

# APLICAÇÃO DE REDES NEURAIS PARA A CLASSIFICAÇÃO AUTOMÁTICA DE EMBARCAÇÕES MERCANTES USANDO SINAIS SONAR PASSIVOS

Application of neural networks for automatic classification of merchant vessels using passive sonar signals

Pedro Mendes Diniz<sup>1</sup> , Rogério de Moraes Calazan<sup>2</sup> 

**Resumo:** A tarefa de classificação de dados é uma das mais amplas áreas de aplicação de algoritmos de aprendizagem de máquina, que está ganhando ainda mais proeminência com a necessidade de rotulação automática de conjuntos de dados cada vez mais abundantes. No ambiente submarino, a aprendizagem de máquina tem servido como importante ferramenta para essa tarefa, em razão, entre outros motivos, da baixa interpretabilidade de dados acústicos pela cognição humana. Entre os algoritmos de aprendizagem de máquina existentes, as redes neurais, em suas diferentes arquiteturas, vêm apresentando excelentes resultados graças à sua capacidade de extrair feições dos dados e aprender modelos não lineares. Este trabalho explora a aplicação de duas arquiteturas de redes neurais, bem como a aplicação da análise de componentes principais para a redução de dimensionalidade dos dados, com vistas à classificação de sinais acústicos de banda larga de embarcações mercantes nas proximidades do Rio de Janeiro. Os resultados apontam para a capacidade das redes neurais de classificar os alvos baseadas no conteúdo do espectro de potência do sinal recebido em um único hidrofone, embora estejam sujeitas às características do conjunto de dados de treinamento.

**Palavras-chave:** Sonar passivo. Perceptron multicamadas. Rede neural convolucional.

**Abstract:** The task of data classification is one of the broadest applications of machine learning algorithms, gaining even greater importance due to the necessity of automatic labelling of increasingly larger data sets. In the underwater environment, machine learning has been used as a powerful tool for this task, among others, due to the low interpretability of acoustic data by human cognition. Among the existing machine learning algorithms, neural networks, in their various architectures, have been displaying excellent results thanks to their capability of extracting features from raw data and learning nonlinear models. This paper explored the application of two neural network architectures, as well as principal component analysis for data dimensionality reduction, for classification of broadband acoustic signals from merchant ships near Rio de Janeiro. Results point to the neural networks' ability of correct target classification based on the power spectrum of a signal from a single hydrophone, although subject to the characteristics of the training data set.

**Keywords:** Passive sonar. Multilayer perceptron. Convolutional neural network.

**1.** Mestrando do Programa de Pós-Graduação em Acústica Submarina (PPGAS) do Instituto de Estudos do Mar Almirante Paulo Moreira (IEAPM), Arraial do Cabo, RJ – Brasil. E-mail: pedro.diniz@marinha.mil.br

**2.** Doutor em Engenharia Eletrônica e Telecomunicações pela Universidade do Algarve. Professor do PPGAS e encarregado da Divisão de Geoacústica do IEAPM, Arraial do Cabo, RJ – Brasil. E-mail: moraes.calazan@marinha.mil.br

## 1. INTRODUÇÃO

A tarefa de classificação de contatos no ambiente submarino é de grande interesse tanto no âmbito militar quanto no civil. Em maior escopo, é uma tarefa relevante na área de aprendizagem de máquina (ML, do inglês *machine learning*) por ser um dos problemas primários de aplicação de algoritmos de aprendizagem estatística na classe de análise de regressão (MATLOFF, 2017). Com o aumento expressivo da quantidade de dados sendo coletados em diversas áreas, tem-se voltado para a ML para a rotulação automática dos conjuntos de dados (FREDRIKSSON; MOSCH; OLSSON, 2022) como forma de processamento ou para torná-los disponíveis para ser usados em aprendizagem supervisionada.

No meio submarino, esse problema é agravado, pois sinais acústicos contêm padrões de difícil reconhecimento pela cognição humana (BIANCO et al., 2019). Além disso, os sinais são contaminados por ruído ou mesmo compostos de ruído (WENZ, 1972). O som irradiado por embarcações é caracterizado por ruído de banda larga causado por cavitação e efeitos hidrodinâmicos, bem como tons discretos e seus harmônicos associados a frequências de operação de maquinário (ZHU et al., 2022).

Por essas razões, a ML tem surgido como ferramenta poderosa na caracterização de dados acústicos. Vários algoritmos de ML, como florestas aleatórias, máquinas de vetor suporte e redes neurais (CHOI; CHOO; LEE, 2019), têm sido empregados. Desses algoritmos, as redes neurais de alimentação direta têm apresentado o melhor desempenho nas tarefas de classificação, com destaque para as redes neurais convolucionais, que empregam camadas de convolução capazes de realizar a extração de feições dos dados por conta própria, reduzindo a carga de trabalho humano e permitindo a aplicação de técnicas mais avançadas, como aprendizagem profunda ou não supervisionada. Redes neurais convolucionais unidimensionais podem ser aplicadas diretamente à forma de onda de um sinal sonar passivo (SOUZA; SEIXAS; MOURA JÚNIOR, 2021) enquanto as bidimensionais, apropriadas para o processamento de imagens, podem ser empregadas para extrair feições de espectrogramas (BORIN et al., 2022).

