

Bearing fault diagnosis using machine learning and a novel set of fault-related spectral features

João Paulo Vieira¹, Rodrigo Kobashikawa Rosa¹, Victor Afonso Bauler², Danilo Braga³,
Danilo Silva¹

¹Departamento de Engenharia Elétrica e Eletrônica
Universidade Federal de Santa Catarina (UFSC)
Caixa Postal 5064 – 88.035-972 – Florianópolis – SC – Brazil

²Departamento de Engenharia Mecânica
Universidade Federal de Santa Catarina (UFSC)

³Dynamox
Caixa Postal 5064 – 88.035-972 – Florianópolis – SC – Brazil

{joaopaulovieirajpv, rodrigokrosa, victor.bauler}@gmail.com

danilo@dynamox.net, danilo.silva@ufsc.br

Abstract. *This article develops methods based on machine learning for fault diagnosis in bearings using data from Case Western Reserve University (CWRU). A multi-label approach proposed in the literature is adopted, with three binary classifiers to identify faults in the inner race, outer race, and rolling element, dividing the dataset by distinct bearings to prevent leakage. Features in the time and frequency domains are used, including a new proposal of spectral features called PFRs (Peak-To-Floor Ratios), calculated over frequencies related to the fault phenomenon. The results, compared with other features from the literature, show improvements in performance and explainability.*

Resumo. *Este artigo desenvolve métodos baseados em aprendizado de máquina para diagnóstico de falhas em rolamentos com dados da Case Western Reserve University (CWRU). Adota-se uma abordagem multirrotulo proposta na literatura, com três classificadores binários para identificar falhas na trilha interna, externa e no elemento rolante, dividindo o conjunto de dados por rolamentos distintos, visando eliminar vazamento. Utilizam-se features no domínio do tempo e frequência, incluindo uma nova proposta de features espectrais, denominadas PFRs (Peak-To-Floor Ratios), calculadas sobre frequências ligadas ao fenômeno da falha. Os resultados, comparados com outras features da literatura, mostram melhorias no desempenho e explicabilidade.*

1. Introdução

O monitoramento da condição de máquinas é essencial na Manutenção baseada em Condição (CBM - *Condition-Based Maintenance*), uma estratégia eficiente para prevenir falhas catastróficas, que podem causar acidentes, perda de produção e altos custos de reparo [Randall 2021]. Dentro desse contexto, a análise de vibração destaca-se como uma solução eficaz, com o advento de sensores IoT, que permitem monitoramento remoto e instantâneo, superando o uso de acelerômetros cabeados [Randall 2021].

Dado o grande volume de medições em ativos industriais, métodos baseados em aprendizado de máquina são uma alternativa para acelerar e facilitar a classificação de sinais de vibração, proporcionando análises rápidas e eficientes [Ahmed and Nandi 2020]. Diversos métodos foram propostos na literatura, desde abordagens tradicionais [Abburi et al. 2023], que envolvem extração de *features* dos sinais, até abordagens de *deep learning*, testadas em diferentes bases de dados [Neupane and Seok 2020, Hendriks et al. 2022].

O enfoque deste trabalho é a detecção de falha em rolamentos através de sinais de vibração, e por isso escolheu-se trabalhar com o conjunto de dados da *Case Western Regional University* (CWRU), analisado a fundo em [Smith and Randall 2015]. Esse conjunto abrange uma série de aquisições com rolamentos em diferentes estágios de falha inseridos artificialmente, com diferentes cargas e localizações. Em alguns trabalhos que utilizam o CWRU, como [Hendriks et al. 2022], [Abburi et al. 2023] e [Rosa et al. 2024], são propostas diferentes divisões do conjunto, com intuito de maximizar o desempenho do modelo enquanto garantem que não haja vazamento de dados, um problema comum no treinamento de modelos de aprendizado de máquina [Kapoor and Narayanan 2023].

