

Classificador do efeito de Lloyd's mirror em tabela de detecção de navios

Fábio Oliveira Baptista da Silva Matheus Ferreira do Espírito Santo Natanael Nunes de Moura Júnior
Grupo de Sistemas Acústicos Submarinos *Laboratório de Processamento de Sinais* *Laboratório de Processamento de Sinais*
Instituto de Pesquisas da Marinha *Universidade Federal de Rio de Janeiro* *Universidade Federal de Rio de Janeiro*
Rio de Janeiro, Brasil Rio de Janeiro, Brasil Rio de Janeiro, Brasil
0000-0001-8410-6475 0009-0009-5672-815X 0000-0003-0828-6085

William Soares Filho Luiz Pereira Calôba
Grupo de Sistemas Acústicos Submarinos *Laboratório de Processamento de Sinais*
Instituto de Pesquisas da Marinha *Universidade Federal de Rio de Janeiro*
Rio de Janeiro, Brasil Rio de Janeiro, Brasil
0000-0002-1026-0573 lcaloba@gmail.com

Resumo—No contexto do Projeto de Monitoramento da Paisagem Acústica Submarina na Baía de Santos (PMPAS-BS), realizado pela PETROBRAS em atendimento a condicionante exigida pelo licenciamento ambiental federal e conduzido pelo IBAMA, surge a oportunidade de separação de dados acústicos para construção de um *dataset* para classificação acústica submarina. Dada a grande quantidade de dados coletados pelo projeto (mais de 66.745 horas), há a necessidade de criação de um classificador para seleção das detecções de embarcações com baixa taxa de falso alarme. Nesse estudo foram comparados os classificadores: *Random Forest*, redes neurais de múltiplas camadas (MLP), classificadores de vetor suporte (SVC) e *ensembles* desses classificadores. Os objetivos deste estudo foram alcançados, obtendo modelos com taxa de falso alarme zero no conjunto de teste e propondo um método de seleção das detecções com base no limiar de decisão da MLP.

Abstract—In the context of the "Projeto de Monitoramento da Paisagem Acústica Submarina na Baía de Santos"(PMPAS-BS), carried out by PETROBRAS in compliance with the condition required by the federal environmental licensing and conducted by IBAMA, the opportunity arises to separate acoustic data for the construction of a dataset for underwater acoustic classification. Given the large amount of data collected by the project (over 66,745 hours), there is a need to develop a classifier for selecting vessel detections with a low false alarm rate. In this study, were compared, the classifiers: Random Forest, Multilayer Perceptron (MLP), Support Vector Classifiers (SVC), and ensembles of these classifiers. The objectives of this study were achieved by obtaining models with a zero false alarm rate in the test set and proposing a method for selecting detections based on the MLP's decision threshold.

Index Terms—MLP, kernel, SVM, ensemble, LLOYD's mirror, sonar

I. INTRODUÇÃO

O Projeto de Monitoramento da Paisagem Acústica Submarina da Baía de Santos (PMPAS-BS) é executado pela PETROBRAS para cumprimento de condicionante ambiental exigida no Licenciamento Ambiental conduzido pelo IBAMA das atividades de produção e escoamento de petróleo e gás natural na Baía de Santos (processo no. 02001.114279/2017-

80). O projeto é executado em ciclos, sendo o primeiro realizado entre novembro de 2017 e dezembro de 2022, e o segundo ciclo teve início em janeiro de 2023. Durante o primeiro ciclo do PMPAS-BS, a coleta de dados foi realizada por meio de monitoramento móvel (perfiladores acústicos e *gliders*) e monitoramento fixo (linhas de fundeio instrumentadas e observatórios submarinos). Para o presente estudo, foram utilizados os dados gravados pelos observatórios submarinos (OS), instalados junto ao leito submarino em regiões costeiras na Baía de Santos, dotados de sistemas de gravação contínua de sinais acústicos[1].

Ao longo do 1º ciclo, os OS gravaram aproximadamente 66.745 horas de dados acústicos processados e qualificados. Uma das análises feitas no projeto é a detecção de embarcações por intensidade sonora, tendo sido detectadas mais de 88 mil embarcações ao longo do projeto, das quais mais de 34 mil foram identificadas[1].

No desenvolvimento de trabalhos de classificação de navios com base em acústica submarina, é evidente a escassez de dados públicos [2], [3]. Entre os conjuntos de dados disponíveis, podemos mencionar: o ShipsEar, composto por 90 gravações de 11 tipos de embarcação [4], e o DeepShip, que contém cerca de 47 horas de gravações de 265 navios de 4 tipos de embarcação [3].

Tendo isso em vista, surge a possibilidade de construção de um dataset com base nos dados coletados no projeto para servir de base para trabalhos na área. Para a separação desses dados acústicos, o primeiro passo consiste em estabelecer uma estratégia de filtragem e validação dos dados relacionados às detecções. É desejável que os dados sejam de embarcações identificadas e que as gravações contenham o ponto de máxima aproximação (PMA) entre a trajetória da embarcação e o local de instalação do OS para garantir a melhor relação sinal-ruído (SNR).

Uma forma de verificar se o sinal acústico que gerou uma detecção corresponde ao PMA de uma embarcação é através

da análise do padrão de interferência causado pelo efeito de Lloyd's mirror [5], que é o efeito de interferência construtiva e destrutiva entre os múltiplos percursos do sinal acústico [6]. Esse efeito pode ser facilmente visualizado em espectrogramas, onde são observadas franjas, conforme ilustrado na Figura 1.

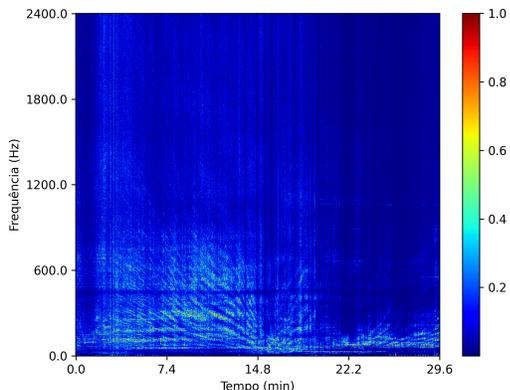


Figura 1. Espectrograma com efeito de interferência desejado

Neste estudo, serão explorados métodos de aprendizado de máquina para prever a ocorrência do fenômeno com base nos dados disponíveis na tabela de detecções. O objetivo é minimizar a necessidade de validação manual, considerando a extensa quantidade de dados envolvida. Para isso, realizou-se a classificação manual de 200 arquivos para identificar a presença do fenômeno.