Domingos et al. (2022) trazem uma extensiva revisão da literatura de classificação acústica de embarcações por métodos de aprendizagem profunda, incluindo aqueles com processamento em frequência, ressaltando que estes demonstradamente

apresentam resultados melhores em grandes distâncias e maior robustez contra ruído quando comparados aos métodos algorítmicos convencionais. Adicionalmente, são listadas técnicas de aumento de conjunto de dados para contornar o principal problema quando se lida com aprendizagem profunda: escassez de dados. Para contornar esse problema, podem-se explorar modelos generativos, sejam eles derivados também de redes adversariais (JAHANIAN et al., 2022), sejam eles modelos clássicos derivados da física (ROSS, 1976) ou da estatística (WALES; HEITMEYER, 2002) para a produção de novos espectros, com técnicas como injeção de ruído (YIN et al., 2015) para o aumento dos conjuntos de dados. O problema da escassez de dados é de particular relevância no campo da acústica submarina, em que há grande dificuldade operacional de se conseguirem dados rotulados de qualidade.

Este trabalho aplica duas arquiteturas distintas de redes neurais para a classificação de sinais sonar passivos gravados em águas rasas nas proximidades do Rio de Janeiro com um único hidrofone, além de aplicar a análise de componentes principais (PCA) para a redução da complexidade de uma das arquiteturas. Foi empregado processamento em banda larga para superar a falta de diversidade espacial nos sinais em decorrência da utilização de um único hidrofone (JESUS et al., 2000). Os sinais gravados são de embarcações mercantes identificadas pelo sistema de identificação automática marítimo (AIS). A classificação dessas embarcações foi feita pela sua classe, entre o total de cinco classes diferentes de embarcações identificadas.

As principais contribuições deste trabalho são: demonstrar que o conteúdo da potência espectral de sinais de um único hidrofone é suficiente para realizar a classificação de navios mercantes; e explorar diferentes abordagens do problema por meio de distintas arquiteturas de redes neurais.

## 2. METODOLOGIA

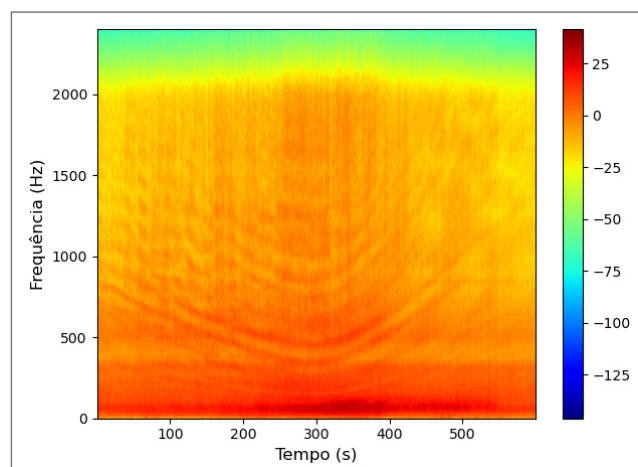
A tarefa de classificação foi realizada por duas arquiteturas de redes neurais: o perceptron multicamadas (MLP) e a rede neural convolucional (CNN). Ambas são redes neurais de alimentação direta acíclicas, ou seja, não é percebida a relação temporal entre as entradas (a rede não possui memória). Desse modo, para a construção dos conjuntos de dados, as gravações são processadas de forma que cada *snapshot*

corresponda a uma amostra independente. Além disso, um segundo teste com o MLP foi feito, no qual a PCA foi usada para reduzir a dimensionalidade do conjunto de dados, possibilitando a redução da complexidade da rede. No total, três algoritmos foram analisados, aqui denominados: MLP, MLP-PCA e CNN. As redes neurais foram construídas usando a interface Tensorflow/Keras.

## 2.1. PROCESSAMENTO DOS SINAIS

Os dados brutos provenientes de 20 gravações, cada uma com uma embarcação identificada, foram decimados por um fator de 10, para uma nova frequência de amostragem de 4.800 Hz. Em seguida, foram recortados os 10 minutos de gravação ao redor do ponto de maior aproximação (PMA) das embarcações, onde o sinal de interesse é mais proeminente, para minimizar a contaminação por ruído de outras embarcações presentes na região durante as gravações. A detecção do PMA foi feita graças ao efeito *Lloyd's mirror*, que produz um distinto padrão de interferência hiperbólico visível no espectrograma, ilustrado na Figura 1, referente à gravação 7.

