram usadas para obter uma previsão final (P_f) através de um classificador de votação.

do aplicativo *Classification Learner*. Uma validação cruzada de 10-*fold* foi empregada em todos os processos de treinamento.

5.2.2.3 Modelo com NARNET e Gráfico de Controle EWMA

Primeiramente, as Funções de Autocorrelação (ACFs) das AOAs durante o período não toxicológico foram consideradas a um nível de significância de 1%. Esta análise objetivou determinar a quantidade necessária de atrasos na entrada de NARNETs para uma adequada modelagem dos dados comportamentais individuais.

As NARNETs contaram, portanto, com sessenta atrasos de entrada para prever a AOA no instante de tempo $t + 1$ ($Y(t + 1)$) com base no comportamento ao longo das últimas 6 horas. Esta arquitetura apresentou uma camada oculta com quinze neurônios cuja função de ativação era tangente hiperbólica e uma camada de saída constituída por um neurônio com uma função de ativação linear. A Figura 31 ilustra a arquitetura empregada.

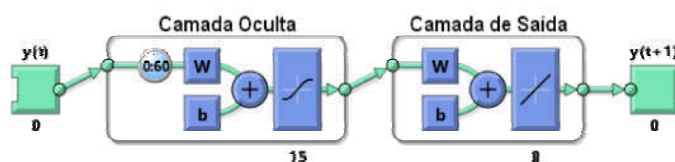


Figura 31: Arquitetura da NARNET para prever $Y(t + 1)$ com base nos dados comportamentais das últimas seis horas (últimas sessenta AOA) de $Y(t)$. A rede possui sessenta atrasos de entrada, uma camada oculta com quinze neurônios cuja função de ativação é uma tangente hiperbólica. A camada de saída abrange um neurônio com uma função de ativação linear.

As AOAs de seis mexilhões selecionados aleatoriamente de cada grupo durante os dois primeiros dias de experimento foram empregadas para treinar NARNETs. Estes dados foram divididos aleatoriamente em treinamento (70%), teste (15%) e validação (15%). O algoritmo Levenberg-Marquardt foi utilizado como método de treinamento juntamente com uma validação cruzada 10-*fold*.

O desempenho da rede neural foi avaliado através do erro quadrático médio (*Mean Squared Error* - MSE) e do coeficiente de correlação de Pearson (r) entre as saídas da rede neural e os dados esperados como saída, isto é, as AOAs a cada intervalo de 0,1 hora. As NARNETs treinadas foram empregadas para prever a AOA de cada animal durante o restante do experimento. Os valores de AOA adquiridos através do experimento (Y) foram subtraídos do sinal previsto (Y_p) para obter resíduos não correlacionados. Como os gráficos de controle EWMA são afetados pela autocorrelação dos dados, os ACFs dos AOAs não-toxicológicos e resíduos foram inspecionados a um nível de significância de 1%.

Os três dias restantes de experimento abrangiam 24 horas de exposição não-toxicológica e 44 horas de exposição à WAF de diesel S-500. Os resíduos de cada animal foram, portanto, utilizados como amostras de um mesmo processo em um gráfico de controle EWMA. Oitenta por cento (80%) dos dados não toxicológicos foram utilizados para estabelecer os limites de controle e os dados restantes para testar o gráfico de controle.

Com o intuito de averiguar a performance do modelo proposto em relação à métodos estatísticos tradicionais, o método estatístico ANOVA de um fator foi empregado para identificar diferenças significativas entre os grupos de controle e expostos à WAF de diesel S-500 a cada intervalo de 0,1 hora. As suposições de normalidade da distribuição e homocedasticidade da variância foram verificados através dos testes de Kolmogov-Smirnov e Bartlett, respectivamente. Caso alguma das suposições não tenha sido verificada, o teste não paramétrico de Kruskal-Wallis foi empregado. Todos os testes estatísticos foram realizados a um nível de significância de 5%.

As saídas fornecidas pelo modelo proposto e pelo emprego destes métodos estatísticos foram comparados com a saída esperada para os períodos de exposição não toxicológica e toxicológica. Por fim, matrizes de confusão foram analisadas a fim de identificar se o modelo baseado na associação de NARNETs e gráfico de controle EWMA ofereceu vantagens em relação aos métodos estatísticos tradicionalmente empregados na construção de biossensores baseados na análise comportamental de bivalves.

6 RESULTADOS E DISCUSSÃO

Biossensores de poluição aquática baseados na análise comportamental de bivalves apresentam como módulos constituintes os próprios animais, as técnicas de valvometria empregadas para mensurar a amplitude de abertura de suas conchas, os parâmetros comportamentais aplicados na análise de seu comportamento e as ferramentas que possibilitam inferir a ocorrência de condições ambientais inóspitas a partir de sua análise (Figura 1). Nesse sentido, a presente dissertação de mestrado propôs um modelo computacional baseado em visão para a construção de uma técnica de valvometria não invasiva bem como três modelos computacionais como ferramentas de análise comportamental. As subseções a seguir versam , portanto, sobre os resultados obtidos através da validação dos modelos propostos.

6.1 Módulo Técnica de Valvometria

As principais fontes de incertezas em sistemas VBM incluem iluminação, ângulo da câmera e equipamento da câmera (SHIRMOHAMMADI; FERRERO, 2014). A metodologia proposta utilizou um sistema de iluminação passiva, considerou os parâmetros intrínsecos e extrínsecos da câmera e posicionou a câmera de forma que o aquário ocupasse toda a altura da imagem. Portanto, os principais efeitos sistemáticos de um sistema VBM podem ser negligenciados. Outros efeitos sistemáticos podem incluir desalinhamento dos animais durante a fixação e falha no processo de pintura dos componentes.

Os sinais visuais resultantes do experimento proposto foram processados em busca das regiões com a coloração de interesse (verde ou laranja) no espaço HSV. Primeiramente, as imagens aquistadas passaram por um pré-processamento a fim de melhorar seu contraste. A Figura 32 ilustra um exemplo de imagem antes e após seu pré-processamento.

Cada animal foi, portanto, representado pela sua região de interesse de modo que cada imagem pudesse ser dividida em sete ROIs. Cada ROI foi processada em busca das regiões laranja e verde com base nos valores HSV médios. Geraram-se, dessa forma, imagens binárias que, a seguir, foram filtradas por meio de um operador morfológico de abertura com elemento estruturante quadrado de tamanho 200 pixels. Esse procedimento

foi necessário a fim de eliminar pixels erroneamente detectados sem comprometer de forma incisiva arestas de interesse.

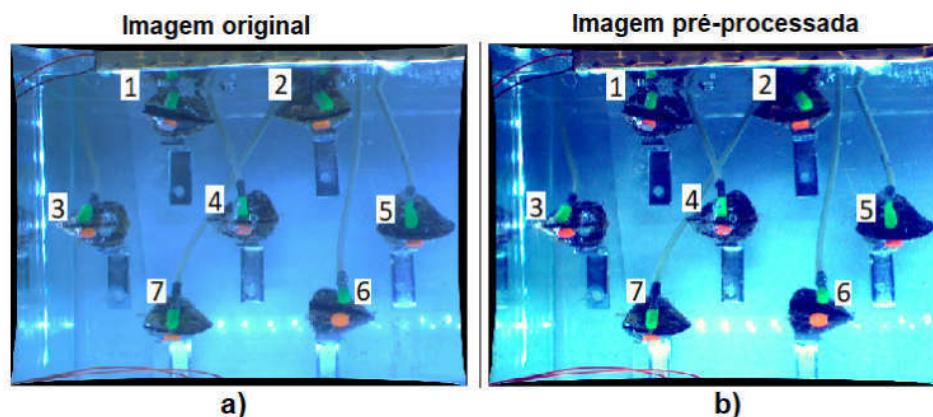


Figura 32: Imagem antes (a) e após (b) melhoria de contraste e a respectiva enumeração dos animais.

Por fim, as imagens binárias filtradas foram utilizadas na detecção dos pontos mais próximos às bordas tanto para a coloração verde quanto laranja. A distância vertical em pixels entre esses pontos foi considerada uma medida da amplitude de abertura das conchas dos animais bivalves. A Figura 33 ilustra um exemplo de ROI, as regiões laranja e verde detectadas a partir do valor HSV médio e a amplitude analisada.

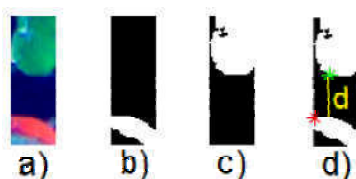


Figura 33: Região de interesse (a), regiões de coloração laranja (b) e verde (c) detectadas a partir dos valores HSV médios. Os pontos verde e vermelhos ilustrados em (c) correspondem aos pontos máximo e mínimo das figuras (b) e (c), respectivamente. A distância d destacada em amarelo corresponde ao valor de amplitude detectado para o caso apresentado.

Nota-se através da Figura 33 que o algoritmo de visão computacional apresentado permitiu a detecção da distância de abertura das conchas dos animais bivalves de forma satisfatória. Dessa maneira, processando todas as imagens geradas pelo experimento obteve-se os dados de abertura e fechamento das conchas de todos os animais empregados ao longo do experimento. A Figura 34 ilustra os gráficos comportamentais normalizados decorrentes do emprego de sensores de efeito Hall e algoritmo de visão computacional para cada molusco bivalve. Devido à falha na conexão do sensor de efeito Hall às conchas do animal 7 (Figura 32), os dados comportamentais referente à este animal foram excluídos da análise.

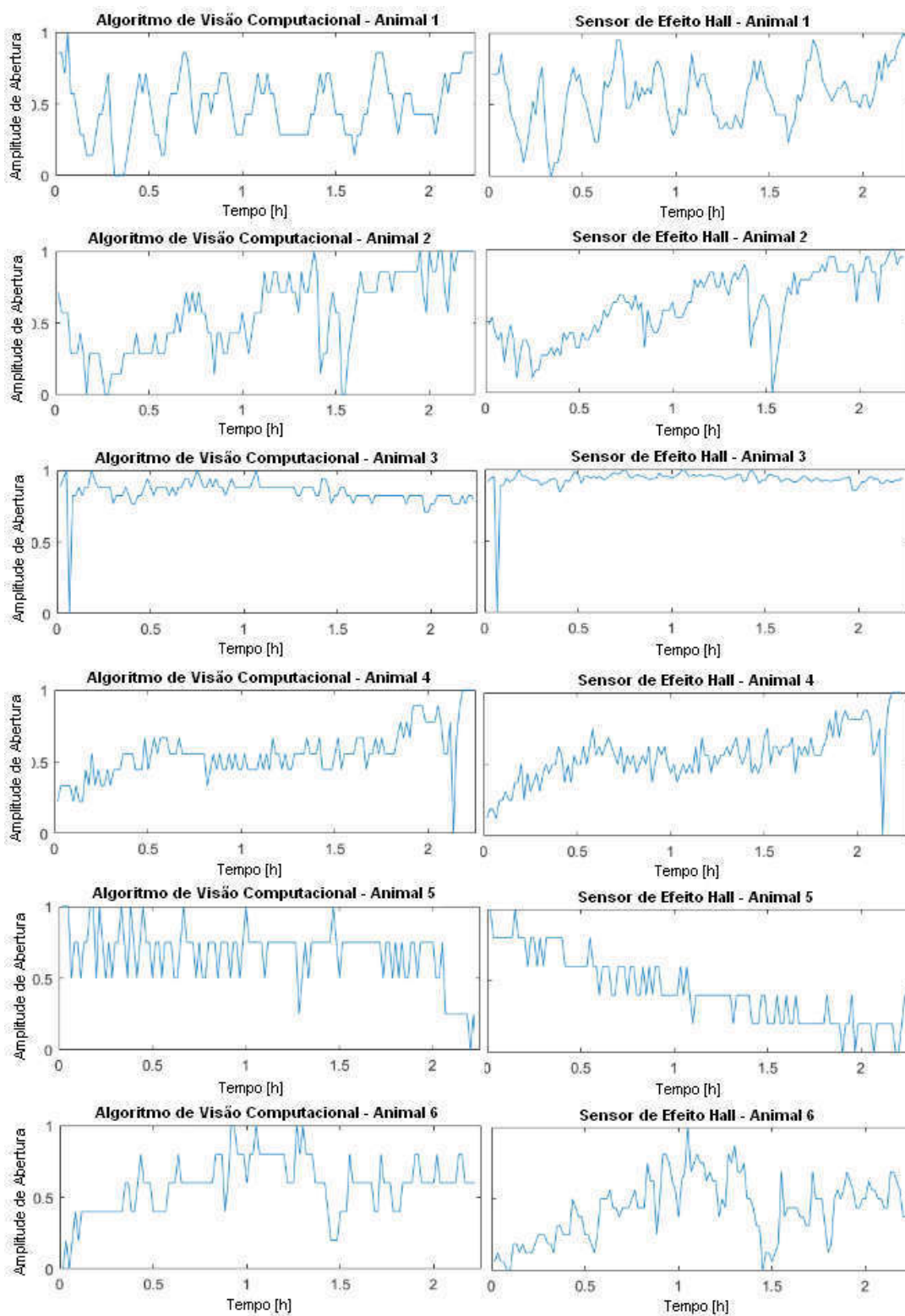


Figura 34: Comportamento de abertura e fechamento de mexilhões *Perna perna* tendo em vista o emprego dos sensores de feito Hall e algoritmo de visão computacional

Os erros entre os valores encontrados através das metodologias propostas foram calculados para cada mexilhão (Figura 35) e proporcionaram, em média, um MSE de $15,35 \cdot 10^{-3}$. A Tabela 7 resume os coeficientes de correlação de Pearson (r), os valores p encontrados em sua análise, as retas de regressão linear ($y = ax + b$), os coeficientes de determinação (r^2) e os erros quadráticos médios entre o sensor de efeito Hall e o algoritmo de visão computacional para cada molusco bivalve.