Em [Abburi et al. 2023], sugere-se uma divisão dos dados com base nos rolamentos para evitar vazamento de informações. O trabalho argumenta que não separar os rolamentos dificulta a generalização dos modelos, pois eles tendem a aprender características específicas de cada rolamento, em vez de focar na detecção de falhas. A solução proposta é dividir os dados pelo tamanho da falha, garantindo que o modelo seja avaliado em rolamentos diferentes dos usados no treinamento. No entanto, essa abordagem é aplicada apenas aos sinais de falha, enquanto as amostras normais são distribuídas entre treino, validação e teste, o que gera resultados excessivamente otimistas, como observado em [Rosa et al. 2024].

Desse modo, [Rosa et al. 2024] propõe um novo *split*, mantendo as amostras normais apenas no conjunto de teste, o que impede o treinamento em uma abordagem de classificação binária com rótulos “falha” e “normalidade”. Ao invés disso, utiliza-se uma abordagem multirrótulo, que detecta independentemente falhas na trilha interna (*inner race*), externa (*outer race*) ou no elemento rolante (*ball*). O problema é reformulado para treinar três modelos binários, cada um focado em um tipo específico de falha. A abordagem de [Rosa et al. 2024] utiliza apenas modelos de *deep learning*, com representações 1D ou 2D dos sinais de vibração, alcançando bons resultados e maior generalização.

Neste artigo, investigamos o uso de métodos tradicionais de aprendizado de máquina sob a metodologia de divisão de conjuntos de dados sem vazamento proposta em [Rosa et al. 2024], o que ainda não foi feito na literatura. Especificamente, avaliamos o desempenho dos modelos *Logistic Regression*, *Support Vector Machines* (SVM), *Random Forest* (RF), *Extreme Gradient Boost* (XGBoost) e *Gaussian Naive Bayes* (GNB), em conjunto com extração de *features* de tempo e frequência.

Inicialmente, o desempenho dos modelos é avaliado usando as *features* de [Abburi et al. 2023], que consistem de métricas estatísticas extraídas do sinal no tempo (como média, mediana absoluta, desvio padrão, curtose, RMS e etc.) e as *frequency bins* da FFT dos sinais. Além delas, utilizam-se outras *features* do tempo (valor pico-a-pico, fator de forma, fator de impulso e etc.), aplicadas à velocidade e o envelope, como feito

em [Barreto et al. 2023]. Na frequência, são consideradas as frequências de falha associadas a rolamentos, segundo [González et al. 2023]. Especificamente, propõem-se novas *features*, as PFRs (*Peak-to-Floor Ratios*), desenvolvidas para expressar o quão ressaltadas estão as frequências de falha no espectro, servindo como indicativo de falha para o modelo. Desse modo, agregam-se as magnitudes e PFRs das frequências de falha.

2. Metodologia

2.1. Conjunto de dados: CWRU

O CWRU é um conjunto de dados que reúne sinais de vibração de rolamentos em diferentes estágios de falha (defeitos de 0.007”, 0.014” e 0.021”) e cargas aplicadas (1HP, 2HP, 3HP), com falhas localizadas na trilha interna, externa ou elemento rolante. O conjunto também inclui sinais de normalidade, obtidos em duas posições do motor: o lado de acionamento (“drive end”) e o lado da ventoinha (“fan end”). Para as configurações em que é utilizado um rolamento com falha, tem-se um rolamento saudável no lado oposto. Desse modo, optou-se por utilizar, neste trabalho, um conjunto de amostras de normalidade compostos pelos sinais das configurações com rolamentos saudáveis (com cargas 1HP, 2HP e 3HP), tão como os sinais de normalidade opostos às aquisições de falha. A Figura 1 mostra a organização dos sinais, onde FE, DE e HP se referem, respectivamente, a *fan end*, *drive end* e *Horsepower*.