II. METODOLOGIA

A Tabela de detecções, que serve como base para este trabalho, contém as seguintes informações:

- **Identificador da Classe:** Número representando as classes:
 - 1) Iate;
 - 2) Navios de Apoio Marítimo para Ancoragem, Reboque e Abastecimento;
 - 3) Embarcação de abastecimento de plataforma;
 - 4) Aliviador; e
 - 5) Graneleiro;
- **Hora:** Hora do dia entre 0 e 24 horas;
- **PMA:** Ponto de máxima aproximação entre a trajetória do navio e a localização de instalação do sensor acústico em metros;
- **Velocidade:** Velocidade, em nós, do navio detectado;
- **Rumo:** Direção de navegação, considerando Norte, Leste, Sul e Oeste, representados por 0°, 90°, 180° e 270°, respectivamente.
- **SPL no PMA:** Intensidade de pressão sonora (Sound pressure level (SPL)) captada pelo sensor durante o PMA em dB re 1 μ Pa;
- **SPL do Ruído:** Intensidade de pressão sonora média, em dB re 1 μ Pa, no sensor estimada ao longo da última uma hora;

- **SNR:** Relação sinal ruído (SPL no PMA - SPL do Ruído), em dB;
- **Calado:** Profundidade máxima da parte do navio abaixo da linha da água, em metros; e
- **Intervalo sem detecções:** Intervalo de tempo, em minutos, entre a detecção do PMA e a detecção de outra embarcação.

A Figura 2 apresenta a visualização t-Distributed Stochastic Neighbor Embedding (t-SNE) [7] para os 200 dados classificados manualmente. As entradas onde o efeito de interferência é visível estão representadas em amarelo, enquanto as entradas onde o efeito não é visível estão representadas em azul. Observa-se uma sobreposição das classes ao longo de todo o domínio, o que ilustra a dificuldade da criação de classificadores.

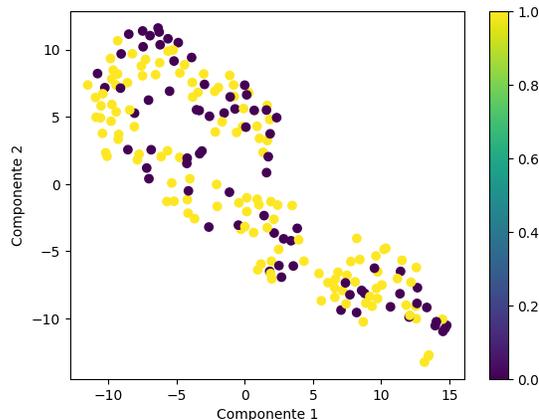


Figura 2. Visualização t-SNE dos dados classificados

A abordagem utilizada neste trabalho consiste em comparar um classificador de floresta aleatória (Random Forest), como *baseline*, com classificadores de redes neurais de múltiplas camadas (MLP) e classificadores de vetor de suporte (SVC). Para isso, foram utilizadas as implementações dos classificadores RandomForestClassifier, SVC e NuSVC disponíveis no pacote Scikit-learn [8], juntamente com a implementação do MLP fornecida pelo Keras [9]. O código de treinamento e avaliação dos resultados foi desenvolvido em Python e está disponibilizado no GitHub em [10].

Este trabalho começa com a identificação das características relevantes para a classificação por meio dos coeficientes de Pearson [11]. Em seguida, foram treinados dois classificadores Random Forest: um utilizando todas as características e outro apenas com as consideradas relevantes. Foram extraídas as métricas de desempenho de referência (*baseline*) para o restante do trabalho e a seleção das características foi validada pelo desempenho dos classificadores e pela importância de Gini obtida durante o treinamento da Random Forest completa [12].

Posteriormente, foram treinadas as MLPs com duas camadas, sendo a camada de saída composta por um único neurônio com função de ativação sigmoideal. Foram realizados

diversos experimentos para determinar o número de neurônios na camada intermediária, escolher a função de ativação, avaliar a necessidade de regularização e *Dropout* [13], além de investigar o impacto de diferentes escalonamentos dos dados de entrada. Foram comparadas técnicas de escalonamento: a normalização z-score [14], o escalonamento Min-Max [15], e a normalização pela transformação de Box-Cox [16], com o auxílio das ferramentas do Scikit-learn. Além disso, para a configuração que obteve os melhores resultados, foi analisada a modificação do limiar de decisão do neurônio de saída.

Em seguida foram treinados os classificadores de vetor de suporte, executando experimentos de busca em grade, tanto para o SVC quanto para o NuSVC do Scikit-learn avaliando os kernels: linear, de base radial (RBF), sigmoidal e polinomial.

Os resultados foram comparados e foram treinados modelos de *ensemble learning* combinando os melhores classificadores, tanto por votação (*majority voting ensemble*) quanto treinando uma MLP com a saída desses modelos (*stacking ensemble*), visando melhorar a performance e a capacidade de generalização dos modelos [17].

As figuras de mérito utilizadas para avaliação e comparação dos modelos foram as seguintes:

- **Taxa de Falso Alarme:** Indica a taxa de classificação incorreta de instâncias negativas [18];
- **Sensibilidade (*Recall*):** Indica a taxa de classificação correta de instâncias positivas [19];
- **F1-score:** É a média harmônica da precisão e da sensibilidade. Fornece uma medida equilibrada do desempenho do modelo, considerando tanto a capacidade de identificar corretamente os casos positivos (sensibilidade) quanto a capacidade de evitar classificar incorretamente os casos negativos (precisão) [19]; e
- **Área sob a Curva ROC¹ (AUC):** Indica a capacidade discriminativa do modelo em classificar corretamente as instâncias positivas e negativas [20].

A taxa de falso alarme foi considerada a figura de mérito mais importante para a aplicação final dos modelos, pois o objetivo é filtrar um grande número de detecções para obter um conjunto com alta confiança de que o efeito de interferência seja visível. Reduzir a taxa de falso alarme significa classificar menos sinais sem o efeito como tendo o efeito, o que pode ser vantajoso mesmo que isso signifique ter mais falsos negativos. No entanto, é importante garantir que o número de verdadeiros positivos seja suficiente para as próximas etapas de criação do *dataset*.

As comparações dos resultados pelas figuras de mérito serão apresentadas em tabelas contendo valores médios e desvio padrão. Nas tabelas de figuras de mérito, os melhores resultados serão identificados em negrito, os demais resultados dentro da dispersão da melhor análise serão apresentados em

¹Curva ROC (*Receiver Operating Characteristic*): apresenta a relação entre a sensibilidade e a especificidade de um classificador binário. A sensibilidade ou taxa de verdadeiros positivos mede a proporção de positivos classificados corretamente; A especificidade ou taxa verdadeiros negativos mede a proporção de negativos classificados corretamente [20].

letra regular e as análises com resultados estatisticamente inferiores à análise destacada serão destacadas em vermelho.