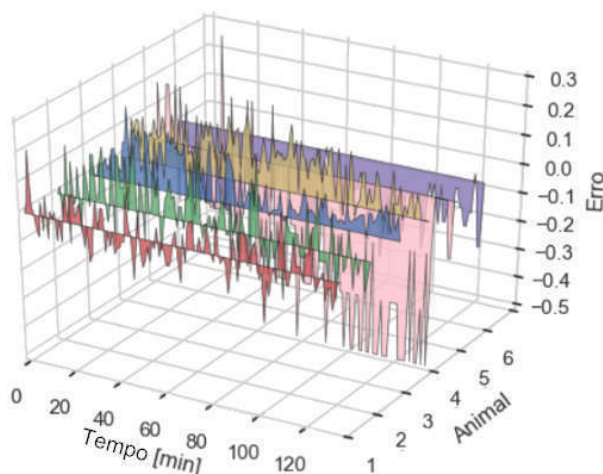


Figura 35: Erro entre sensores de efeito Hall e algoritmo de visão computacional para cada animal ao longo de duas horas de experimento.

Tabela 7: Coeficiente (r) e valor p resultante da análise de correlação de Pearson para cada animal inspecionado. A reta de regressão linear ($y = ax + b$) que minimiza a soma residual dos quadrados entre os dados obtidos pelos sensores de efeito hall (y) e algoritmo de visão computacional (x) de acordo com os coeficientes a e b . Também são apresentados o coeficiente de determinação (r^2) e o erro quadrático médio (MSE) entre o sensor de efeito Hall e o algoritmo de visão computacional.

Animal	r	r^2	valor p	a	b	MSE
1	0,96	0,92	$2,17 \times 10^{-78}$	0,84	0,09	$4,44 \times 10^{-3}$
2	0,94	0,88	$3,40 \times 10^{-62}$	0,83	0,13	$8,68 \times 10^{-3}$
3	0,88	0,64	$5,90 \times 10^{-31}$	0,68	0,35	$3,64 \times 10^{-3}$
4	0,93	0,86	$1,49 \times 10^{-59}$	0,86	0,15	$5,09 \times 10^{-3}$
5	0,30	0,09	$3,66 \times 10^{-4}$	0,36	0,19	$63,67 \times 10^{-3}$
6	0,93	0,86	$5,44 \times 10^{-58}$	1,04	-0,18	$6,60 \times 10^{-3}$

A Figura 36 ilustra as retas de regressão linear modeladas para relacionar os sinais obtidos através dos sensores de efeito Hall e algoritmo de visão computacional para cada mexilhão *Perna perna*. Além disso, encontram-se demonstrados os coeficientes de correlação de Pearson.

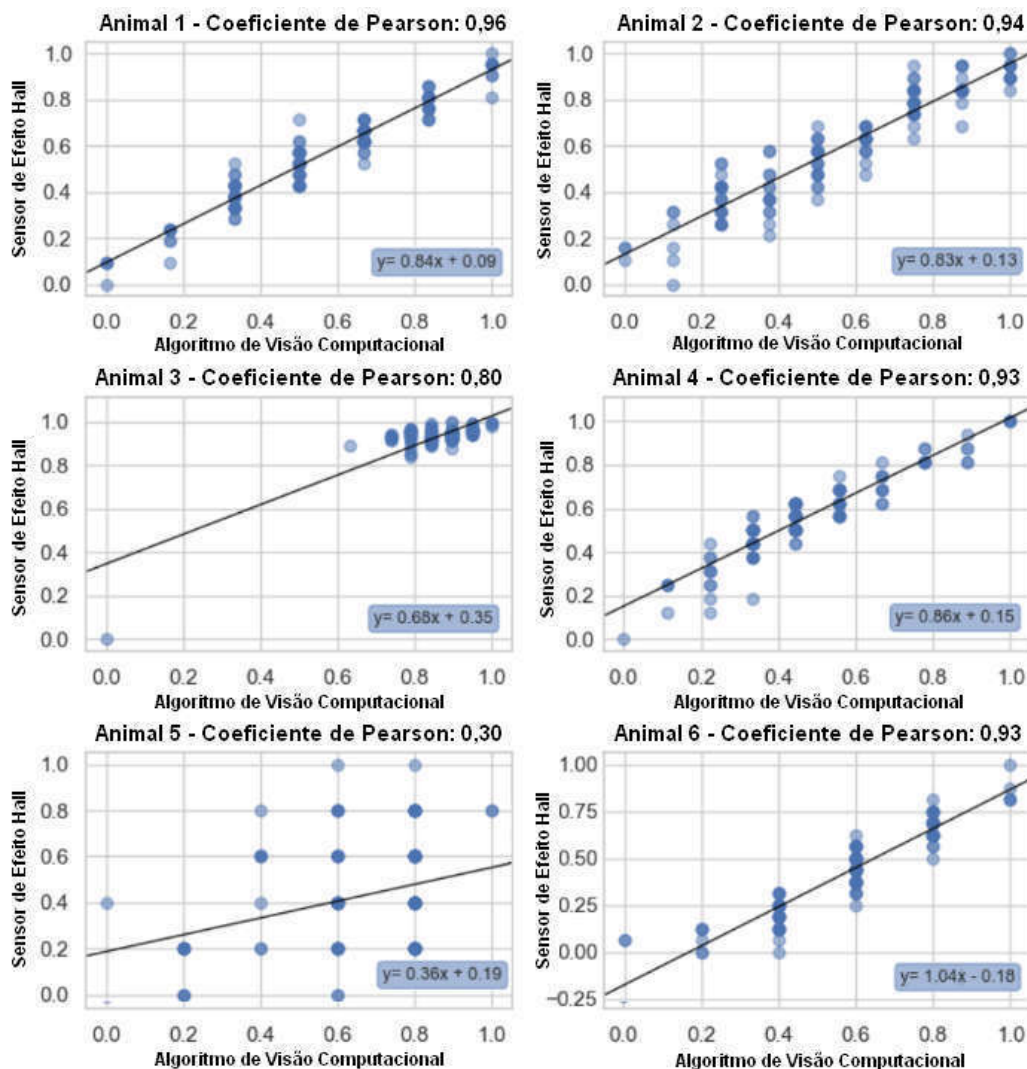


Figura 36: Gráficos de dispersão do comportamento de mexilhões *Perna perna* de acordo com sensores de efeito Hall e algoritmo de visão computacional. A linha de regressão linear ($y = ax + b$) minimiza a soma residual dos quadrados entre os dados obtidos pelos sensores de efeito hall (y) e o algoritmo de visão computacional (x) de acordo com os coeficientes a e b . Os coeficientes de correlação de Pearson encontram-se ilustrados para cada bivalve.

Para todos os mexilhões investigados, foram encontradas correlações positivas estatisticamente significantes ($p < 0,05$) entre os sinais provenientes dos sensores de efeito Hall e algoritmo de visão computacional. De acordo com os coeficientes de correlação, houve correlação muito forte ($0,8 \leq |r| < 1$) em 83,34% dos casos investigados. Somente para o mexilhão 5 foi observada uma correlação fraca ($0,2 \leq |r| < 0,4$).

Com exceção do mexilhão 5, de acordo com os coeficientes de determinação (r^2) os modelos de regressão linear foram capazes de explicar, em média, 83,2% da variância nos

dados dos sensores de efeito Hall com base nos valores resultantes do processamento de imagens. O processamento dos dados referentes ao animal 5 proporcionou coeficientes de correlação e determinação bem como MSE discrepantes em relação aos demais casos investigados que, em média, apresentaram um MSE de $5,60 \times 10^{-3}$. Uma análise visual *frame-a-frame* do desempenho do algoritmo proposto demonstrou que ele foi capaz de corretamente detectar os pontos mais próximos à região de abertura do animal 5 e, portanto, a amplitude de abertura.

Dessa forma, a *performance* referente ao animal 5 pode estar associada ao curto período de tempo analisado (2 horas) e que, diferente dos demais animais, o animal 5 permaneceu com amplitude de abertura praticamente constante (3 pixels de diferença entre as amplitudes máxima e mínima registradas). Dessa forma, como os dados foram normalizados de 0 à 1, e houve pouca diferença entre os valores máximo e mínimo de abertura isto pode ter afetado a análise proposta.

A FDP foi estimada para cada animal tendo em vista o emprego da EDK baseada em uma função Gaussiana. Nesse sentido, as Figuras 37 e 38 ilustram os histogramas e as FDP estimadas com base no comportamento individual e geral dos mexilhões investigados, respectivamente.

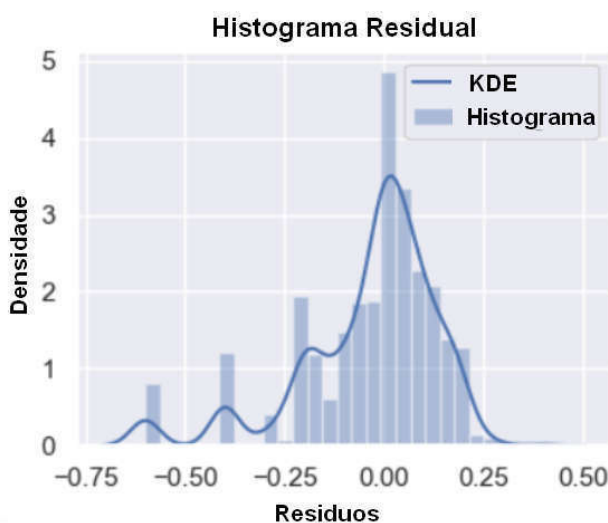


Figura 38: Estimativa de densidade por kernel (KDE) de histogramas de erro entre sensores de efeito hall e algoritmo de visão computacional. As funções densidade de probabilidade e os histogramas de erro são demonstradas como linhas e barras azuis, respectivamente.

As PDF resultantes indicaram uma maior probabilidade de obter erros nulos. Além disso, a assimetria negativa ilustrada na Figura 38 demonstrou uma maior probabilidade de obter erros menores ou iguais a zero. Portanto, a abordagem de visão computacional proposta tende a fornecer amplitudes de abertura maiores que aquelas obtidas pela

técnica clássica de valvometria. Em geral, uma redução na amplitude média de abertura das conchas bivalves está associada à exposição dos bivalves a contaminantes. Nesse sentido, a metodologia proposta reduziria taxas de falsos positivos no biomonitoramento de ambientes aquáticos.

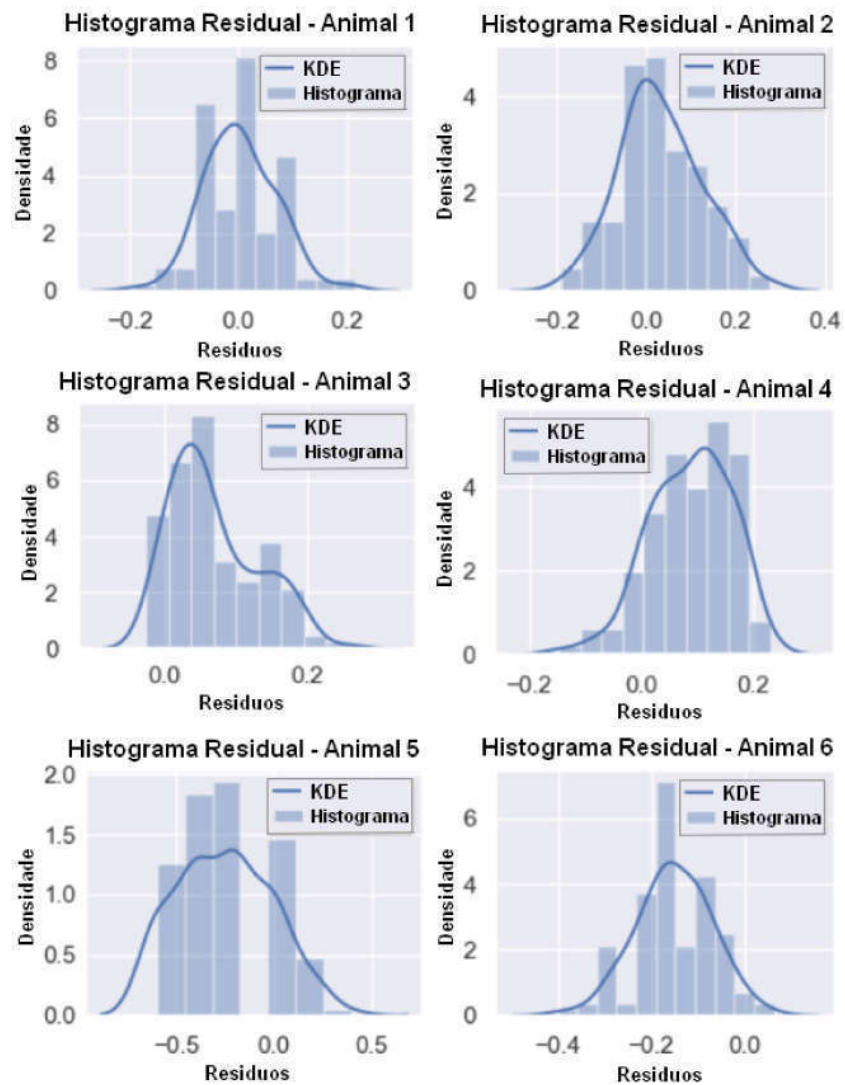


Figura 37: Estimativa de densidade por kernel (KDE) de histogramas de erro entre sensores de efeito hall e algoritmo de visão computacional para cada mexilhão *Perna perna*. As funções densidade de probabilidade e os histogramas de erro são demonstradas como linhas e barras azuis, respectivamente.