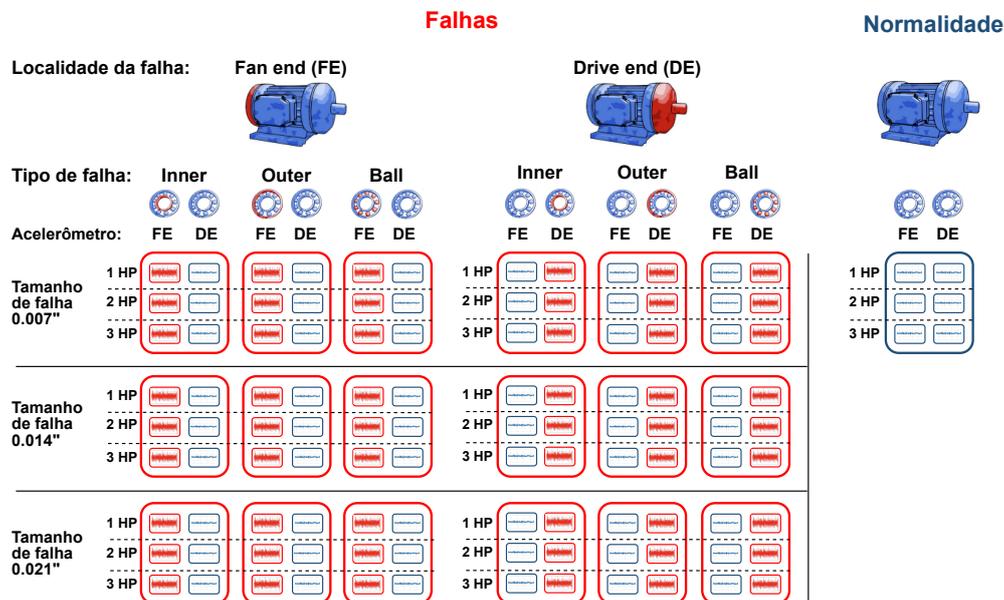


Figura 1. Divisão das aquisições do CWRU por localização, tipo e tamanho de falha e cargas distintas, em que os sinais em azul representam normalidade, enquanto os em vermelho, falha. Adaptado de [Rosa et al. 2024].

2.2. Pré-processamento dos sinais de entrada

Como cada configuração do CWRU tem uma aquisição de cerca de 10 segundos, os sinais foram segmentados para ampliar o conjunto de dados. Seguindo [Abburi et al. 2023], cada sinal foi dividido em segmentos de 2048 amostras com 25% de sobreposição, expandindo o conjunto em 78 vezes. Considerando a rotação em torno de 1800 RPM, trabalhar

com segmentos de 170ms (2048 amostras) preserva as informações espectrais associadas às frequências de falha. Para sinais de baixa rotação, segmentos mais longos seriam necessários para capturar períodos completos de rotação.

2.3. Extração de features

Nesta seção será feita uma revisão de *features* já apresentadas na literatura. Em sequência, será definida uma nova *feature* desenvolvida neste trabalho, chamada *Peak-to-Floor Ratio* (PFR).

2.3.1. Revisão da Literatura

Na revisão realizada exploraram-se trabalhos associados a detecção de falha em rolamentos com modelos de aprendizado de máquina, em que são discutidos conceitos associados à sinais de vibração e apresentadas *features* do domínio do tempo e da frequência.

2.3.1.1 Features do domínio do tempo

Em [Abburi et al. 2023] são utilizadas métricas estatísticas retiradas dos sinais no domínio do tempo, como a média, mediana absoluta, desvio padrão, assimetria, curtose, fator de crista, energia, valor RMS, número de picos, número de *zero-crossings*, teste de Shapiro e divergência KL, descritas em [Juodelyte et al. 2022]. Em outros trabalhos, como o de [Barreto et al. 2023], incluem-se *features* como valor pico-a-pico, fator de impulso, forma e folga e fator de crista *plus*, aplicados à outras representações de sinal além da aceleração, como velocidade e envelope. Observam-se resultados positivos com essas *features* em ambos os trabalhos, e por isso optou-se por utilizá-las no treinamento de modelos.