III. RESULTADOS

A. Seleção de características

A Tabela I apresenta os coeficientes de correlação de Pearson para as características, sendo destacadas em negrito aquelas que apresentam uma correlação linear mais significativa. Com base nesse critério, foram selecionadas as seguintes características como relevantes: SNR, Hora, SPL PMA e Intervalo sem detecções. A correlação entre as características relevantes não foi apresentada, uma vez que, para o escopo deste trabalho, não há benefício em reduzir ainda mais o número de características.

Tabela I
RELEVÂNCIA DAS CARACTERÍSTICAS - COEFICIENTE DE PEARSON

| | Efeito Visível |
|-------------------------|----------------|
| Identificador da Classe | 0,12 |
| Hora | -0,27 |
| PMA | 0,03 |
| Velocidade | -0,09 |
| Rumo | -0,07 |
| SPL PMA | 0,26 |
| SPL Noise | -0,00 |
| SNR | 0,39 |
| Calado | 0,13 |
| Intervalo sem detecções | 0,18 |

B. Validação cruzada

A estratégia de validação cruzada utilizada nos experimentos deste estudo foi a Repeated Stratified K-fold [8], com 3 folds diferentes e 4 repetições por experimento. Essa estratégia foi adotada para aumentar a robustez e a confiabilidade dos resultados, evitando uma redução excessiva do conjunto de validação devido à limitação da quantidade de dados etiquetados. Das 200 amostras utilizadas neste trabalho, 20 foram separadas para formar o conjunto de teste, enquanto as outras 180 amostras foram destinadas ao treinamento e validação durante o processo de validação cruzada. Isso resultou em conjuntos de validação de 60 amostras e conjuntos de treinamento com 120 amostras.

C. Baseline

Foram treinados classificadores *Random Forest* considerando todas as características da tabela, bem como apenas as características consideradas relevantes na Seção III-A, com o objetivo de estabelecer uma referência para os resultados dos experimentos das seções subsequentes. Para ambos os casos, foram utilizados os parâmetros padrões fornecidos pelo algoritmo disponibilizado pela biblioteca Scikit-learn [8].

A Tabela II apresenta as figuras de mérito para o conjunto de validação. Observa-se que, tanto na abordagem que considera as variáveis relevantes de acordo com o coeficiente de Pearson quanto na abordagem sem essa seleção, ambos os modelos apresentaram desempenhos semelhantes em todas as métricas.

Além disso, para reforçar a confiabilidade da seleção das características, a Figura 3 apresenta a distribuição da importância

Tabela II
FIGURAS DE MÉRITO NO CONJUNTO DE VALIDAÇÃO - RANDOM FOREST

| | Falso Alarme | Sensibilidade | F1-score | AUC |
|------------|---------------------|---------------------|---------------------|---------------------|
| relevantes | 45,59 ± 8,69 | 72,78 ± 5,24 | 72,82 ± 3,59 | 63,60 ± 4,80 |
| todas | 48,22 ± 8,68 | 76,55 ± 6,75 | 74,55 ± 4,34 | 64,16 ± 5,24 |

das características (*Gini importance*). Observa-se que as características mais relevantes de acordo com esse critério são as mesmas da Seção III-A, o que reforça a seleção realizada. Nos demais experimentos, os classificadores serão treinados apenas com as características relevantes e para efeitos de comparação, o *baseline* foi o resultado do treinamento utilizando apenas as características relevantes.

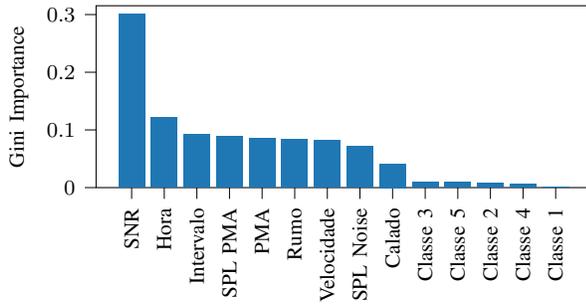


Figura 3. Relevância das características - GINI IMPORTANCE

D. MLP

Os experimentos realizados nesta Seção foram conduzidos de forma sequencial, ou seja, a cada Seção as configurações da rede foram incrementadas. Durante as avaliações iniciais da curva de treinamento, foram definidos os seguintes parâmetros básicos: fator de aprendizado de 0,01, treinamento em lotes de 20 amostras e 1200 épocas.

1) *Configuração da camada intermediária*: A fim de determinar a melhor configuração para a camada intermediária, foi realizada uma busca em grade variando o número de neurônios de 4 a 8 e a função de ativação entre tangente hiperbólica (tanh) e a unidade linear retificada (ReLU).

A Tabela III apresenta os resultados para o conjunto de validação. Observa-se que, para toda a faixa utilizada na busca, os resultados estão dentro da mesma faixa para todas as figuras de mérito. Portanto, qualquer uma das combinações poderia ser utilizada para obter resultados similares. Nas próximas seções, em adição aos parâmetros já utilizados durante este experimento, a camada intermediária foi construída com quatro neurônios e função de ativação tangente hiperbólica, pois essa configuração apresentou a melhor média de falso alarme.

Durante as análises, foi identificado que os resultados das figuras de mérito para o conjunto de treinamento, são significativamente melhores do que as figuras de mérito para o conjunto de validação, indicando um provável problema de sobreajuste (*overfitting*) e a consequente perda da capacidade de generalização do modelo. Essa disparidade entre os conjuntos de treinamento e validação sugere a necessidade de

Tabela III
CONFIGURAÇÃO DA CAMADA INTERMEDIÁRIA, RESULTADOS PARA O CONJUNTO DE VALIDAÇÃO - MLP

| Ativação | Neurônios | Falso Alarme | Sensibilidade | F1-score | AUC |
|----------|-----------|---------------------|---------------------|---------------------|---------------------|
| tanh | 4 | 30,52 ± 8,83 | 68,80 ± 6,36 | 72,30 ± 3,68 | 72,21 ± 3,93 |
| | 5 | 31,04 ± 7,96 | 66,10 ± 7,04 | 71,50 ± 4,53 | 71,98 ± 4,15 |
| | 6 | 34,03 ± 9,00 | 67,03 ± 6,69 | 71,49 ± 4,08 | 71,41 ± 4,69 |
| | 7 | 35,08 ± 5,77 | 66,16 ± 6,73 | 70,61 ± 3,84 | 70,77 ± 3,60 |
| ReLU | 8 | 33,96 ± 8,34 | 68,14 ± 5,90 | 72,28 ± 3,12 | 71,25 ± 4,03 |
| | 4 | 31,04 ± 7,50 | 65,05 ± 7,80 | 70,73 ± 5,40 | 72,14 ± 5,41 |
| | 5 | 30,68 ± 8,83 | 65,27 ± 7,03 | 71,04 ± 5,21 | 73,12 ± 6,21 |
| | 6 | 32,46 ± 5,84 | 66,83 ± 6,44 | 71,69 ± 4,40 | 72,73 ± 4,36 |
| | 7 | 31,39 ± 8,46 | 67,04 ± 7,02 | 72,09 ± 4,82 | 72,35 ± 4,17 |
| | 8 | 37,37 ± 8,24 | 66,84 ± 7,56 | 70,55 ± 4,73 | 70,26 ± 5,10 |

aplicar técnicas de regularização para melhorar o desempenho do modelo em dados não vistos anteriormente.