6.2 Módulo Ferramentas de Análise Comportamental

Com o intuito de validar os modelos de ferramentas de análise comportamental, um experimento toxicológico empregando sensores de efeito Hall em conjunto com ímãs como técnica de valvometria foi realizado. O monitoramento comportamental dos mexilhões *Perna perna* ao longo deste experimento forneceu respostas idiossincráticas e demonstrou a complexidade da avaliação comportamental conforme exemplificado por meio da Figura 39. O fechamento das conchas ocorreu de forma assíncrona entre mexilhões individuais durante tanto o período de exposição toxicológica quanto não toxicológica conforme também relatado nos trabalhos Comeau et al. (2018) e Hartmann et al. (2016). O comportamento individual dos bivalves, principalmente em ambiente contaminado, diferiu e demonstrou a plasticidade fenotípica para tentativas individuais de diminuir o contato com meios desagradáveis (DINGEMANSE et al., 2010).

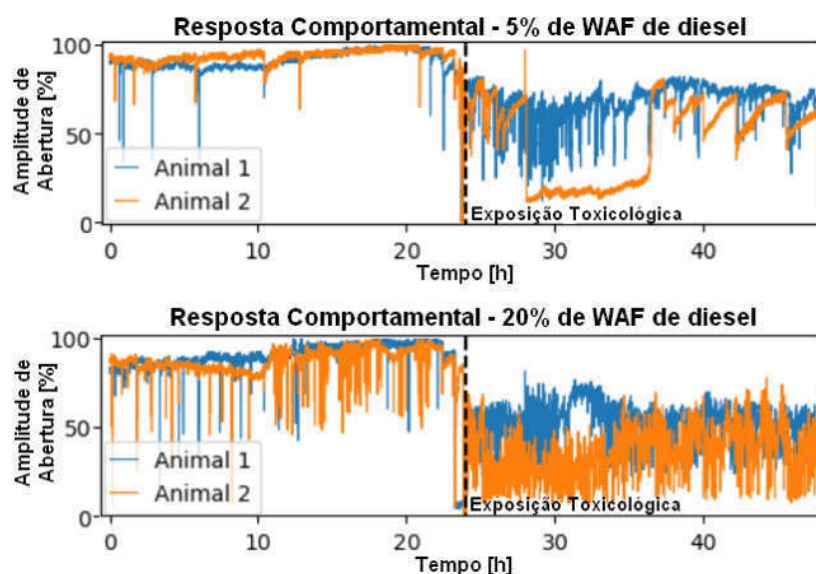


Figura 39: Exemplos de respostas comportamentais de mexilhões *Perna perna* expostos a 5% e 20% de WAF de diesel. Sensores de efeito Hall foram empregados para adquirir dados comportamentais em um experimento que compreendeu 24 horas de exposição não toxicológica e toxicológica, como demonstrado pela linha tracejada preto.

Os dados obtidos por meio deste experimento utilizando WAF de diesel S-500 possibilitou a construção e validação de três modelos como ferramentas de análise comportamental. Nesse sentido, as subseções 6.2.1, 6.2.2 e 6.2.3 abordam os modelos propostos tendo em vista o emprego de um classificador baseado em grupo, uma combinação de classificadores e uma associação de NARNETs com gráfico de controle EWMA, respectivamente.

6.2.1 Modelo com Classificador Baseado em Grupo

O aplicativo *Classification Learner* do *software* Matlab foi empregado para treinar de forma supervisionada modelos clássicos de classificação capazes de inferir o *status* de contaminação ambiente com base em parâmetros comportamentais horários de cinco milhões *Perna perna*. O classificador *ensemble bagged* forneceu maior acurácia (99,97%) e AUC (0,99) durante o processo de treinamento sendo, portanto, selecionado para validar o modelo proposto. A Figura 40 ilustra a matriz de confusão, curva ROC e AUC resultantes de seu processo de treinamento.

O emprego do classificador *ensemble bagged* para inferir a presença de WAF de diesel S-500 no ambiente aquático com base no conjunto de teste composto pelo comportamento de bivalves não empregados durante o processo de treinamento forneceu uma acurácia, taxa de falsos positivos e sensibilidade de 93,15%, 3,35% e 89,66%, respectivamente. A Figura 41 (A) ilustra a matriz de confusão resultante do teste do modelo em questão.

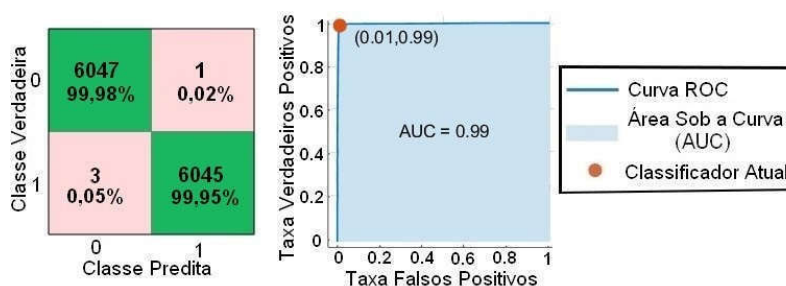


Figura 40: Matriz de confusão, curva ROC e AUC referentes ao processo de treinamento de um classificador *ensemble bagged* para inferir o *status* de contaminação ambiente a partir de parâmetros comportamentais horários de um grupo de cinco moluscos bivalves.

Por fim, o método estatístico ANOVA de múltiplos fatores ou o método não paramétrico PERMANOVA foram empregados para verificar diferenças significativas entre os grupos de controle e de teste a um nível de significância de 5%. Este processo forneceu uma acurácia, taxa de falsos positivos e sensibilidade de 60,83%, 2,38% e 24,03%, respectivamente. A Figura 41 (B) ilustra a matriz de confusão resultante deste processo.

O modelo proposto apresentou desempenho satisfatório de acordo com todas as métricas de avaliação bem como capacidade de generalização para dados e comportamentos não empregados durante o processo de treinamento. Além disso, quando comparados com o emprego de métodos estatísticos, sua utilização apresentou vantagens em termos de acurácia e sensibilidade sugerindo, portanto, o potencial de modelos computacionais de inteligência artificial na construção de biossensores de poluição aquática.

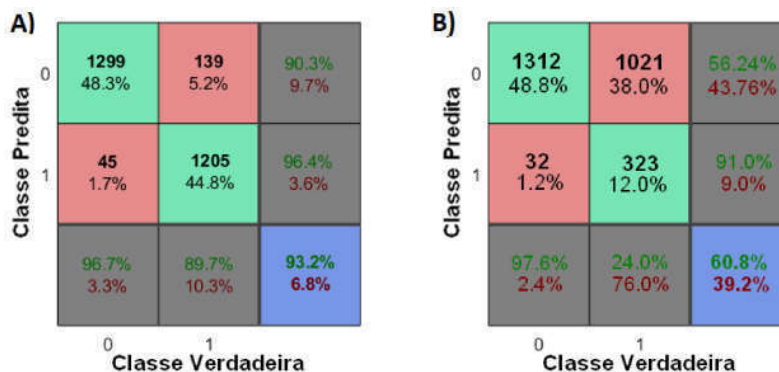


Figura 41: Matrizes de confusão referentes ao processo de teste do classificador *bagged trees* (A) e ao emprego de métodos estatísticos na identificação de diferenças significativas entre os grupos de controle e de teste ($p < 0,05$) (B)

6.2.2 Modelo com Combinação de Classificadores

A validação proposta para o modelo com combinação de classificadores tinha como objetivo verificar se a combinação de diferentes métodos de classificação forneceria vantagens em relação ao emprego de um método único. Ademais, objetivou-se verificar a aplicabilidade do modelo proposto na monitoração aquática a partir da análise comportamental de um molusco bivalve por diferentes métodos de classificação e sua combinação na determinação do *status* de contaminação ambiente.

Diferentes modelos de classificação foram treinados através do aplicativo *Classification Learner* do *software* MATLAB. Os classificadores resultantes com maior acurácia e valor AUC durante os processos de treinamento foram selecionados para validar a combinação de classificadores proposta. Dessa forma, a primeira etapa de treinamento para o classificador de votação forneceu, em média, acurácias 86,5%, 85,9%, 85,7% e 85,3% para os classificadores *bagged tree*, SVM cúbico, SVM quadrático e SVM gaussiano, respectivamente.

A Figura 42 ilustra as matrizes de confusão, curvas ROC e valores de AUC de cada classificador obtido. Eles atingiram valores de AUC maiores ou iguais a 0,9, indicando uma alta qualidade de classificação. Em geral, os classificadores tenderam a fornecer maiores taxas de falsos negativos em comparação com as taxas de falsos positivos.

Os quatro classificadores selecionados a partir de suas acurácias e valores AUC foram utilizados para inferir o *status* de contaminação ambiente (classificação binária acerca da presença ou ausência de contaminantes) de acordo com o segundo conjunto de dados de treinamento.

As classificações adquiridas foram empregadas para novamente treinar diferentes modelos de classificação responsáveis por combinar estes dados e determinar uma classificação final acerca da condição ambiente. Dentre os classificadores resultantes, o classificador SVM Gaussiano foi selecionado como classificador de votação em virtude

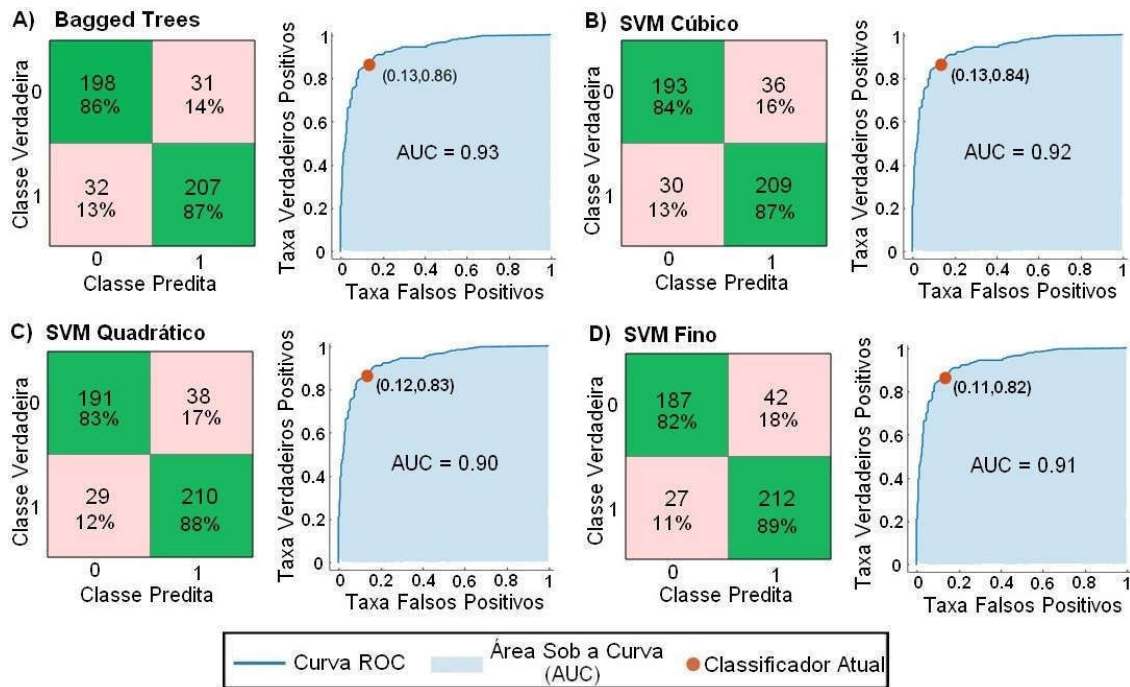


Figura 42: Matrizes de confusão, curvas ROC e AUC dos classificadores *bagged trees* (A), SVM cúbica (B), SVM quadrática (C) e SVM gaussiana (D) obtidas durante a primeira etapa de treinamento do modelo baseado na combinação de classificadores para detecção de compostos petroquímicos.

de sua maior acurácia (84,6 %) e AUC (0,88) conforme ilustrado na Figura 43.

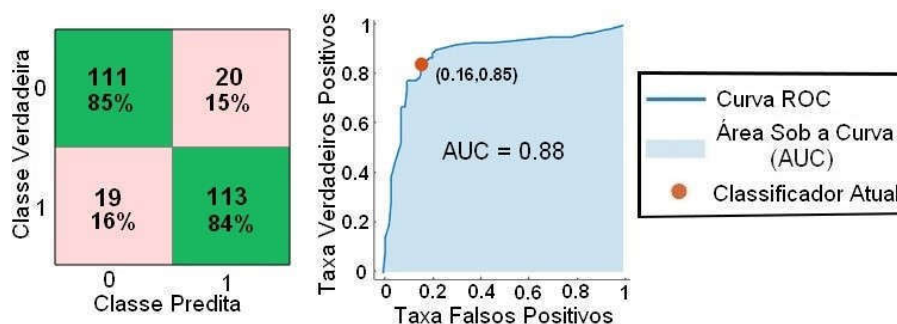


Figura 43: Matriz de Confusão, curva ROC e valor AUC obtidos durante o treinamento do classificador de votação para detecção de compostos petroquímicos.

Por fim, a combinação destes classificadores foi empregada para inferir o *status* de contaminação com base em dados comportamentais de teste e forneceu uma precisão de 95,8% para ambas concentrações de WAF de diesel. A Figura 44 ilustra as matrizes de confusão resultantes deste processo considerando grupos expostos à 5% e 20% de WAF.