2.3.1.2 Features do dominio da frequencia

Quanto ao domínio da frequência, [Abburi et al. 2023] usa as *frequency bins* retiradas da FFT dos segmentos de 2048 amostras. Já [González et al. 2023] utiliza as frequências de falha dos rolamentos, como a BPFi (*Ballpass frequency, inner race*), BPFO (*Ballpass frequency, outer race*) e BSF (*Ball spin frequency*). Igualmente, ambos os trabalhos apresentaram bons resultados para as *features* apresentadas, o que motivou a sua incorporação.

2.3.1.3 Frequências de falha de um rolamento

As frequências de falha de um rolamento são informações que podem ser calculadas a partir de sua geometria e velocidade de rotação, aparecendo como impulsos no espectro, com cálculos e conceito descritos detalhadamente em [Randall 2021]. A Figura 2 relaciona os sinais no tempo associados a cada parte do rolamento com suas respectivas frequências de falha.

Considerando a presença de falhas moduladas ao longo do espectro, faz-se a extração do espectro do envelope para trazê-las à banda base - como explicado em

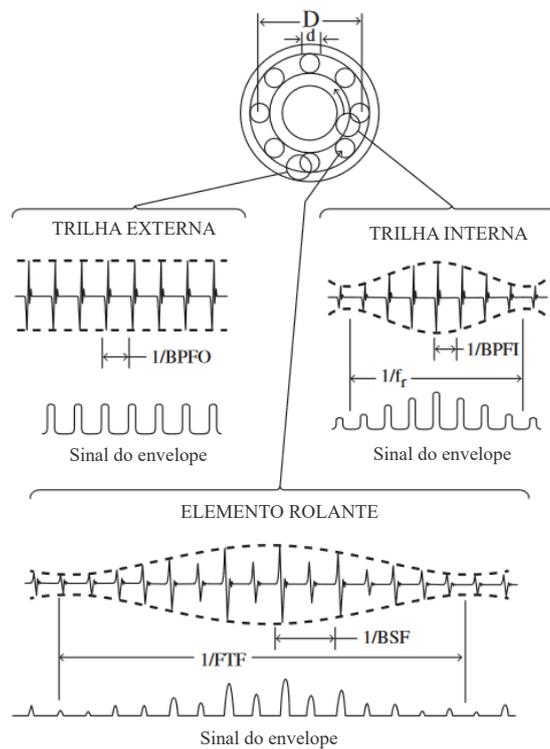


Figura 2. Sinais típicos extraídos de localidades de falha em rolamentos. Adaptado de [Randall 2021].

[Smith and Randall 2015] - e então aplicar algoritmos para identificar magnitudes de picos no espectro. Neste trabalho optou-se por fazer a extração em duas bandas, de 5 Hz - 3kHz e 3kHz - 6kHz. A Figura 3 ilustra a ideia proposta.

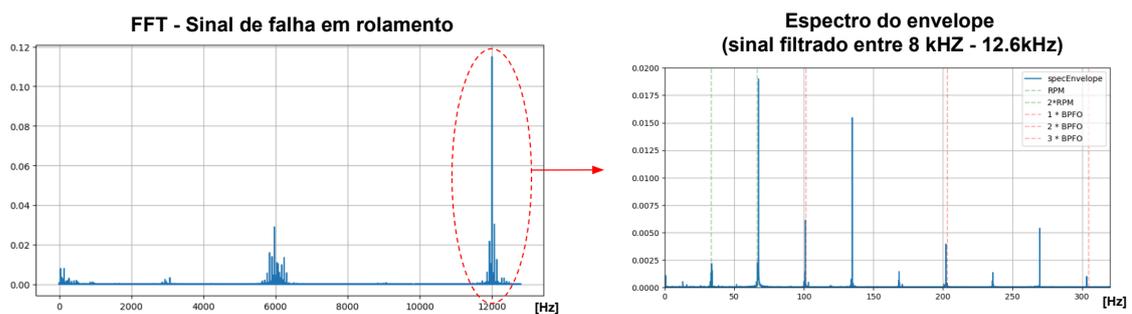


Figura 3. Exemplo de um sinal de falha de rolamento em que é possível extrair as frequências de falha através da retirada do envelope.