2) *Regularização*: Neste experimento, foram comparados os resultados do treinamento da rede neural para diferentes valores de regularização na camada intermediária. Os valores testados foram {0,0500; 0,0100; 0,0050; 0,0010; e 0,0005}. A Tabela IV apresenta os resultados para o conjunto de validação.

Tabela IV
REGULARIZAÇÃO, RESULTADOS PARA O CONJUNTO DE VALIDAÇÃO - MLP

| | Falso Alarme | Sensibilidade | F1-score | AUC |
|-------|----------------------|---------------------|---------------------|---------------------|
| 00500 | 30,65 ± 10,64 | 68,34 ± 6,63 | 73,23 ± 5,55 | 77,48 ± 6,38 |
| 00100 | 29,15 ± 10,45 | 67,48 ± 7,16 | 72,93 ± 5,50 | 76,62 ± 5,75 |
| 00050 | 29,91 ± 9,63 | 65,27 ± 8,61 | 71,12 ± 6,23 | 74,86 ± 5,09 |
| 00010 | 31,06 ± 11,33 | 67,70 ± 5,07 | 72,76 ± 3,81 | 73,19 ± 4,66 |
| 00005 | 34,70 ± 7,11 | 68,17 ± 6,10 | 72,13 ± 4,24 | 71,58 ± 5,40 |

Observa-se que a regularização começa a ter um efeito significativo a partir do valor de 0,0050, onde os resultados para o conjunto de treinamento se aproximam mais dos resultados para o conjunto de validação. Para valores menores que 0,0050, os resultados para o conjunto de treinamento ainda são significativamente superiores em relação ao conjunto de validação. No entanto, para dar continuidade ao trabalho, foi escolhida a regularização de 0,0100 por ter obtido a melhor média de falso alarme.

Além disso, foi testada a inserção de uma camada de *Dropout* com valores de {0,1; 0,2; 0,3; 0,4; e 0,5}, porém não foi observada nenhuma diferença significativa nas figuras de mérito. Como os treinamentos sem *Dropout* apresentaram a melhor média essa camada não será utilizada nos próximos experimentos.

3) *Escalonamento*: Nesta Seção, foi realizado um experimento para determinar a melhor técnica de escalonamento a ser utilizada no conjunto de dados, com o objetivo de subsidiar a escolha da técnica a ser empregada na Seção de classificadores de vetor de suporte. Foram testadas três diferentes técnicas de escalonamento disponíveis na biblioteca Scikit-learn [8]: a normalização z-score [14], o escalonamento Min-Max [15] e a normalização pela transformação de Box-Cox [16].

A Tabela V apresenta os resultados das figuras de mérito para as diferentes técnicas de escalonamento. Podemos observar que os resultados estão sobrepostos dentro da margem de

erro para todas as técnicas testadas. Portanto, para dar continuidade às análises, utilizaremos a técnica de normalização por z-score, considerando que ela proporcionou o melhor desempenho em relação às outras técnicas testadas.

Tabela V
ESCALONAMENTO, RESULTADOS PARA O CONJUNTO DE VALIDAÇÃO - MLP

| | Falso Alarme | Sensibilidade | F1-score | AUC |
|---------|---------------------|---------------------|---------------------|---------------------|
| z-score | 30,27 ± 9,37 | 67,48 ± 7,63 | 72,62 ± 5,57 | 76,74 ± 5,63 |
| Min-Max | 31,79 ± 9,81 | 68,36 ± 6,79 | 72,94 ± 5,12 | 77,65 ± 6,23 |
| Box-Cox | 32,89 ± 9,61 | 68,36 ± 7,06 | 72,66 ± 4,74 | 76,33 ± 5,91 |

4) *Definição do limiar de detecção*: Neste experimento, foi analisado o impacto da modificação do limiar de decisão da saída da rede. Normalmente, em uma rede de saída sigmoidal, onde os valores estão limitados ao intervalo de 0 a 1, o limiar de decisão padrão é definido como 0,5, que serve como fronteira para a decisão entre as classes. No entanto, é possível ajustar esse limiar para melhorar o desempenho da rede em relação à taxa de falso alarme.

Quanto mais próximo de 1 for o limiar de decisão, maior será a confiança necessária para que uma classificação seja considerada como positiva. Portanto, aumentando o limiar, é possível reduzir a taxa de falso alarme da rede a custo de perda da sensibilidade.

A Tabela VI apresenta os resultados das figuras de mérito quando os limiares de decisão foram analisados na faixa de 0,5 a 0,9. Note que a AUC não está incluída na tabela, pois ela não é afetada pela modificação do limiar de corte.

Tabela VI
DEFINIÇÃO DO LIMIAR DE DETECÇÃO, RESULTADOS PARA O CONJUNTO DE VALIDAÇÃO - MLP

| | Falso Alarme | Sensibilidade | F1-score |
|---------|---------------------|---------------------|---------------------|
| MLP(05) | 30,27 ± 9,37 | 67,48 ± 7,63 | 72,62 ± 5,57 |
| MLP(06) | 17,52 ± 5,86 | 60,39 ± 6,08 | 70,59 ± 5,23 |
| MLP(07) | 10,03 ± 6,25 | 45,35 ± 6,85 | 59,65 ± 5,60 |
| MLP(08) | 5,58 ± 4,09 | 28,53 ± 9,04 | 42,53 ± 9,92 |
| MLP(09) | 2,22 ± 2,85 | 16,61 ± 7,05 | 27,55 ± 9,92 |

Como esperado, ao aumentar o limiar de corte, houve uma redução significativa no falso alarme, porém, houve uma consequente redução na sensibilidade e no f1-score. Não há uma definição precisa do limiar de corte a ser utilizado com base nas análises realizadas. Para fins de comparação futura neste trabalho, foram considerados três limiares: 0,5; 0,7; e 0,9.

E. Classificadores de vetor de suporte(SVC)

Ao longo desta Seção foram realizados experimentos de busca em grade variando os parâmetros para os modelos SVC e NuSVC. Para o SVC, o parâmetro C , inversamente proporcional à regularização do modelo, foi testado nos valores $\{0,1; 0,5; e 1,0\}$. Para o NuSVC, o parâmetro ν foi testado nos valores $\{0,3; 0,5; e 0,7\}$. Além destes, para cada kernel foram variados com parâmetros específicos, descritos ao longo do trabalho.