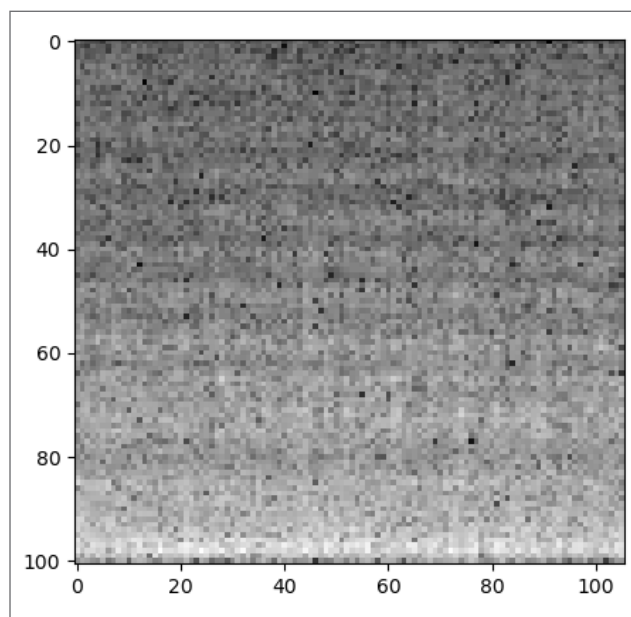
Para o MLP, as gravações foram subsequentemente divididas em 600 *snapshots* de 1 segundo, para os quais se calculou a densidade espectral de potência (PSD) pelo método de Welch, com 10 segmentos e *overlap* de 50%. Um vetor  $\mathbf{x}$  com os *bins* de frequência de 10 a 2.000 Hz, com espaçamento de 10 Hz, totalizando 200 pontos, foi extraído para servir de entrada à rede neural e então normalizado pelo método min-max (Equação 1), o que resultou em um conjunto de dados com o total de 12 mil amostras para todas as 20 gravações.



**Figura 1.** Espectrograma dos 10 minutos selecionados da gravação 7.

$$x'_i = \frac{x_i - \min \mathbf{x}}{\max \mathbf{x} - \min \mathbf{x}} \quad (1)$$

Já para a CNN, a entrada também usou informação da potência espectral das gravações, mas no formato de espectrograma, para aproveitar a capacidade de extração de feições e processamento de imagens característica das CNN bidimensionais. Há duas relações de concessão em jogo no processamento dos dados dessa forma: a segmentação no tempo impacta diretamente o tamanho do conjunto de dados, pois *snapshots* mais longos são necessários para que as feições sejam distinguíveis, mas aumentar a duração reduz a quantidade de amostras totais disponíveis; e a resolução, tanto em frequência quanto no tempo, impacta as dimensões do espectrograma, afetando a nitidez das feições e o custo computacional no treinamento da CNN. Decidiu-se por produzir o espectrograma de um *snapshot* de 4 segundos, com 20 segmentos, *overlap* de 60%, usando 240 pontos na *fast Fourier Transform* (FFT). O espectrograma foi convertido para escala logarítmica, extraída a banda de 10 a 2.000 Hz, e normalizado, resultando num conjunto com o total de 3 mil amostras constituídas de matrizes 101×106 com um único canal, a magnitude do PSD. Uma dessas amostras está ilustrada na Figura 2.



**Figura 2.** Uma das amostras da gravação 7 pré-processada para a rede neural convolucional. Os eixos referem-se ao número de pontos no tempo (106) e na frequência (101).

O rótulo de cada amostra, em ambos os conjuntos de dados, é um vetor binário com cinco elementos que correspondem às cinco classes identificadas, com valor 1 somente no elemento correspondente à classe correta. As classes identificadas, bem como a quantidade de cada uma constante no conjunto de dados, são:

- 1: Offshore Tug/Supply (7);
- 2: Container Ship (4);
- 3: Vehicles Carrier (2);
- 4: Fishing Vessel (3);
- 5: Crude Oil Tanker (4).

## 2.2. PERCEPTRON MULTICAMADAS

O MLP teve uma arquitetura com uma única camada escondida de 800 neurônios, completamente conectada entre a entrada (200 neurônios) e a saída (5 neurônios). A função de ativação dessa camada é a *rectified linear unit* (ReLU), pela qual a saída do neurônio não é modificada, se for positiva, ou é retificada para zero, se for negativa. A camada de saída usou a função *softmax*, que converte o vetor de saída em uma distribuição de probabilidades. A estimativa da rede é a classe (neurônio) com maior probabilidade.

## 2.3. ANÁLISE DE COMPONENTES PRINCIPAIS

A PCA consiste na transformação do conjunto de dados para um novo conjunto de variáveis (os componentes principais) de maneira que cada coordenada seja descorrelacionada com as demais, e uma pequena quantidade de coordenadas seja capaz de reter a maior parte da variância do conjunto original (JOLLIFFE, 2002). O  $k$ -ésimo componente principal pode ser computado pela Equação 2:

$$\mathbf{t}_k = \mathbf{v}_k^T \mathbf{X} \quad (2)$$

Em que:

$\mathbf{v}_k$  = o autovetor normalizado com o  $k$ -ésimo maior autovalor da matriz de covariância da matriz de dados  $\mathbf{X}$  (na qual cada coluna é um vetor de amostra).

A variância do conjunto de dados que é explicada por um componente principal é dada por seu autovalor associado. Dessa forma, é possível reduzir a dimensionalidade do problema ao utilizar somente aqueles componentes que explicam

a maior parte da variância. Para o conjunto completo de dados processados para o MLP, 19, 27 e 46 componentes explicam, respectivamente, 90, 95 e 99% da variância.

Decidiu-se por utilizar os 46 primeiros componentes principais, de modo que a camada de entrada do MLP-PCA passou a ter número equivalente de neurônios. A complexidade da rede foi reduzida de acordo: a camada escondida passou a ter somente 200 neurônios.