O desempenho de todos os classificadores que compuseram esta combinação foram avaliados através dos valores de acurácia, curvas PR, AUC-PR e *F-score*. A Figura 45 ilustra as curvas PR destes classificadores para cada concentração de WAF de diesel e a

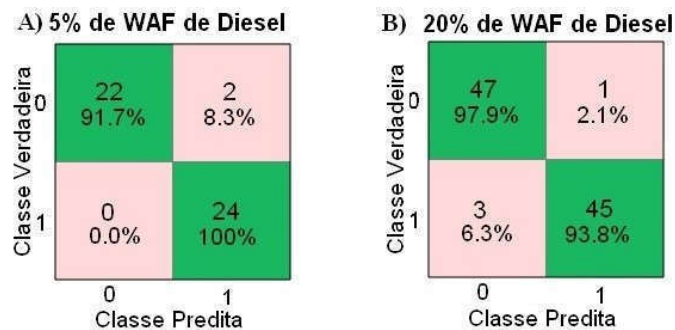


Figura 44: Matrizes de Confusão, curvas ROC e valores AUC obtidos durante o teste do classificador de votação tendo em vista 5% (A) e 20% (B) de WAF de diesel.

Tabela 8 resume os desempenhos de teste.

Embora o classificador *bagged trees* tenha apresentado a maior precisão durante a primeira etapa do treinamento, ele teve um menor desempenho em comparação ao classificador de votação de acordo com todas as métricas avaliadas. O classificador SVM quadrático apresentou maior precisão quando testado para detectar a maior concentração do composto petroquímicos.

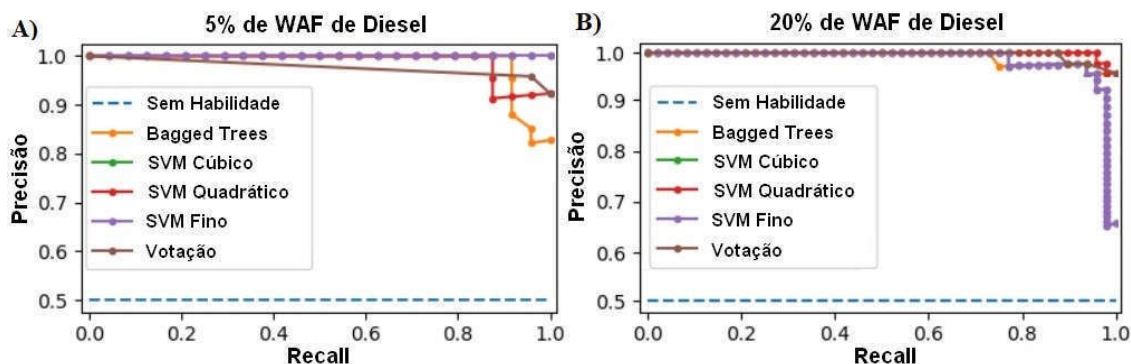


Figura 45: Curvas Precisão-Revocação (PR) para 5% (A) e 20% (B) de WAF de diesel.

No entanto, ao monitorar ambientes aquáticos, pode ser mais vantajoso um biossensor capaz de detectar concentrações mais baixas de contaminantes. Dessa forma, o classificador de votação foi capaz de fornecer desempenho adequado para ambas as concentrações de compostos petroquímicos e apresentou vantagens em comparação com o SVM quadrático na detecção de 5% de WAF de diesel de acordo com as métricas *F-score* e AUC-PR. Embora o SVM cúbico não tenha fornecido as melhores métricas de desempenho durante o estágio de treinamento, ele obteve desempenho similar ao classificador de votação.

Torna-se evidente, portanto, que o uso da combinação de classificadores foi propício à detecção de compostos petroquímicos em ambientes aquáticos. Por considerar o comportamento horário dos bivalves no monitoramento de ambientes aquáticos, ele contribui

Tabela 8: Performance dos classificadores durante o teste tendo em vista a exposição à 5% e 20% de WAF de diesel.

Classificador	5% de WAF de Diesel			20% de WAF de Diesel		
	Acurácia (%)	<i>F-Score</i>	AUC-PR	Acurácia (%)	<i>F-Score</i>	AUC-PR
Bagged Trees	87,5	0,89	0,99	94,8	0,95	0,99
SVM Cúbico	95,8	0,96	1,00	95,8	0,96	0,99
SVM Quadrático	91,7	0,92	0,99	97,9	0,98	1,00
SVM Gaussiano	89,6	0,96	1,00	95,8	0,957	0,99
Votação	95,8	0,96	0,98	95,8	0,957	1,00

para o desenvolvimento de um biossensor rápido, econômico e eficiente. Além disso, o sistema de múltiplos classificadores forneceu acurácia adequada quando utilizado para inferir o *status* de contaminação com base no comportamento de mexilhões não utilizados durante os processos de treinamento. A metodologia proposta é, portanto, adequada para lidar com a natureza adaptativa dos moluscos bivalves.

O uso do sistema baseado em múltiplos classificadores forneceu classificações mais confiáveis em relação a classificadores individuais para todos os casos investigados. Além disso, a validação proposta baseia-se na análise comportamental de mexilhões individuais. Tendo em vista que esses animais vivem em grupos e que sistemas baseados na análise comportamental de pequenos grupos encontram-se em soluções comerciais efetivas, uma abordagem baseada em um conjunto de animais e combinação de classificadores pode contribuir ainda mais para o desenvolvimento de biossensores de poluição aquática.

6.2.3 Modelo com NARNETs e Gráfico de Controle EWMA

Com o intuito de considerar a natureza adaptativa individual de moluscos bivalves na construção de biossensores de poluição aquática, a presente dissertação propôs a construção de um módulo ferramenta de análise comportamental baseado na associação de NARNETs e gráfico de controle EWMA.

O comportamento de cada mexilhão foi empregado no treinamento de uma NARNET dedicada responsável por prever sua amplitude de abertura em um ambiente não contaminado com base em seu comportamento ao longo das últimas 6 horas. Nesse sentido, a Figura 46 ilustra a resposta de uma NARNET treinada para prever a AOA de um mexilhão durante os estágios de treinamento, validação e teste com base nos dois primeiros dias de experimento.

Conforme Figura 46, houve dificuldade em prever variações abruptas na amplitude de abertura das conchas bivalves. A Figura 47 ilustra o desempenho da rede neural e a correlação entre seus dados de saída e as amplitudes esperadas. Não houve indicação de *overfitting* já que os erros de treinamento e validação diminuíram até a época destacada.

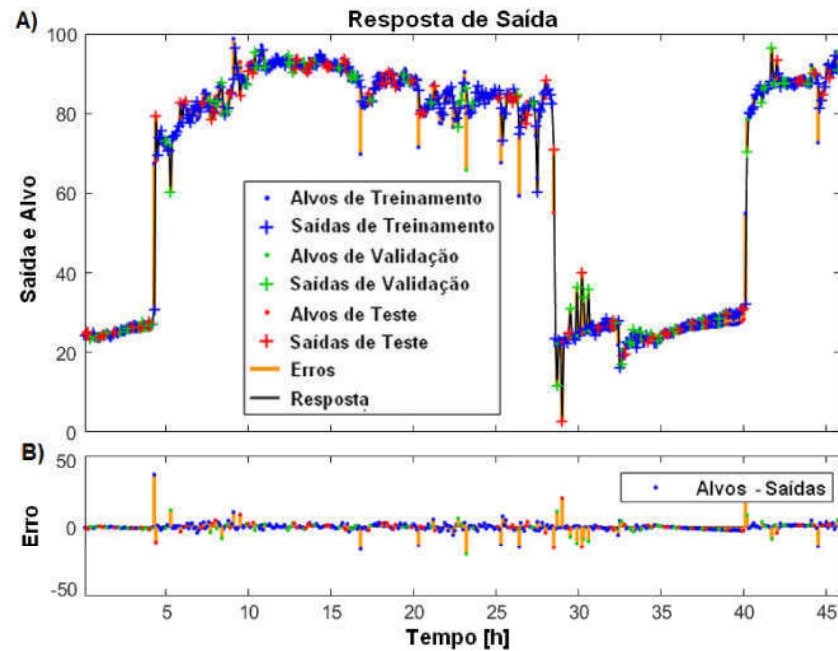


Figura 46: Exemplo de resposta obtida através uma NARNET durante o treinamento, validação e teste (A) e o respectivo erro ao longo do tempo (B).

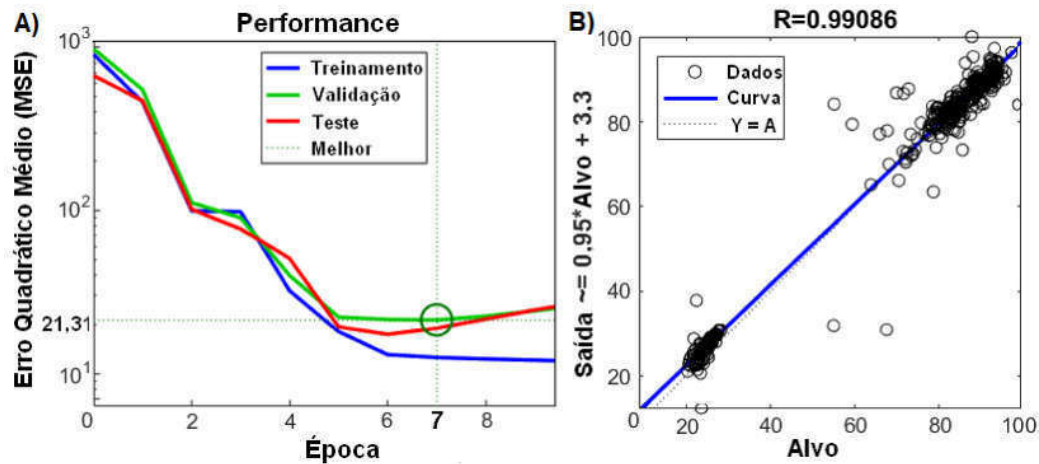


Figura 47: Exemplo de desempenho de uma NARNET ao longo das épocas de treinamento (A) e respectivo gráfico de dispersão entre sua saída e a amplitude média esperada (B). Neste exemplo é possível verificar uma correlação muito forte em um intervalo de confiança de 95% de acordo com o coeficiente de correlação de Pearson (r).

As saídas da rede neural foram, conforme indicado pelo coeficiente de correlação de Pearson (r), fortemente correlacionadas com os dados almejados em sua saída a um nível de significância de 5% (CHRISTMANN; BADGETT, 2009). As ACFs de AOA experimentais e residuais indicaram uma autocorrelação significativa em até, em média, 50 lags (intervalo de 5 horas) e 2 lags, respectivamente. Portanto, o emprego de NARNETs di-

minuiu consideravelmente a autocorrelação residual conforme exemplificado pela Figura 48.

Após o treinamento, validação e teste das NARNETs, elas foram empregadas para prever as amplitudes de abertura durante as próximas 68 horas de experimento. As últimas 44 horas deste período abrangiam a exposição toxicológica dos mexilhões *P. perna* que foram expostos à WAF de diesel. Para esse fim, o comportamento de 48 à 54 horas de experimento foi utilizado para prever a AOA no próximo intervalo de 0,1 hora. O valor previsto foi empregado para prever o próximo e assim sucessivamente. A Tabela 9 resume os valores médios dos coeficientes de correlação de Pearson (r), do desempenho durante o treinamento, validação e teste e dos erros quadráticos médios (MSE) referentes aos períodos exposição não toxicológica e toxicológica.

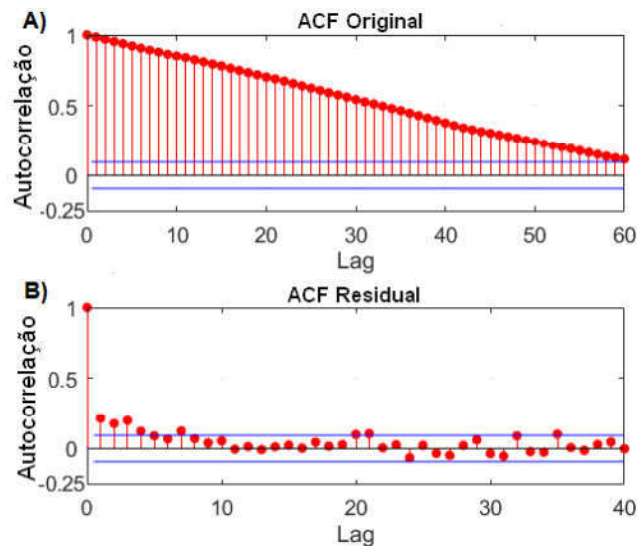


Figura 48: Exemplo de funções de autocorrelação (ACF) para a amplitude média de abertura (A) e para os resíduos decorrentes do uso da NARNET (B) a um nível de significância de 5%.

Com base na Tabela 9, houve um aumento inegável do MSE sob exposição toxicológica para ambas as concentrações de WAF de diesel. Nesse sentido, a análise dos resíduos pareceu adequada para investigar a discrepância entre o comportamento previsto e o comportamento real como biomarcador de poluição aquática. Isto corroborou com a afirmação de que bivalves alteram seu comportamento para diminuir o contato de seus tecidos moles com um ambiente desagradável (TRAN et al., 2003).

Nesse sentido, os resíduos entre o comportamento previsto e o comportamento real a cada intervalo de 0,1 hora foram utilizados como amostras de um gráfico de controle EWMA considerando tanto o grupo de controle quanto os animais expostos à 20% e 5% de WAF de diesel. Os primeiros 144 resíduos (80% das amplitudes médias de abertura referentes ao período de exposição não toxicológica) foram empregados para estabelecer

Tabela 9: Resumo da performance resultante do emprego de NARNETs na previsão da amplitude média de abertura de mexilhões *P. perna* a cada intervalo de 0,1 hora. A performance foi avaliada com base nos valores de acurácia, coeficiente de correlação de Pearson (r) e erro quadrático médio (MSE) referentes aos períodos de exposição não toxicológica e toxicológica.