1) *Kernel linear*: O experimento conduzido nesta Seção consistiu em uma busca em grade básica para os modelos SVC e NuSVC. Os resultados obtidos são apresentados na Tabela VII. Observa-se que os resultados para SVC são praticamente idênticos, com todos os valores dentro da mesma margem de erro para todas as métricas avaliadas. Por outro lado, os resultados para NuSVC mostram que apenas o valor de $\nu = 0,7$ está dentro da margem dos resultados do SVC, porém, em média, apresenta desempenho inferior em todas as métricas avaliadas.

Tabela VII
SVC LINEAR, RESULTADOS PARA O CONJUNTO DE VALIDAÇÃO

| | Falso Alarme | Sensibilidade | F1-score | AUC |
|--------------------|----------------------|---------------------|---------------------|----------------------|
| SVC(C=0,1) | 31,39 ± 10,94 | 69,24 ± 6,11 | 73,70 ± 5,34 | 68,92 ± 6,98 |
| SVC(C=0,5) | 32,13 ± 8,47 | 69,24 ± 6,47 | 73,47 ± 5,23 | 68,55 ± 6,02 |
| SVC(C=1,0) | 30,99 ± 9,49 | 69,90 ± 6,31 | 74,18 ± 4,84 | 69,45 ± 5,84 |
| NuSVC($\nu=0,3$) | 57,54 ± 11,82 | 57,85 ± 24,64 | 57,73 ± 18,79 | 50,16 ± 11,58 |
| NuSVC($\nu=0,5$) | 47,64 ± 15,66 | 59,69 ± 21,72 | 61,51 ± 18,41 | 56,02 ± 12,98 |
| NuSVC($\nu=0,7$) | 52,24 ± 8,45 | 86,50 ± 5,71 | 79,51 ± 3,82 | 67,13 ± 5,28 |

Para fins de comparação com as demais análises, a melhor configuração desta Seção é o SVC com um valor de C igual a 1,0, que apresentou a melhor média de falso alarme para o conjunto de validação.

2) *Kernel RBF*: O experimento realizado nesta Seção incluiu uma busca em grade básica combinada com a variação do parâmetro γ , que controla a largura do kernel. Os valores testados foram $\{0,0625; 0,1250; 0,2500; 0,5000; e 1,0000\}$. Devido ao grande número de combinações geradas pela busca em grade, as tabelas de resultados completos não foram incluídas neste documento devido ao seu tamanho, o mesmo ocorreu para os próximos kernels.

A Tabela VIII apresenta a comparação entre o melhor resultado, de falso alarme, obtido para SVC ($C = 1$ e $\gamma = 0,625$) e o melhor resultado obtido para NuSVC ($\nu = 0,3$ e $\gamma = 1$). É possível observar que os valores das figuras de mérito são marginalmente iguais, estando todos os resultados dentro da mesma margem de erro, indicando que ambos os modelos têm desempenho semelhante para as métricas avaliadas.

Tabela VIII
SVC RBF, RESULTADOS PARA O CONJUNTO DE VALIDAÇÃO

| | Falso Alarme | Sensibilidade | F1-score | AUC |
|-----------|---------------------|---------------------|---------------------|---------------------|
| SVC RBF | 28,41 ± 8,78 | 68,57 ± 5,69 | 73,92 ± 4,30 | 70,08 ± 5,19 |
| NuSVC RBF | 43,66 ± 7,46 | 71,90 ± 5,65 | 72,63 ± 3,70 | 64,12 ± 4,17 |

Com base nos resultados obtidos para o conjunto de validação, considera-se que a melhor configuração desta Seção é o SVC com os parâmetros $C = 1$ e $\gamma = 0,625$, que apresentou a melhor média de falso alarme para o conjunto de validação.

3) *Kernel Sigmoidal*: Neste experimento, foi realizada uma busca em grade básica para toda a Seção, variando os parâmetros γ nos valores $\{0,0625; 0,1250; 0,2500; 0,5000; e 1,0000\}$, bem como o parâmetro coeficiente, termo independente, nos valores $\{-1; 0; e 1\}$.

A Tabela VIII apresenta a comparação do melhor resultado para SVC com $C = 0,1$, $\gamma = 0,1250$ e coeficiente de 1 e

o melhor resultado para NuSVC com $\nu = 0,5$, $\gamma = 0,0625$ e coeficiente de 0. Observa-se que, embora haja similaridade nos resultados de f1-score e AUC, o SVC é superior em termos de falso alarme enquanto o NuSVC é superior na sensibilidade.

Tabela IX
SVC SIGMOIDAL, RESULTADOS PARA O CONJUNTO DE VALIDAÇÃO

| | Falso Alarme | Sensibilidade | F1-score | AUC |
|-----------------|----------------------|----------------------|---------------------|---------------------|
| SVC sigmoidal | 11,89 ± 4,78 | 54,61 ± 12,17 | 66,80 ± 10,05 | 71,36 ± 6,39 |
| NuSVC sigmoidal | 42,23 ± 11,96 | 75,87 ± 6,54 | 75,49 ± 4,44 | 66,82 ± 6,25 |

Para fins de comparação com as demais análises, foi considerado o SVC com $C = 0,1$, $\gamma = 0,1250$ e coeficiente de 1 que apresentou um desempenho melhor na taxa de falso alarme. O resultado para NuSVC foi desconsiderado por estar dentro da mesma faixa do baseline em todas as figuras de mérito.

4) *Kernel Polinomial*: A busca em grade realizada para a Seção incluiu a variação dos parâmetros de grau do polinômio nos valores {3; 4; 5; e 6} e o parâmetro coeficiente nos valores {-1; 0; e 1}.

A Tabela X apresenta a comparação do melhor resultado para SVC ($C = 0,1$, grau= 4 e coeficiente = 0) e o melhor resultado para NuSVC ($\nu = 0,5$, grau= 4 e coeficiente = -1). Nota-se que todos os resultados são significativamente diferentes, sendo o SVC melhor para falso alarme, enquanto o NuSVC para as demais figuras de mérito.

Tabela X
SVC POLINOMIAL, RESULTADOS PARA O CONJUNTO DE VALIDAÇÃO

| | Falso Alarme | Sensibilidade | F1-score | AUC |
|------------------|---------------------|---------------------|---------------------|---------------------|
| SVC polinomial | 3,74 ± 4,07 | 11,75 ± 6,26 | 19,99 ± 9,68 | 54,00 ± 2,77 |
| NuSVC polinomial | 19,78 ± 8,65 | 60,15 ± 7,87 | 69,74 ± 6,44 | 70,18 ± 5,47 |

Para fins de comparação com as demais análises, é válido considerar ambas as técnicas, uma vez que o SVC com $C = 0,1$, grau= 4 e coeficiente = 0, $\gamma = 0,1250$ e coeficiente de 1 mostrou-se significativamente superior de falsos alarmes. Enquanto o NuSVC com $\nu = 0,5$, grau= 4 e coeficiente = -1 obteve figuras de mérito melhores e uma taxa de falso alarme significativamente inferior aos resultados da baseline. Portanto, ao considerar múltiplas figuras de mérito, ambas as técnicas podem ser relevantes para análises e comparações futuras.