## 2.4. REDE NEURAL CONVOLUCIONAL

A CNN dispôs de uma arquitetura mais complexa, com três camadas convolucionais para a extração de feições do espectrograma, seguidas de camadas de *max-pooling*. As duas primeiras camadas convolucionais contaram com 32 filtros com *kernel*  $3 \times 3$ , enquanto a terceira usou 64 filtros  $2 \times 2$ . O *kernel* das camadas de *pooling* usou a mesma dimensão dos filtros das camadas convolucionais que as antecedem. A saída da rede convolucional foi linearizada e conectada a uma camada comum com 32 neurônios antes da saída. Mais uma vez, a função de ativação de todas as camadas intermediárias foi a ReLU, enquanto a camada de saída usou a *softmax*.

## 2.5. TREINAMENTO E MÉTRICAS

O otimizador usado foi o Adam, com taxa de aprendizagem de 0,01 para o MLP e 0,004 para a CNN, e o treinamento ocorreu em *batches* de 100 amostras ao longo de 100 iterações. A função de custo a ser minimizada foi a entropia cruzada categórica, que mede a diferença entre duas distribuições de probabilidade. Não foram avaliadas técnicas de regularização no treinamento.

O treinamento e a avaliação das redes neurais, neste caso, não puderam seguir o paradigma tradicional de divisão do conjunto de dados em treinamento, validação e teste, pois as amostras originárias de uma mesma gravação, embora distintas, são altamente correlacionadas entre si. Uma avaliação baseada em uma segmentação aleatória do conjunto de dados forneceria uma estimativa deveras otimista da capacidade de generalização da rede graças à similaridade entre amostras de treinamento e teste. Tampouco separar um único subconjunto referente a uma das gravações seria adequado, pois a estimativa de desempenho da rede teria forte viés por conta da uniformidade dos dados de teste.

A metodologia de validação mais adequada, neste caso, é a validação cruzada. Foram treinados 20 modelos de cada

tipo. Para cada um deles, uma das gravações foi separada como conjunto de testes e a rede foi treinada com os dados das outras 19, permitindo que as características de cada subconjunto, individualmente, pudessem ser avaliadas e reduzindo o viés na estimativa final de desempenho dos modelos.

Para a PCA, a matriz de transformação linear contendo os 46 componentes foi definida com base nos dados das 19 gravações de treinamento e então aplicada aos dados da gravação remanescente.

A métrica usada para a avaliação dos modelos treinados foi a acurácia, ou a frequência com que o classificador

fez estimativas corretas. O processo foi repetido três vezes, dada a natureza estocástica da ML, e os melhores resultados estão reportados.

### 3. RESULTADOS E DISCUSSÃO

A Tabela 1 traz a acurácia percentual dos três algoritmos quando testados em cada um dos 20 conjuntos de dados de teste. Cada resultado corresponde a um modelo distinto que foi treinado independentemente nos 19 conjuntos de dados

**Tabela 1. Acurácia (%) dos três algoritmos testados para cada conjunto de dados de teste.**

Dado	Embarcação	Classe	MLP	MLP-PCA	CNN
1	Bram Bravo	Offshore Tug/Supply	1,67	7,83	0,67
2	Far Sagaris	Offshore Tug/Supply	98,83	98,83	100
3	Far Sagaris	Offshore Tug/Supply	99,17	96,83	89,33
4	Skandi Angra	Offshore Tug/Supply	99,33	94,33	98,00
5	MSC Cadiz	Container Ship	29,50	21,33	11,33
6	MSC Cadiz	Container Ship	89,17	84,17	52,67
7	Arc Resolve	Vehicles Carrier	70,83	56,17	73,33
8	Arc Resolve	Vehicles Carrier	67,11	48,50	70,67
9	Uluwatu I1	Fishing Vessel	53,33	51,83	43,33
10	Uluwatu I1	Fishing Vessel	59,50	76,17	78,00
11	MSC Barcelona	Container Ship	47,00	64,33	42,67
12	MSC Barcelona	Container Ship	58,33	60,33	58,67
13	Far Sagaris	Offshore Tug/Supply	67,00	74,33	87,33
14	Uluwatu I1	Fishing Vessel	17,50	26,00	44,00
15	Bram Brasil	Offshore Tug/Supply	10,17	18,33	4,67
16	Barcelona Spirit	Crude Oil Tanker	50,67	45,33	27,33
17	Bram Brasil	Offshore Tug/Supply	62,67	83,50	65,33
18	Barcelona Spirit	Crude Oil Tanker	52,33	58,33	74,00
19	Garrincha	Crude Oil Tanker	66,00	54,00	57,33
20	Fortaleza Knutsen	Crude Oil Tanker	24,50	20,50	41,33
$\mu$	-	-	56,23	57,05	55,90
$\sigma$	-	-	27,98	27,04	28,71

MLP: perceptron multicamadas; MLP-PCA: perceptron multicamadas precedido pela análise de componentes principais; CNN: rede neural convolucional.

remanescentes. Percebe-se que a acurácia do modelo de ML variou amplamente entre o conjunto de dados. Uma das possíveis razões para a grande variância no desempenho é a qualidade dos dados. Uma vez que os dados foram coletados em condições naturais por passagens oportunas de embarcações, não houve como controlar os diferentes regimes de operação representados no conjunto de dados. A Figura 3 traz os espectrogramas das gravações 5 e 6 para exemplificar essa questão. Ambas são referentes ao navio contêiner MSC Cadiz, mas foram gravadas em dias distintos.