Dados	Grupo de mexilhões		
	20% de WAF	5% de WAF	Controle
Acurácia de treinamento	9,00	10,22	17,12
Acurácia de validação	69,57	24,35	60,37
Acurácia de teste	83,73	44,45	78,61
r	0,96	0,95	0,89
MSE durante exposição não toxicológica (24 horas)	42,59	17,06	112,77
MSE durante exposição toxicológica (44 horas)	373,15	428,67	-

os limites central, superior e inferior do gráfico de controle (CL, UCL e LCL, respectivamente).

As Figuras 49, 50 e 51 ilustram, portanto, os gráficos de controle EWMA referentes aos grupos de animais não expostos à contaminante e submetidos à 5% e 20% de WAF de diesel S-500, respectivamente. A inserção das WAFs nos aquários ocorreu no intervalo de tempo 181 conforme indicado pelas linhas tracejadas em azul.

Tendo em vista que o grupo de controle não foi exposto a nenhum contaminante, esperava-se que as médias móveis exponencialmente ponderadas iriam se manter dentro dos limites de controle estabelecidos no processo de calibração. Analisando-se, entretanto, a performance geral do modelo e considerando que cada ponto refere-se a intervalos de tempo de 0,1 hora, pode-se afirmar que a ocorrência de falsos positivos foram esporádicas e de curta duração. Considerando como classe positiva a presença de contaminantes, o modelo apresentou uma taxa de falsos positivos de 4,03%.

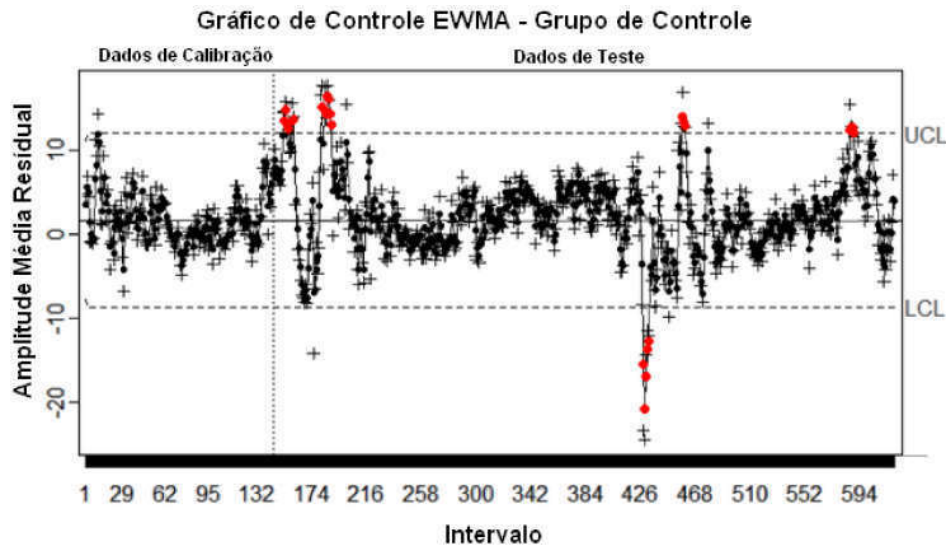


Figura 49: Gráfico de controle EWMA para mexilhões não expostos à WAF de diesel S-500. Dados de calibração foram empregados para estabelecer os limites do gráfico de controle. As cruces representam as amplitude média de abertura e pontos vermelhos e pretos indicam os valores da média móvel exponencialmente ponderada que ultrapassam ou não os limites de controle superior (UCL) e inferior (LDL) a cada intervalo de 0,1 hora, respectivamente.

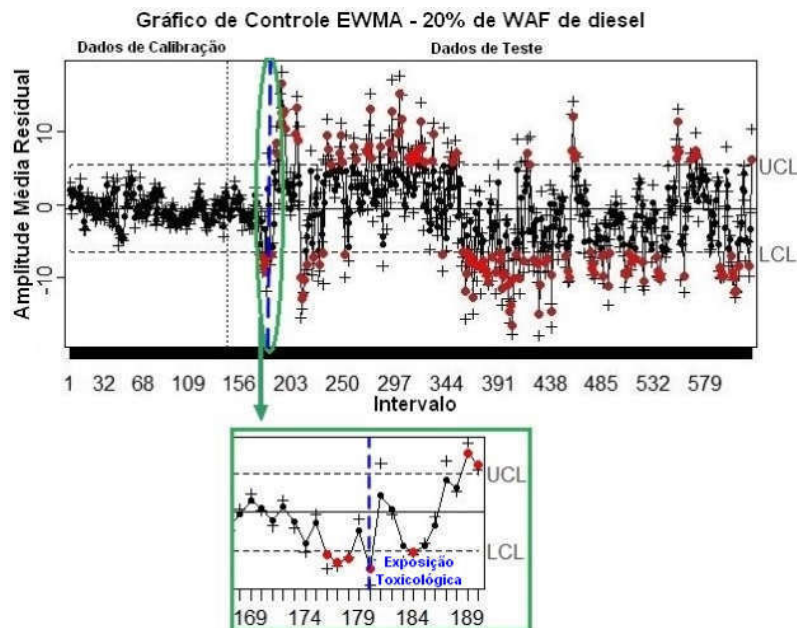


Figura 51: Gráfico de controle EWMA para mexilhões expostos a 20% de WAF de diesel. Dados de calibração foram empregados para estabelecer seus limites. A linha tracejada azul indica o início da exposição toxicológica e as cruces representam as amplitudes médias de abertura. Pontos vermelhos e pretos indicam os valores EWMA que ultrapassaram ou não os limites de controle superior (UCL) e inferior (LDL), respectivamente.

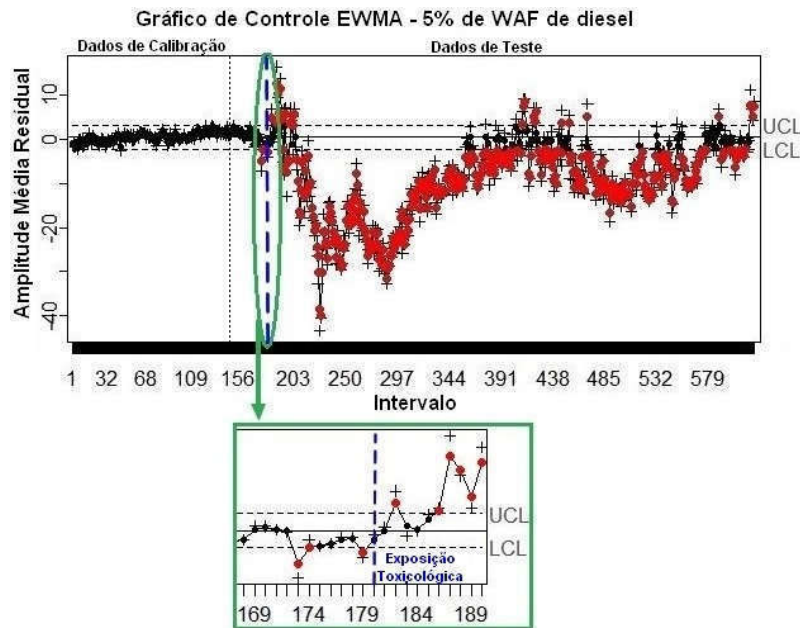


Figura 50: Gráfico de controle EWMA para mexilhões expostos a 5% de WAF de diesel. Dados de calibração foram empregados para estabelecer seus limites. A linha tracejada azul indica o início da exposição toxicológica e as cruzes representam as amplitudes médias de abertura. Pontos vermelhos e pretos indicam os valores EWMA que ultrapassaram ou não os limites de controle superior (UCL) e inferior (LDL), respectivamente.

Ambos gráficos de controle referentes à grupos de animais expostos à contaminante não cruzaram os limites superior ou inferior ao longo do período de calibração. Dessa forma, nenhum resíduo foi significativamente maior ou menor que o esperado durante este período não toxicológico. Durante a exposição toxicológica à ambas as concentrações de WAF de diesel, os gráficos de controle EWMA indicaram mudanças significativas nos AOA residuais e apresentaram pequenos desvios em relação à UCL ou LCL.

Considerando como classe verdadeira positiva a presença de contaminantes no ambiente bem como os períodos de exposição não toxicológica e toxicológica, as matrizes de confusão ilustradas na Figura 52 foram construídas. Nota-se que o desempenho do modelo proposto foi melhor considerando o grupo de animais expostos à 5% de contaminante em relação ao grupo exposto à maior concentração de WAF de diesel. A acurácia, taxa de falsos positivos e sensibilidade para o grupo exposto à 5% de WAF foram de 89,03%, 16,81% e 92,25%, respectivamente. Da mesma forma, nota-se que considerando o grupo de bivalves expostos à 20% de WAF de diesel obtiveram-se acurácia, taxa de falsos positivos e sensibilidade de 66,61%, 12,27% e 55%, respectivamente.

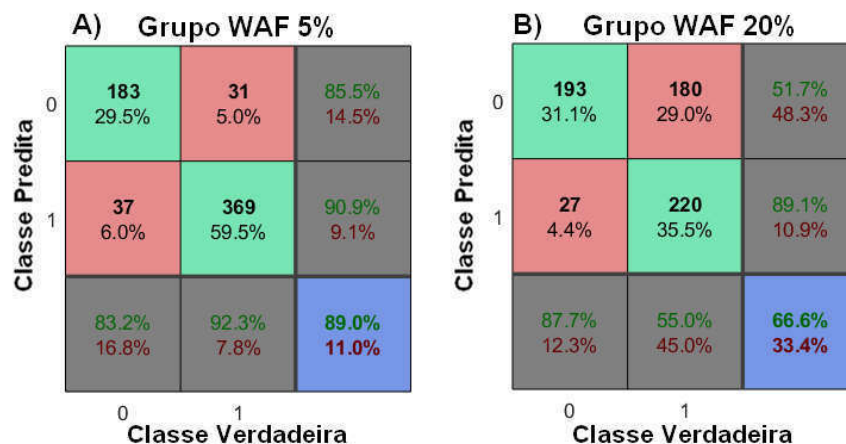


Figura 52: Matrizes de confusão referentes às saídas do modelo baseado na associação de NARNETs e gráfico de controle EWMA para os grupos expostos à 5% (A) e 20% de contaminante.

Como o gráfico de controle da EWMA pode ser definido como uma média ponderada de todas as observações passadas e atuais (MONTGOMERY, 2009), os valores mais distantes dos limites de controle tenderam a estar defasados em relação ao início da exposição toxicológica. Os valores de picos da média móvel exponencialmente ponderada ficaram defasados em relação ao início da exposição toxicológica em aproximadamente 47 (4,7 horas) e 16 (1,6 horas) intervalos para concentrações de 5% e 20%, respectivamente. Isso pode ser explicado porque, embora estes gráficos apresentem um bom desempenho em relação à pequenos desvios, eles demandam alguns intervalos de tempo para responder a variações no processo sob análise (MONTGOMERY, 2009).

Embora a exposição toxicológica tenha permanecido até o final do experimento, os dados de EWMA tenderam a retornar aos limites de controle ao longo do tempo, principalmente quando a exposição à menor concentração de WAF de diesel foi investigada (Figura 50). Em geral, moluscos bivalves sob exposição toxicológica podem adotar um comportamento evasivo em detrimento de suas necessidades nutricionais, a fim de reduzir seu contato com um ambiente hostil (TRAN et al., 2010). Este comportamento está relacionado à alterações em parâmetros comportamentais importantes como AOA e frequência de transição (TRAN et al., 2010; BASTI et al., 2009; TRAN et al., 2003). No entanto, a adoção de um comportamento evasivo pronunciado por longos períodos de tempo é inviável devido às necessidades nutricionais do animal e ao aumento do Ph no interior de sua concha decorrente de uma respiração anaeróbica prolongada.

Para ambas as concentrações de contaminante, houve algumas observações fora dos limites de controle imediatamente antes do início da exposição toxicológica. Como mencionado anteriormente, o desempenho das NARNETs demonstrou certa dificuldade em prever variações abruptas na AOA. Além disso, Wilson, Reuter e Wahl (2005) relatou

uma forte correlação entre a iluminação ambiente e o ângulo de abertura das conchas indicando uma tendência geral de fechamento durante condições de luz ambiente .

A WAF de diesel foi inserida nos aquários no início da manhã e o fotoperíodo do laboratório mudou automaticamente a iluminação de escura para clara às 8 horas da manhã. Portanto, os valores observados podem ter sido causados por uma mudança na iluminação ambiente associada à incapacidade da rede neural de prever valores próximos às regiões de abertura e fechamento abruptos. A Figura 53 ilustra a AOA, a saída da NARNET os resíduos ao longo do tempo para animais expostos a 5% (A) e 20% (B) de WAF de diesel.

A exposição de bivalves a condições ambientais desagradáveis do ambiente forneceu valores de EWMA sistematicamente fora dos limites de controle, em vez de pequenos desvios pontuais devido ao seu comportamento natural. Portanto, o uso dos gráficos de controle EWMA juntamente com NARNETs provou ser uma ferramenta potencial para o monitoramento automático e online de ambientes aquáticos. Além disso, o uso de dados de iluminação ambiente, juntamente com outras variáveis ambientais como entradas exógenas pode auxiliar no desenvolvimento de redes neurais mais eficazes.

Com o intuito de comparar o desempenho de métodos estatísticos tradicionais em relação ao modelo proposto, os métodos estatísticos ANOVA ou Kruskal-Wallis de um fator foram utilizados para identificar diferenças significativas entre os grupos de controle e expostos à WAF de diesel S-500 a cada intervalo de 0,1 hora. As Figuras 54, 55 e 56 ilustram, portanto, as matrizes de confusão resultantes deste emprego ($p < 0,05$) e sua performance longo do tempo considerando o comportamento médio de bivalves expostos à 5% e 20% de WAF, respectivamente.