F. Comparação de resultados

A Tabela XI apresenta uma comparação dos resultados obtidos pelos modelos selecionados ao longo das seções anteriores deste trabalho para o conjunto de validação.

Ao analisar os resultados, é possível identificar três grupos distintos de modelos.

O primeiro grupo consiste nos modelos que apresentaram desempenho similar ao baseline. Esses modelos incluem o MLP com limiar de corte de 0,5, o SVC linear e o SVC RBF. Embora o MLP tenha obtido uma pontuação AUC significativamente melhor do que o baseline, os demais resultados não foram significativamente diferentes.

O segundo grupo é composto pelos modelos que alcançaram taxas muito baixas de falsos alarmes (em média, menos de

Tabela XI
COMPARAÇÃO DE RESULTADOS PARA O CONJUNTO DE VALIDAÇÃO

| | Falso Alarme | Sensibilidade | F1-score | AUC |
|------------------|---------------------|----------------------|---------------------|---------------------|
| baseline | 45,59 ± 8,69 | 72,78 ± 5,24 | 72,82 ± 3,59 | 63,60 ± 4,80 |
| MLP(05) | 30,27 ± 9,37 | 67,48 ± 7,63 | 72,62 ± 5,57 | 76,74 ± 5,63 |
| MLP(07) | 10,03 ± 6,25 | 45,35 ± 6,85 | 59,65 ± 5,60 | 76,74 ± 5,63 |
| MLP(09) | 2,22 ± 2,85 | 16,61 ± 7,05 | 27,55 ± 9,92 | 76,74 ± 5,63 |
| SVC linear | 30,99 ± 9,49 | 69,90 ± 6,31 | 74,18 ± 4,84 | 69,45 ± 5,84 |
| SVC RBF | 28,41 ± 8,78 | 68,57 ± 5,69 | 73,92 ± 4,30 | 70,08 ± 5,19 |
| SVC sigmoidal | 11,89 ± 4,78 | 54,61 ± 12,17 | 66,80 ± 10,05 | 71,36 ± 6,39 |
| SVC polinomial | 3,74 ± 4,07 | 11,75 ± 6,26 | 19,99 ± 9,68 | 54,00 ± 2,77 |
| NuSVC polinomial | 19,78 ± 8,65 | 60,15 ± 7,87 | 69,74 ± 6,44 | 70,18 ± 5,47 |

4%) e baixa sensibilidade (em média, menor que 20%). Esses modelos são o MLP com limiar de corte de 0,9 e o SVC polinomial. Esses modelos se destacam por sua capacidade de minimizar os falsos alarmes, mas, em contrapartida, apresentam uma menor sensibilidade na detecção de amostras positivas.

O terceiro grupo é composto pelo MLP com limiar de corte de 0,7, o SVC sigmoidal e o NuSVC polinomial. Esses modelos apresentam baixas taxas de falsos alarmes (média entre 10% e 20%) e uma sensibilidade moderada (na faixa de 45% a 60% em média). Esses modelos conseguem um equilíbrio entre a taxa de falsos alarmes e a capacidade de detecção das amostras positivas. Vale destacar que o NuSVC está nesse grupo mas poderia ser considerado como a transição do terceiro e do primeiro grupo.

A Figura 4 exibe a análise da taxa de falso alarme no conjunto de validação utilizando um gráfico de violino. É possível identificar as diferenças nos desempenhos e nas distribuições dos modelos, contribuindo para a visualização dos grupos.

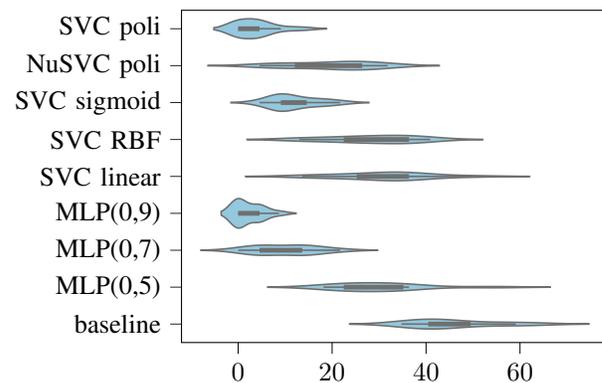


Figura 4. Comparação de resultados, gráfico de violino para o conjunto de validação da figura de mérito falso alarme

G. Ensemble

Na busca por melhorar a performance e a capacidade de generalização dos modelos, foram realizados experimentos com modelos de *ensemble learning*, que combinam os melhores classificadores obtidos anteriormente. Duas abordagens de *ensemble* foram testadas: votação majoritária (*majority voting ensemble*) e empilhamento de modelos (*stacking ensemble*).

No caso da votação majoritária, foram testadas duas combinações de modelos. A primeira, denominada votação A,

consiste na combinação dos modelos com melhores resultados de falsos alarmes: o MLP com limiar de decisão de 0,9, o SVC polinomial e o SVC sigmoidal. A segunda, chamada de votação B, utiliza os resultados do terceiro grupo de análises da Seção III-F, ou seja, o MLP com limiar de decisão de 0,7, o NuSVC polinomial e o SVC sigmoidal.

Já no *stacking ensemble*, foram empregadas as mesmas combinações de modelos utilizadas na votação majoritária, porém, em vez de combinar as saídas dos modelos após o limiar de decisão, os modelos foram treinados utilizando as saídas originais da MLP antes da aplicação do limiar de decisão. Essas combinações foram denominadas empilhamento A e empilhamento B. Além disso, os resultados das MLPs de saída foram avaliados para diferentes limiares de decisão.

A Tabela XII apresenta os resultados obtidos por esses modelos aplicados ao conjunto de validação. Essa tabela permite avaliar o desempenho dos modelos de *ensemble* em relação às métricas de avaliação para cada uma das combinações mencionadas.