Esse é apenas um aspecto do problema maior relacionado à quantidade dos dados, que estão claramente desbalanceados: enquanto a classe 1 possui sete gravações, a classe 3 possui somente duas e a classe 4 três. Isso implica que, quando o dado pertencente a uma dessas classes de menor volume é usado no teste, a rede neural é exposta a poucos exemplos dela durante o treinamento. Além disso, os dados que estão presentes pertencem a uma quantidade limitada de navios e, como já foi notado, a uma quantidade limitada de regimes de máquinas, então mesmo nos casos em que os classificadores tiveram bom desempenho não há garantia de capacidade de generalização.

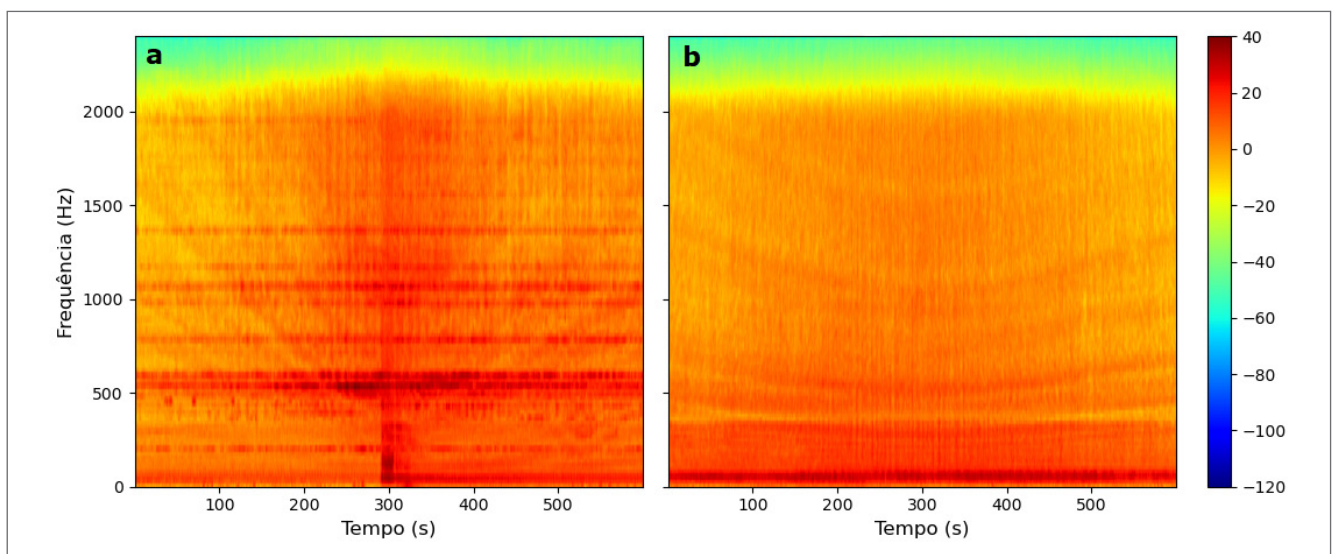
Qualidade e quantidade dos dados são problemas fundamentais quando se trabalha com ML. Para contorná-los, pode-se realizar coleta extensiva de dados em diversas condições ou aplicar técnicas de aumento de conjuntos de dados, a fim de treinar uma rede neural robusta. O presente estudo

não dispôs de nenhuma dessas abordagens, então os resultados apresentados são fundamentalmente limitados pelos dados usados.

Mesmo com essas considerações, os resultados obtidos pelos classificadores empregados foram satisfatórios. É preciso lembrar que todas as amostras de teste, em cada caso, pertencem a uma única gravação contínua. Assim, uma forma melhor de enxergar a acurácia na Tabela 1 é como a probabilidade acumulada ao longo dos 10 minutos de que aquele navio pertence a determinada classe. Faz-se necessário observar essa probabilidade com respeito às das demais classes, e não apenas o valor da acurácia. A classe estimada será aquela com maior probabilidade, mesmo que seu valor absoluto seja pequeno. Esse processo é análogo ao que é feito na camada de saída da rede neural, mas leva em conta o efeito cumulativo de estimativas sucessivas. A Figura 4 traz as distribuições de probabilidade normalizadas para todos os testes.

O que se tem é o efeito de uma janela temporal que limita o efeito de perturbações transientes no sinal que levem a classificações errôneas, momentaneamente, na estimativa final. Na Figura 4, a janela compreende os 10 minutos de dados.