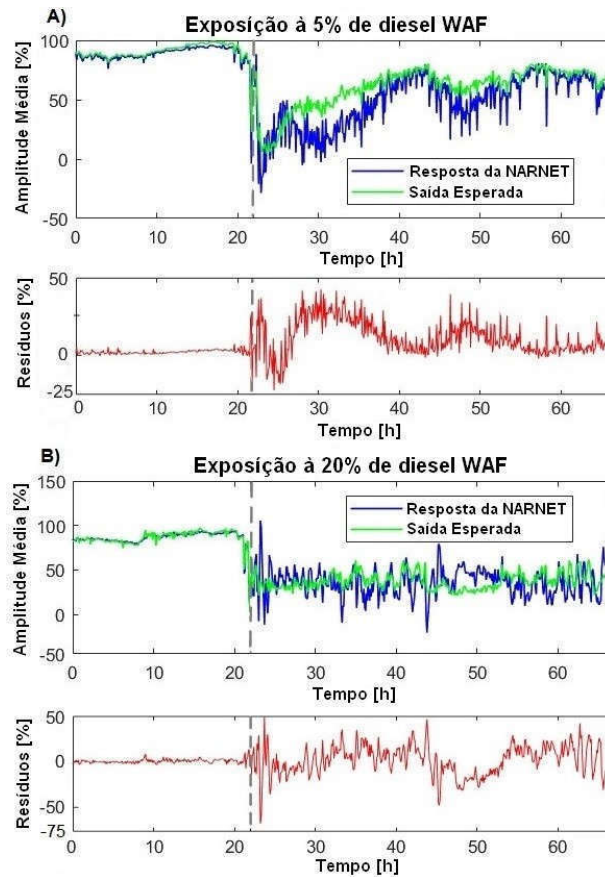


Figura 53: Resposta da NARNET, valores almejados de amplitude de abertura e resíduos ao longo do tempo para os grupos de animais expostos a 5% (A) e 20% (B) de WAF de diesel. Linhas pontilhadas em cinza indicam o início do período de contaminação.

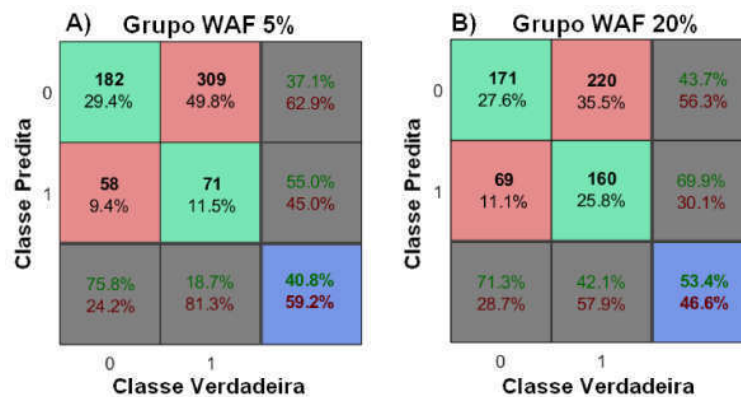


Figura 54: Matrizes de confusão referentes ao emprego dos testes estatísticos na comparação das amplitudes de abertura de mexilhões *Perna perna* expostos à 5% (A) e 20% (B) de WAF de diesel S-500 em relação ao grupo de controle a cada 0,1 hora.

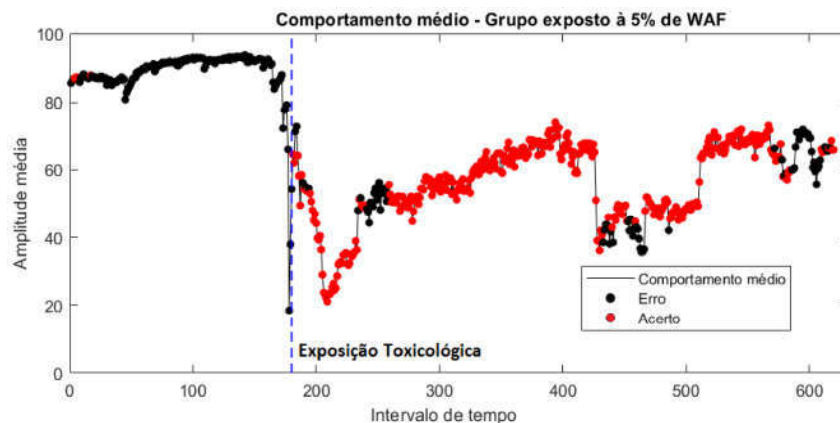


Figura 55: Comportamento médio de moluscos bivalves expostos à 5% de WAF de diesel S-500 a cada intervalo de 0,1 hora. A linha tracejada em azul indica o início da exposição toxicológica na qual a WAF de diesel foi diluída nos aquários. Pontos em vermelho e preto identificam diferenças significativas erroneamente e corretamente identificadas a um nível de significância de 5%, respectivamente.

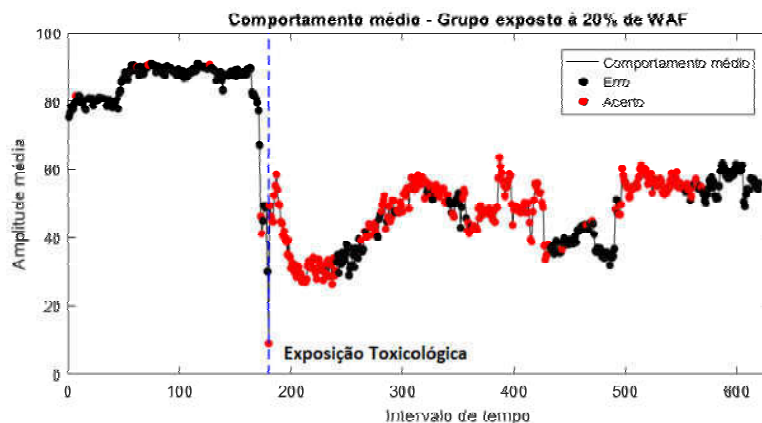


Figura 56: Comportamento médio de moluscos bivalves expostos à 20% de WAF de diesel S-500 a cada intervalo de 0,1 hora. A linha tracejada em azul indica o início da exposição toxicológica na qual a WAF de diesel foi diluída nos aquários. Pontos em vermelho e preto identificam diferenças significativas erroneamente e corretamente identificadas a um nível de significância de 5%, respectivamente.

Comparando-se o comportamento do grupo de animais exposto à 20% de WAF de diesel em relação ao grupo de controle através de métodos estatísticos, obteve-se uma acurácia, taxa de falsos positivos e sensibilidade de 53,38%, 28,75% e 42,11%, respectivamente. Da mesma forma, a comparação do comportamento do grupo de animais exposto à 5% de WAF de diesel em relação ao grupo de controle por meio de métodos estatísticos tradicionais forneceu acurácia, taxa de falsos positivos e sensibilidade de 40,81%, 24,17% e 18,68%, respectivamente.

O emprego do modelo proposto como ferramenta de análise comportamental apresen-

tou maior acurácia e sensibilidade em relação ao uso de técnicas estatísticas tradicionais. Nota-se, portanto, que o modelo proposto como ferramenta de análise comportamental baseado na associação de NARNETs e gráfico de controle EWMA tem potencial aplicabilidade na construção de biossensores de poluição aquática.

7 CONCLUSÃO

A presente dissertação de mestrado teve como objetivo geral propor modelos computacionais de inteligência artificial e visão para construir biossensores de poluição aquática baseados na análise comportamental de mexilhões *Perna perna*. Tendo em vista os módulos constituintes destas soluções, um modelo baseado em visão computacional como técnica de valvometria e sua validação através de um comparativo com sensores de efeito Hall foi apresentado. Da mesma forma, três modelos baseados em grupo, na combinação de classificadores e na associação de NARNETs e gráfico de controle EWMA foram investigados como ferramentas de análise comportamental.

O modelo baseado em visão como técnica de valvometria demonstrou empregabilidade satisfatória tendo em vista sua comparação com sensores de efeito Hall. Nesse sentido, um módulo técnica de valvometria não invasivo foi proposto e validado e pode contribuir para o projeto de novos modelos através de melhorias e adaptações.

O experimento toxicológico proposto para a validação dos modelos de ferramentas de análise demonstrou a complexidade da avaliação comportamental de mexilhões *Perna perna* devido à respostas individuais assíncronas durante tanto o período de exposição toxicológica quanto não toxicológica. Dessa forma, o comportamento individual dos bivalves, principalmente em ambiente contaminado, diferiu e demonstrou a plasticidade fenotípica para uma tentativa individual de diminuir seu contato com meios desagradáveis (DINGEMANSE et al., 2010).

Ambos modelos computacionais baseados em classificadores para inferir o *status* de contaminação ambiente apresentaram resultados satisfatórios (acurácias maiores que 90%) e capacidade de lidar com o comportamento de animais não empregados nos processos de treinamento. Ademais, o modelo computacional baseado na associação de NARNETs e gráfico de controle EWMA foi capaz de considerar a natureza adaptativa dos moluscos bivalves.

Tanto a associação de NARNETs e gráfico de controle EWMA quanto o classificador baseado em grupo apresentaram vantagens na identificação de WAF de diesel S-500 em relação ao emprego de métodos estatísticos tradicionais. Dessa forma, a presente dissertação de mestrado contribuiu com o desenvolvimento e validação de métodos com-

putacionais relevantes e com potencial empregabilidade no monitoramento de ambientes aquáticos.

8 PUBLICAÇÕES

Algumas publicações foram realizadas ao longo do desenvolvimento da presente dissertação de mestrado. Dessa forma, a seguir encontram-se relacionadas a publicações realizadas e os respectivos *qualis* em ciência da computação com base no último quadriênio.

- GUTERRES, Bruna et al. Feasibility of visual signals on the construction of biosensors based on behavioral analysis of *Perna perna* mussels. *Ecological Informatics*, p. 101118, 2020. Qualis A2 - Apresenta o modelo baseado em visão computacional como técnica de valvometria não invasiva e sua comparação com sensores de efeito Hall;
- GUTERRES, Bruna et al. Mussels as Aquatic Pollution Biosensors using Neural Networks and Control Charts. *IEEE International Conference on Industrial Informatics (INDIN)*. Qualis B1 - Apresenta a associação de NARNETs e gráfico de controle EWMA como modelo computacional na detecção de poluentes e o potencial emprego da luminosidade ambiente na melhoria da modelagem comportamental de moluscos bivalves. Neste artigo a comparação com métodos estatísticos tradicionais não é apresentada;
- GUTERRES, Bruna et al. Intelligent Classifiers on the Construction of Pollution Biosensors Based on Bivalves Behavior. *Brazilian Conference on Intelligent Systems (BRACIS)*. Qualis B2 - Apresenta o modelo baseado na combinação de classificadores como ferramenta de análise comportamental capaz de inferir o *status* de contaminação ambiente. Neste artigo a comparação com métodos estatísticos tradicionais não é apresentada.

REFERÊNCIAS

ALI AMIN ALI, A. U. M. H. J. A. K. Hardware and software components of an oysters' gape measurement system. In: IEEE. *American Journal of Engineering and Applied Sciences*. [S.l.], 2019. p. 319–328.

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.

ANDREWARTHA, S. et al. Aquaculture sentinels: smart-farming with biosensor equipped stock. *J. Aquac. Res. Dev*, v. 7, p. 1–4, 2015.

AQUADECT. *MOSELMONITOR*. 2020. Disponível em: <http://www.mosselmonitor.nl/>.

BAE, M.-J.; PARK, Y.-S. Biological early warning system based on the responses of aquatic organisms to disturbances: a review. *Science of the Total Environment*, Elsevier, v. 466, p. 635–649, 2014.

BAMBER, S. D.; WESTERLUND, S. Behavioral responses of arctica islandica (bivalvia: Arcticidae) to simulated leakages of carbon dioxide from sub-sea geological storage. *Aquatic Toxicology*, Elsevier, v. 180, p. 295–305, 2016.

BARELLI, F. *Introdução à Visão Computacional: Uma abordagem prática com Python e OpenCV*. Casa do Código, 2018. ISBN 9788594188588. Disponível em: <https://books.google.com.br/books?id=CA5ZDwAAQBAJ>.

BASTI, L. et al. Effects of the toxic dinoflagellate heterocapsa circularisquama on the valve movement behaviour of the manila clam ruditapes philippinarum. *Aquaculture*, Elsevier, v. 291, n. 1-2, p. 41–47, 2009.

BENMOUIZA, K.; CHEKNANE, A. Forecasting hourly global solar radiation using hybrid k-means and nonlinear autoregressive neural network models. *Energy Conversion and Management*, Elsevier, v. 75, p. 561–569, 2013.

BLACK, K. *Business for Contemporary decision-making Statistics*. 2008.

BLACK, K. *Business Statistics: Contemporary Decision Making*. John Wiley & Sons, 2009. (Wiley Plus Products Series). ISBN 9780470409015. Disponível em: <https://books.google.com.br/books?id=KQ25WExx5usC>.