Tabela XII
ENSEMBLE, RESULTADOS PARA O CONJUNTO DE VALIDAÇÃO

| | Falso Alarme | Sensibilidade | FI-score | AUC |
|--------------------|---------------|---------------|---------------|--------------|
| MLP(07) | 10,03 ± 6,25 | 45,35 ± 6,85 | 59,65 ± 5,60 | 76,74 ± 5,63 |
| MLP(09) | 2,22 ± 2,85 | 16,61 ± 7,05 | 27,55 ± 9,92 | 76,74 ± 5,63 |
| SVC sigmoidal | 11,89 ± 4,78 | 54,61 ± 12,17 | 66,80 ± 10,05 | 71,36 ± 6,39 |
| SVC polinomial | 3,74 ± 4,07 | 11,75 ± 6,26 | 19,99 ± 9,68 | 54,00 ± 2,77 |
| NuSVC polinomial | 19,78 ± 8,65 | 60,15 ± 7,87 | 69,74 ± 6,44 | 70,18 ± 5,47 |
| votação A | 3,36 ± 3,22 | 19,48 ± 7,99 | 31,29 ± 10,69 | 58,06 ± 3,48 |
| votação B | 12,25 ± 5,32 | 55,28 ± 8,85 | 67,61 ± 7,41 | 71,51 ± 4,96 |
| empilhamento A(05) | 24,29 ± 12,78 | 65,68 ± 5,04 | 72,97 ± 4,37 | 76,77 ± 5,24 |
| empilhamento A(07) | 12,24 ± 6,91 | 47,61 ± 12,21 | 60,45 ± 10,67 | 76,77 ± 5,24 |
| empilhamento A(09) | 1,88 ± 2,89 | 8,85 ± 5,02 | 15,71 ± 8,50 | 76,77 ± 5,24 |
| empilhamento B(05) | 23,16 ± 10,11 | 65,92 ± 5,10 | 73,37 ± 4,38 | 76,39 ± 5,70 |
| empilhamento B(07) | 10,75 ± 5,75 | 49,80 ± 10,66 | 63,08 ± 9,24 | 76,39 ± 5,70 |
| empilhamento B(09) | 0,00 ± 0,00 | 0,45 ± 1,49 | 0,85 ± 2,83 | 76,39 ± 5,70 |

Para o *majority voting ensemble*, observa-se que os resultados para a combinação votação A estão dentro da faixa dos resultados do segundo grupo de análises, apresentando, em média, o melhor desempenho em sensibilidade. Já para a combinação votação B, os resultados ficaram dentro da faixa do terceiro grupo.

Ao analisar os resultados do *stacking ensemble*, nota-se que tanto para a abordagem A quanto para a abordagem B, os resultados estão dentro da faixa ou piores em comparação com a MLP simples. Isso indica que o *stacking ensemble* não proporcionou uma melhoria significativa para o conjunto de validação avaliado.

A Tabela XII apresenta os resultados para esses modelos aplicados ao conjunto de teste. Observa-se que os melhores resultados foram obtidos pelas MLPs com limiar de decisão de 0,7, tanto para a MLP simples quanto para as abordagens de *stacking*, tendo métricas na mesma faixa que o SVC sigmoidal e a votação B.

IV. DISCUSSÃO

Os resultados obtidos estão alinhados com os objetivos do estudo, que visam obter uma baixa taxa de falso alarme para filtrar as detecções do projeto PMPAS-BS e selecionar apenas aquelas com alta confiança de que representam o efeito

Tabela XIII
ENSEMBLE, RESULTADOS PARA O CONJUNTO DE TESTE

| | Falso Alarme | Sensibilidade | FI-score | AUC |
|--------------------|---------------|---------------|---------------|--------------|
| MLP(07) | 0,00 ± 0,00 | 32,69 ± 3,33 | 49,18 ± 3,68 | 64,29 ± 1,59 |
| MLP(09) | 0,00 ± 0,00 | 17,95 ± 7,90 | 29,66 ± 11,64 | 64,29 ± 1,59 |
| SVC sigmoidal | 2,38 ± 5,32 | 35,26 ± 3,79 | 51,53 ± 3,91 | 66,44 ± 2,62 |
| SVC polinomial | 0,00 ± 0,00 | 23,72 ± 2,13 | 38,30 ± 2,64 | 61,86 ± 1,06 |
| NuSVC polinomial | 11,90 ± 5,32 | 38,46 ± 0,00 | 53,12 ± 1,09 | 63,28 ± 2,66 |
| votação A | 0,00 ± 0,00 | 21,79 ± 4,25 | 35,59 ± 5,77 | 60,90 ± 2,13 |
| votação B | 2,38 ± 5,32 | 35,26 ± 3,79 | 51,53 ± 3,91 | 66,44 ± 2,62 |
| empilhamento A(05) | 9,52 ± 12,14 | 47,44 ± 6,15 | 62,00 ± 4,46 | 67,67 ± 2,48 |
| empilhamento A(07) | 0,00 ± 0,00 | 34,62 ± 4,97 | 51,22 ± 5,69 | 67,67 ± 2,48 |
| empilhamento A(09) | 0,00 ± 0,00 | 16,03 ± 3,79 | 27,44 ± 5,63 | 67,67 ± 2,48 |
| empilhamento B(05) | 13,10 ± 12,31 | 44,23 ± 5,55 | 58,31 ± 3,98 | 64,65 ± 1,79 |
| empilhamento B(07) | 0,00 ± 0,00 | 32,05 ± 2,87 | 48,47 ± 3,17 | 64,65 ± 1,79 |
| empilhamento B(09) | 0,00 ± 0,00 | 1,28 ± 4,25 | 2,22 ± 7,37 | 64,65 ± 1,79 |

de interferência. Dentre os classificadores propostos alguns obtiveram zero falsos negativos com diferentes resultados de sensibilidade.

A próxima etapa natural do trabalho é aplicar o modelo desenvolvido nas detecções do projeto PMPAS-BS para verificar se a quantidade de detecções classificadas como positivas atende às necessidades do projeto na construção do *dataset*. O melhor resultado obtido sugere a utilização direta do MLP simples, ajustando um limiar de decisão alto (por exemplo, 0,9) e reduzindo gradualmente esse limiar até que o número de detecções classificadas como positivas seja suficiente para a construção do *dataset*.

Para a construção do *dataset*, ainda é necessária uma etapa de confirmação da identificação do navio. Isso pode ser feito utilizando métodos de redução de dimensionalidade, como o t-SNE, para verificar se os espectrogramas gerados a partir do PMA dos navios geram espaços de classes separáveis. Outra opção é a utilização de um classificador, onde algumas das detecções teriam que ser manualmente classificadas para a construção de um modelo de classificação de navios que possa auxiliar na decisão de inclusão ou não das detecções no *dataset*.

É importante ressaltar que os resultados obtidos neste estudo foram limitados pela baixa quantidade de dados manualmente classificados, o que resultou em um conjunto de teste muito pequeno, com apenas 20 amostras. Portanto, estudos futuros podem ser realizados com uma maior quantidade de dados classificados manualmente, visando melhorar o desempenho dos modelos desenvolvidos.

V. CONCLUSÃO

Nesse estudo foram treinados diversos classificadores para identificação da existência do efeito de interferência de Lloyd's mirror com base na tabela de detecções do projeto PMPAS-BS. Objetivando a construção de modelos capazes de filtrar a tabela com mais de 88 mil detecções do projeto, aumentando a confiança de que a detecção equivalha ao PMA de um navio.