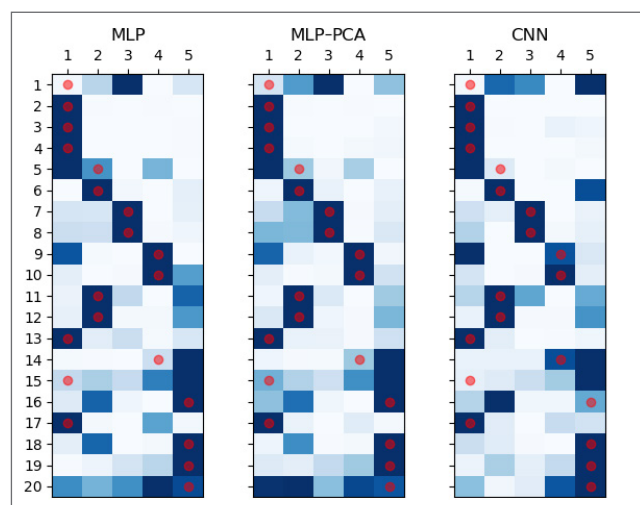
Com isso, pode-se perceber que o desempenho dos três classificadores usados não apresentou grandes variações, com 15 acertos cada um, ou taxa de acerto de 75%. A acurácia média e o desvio padrão (Tabela 1) de todos os modelos também



**Figura 3.** Espectrogramas das gravações (A) 5 e (B) 6.

foram bastante semelhantes, resultado do fato de que, enquanto alguns dados obtiveram resultados melhores com um dado algoritmo, outros pioraram, de modo que nenhum algoritmo foi preponderante sobre os demais. Isso se reflete nas matrizes de confusão da Figura 5, que apresentam o somatório das matrizes de confusão de cada teste, com o total de amostras e o percentual (por linha) que corresponde a cada categoria.

A distribuição das três matrizes é bastante semelhante, com a maior parte dos valores concentrada na diagonal principal,



**Figura 4.** Distribuições de probabilidades normalizadas para cada um dos 20 dados de teste. O bloco mais escuro em cada linha é a classe de maior probabilidade. O círculo vermelho sinaliza a classe correta.

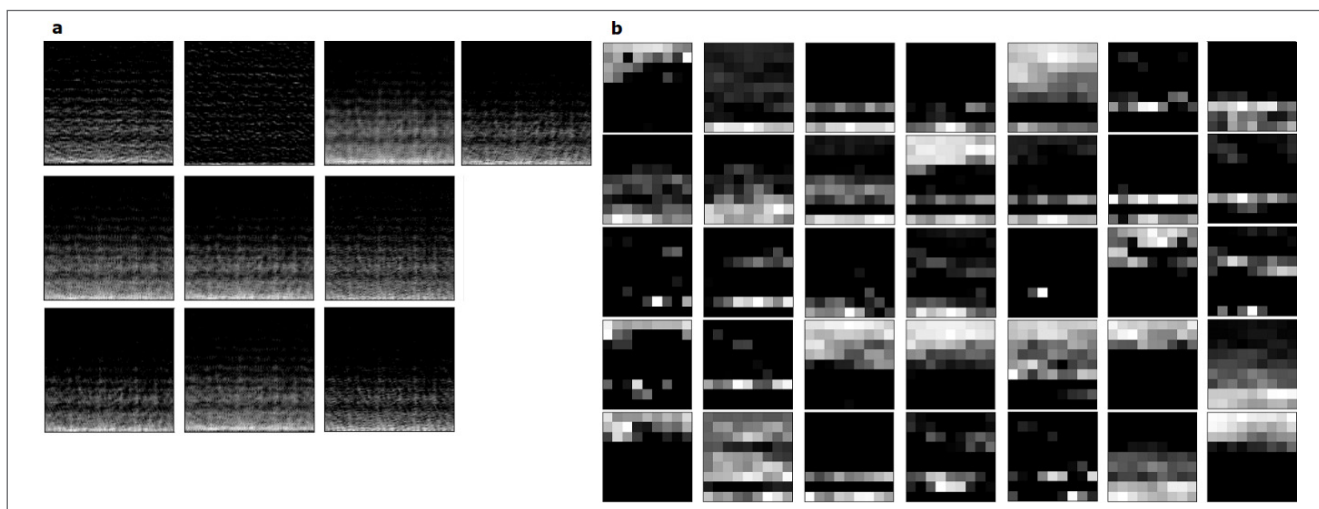
o que é um bom indicativo para um modelo de classificação. O MLP-PCA e a CNN conseguiram melhor discriminação da diagonal principal em relação ao MLP regular, particularmente na classe 4, mas nenhum dos algoritmos apresentou melhora definitiva em todas as classes, o que se reflete em valores muito próximos para o traço (percentual) dessas matrizes: 2,80, 2,73 e 2,81, respectivamente.

Há outros ganhos que não somente em acurácia, no entanto. Comparado ao MLP regular, o MLP-PCA alcançou resultados equivalentes com um modelo muito menos complexo: 10.405 parâmetros contra 164.805 do MLP regular. Essa simplicidade do modelo reverte-se em redução no custo computacional para novas estimativas e diminuição apreciável no tempo de treinamento da rede, o que é importante em uma aplicação de aprendizagem contínua *online*, na qual o modelo continua a ser ajustado com base em novos dados encontrados após a implantação (PARISI et al., 2019).

Com a CNN, é possível analisar as saídas das camadas convolucionais, resultantes da convolução dos filtros com a entrada, chamadas mapas de feições, para entender quais feições do conjunto de dados a rede neural está considerando mais relevantes. A Figura 6 mostra alguns dos mapas de feições da primeira e última camadas convolucionais de um dos modelos treinados. A capacidade da CNN de encontrar padrões e feições em espectrogramas por conta própria torna-a apropriada para aplicações de aprendizagem semi ou não supervisionada.