- BRADSKI, G.; KAEHLER, A. *Learning OpenCV: Computer Vision with the OpenCV Library*. O'Reilly Media, 2008. ISBN 9780596554040. Disponível em: <https://books.google.com.br/books?id=seAgiOfu2EIC>).
- BUSSAB, W. d. O. *Análise de variância e regressão, 2a. edição*. [S.l.]: Editora Saraiva, 1988.
- CANELAS, A. d. S. Evolução da importância econômica da indústria de petróleo e gás natural no Brasil: contribuição a variáveis macroeconômicas. *Rio de Janeiro (RJ): Universidade Federal do Rio de Janeiro. Coordenação dos Programas de Pós-Graduação de Engenharia*, 2007.
- CHICCO, D. Ten quick tips for machine learning in computational biology. *BioData mining*, Springer, v. 10, n. 1, p. 35, 2017.
- CHRISTMANN, E.; BADGETT, J. *Interpreting Assessment Data: Statistical Techniques You Can Use*. NSTA Press, 2009. ISBN 9781933531366. Disponível em: <https://books.google.com.br/books?id=3IUyYvGQsdYC>).
- CLEMENTS, J. C. et al. Behavioural responses to predators in mediterranean mussels (*mytilus galloprovincialis*) are unaffected by elevated pCO₂. *Marine Environmental Research*, Elsevier, v. 161, p. 105148, 2020.
- COCKBURN, I. M.; HENDERSON, R.; STERN, S. *The impact of artificial intelligence on innovation*. [S.l.], 2018.
- COMEAU, L. A. et al. Valve-gaping behavior of raft-cultivated mussels in the ria de Arousa, Spain. *Aquaculture Reports*, Elsevier, v. 9, p. 68–73, 2018.
- COMEAU, L. A. et al. Psp-producing dinoflagellate *Alexandrium minutum* induces valve microclosures in the mussel *mytilus galloprovincialis*. *Aquaculture*, Elsevier, v. 500, p. 407–413, 2019.
- COQUEREAU, L. et al. Short-term behavioural responses of the great scallop *Pecten maximus* exposed to the toxic alga *Alexandrium minutum* measured by accelerometry and passive acoustics. *PloS one*, Public Library of Science San Francisco, CA USA, v. 11, n. 8, p. e0160935, 2016.
- CORRÊA¹, J. H. G. et al. Um modelo simples e parametrizável para classificação de cores no sistema rgb. 2014.
- 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.
- COUTO, M. R. M. *Análise de variância multivariada aplicada a dados com medidas repetidas*. Universidade Federal de Santa Maria, 2006.
- DAVIS, J.; GOADRICH, M. The relationship between precision-recall and roc curves. In: *Proceedings of the 23rd international conference on Machine learning*. [S.l.: s.n.], 2006. p. 233–240.
- DAWSON-HOWE, K. *A Practical Introduction to Computer Vision with OpenCV*. Wiley, 2014. ISBN 9781118848739. Disponível em: <https://books.google.com.br/books?id=F9MsAwAAQBAJ>).

- DEMÉTRIO, B. G. *RAnálise multidimensional para dados de cana-de-açúcar*. Tese (Doutorado Estatística e Experimentação Agrônômica) — Unidersidade de São Paulo, Piracicaba - SP, 1995.
- DINGEMANSE, N. J. et al. Behavioural reaction norms: animal personality meets individual plasticity. *Trends in ecology & evolution*, Elsevier, v. 25, n. 2, p. 81–89, 2010.
- DORST, B. V. et al. Recent advances in recognition elements of food and environmental biosensors: a review. *Biosensors and Bioelectronics*, Elsevier, v. 26, n. 4, p. 1178–1194, 2010.
- FACHADA, N. et al. micompr: An r package for multivariate independent comparison of observations. *arXiv preprint arXiv:1603.06907*, 2016.
- FORSYTH, D.; PONCE, J. *Computer Vision: A Modern Approach*. Prentice Hall, 2003. (Alan R. Apt book). ISBN 9780131911932. Disponível em: <https://books.google.com.br/books?id=a3-TQgAACAAJ>.
- FOURNIER, E. et al. Valve closure response to uranium exposure for a freshwater bivalve (corbicula fluminea): Quantification of the influence of ph. *Environmental Toxicology and Chemistry: An International Journal*, Wiley Online Library, v. 23, n. 5, p. 1108–1114, 2004.
- FUJII, T. Measurement of periodic open and shut shell movement of bivalves by the strain-gauge method. *bull. Jap. Soc. Sci. Fish.*, v. 43, p. 901, 1977.
- FUJII, T. The study for periodic behaviour of bivalves, 1: Periodicity observed in short-necked clam tapes japonica deshyes put in natural environment. *Bulletin of Tohoku Regional Fisheries Research Laboratory*, 1979.
- FURLAN, É. et al. *Qualidade e Processamento de Pescado*. [S.l.]: Elsevier Brasil, 2015.
- GAVRILESCU, M. et al. Emerging pollutants in the environment: present and future challenges in biomonitoring, ecological risks and bioremediation. *New biotechnology*, Elsevier, v. 32, n. 1, p. 147–156, 2015.
- GIVENS, G.; HOETING, J. *Computational Statistics*. Wiley, 2012. (Wiley Series in Computational Statistics). ISBN 9780470533314. Disponível em: <https://books.google.com.br/books?id=bCJx53VQS7IC>.
- GREKOV, A. N.; KUZMIN, K. A.; MISHUROV, V. Z. Automated early warning system for water environment based on behavioral reactions of bivalves. In: *IEEE. 2019 International Russian Automation Conference (RusAutoCon)*. [S.l.], 2019. p. 1–5.
- HABERKORN, H. et al. Relationship between valve activity, microalgae concentration in the water and toxin accumulation in the digestive gland of the pacific oyster crassostrea gigas exposed to alexandrium minutum. *Marine pollution bulletin*, Elsevier, v. 62, n. 6, p. 1191–1197, 2011.
- HARTMANN, J. T. et al. Establishing mussel behavior as a biomarker in ecotoxicology. *Aquatic Toxicology*, Elsevier, v. 170, p. 279–288, 2016.

HAYKIN, S. *Redes Neurais: Princípios e Prática*. Artmed, 2007. ISBN 9788577800865. Disponível em: <https://books.google.com.br/books?id=bhMwDwAAQBAJ>).

HENRIQUES, M. B. *RESISTÊNCIA DO MEXILHÃO Perna perna (LINNAEUS, 1758) PROVENIENTE DE BANCOS NATURAIS DA BAIXADA SANTISTA, A VARIAÇÕES DE TEMPERATURA, SALINIDADE, TEMPO DE EXPOSIÇÃO AO AR E DETERMINAÇÃO DA INCIDÊNCIA DE PARASITISMO*. Tese (Doutorado em Ciências Biológicas) — Universidade Estadual Paulista “Julio de Mesquita Filho”, Rio Claro - SP, 2004.

HERCULANO-HOUZEL, S. The human brain in numbers: a linearly scaled-up primate brain. *Frontiers in human neuroscience*, Frontiers, v. 3, p. 31, 2009.

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.

HUMOS, A. A.-E. et al. Real time environmental/biological monitoring system. In: IEEE. *2019 International Conference on Computational Science and Computational Intelligence (CSCI)*. [S.l.], 2019. p. 1179–1183.

IBRAHIM, M. et al. Nonlinear autoregressive neural network in an energy management strategy for battery/ultra-capacitor hybrid electrical vehicles. *Electric Power Systems Research*, Elsevier, v. 136, p. 262–269, 2016.

JIN, X. et al. Artificial intelligence biosensors: Challenges and prospects. *Biosensors and Bioelectronics*, Elsevier, p. 112412, 2020.

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.

JR, C. R.; BRASIL, E. d. S.; SALOMÃO, L. C. The effect of temperature and salinity on the physiological rates of the mussel *perna perna* (linnaeus 1758). *Brazilian archives of biology and technology*, SciELO Brasil, v. 50, n. 3, p. 543–556, 2007.

KRAMER, K. J.; JENNER, H. A.; ZWART, D. de. The valve movement response of mussels: a tool in biological monitoring. *Hydrobiologia*, Springer, v. 188, n. 1, p. 433–443, 1989.

KUWATANI, Y. Effect of photo-illumination on rhythmical shell movement of pearl oyster, *pinctada martensii* (dunker). *Bull. Jpn. Soc. Fish*, v. 29, p. 1064–1070, 1963.

LASSOUED, J. et al. Behavioural and eco-physiological responses of the mussel *mytilus galloprovincialis* to acidification and distinct feeding regimes. *Marine Ecology Progress Series*, v. 626, p. 97–108, 2019.

LDA, R. F. Q. *O que são biomarcadores?* 2020. <https://www.roche.pt/corporate/index.cfm/farmacutica/ensaios-clinicos-profissionais-de-saude/biomarcadores/o-que-sao-biomarcadores/>. [Online; accessed 23-July-2020].

LEME, D. M.; MARIN-MORALES, M. A. Avaliação da qualidade de águas impactadas por petróleo por meio de sistema-teste biológico (allium cepa)-um estudo de caso. In: *IN: 4 Congresso Brasileiro de Pesquisa e Desenvolvimento em Petróleo e Gas, Campinas-SP*. [S.l.: s.n.], 2007.

LIAO, C.-M. et al. Valve movement response of the freshwater clam *Corbicula fluminea* following exposure to waterborne arsenic. *Ecotoxicology*, Springer, v. 18, n. 5, p. 567–576, 2009.

LÜCHMANN, K. H. et al. Biochemical biomarkers and hydrocarbons concentrations in the mangrove oyster *Crassostrea brasiliana* following exposure to diesel fuel water-accommodated fraction. *Aquatic toxicology*, Elsevier, v. 105, n. 3-4, p. 652–660, 2011.

LUMMER, E.-M.; AUERSWALD, K.; GEIST, J. Fine sediment as environmental stressor affecting freshwater mussel behavior and ecosystem services. *Science of the Total Environment*, Elsevier, v. 571, p. 1340–1348, 2016.

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.

MARKICH, S. Behavioural responses of the tropical freshwater bivalve *Velesunio angasi* exposed to uranium. *Wetland Research in the Wet–Dry Tropics of Australia. Supervising Scientist. Canberra, Australia*, p. 247–257, 1995.

MARQUES, H. *Criação comercial de mexilhões*. Nobel, 1997. ISBN 9788521309253. Disponível em: <https://books.google.com.br/books?id=ymrhVYbsUp8C>.

MARTINO, M. M. F. D.; NETO, J. C. Variabilidade circadiana da temperatura oral e do ciclo vigília-sono em enfermeiras de diferentes turnos de trabalho. *Revista de Ciências Médicas*, v. 10, n. 3, 2012.

MASSABUAU, J.-C. et al. Environmental monitoring of arctic waters with unmanned bivalve biosensor technology: one year of background data acquisition in the barents sea. In: SOCIETY OF PETROLEUM ENGINEERS. *SPE Russian Petroleum Technology Conference*. [S.l.], 2015.

MATHWORKS. *Single Camera Calibrator App*. 2019. <https://www.mathworks.com/help/vision/ug/single-camera-calibrator-app.html>. Acessado em: 20 set. 2019.

MCARDLE, B. H.; ANDERSON, M. J. Fitting multivariate models to community data: a comment on distance-based redundancy analysis. *Ecology*, Wiley Online Library, v. 82, n. 1, p. 290–297, 2001.

MOLLUSCAN. *MolluSCAN EYE*. 2020. Disponível em: <https://molluscan-eye.epoc.u-bordeaux.fr/>.

MONTGOMERY, D. *Introduction to Statistical Quality Control*. Wiley, 2009. ISBN 9780470233979. Disponível em: <https://books.google.com.br/books?id=oG1xPgAACAAJ>.

- MOULLAC, G. L. et al. Non-invasive functional exploration techniques for bivalves with applications to pearl oyster *pinctada margaritifera*. *Reviews in Aquaculture*, Wiley Online Library, v. 12, n. 3, p. 1783–1791, 2020.
- 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.
- NARCHI, W.; GALVÃO-BUENO, M. S. Anatomia funcional de perna perna (11nné) (bivalvia, mytilidae). *Revista Brasileira de Zoologia*, v. 14, p. 135–168, 1997.
- NEFF, J. M. *Bioaccumulation in marine organisms: effect of contaminants from oil well produced water*. [S.l.]: Elsevier, 2002.
- NEWTON, T. J.; COPE, W. G. 10 biomarker responses of unionid mussels to environmental contaminants. 2007.
- NOGUEIRA, L. et al. Oxidative stress in nile tilapia (*oreochromis niloticus*) and armored catfish (*pterygoplichthys anisitsi*) exposed to diesel oil. *Environmental monitoring and assessment*, Springer, v. 180, n. 1-4, p. 243–255, 2011.
- NYCE, D. S. *Linear position sensors: theory and application*. [S.l.]: John Wiley & Sons, 2004.
- OOST, R. Van der; BEYER, J.; VERMEULEN, N. P. Fish bioaccumulation and biomarkers in environmental risk assessment: a review. *Environmental toxicology and pharmacology*, Elsevier, v. 13, n. 2, p. 57–149, 2003.
- OSHIRO, T. M. *Uma abordagem para a construção de uma única árvore a partir de uma Random Forest para classificação de bases de expressão gênica*. Tese (Dissertação) — Universidade de São Paulo, 2013.
- OTT, R. L.; LONGNECKER, M. T. *An introduction to statistical methods and data analysis*. [S.l.]: Nelson Education, 2015.
- PAULA, M. B. de. *Indução automática de árvores de decisão*. Tese (Dissertação) — Universidade Federal de Santa Catarina, 2002. Disponível em: <https://repositorio.ufsc.br/bitstream/handle/123456789/82465/189971.pdf?sequence=1>).
- PHILLIS, D. *Quantitative aquatic biological indicators. Pollution monitoring series*. [S.l.]: London: Applied Science, 1980.
- PORTER, E. T.; BREITBURG, D. L. Eastern oyster, *crassostrea virginica*, valve gape behavior under diel-cycling hypoxia. *Marine biology*, Springer, v. 163, n. 10, p. 1–12, 2016.
- PUBLISHERS, E. *Estatística Basica, 6a Edicao, O. Bussab & A. Morettin, 2010: Estatística Basica.* [s.n.], 2010. (Estatística Basica,). Disponível em: <https://books.google.com.br/books?id=FK1mDwAAQBAJ>).
- RAMSDEN, E. *Hall-effect sensors: theory and application*. [S.l.]: Elsevier, 2011.
- REIS, E. *Estatística multivariada aplicada. Edições Sílabo*, 2001.