Apresentou-se os resultados e as comparações entre os modelos propondo a utilização do MLP variando o limiar de decisão para atender a necessidade do projeto. Esse trabalho contribuiu para a separação de dados para a construção de um *dataset* confiável, tendo atingido resultados satisfatórios para o projeto.

VI. AGRADECIMENTO

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001

REFERÊNCIAS

- [1] W. S. Filho, C. Galotta e A. Bazyl, “6º Relatório Técnico do PMPAS”, PETROBRAS e IPQM, Relatório Técnico 6, ago. de 2022, p. 242. URL: https://comunicaciadesantos.petrobras.com.br/sites/default/files/6RT_PMPAS_Final_Assinado_17.11.22.pdf (acedido em 11/06/2023).
- [2] J. d. C. V. Fernandes, N. N. de Moura Junior e J. M. de Seixas, “Deep Learning Models for Passive Sonar Signal Classification of Military Data”, en, *Remote Sensing*, vol. 14, n.º 11, p. 2648, jan. de 2022, Number: 11 Publisher: Multidisciplinary Digital Publishing Institute, ISSN: 2072-4292. DOI: 10.3390/rs14112648. URL: <https://www.mdpi.com/2072-4292/14/11/2648> (acedido em 04/05/2023).
- [3] M. Irfan, Z. Jiangbin, S. Ali, M. Iqbal, Z. Masood e U. Hamid, “DeepShip: An underwater acoustic benchmark dataset and a separable convolution based autoencoder for classification”, en, *Expert Systems with Applications*, vol. 183, p. 115270, nov. de 2021, ISSN: 0957-4174. DOI: 10.1016/j.eswa.2021.115270. URL: <https://www.sciencedirect.com/science/article/pii/S0957417421007016> (acedido em 20/03/2023).
- [4] D. Santos-Domínguez, S. Torres-Guijarro, A. Cardenal-López e A. Pena-Gimenez, “ShipsEar: An underwater vessel noise database”, en, *Applied Acoustics*, vol. 113, pp. 64–69, dez. de 2016, ISSN: 0003-682X. DOI: 10.1016/j.apacoust.2016.06.008. URL: <https://www.sciencedirect.com/science/article/pii/S0003682X16301566> (acedido em 20/03/2023).
- [5] D. F. G. D. de Andrade, *Determinação de parâmetros de alvos submarinos a partir de interferências devidas ao efeito lloyd-mirror*, abr. de 2020. URL: https://w1files.solucaoatrio.net.br/atrio/ufrij-peno_upl/THESIS/10001357/2020_diego_felipe_gimezes_de_andrade_dissert_vf_24042020_2021022321401691.pdf (acedido em 23/05/2023).
- [6] F. B. Jensen, ed., *Computational ocean acoustics*, Corr. 2nd print, AIP series in modern acoustics and signal processing, OCLC: ocm44623053, New York: AIP Press : Springer, 2000, 578 pp.
- [7] L. van der Maaten e G. Hinton, “Visualizing data using t-SNE”, *Journal of Machine Learning Research*, vol. 9, pp. 2579–2605, nov. de 2008.
- [8] F. Pedregosa, G. Varoquaux, A. Gramfort et al., “Scikit-learn: Machine Learning in Python”, *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [9] F. Chollet et al., *Keras*, <https://keras.io>, 2015.
- [10] F. Oliveira, *Biblioteca de treinamento para machine learning*, <https://github.com/obs-fabio/ml>, Acesso em: 11 de junho de 2023.
- [11] S. Kumar, B. Bhushan, L. Bhambhu, M. Thakur, U. M. Mohapatra e D. K. Choubey, “Medical Datasets Classification using a Hybrid Genetic Algorithm for Feature Selection based on Pearson Correlation Coefficient”, em *2022 International Conference on Machine Learning, Computer Systems and Security (MLCSS)*, ago. de 2022, pp. 214–218. DOI: 10.1109/MLCSS57186.2022.00047.
- [12] Y. Fan, L. Xuan, Z. Qifeng e L. Linkai, “Margin based variable importance for random forest”, em *2011 6th International Conference on Computer Science & Education (ICCSE)*, ago. de 2011, pp. 1361–1366. DOI: 10.1109/ICCSE.2011.6028885.
- [13] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever e R. Salakhutdinov, “Dropout: a simple way to prevent neural networks from overfitting”, *The Journal of Machine Learning Research*, vol. 15, n.º 1, pp. 1929–1958, 1 de jan. de 2014, ISSN: 1532-4435.
- [14] D. Singh e B. Singh, “Investigating the impact of data normalization on classification performance”, *Applied Soft Computing*, vol. 97, p. 105524, 1 de dez. de 2020, ISSN: 1568-4946. DOI: 10.1016/j.asoc.2019.105524. URL: <https://www.sciencedirect.com/science/article/pii/S1568494619302947> (acedido em 17/06/2023).
- [15] S. K. Patro e K. K. Sahu, “Normalization: A preprocessing stage”, *IARJSET*, pp. 20–22, 20 de mar. de 2015, ISSN: 23938021. DOI: 10.17148/IARJSET.2015.2305. URL: <http://www.iarjset.com/upload/2015/march-15/IARJSET%205.pdf> (acedido em 17/06/2023).
- [16] F. Malik, “BOX-COX TRANSFORMATION APPROACH FOR DATA NORMALIZATION: A STUDY OF NEW PRODUCT DEVELOPMENT IN MANUFACTURING SECTOR OF PAKISTAN”, *IBT Journal of Business Studies*, vol. 14, pp. 110–119, 1 de jan. de 2018. DOI: 10.46745/ilma.jbs.2018.14.01.09.
- [17] D. Mienye e Y. Sun, “A Survey of Ensemble Learning: Concepts, Algorithms, Applications, and Prospects”, *IEEE Access*, vol. PP, pp. 1–1, 16 de set. de 2022. DOI: 10.1109/ACCESS.2022.3207287.
- [18] T. Fawcett, “An introduction to ROC analysis”, *Pattern Recognition Letters*, vol. 27, n.º 8, pp. 861–874, jun. de 2006, ISSN: 01678655. DOI: 10.1016/j.patrec.2005.10.010. URL: <https://linkinghub.elsevier.com/retrieve/pii/S016786550500303X> (acedido em 18/06/2023).
- [19] A. Géron, *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: concepts, tools, and techniques to build intelligent systems*, Second edition. Beijing [China] ; Sebastopol, CA: O’Reilly Media, Inc, 2019, 819 pp., ISBN: 978-1-4920-3264-9.
- [20] C. Sammut e G. I. Webb, eds., *Encyclopedia of machine learning*, OCLC: ocn651073009, New York ; London: Springer, 2010, 1030 pp., ISBN: 978-0-387-30768-8 978-0-387-34558-1 978-0-387-30164-8.