Classe verdadeira	a					b					c				
	Offshore Tug/Supply	Container Ship	Vehicles Carrier	Fishing Vessel	Crude Oil Tanker	Offshore Tug/Supply	Container Ship	Vehicles Carrier	Fishing Vessel	Crude Oil Tanker	Offshore Tug/Supply	Container Ship	Vehicles Carrier	Fishing Vessel	Crude Oil Tanker
Offshore Tug/Supply	2633.0 62.69%	251.0 5.98%	554.0 13.19%	381.0 9.07%	381.0 9.07%	2844.0 67.71%	308.0 7.33%	354.0 8.43%	220.0 5.24%	474.0 11.29%	668.0 63.62%	83.0 7.9%	59.0 5.62%	63.0 6.0%	177.0 16.86%
Container Ship	335.0 13.96%	1344.0 56.0%	95.0 3.96%	141.0 5.88%	485.0 20.21%	435.0 18.12%	1381.0 57.54%	115.0 4.79%	123.0 5.12%	346.0 14.42%	156.0 26.0%	248.0 41.33%	35.0 5.83%	4.0 0.67%	157.0 26.17%
Vehicles Carrier	162.0 13.5%	154.0 12.83%	828.0 69.0%	4.0 0.33%	52.0 4.33%	214.0 17.83%	278.0 23.17%	628.0 52.33%	5.0 0.42%	75.0 6.25%	56.0 18.67%	10.0 3.33%	216.0 72.0%	2.0 0.67%	16.0 5.33%
Fishing Vessel	316.0 17.56%	8.0 0.44%	5.0 0.28%	782.0 43.44%	689.0 38.28%	314.0 17.44%	27.0 1.5%	24.0 1.33%	924.0 51.33%	511.0 28.39%	100.0 21.98%	5.0 1.1%	6.0 1.32%	247.0 54.29%	97.0 21.32%
Crude Oil Tanker	179.0 7.46%	591.0 24.62%	186.0 7.75%	283.0 11.79%	1161.0 48.38%	315.0 13.12%	604.0 25.17%	156.0 6.5%	256.0 10.67%	1069.0 44.54%	80.0 13.33%	123.0 20.5%	19.0 3.17%	78.0 13.0%	300.0 50.0%

**Figura 5.** Somatório das matrizes de confusão para (A) perceptron multicamadas, (B) perceptron multicamadas precedido pela análise de componentes principais e (C) rede neural convolucional.



**Figura 6.** Mapas de feições da (A) primeira e (B) última camadas convolucionais da rede neural convolucional para amostra ilustrada na Figura 2.

Nota-se que a primeira camada foi capaz de salientar o padrão de interferência existente no espectrograma, que aparece de alguma forma em todos os mapas, embora a rede tenha privilegiado a seção de baixa frequência dos dados. A última camada realça bandas de frequências inteiras, representadas pelas barras horizontais, e feições locais mais abstratas. Os mapas exibidos são todos os que foram produzidos nas respectivas camadas, pois os demais produziram uma matriz repleta de zeros em razão dos pesos majoritariamente negativos dos respectivos filtros. Esse é um problema conhecido da função de ativação ReLU, conhecido como “*dying ReLU*” (LU et al., 2020) — consequência da dissipação do gradiente que, durante o treinamento, leva um neurônio (ou um filtro) para uma região da função em que ele jamais ficará ativo e produzirá 0 para qualquer entrada. Esse problema mostra que a rede se beneficiaria de regularização no treinamento e ajuste dos hiperparâmetros.

As informações obtidas por esta análise podem ser usadas para reavaliar o pré-processamento dos dados, modificar as funções de ativação e topologia da rede ou outros hiperparâmetros, como taxa de aprendizagem e inicialização dos pesos. Esse é um processo de metaotimização que pode ser encarado como parte do treinamento da rede e do processo de seleção do modelo (RASCHKA, 2020). Esse processo iterativo de ajuste do modelo, embora não tenha sido feito neste estudo, é fundamental quando da seleção de um modelo final para aplicação.

## 4. CONCLUSÃO

Este trabalho apresentou uma metodologia de classificação de embarcações baseada em redes neurais usando o conteúdo da potência espectral do sinal gravado por um sonar passivo. O problema foi abordado por três algoritmos distintos de redes neurais, os quais obtiveram resultados satisfatórios, mas que podem ser subsequentemente refinados a fim de se obterem resultados melhores em trabalhos futuros.

A acurácia média dos três algoritmos foi bastante próxima, e todos os três classificaram corretamente 15 das 20 embarcações. Apesar dos bons resultados alcançados, a análise aqui feita sugere que as arquiteturas usadas são subótimas e podem ser melhoradas. Como qualquer estudo de ML, os resultados são fundamentalmente limitados pela quantidade e qualidade dos dados usados. Para trabalhos futuros, sugere-se a investigação da metodologia em conjuntos de dados mais abrangentes, com mais exemplos de navios de cada classe, associada a um processo de otimização de hiperparâmetros e regularização.