- SANTOS, E. M. dos. *Teoria e Aplicação de Support Vector Machines à Aprendizagem e Reconhecimento de Objetos Baseado na Aparência*. Tese (Dissertação) — Universidade Federal da Paraíba, 2002. Disponível em: http://docs.computacao.ufcg.edu.br/posgraduacao/dissertacoes/2002/Dissertacao_EulandaMirandadosSantos.pdf.
- SERVICE, R. F. *'The game has changed.'* *AI triumphs at protein folding*. [S.l.]: American Association for the Advancement of Science, 2020.
- SHIRMOHAMMADI, S.; FERRERO, A. Camera as the instrument: the rising trend of vision based measurement. *IEEE Instrumentation & Measurement Magazine*, IEEE, v. 17, n. 3, p. 41–47, 2014.
- SILVA, F. C. da. *Análise ROC*. 2006. https://www.cin.ufpe.br/~jead/bioInformatica/roc_analyzes.pdf.
- SILVERMAN, B. *Density Estimation for Statistics and Data Analysis*. Taylor & Francis, 1986. (Chapman & Hall/CRC Monographs on Statistics & Applied Probability). ISBN 9780412246203. Disponível em: <https://books.google.com.br/books?id=e-xsrjsL7WkC>.
- SINGER, M. et al. Standardization of the preparation and quantitation of water-accommodated fractions of petroleum for toxicity testing. *Marine Pollution Bulletin*, Elsevier, v. 40, n. 11, p. 1007–1016, 2000.
- SIREGAR, S. et al. Statistical methods to monitor risk factors in a clinical database: example of a national cardiac surgery registry. *Circulation: Cardiovascular Quality and Outcomes*, Am Heart Assoc, v. 6, n. 1, p. 110–118, 2013.
- SOW, M. et al. Water quality assessment by means of hfni valvometry and high-frequency data modeling. *Environmental monitoring and assessment*, Springer, v. 182, n. 1-4, p. 155–170, 2011.
- STARODUB, N. et al. Biosensors for water quality monitoring. In: *Modern Tools and Methods of Water Treatment for Improving Living Standards*. [S.l.]: Springer, 2005. p. 51–70.
- SUN, J. *The Microbiome in Health and Disease*. Elsevier Science, 2020. (ISSN). ISBN 9780128200018. Disponível em: <https://books.google.com.br/books?id=kiToDwAAQBAJ>.
- TAGUE, N. R. *The Quality Toolbox*. [S.l.]: ASQ Quality Press, 2005. ISBN 9780873896399.
- TIERNEY, K. B.; SHAMCHUK, A. L. Phenotyping stimulus evoked responses in larval zebrafish. *Behaviour*, Brill, v. 149, n. 10-12, p. 1177–1203, 2012.
- TRAN, D.; BOUDOU, A.; MASSABUAU, J. Mechanism of oxygen consumption maintenance under varying levels of oxygenation in the freshwater clam corbicula fluminea. *Canadian Journal of Zoology*, v. 78, p. 2027–2036, 2000.
- TRAN, D. et al. Modulation of uranium bioaccumulation by hypoxia in the freshwater clam corbicula fluminea: induction of multixenobiotic resistance protein and heat shock protein 60 in gill tissues. *Environmental Toxicology and Chemistry: An International Journal*, Wiley Online Library, v. 24, n. 9, p. 2278–2284, 2005.

TRAN, D. et al. Estimation of potential and limits of bivalve closure response to detect contaminants: application to cadmium. *Environmental Toxicology and Chemistry*, Wiley Online Library, v. 22, n. 4, p. 914–920, 2003.

TRAN, D. et al. The toxic dinoflagellate alexandrium minutum disrupts daily rhythmic activities at gene transcription, physiological and behavioral levels in the oyster crassostrea gigas. *Aquatic Toxicology*, Elsevier, v. 158, p. 41–49, 2015.

TRAN, D. et al. Behavioral responses of crassostrea gigas exposed to the harmful algae alexandrium minutum. *Aquaculture*, Elsevier, v. 298, n. 3-4, p. 338–345, 2010.

TRAN, D. et al. Bivalve mollusc circadian clock genes can run at tidal frequency. *Proceedings of the Royal Society B*, The Royal Society, v. 287, n. 1918, p. 20192440, 2020.

UNEP CHEMICALS. *Guidance for a Global Monitoring Programme for Persistent Organic Pollutants (POPs)*. 2004. Geneva, Suíça.

VERNBERG, W. et al. *Marine pollution: functional responses*. Academic Press, 1979. ISBN 9780127182605. Disponível em: <https://books.google.com.br/books?id=j2zwAAAAMAAJ>.

WILSON, R.; REUTER, P.; WAHL, M. Muscling in on mussels: new insights into bivalve behaviour using vertebrate remote-sensing technology. *Marine Biology*, Springer, v. 147, n. 5, p. 1165–1172, 2005.

APÊNDICES

A construção de biossensores de poluição aquática baseados na análise comportamental de molusco bivalves através de métodos computacionais de visão e inteligência artificial engloba uma temática abrangente e multidisciplinar. Nesse sentido, os apêndices a seguir englobam assuntos pertinentes ao entendimento da construção destes biossensores.

Os apêndices A e B apresentam de forma breve informações necessárias à compreensão de moluscos bivalves bem como o funcionamento de sensores de efeito Hall que constituem uma das principais técnicas de valvometria empregadas na monitoração de bivalves. Da mesma forma, o apêndice C apresenta a fração acomodada em água como meio de exposição padronizado capaz de tornar os experimentos processos reprodutíveis e comparáveis com outros experimentos de mesma natureza.

A Moluscos Bivalves

De acordo com HENRIQUES (2004), moluscos são animais invertebrados de corpo não articulado que possuem simetria bilateral. Os moluscos da classe bivalvia (moluscos bivalves) vivem exclusivamente em água, possuem uma concha formada por duas valvas unidas dorsalmente por um ligamento e geralmente apresentam sexos separados.

Os moluscos bivalves foram indicados como possíveis animais em programas de biomonitoramento de contaminantes marinhos em 2004 pelo Programa de Meio Ambiente das Nações Unidas (*United Nations Environment Programme - UNEP*) (UNEP CHEMICALS, 2004). Como esses animais apresentam baixa mobilidade, hábito filtrador, capacidade de acumular substâncias em seus tecidos e ampla distribuição geográfica bem como alta densidade populacional, eles são apontados como organismos de importante interesse pelos pesquisadores em estudos ecotoxicológicos (VERNBERG et al., 1979).

A.1 Mexilhões *Perna perna*

O termo mexilhão é comumente utilizado na denominação de diversas espécies de moluscos bivalves pertencente à família Mytilidae, sendo comumente mais empregado para se referir aquelas empregadas em larga escala como fonte de proteína animal de baixo custo e alto valor nutricional (HENRIQUES, 2004). A criação de mexilhões (mitilicultura) é uma das atividades mais produtivas e prósperas da aquicultura em diversas regiões do mundo, inclusive no Brasil, alcançando até 30 t de carne por ha/ano (FURLAN et al., 2015).

Segundo Furlan et al. (2015) dentre as espécies de interesse comercial que ocorrem no Brasil, a espécie *Perna perna* foi apontada pela Organização das Nações Unidas para a Alimentação e a Agricultura (*Food and Agriculture Organization of the United Nations - FAO*) como uma das espécies promissoras dentre as demais emergentes na aquicultura mundial. Esses mexilhões são os maiores dos mitilídeos brasileiros possuindo um tamanho médio de 5-8 cm e comprimento de 3-4 cm (NARCHI; GALVÃO-BUENO, 1997). Algumas das estruturas constituintes desses animais encontram-se ilustradas através da Figura 57.

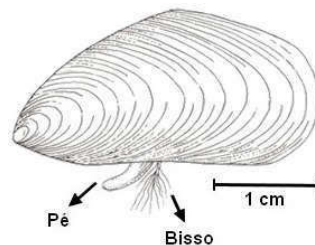


Figura 57: Algumas estruturas de mexilhões *Perna perna* (NARCHI; GALVÃO-BUENO, 1997, adaptado)

O pé, de acordo com Marques (1997), vem a ser o órgão de locomoção, por raspejamento, dos animais sendo principalmente utilizado em casos de desprendimento do animal do substrato original. Ademais, ele auxilia na alimentação retirando partículas aderidas sobre o substrato e inserindo-os em sua cavidade denominada manto. O bisso, por sua vez, corresponde ao órgão de fixação do animal sendo constituído por um conjunto de fibras proteicas de alta resistência secretadas por glândulas especiais localizadas na base do pé. Os mexilhões possuem um regime alimentar exclusivamente filtrador, ou seja, retiram seu alimento da água utilizada no processo de respiração uma vez que suas lâminas branquiais absorvem oxigênio e atuam na seleção de partículas alimentares constituídas por algas microscópicas (MARQUES, 1997).

Mexilhões *Perna perna* são encontrados desde a Venezuela até o Uruguai, sendo muito abundantes entre o Rio de Janeiro e Santa Catarina (HENRIQUES, 2004), vivem presos pelo bisso a substratos duros da zona interdital são mais abundantes em costões rochosos expostos à ação das ondas. Como vivem principalmente na região entre marés esses animais estão adaptados a permanecer por longos períodos expostos ao ar e ao sol (HENRIQUES, 2004).

B Sensores de efeito Hall

O efeito responsável pelo surgimento dos sensores de efeito Hall consiste no surgimento de uma tensão potencial através de um condutor percorrido por corrente quando submetido a um campo magnético (NYCE, 2004) conforme exemplificado pela a Figura 58. A tensão de efeito Hall apresenta amplitude proporcional à corrente que circula pelo elemento sensor e à intensidade do campo magnético ao qual está submetido e possui direção perpendicular à ambos (RAMSDEN, 2011). Desta forma, tem-se um dispositivo capaz de medir a intensidade e a polaridade de campos magnéticos estáticos e dinâmicos (NYCE, 2004).

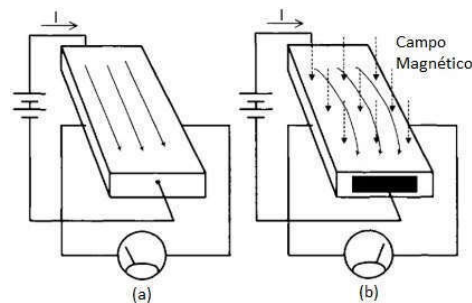


Figura 58: Efeito Hall em uma folha condutora percorrida por uma corrente elétrica (I). Ponteiros de prova de um voltímetro são conectadas com orientação transversal ao sentido da corrente medindo uma tensão induzida não nula quando sujeita a um campo magnético (RAMSDEN, 2011, Adaptado).

Estes dispositivos podem ser aplicados como sensores/transdutores de posição nos quais mudanças na posição de um ímã preso a um alvo resulta em alterações na densidade de fluxo magnético na localização do elemento sensor conforme ilustrado através da Figura 59. Segundo Nyce (2004), nessa configuração a intensidade de campo magnético no elemento sensor varia com o quadrado da distância em relação ao ímã.

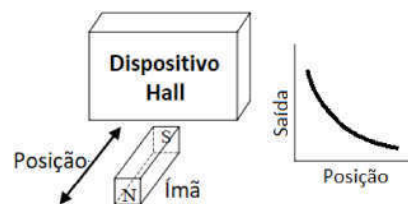


Figura 59: Sensor de posição baseado em dispositivo Hall (NYCE, 2004)[p. 113] (Adaptado)

C Fração Acomodada em Água (WAF)

De acordo com Leme e Marin-Morales (2007), a importância da indústria petrolífera, dentro de toda a cadeia produtiva do petróleo, não se resume apenas à produção final de derivados, ela também se dá na proteção do meio ambiente. Isto porque suas atividades podem gerar danos ambientais em virtude da ocorrência de vazamentos acidentais de óleo e efluentes de refinarias. Dentro seus derivados, o óleo diesel é um importante combustível utilizado em motores de veículos rodoviários, ferroviários e marítimos e um dos contaminantes aquáticos mais comuns (NOGUEIRA et al., 2011).

De acordo com Singer et al. (2000), avaliar toxicologicamente misturas com petróleo e derivados é uma tarefa complexa. Existe uma variabilidade substancial na coleta, interpretação e uso de dados toxicológicos decorrentes do emprego destas substâncias tornando a comparação de resultados discutível mesmo considerando diferenças experimentais. Logo, a variabilidade metodológica e a falta de padronização afetam a reprodutibilidade e confiabilidade dos dados. Torna-se mandatório, portanto, a padronização de métodos a fim de produzir estimativas de toxicidade confiáveis (SINGER et al., 2000), maior coerência na comparação de resultados experimentais e assegurar dados utilizáveis na tomada de decisões quando o petróleo vaza no ambiente aquático.

Dentre os aspectos experimentais que necessitam padronização, destaca-se o preparo do meio de exposição (SINGER et al., 2000). Nesse sentido, a fração acomodada em água (*Water-accommodated Fraction* - WAF) consiste em um meio de exposição preparado em laboratório decorrente da mistura em baixa energia (sem formação de vórtex) de um material pouco solúvel (como petróleo e derivados) essencialmente sem partículas por um período não superior a 24 horas. Sua preparação requer a filtragem da água do mar através de filtros de espessura $0,45 \mu\text{m}$ e sua mistura com o poluente escolhido utilizando agitadores magnéticos a uma energia suficientemente baixa para não produzir vórtex (SINGER et al., 2000).