Para o cálculo das frequências de falha, no contexto deste trabalho, utilizam-se as informações fornecidas na Tabela 1, que representam os parâmetros geométricos associados a cada rolamento do conjunto de dados, disponibilizadas virtualmente pelos autores ¹.

Com isso, basta-se multiplicar a constante geométrica de cada rolamento pela sua frequência de operação, obtendo-se as frequências de falha, conforme a Tabela 2.

¹CWRU Bearing Data Center: <https://engineering.case.edu/bearingdatacenter>.

Tabela 1. Constantes geométricas para cálculo da frequência de falha.

| Localização | Anel Interno | Anel Externo | Elemento rolante |
|---------------------|--------------|--------------|------------------|
| Lado do acionamento | 5.4152 | 3.5848 | 4.7135 |
| Lado da ventoinha | 4.9469 | 3.0530 | 3.9874 |

Tabela 2. Exemplo de frequências de falha calculadas para uma frequência de operação de 1800 RPM e rolamento localizado no lado do acionamento.

| Frequência de operação | BPFI [Hz] | BPFO [Hz] | BSF [Hz] |
|------------------------|-----------|-----------|----------|
| 1800 RPM (30 Hz) | 162.45 | 107.54 | 141.405 |

2.3.2. Feature proposta: Peak-to-Floor Ratio (PFR)

Com o intuito de incorporar mais informações associadas à frequências de falha no modelo, desenvolveu-se uma métrica denominada PFR, que expressa a proeminência de uma frequência de falha em relação a sua banda lateral. Assim, aplica-se a PFR em múltiplos harmônicos da BPFI, BPFO e BSF, obtendo-se um conjunto de *features* relacionado a três tipos de falha.

Considerando que deseja-se medir essa proeminência localmente, ou seja, apenas em relação a uma banda curta próxima ao pico dos harmônicos, utiliza-se um tamanho de janela limitado, múltiplo da frequência em que está localizado. O tamanho da banda é definido na Equação 1, em que B representa a banda, *window_size* o percentual em relação à fundamental e f_{peak} a fundamental de uma frequência de falha.

$$B = window_size * f_{peak} \quad (1)$$

Na Figura 4, é dado um exemplo visual de um pico e a banda selecionada para cálculo da métrica.

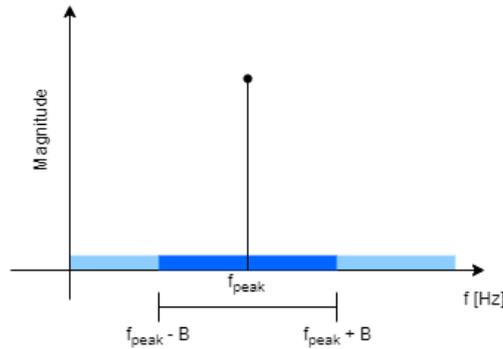


Figura 4. Exemplo da seleção de banda para cálculo da PFR, em que *window_size* = 0.5.