## 5. AGRADECIMENTOS

Os autores agradecem ao Instituto de Pesquisas da Marinha e à Petrobras a disponibilização dos dados acústicos analisados neste trabalho por ocasião do XIV Encontro de Tecnologia em Acústica Submarina.



## REFERÊNCIAS

- BIANCO, M. J.; GERSTOFT, P.; TRAER, J.; OZANICH, E.; ROCH, M. A.; GANNOT, S.; DELEDALLE, C.-A. Machine learning in acoustics: theory and applications. *Journal of the Acoustical Society of America*, v. 146, n. 5, p. 3590-3628, 2019. <https://doi.org/10.1121/1.5133944>
- BORIN, M.; HOLTZ, M.; JACOMUSSI, L.; MONTEIRO, C.; WEITBRECHT, P.; JARDIM, C. Automatic identification of city intrusive noises using Deep Learning. *In: CONGRESSO IBEROAMERICANO DE ACÚSTICA*, 12., 2022. *Anais [...]*. Florianópolis: FIA, 2022.
- CHOI, J.; CHOO, Y.; LEE, K. Acoustic classification of surface and underwater vessels in the ocean using supervised machine learning. *Sensors*, v. 19, n. 16, 3492, 2019. <https://doi.org/10.3390/s19163492>
- DOMINGOS, L. C. F.; SANTOS, P. E.; SKELTON, P. S. M.; BRINKWORTH, R. S. A.; SAMMUT, K. A survey of underwater acoustic data classification methods using deep learning for shoreline surveillance. *Sensors*, v. 22, n. 6, 2181, 2022. <https://doi.org/10.3390/s22062181>
- FREDRIKSSON, T.; MOSCH, J.; OLSSON, H. H. Machine learning models for automatic labeling: A systematic literature review. *In: INTERNATIONAL CONFERENCE ON SOFTWARE TECHNOLOGIES*, 15., 2020. *Anais [...]*. 2020. p. 552-561.
- JAHANIAN, A.; PUIG, X.; TIAN, Y.; ISOLA, P. Generative models as a data source for multiview representation learning. *In: INTERNATIONAL CONFERENCE ON LEARNING REPRESENTATIONS*, 2022. *Anais [...]*. 2022
- JESUS, S. M.; PORTER, M. B.; STÉPHAN, Y.; DÉMOULIN, X.; RODRÍGUEZ, O. C.; COELHO, E. M. M. Ferreira. Single hydrophone source localization. *IEEE Journal of Oceanic Engineering*, v. 25, n. 3, p. 337-346, 2000. <https://doi.org/10.1109/48.855379>
- JOLLIFFE, I. T. *Principal Component Analysis*. 2. ed. Nova York: Springer-Verlag, 2002.
- LU, L.; SHIN, Y.; SU, Y.; KARNIADAKIS, G. E. Dying ReLU and initialization: Theory and numerical examples. *Communications in Computational Physics*, v. 28, p. 1671-1706, 2020. <https://doi.org/10.4208/cicp.OA-2020-0165>
- MATLOFF, N. *Statistical regression and classification: from linear models to machine learning*. Boca Raton: CRC Press, 2017.
- PARISI, G. I.; KEMKER, R.; PART, J. L.; KANAN, C.; WERMTER, S. Continual lifelong learning with neural networks: a review. *Neural Networks*, v. 113, p. 54-71, 2019. <https://doi.org/10.1016/j.neunet.2019.01.012>
- RASCHKA, S. Model evaluation, model selection, and algorithm selection in machine learning. *arXiv:1811.12808*, 2020. <https://doi.org/10.48550/arXiv.1811.12808>
- ROSS, D. *Mechanics of underwater noise*. Nova York: Pergamon Press, 1976.
- SOUZA, M. J. de; SEIXAS, J. M. de; MOURA JÚNIOR, N. N. de. Classificação de sinais temporais de sonar passivo utilizando redes neurais convolucionais. *In: CONGRESSO BRASILEIRO DE INTELIGÊNCIA COMPUTACIONAL*, 15., 2021. *Anais [...]*. Joinville: SBIC, 2021.
- WALES, S. C.; HEITMEYER, R. M. An ensemble source spectra model for merchant ship-radiated noise. *Journal of the Acoustical Society of America*, v. 111, n. 3, p. 1211-1231, 2002. <https://doi.org/10.1121/1.1427355>
- WENZ, G. M. Review of underwater acoustics research: noise. *Journal of the Acoustical Society of America*, v. 51, n. 3B, p. 1010-1024, 1972. <https://doi.org/10.1121/1.1912921>
- YIN, S.; LIU, C.; ZHANG, Z.; LIN, Y.; WANG, D.; TEJEDOR, J.; ZHENG, T. F.; LI, Y. Noisy training for deep neural networks in speech recognition. *Journal on Audio, Speech, and Music Processing*, v. 2015, n. 2, p. 1-14, 2015. <https://doi.org/10.1186/s13636-014-0047-0>
- ZHU, C.; GAGGERO, T.; MAKRIS, N. C.; RATILAL, P. Underwater sound characteristics of a ship with controllable pitch propeller. *Journal of Marine Science and Engineering*, v. 10, n. 3, 328, 2022. <https://doi.org/10.3390/jmse10030328>