Observa-se que ao se calcular a PFR para múltiplos harmônicos, o tamanho da banda lateral permanece fixo. Calcula-se, então, a energia associada ao pico e a banda, seguindo a Equação 2, em que x representa o vetor de magnitudes obtido via FFT.

$$peak_{energy} = x[f_{peak}]^2, band_{energy} = \sum_{n=f_{peak}-B}^{f_{peak}+B} x[n]^2 \quad (2)$$

Com as energias, calcula-se a PFR segundo a Equação 3, fazendo uma relação que expressa a proeminência do pico em relação a banda numa faixa de valores entre 0 e 1.

$$PFR = \frac{peak_{energy}}{band_{energy}} \quad (3)$$

Através das informações disponibilizadas na Tabela 1, estimaram-se as fundamentais das frequências de falha, utilizadas para identificar as magnitudes e calcular as PFRs. Optou-se por buscar cinco harmônicos da BPFi, BPFO e BSF no espectro do envelope do sinal. Considerando a baixa a resolução espectral dos segmentos de 2048 amostras, usaram-se dois pontos para representar cada frequência de falha, nos casos em que o valor calculado está entre dois *frequency bins* no espectro. Para obter um valor de magnitude, foi retirado o valor RMS entre os pontos, fazendo $peak_{mag} = \sqrt{p_1^2 + p_2^2}$. Com as magnitudes obtidas, passou-se ao cálculo das PFRs.

2.3.3. Features utilizadas

A seguir, na Tabela 3, são dadas todas as *features* utilizadas neste trabalho.

Tabela 3. Features utilizadas no treinamento dos modelos.

| Conjunto de features | Descrição |
|----------------------|--|
| Abburri - Tempo | média, mediana absoluta, desvio padrão, assimetria, curtose, fator de crista, energia, RMS, número de picos ¹ , número de <i>zero-crossings</i> , teste de Shapiro, divergencia KL ¹ |
| Abburri - Freq. | Lado direito do espectro de magnitude - 1024 <i>bins</i> |
| A (Tempo) | RMS, pico-a-pico, curtose, assimetria, fator de crista, fator de crista plus, fator de forma, fator de impulso, fator de folga; todas aplicados à aceleração, envelope e velocidade. |
| B (Freq.) | magnitudes da BPFi/BPFO/BSF do primeiro ao quinto harmônico e PFRs aplicadas à BPFi/BPFO/BSF do primeiro ao quinto harmônico; todas extraídas de espectros do envelope filtrados em 5Hz-3kHz e 3kHz -6kHz. |
| C (Tempo + Freq.) | A + B |

¹ *Features* não incluídas nos treinamentos devido à falta de detalhes em [Abburri et al. 2023] e [Juodelyte et al. 2022] sobre como foram calculadas.

2.4. Divisão de dados

Seguiu-se a metodologia de [Rosa et al. 2024] criando *splits* de treino e teste separando com base no rolamento, e mantendo as amostras de normalidade apenas no conjunto de teste. Essa formulação é viável para o problema multirrótulo dado, considerando que cada modelo buscará identificar um tipo de falha específica. Ou seja, os rótulos para o modelo de *inner* (trilha interna) serão a presença ou não de falha na trilha interna, e o mesmo procedimento é repetido para os outros modelos.

Para otimização de hiperpâmetros e avaliação do modelo, utilizou-se a CVM-CV (*Double Cross-Validation Method*) [Tsamardinos et al. 2015]. A técnica foi proposta para

reduzir o viés da avaliação e consiste em duas etapas. Na primeira, realiza-se uma divisão aleatória em todo o conjunto de dados, e faz-se validação cruzada para otimização de hiperparâmetros. Na segunda etapa, são feitas M divisões aleatórias (distintas da que foi usada na validação cruzada), e são treinados e avaliados modelos usando os melhores hiperparâmetros obtidos anteriormente, produzindo como resultado final uma média e desvio padrão associado aos M modelos. Neste artigo, realizou-se a otimização K -fold com $K = 3$.

A Figura 5 exibe um exemplo das divisões de conjuntos em treino-teste e das *folders* para validação cruzada, em que cada elemento representa um rolamento com um tamanho da falha (0.007”, 0.014” ou 0.021”), tipo de falha (*inner*, *outer* ou *ball*) e localização (*fan end* ou *drive end*) distintos.

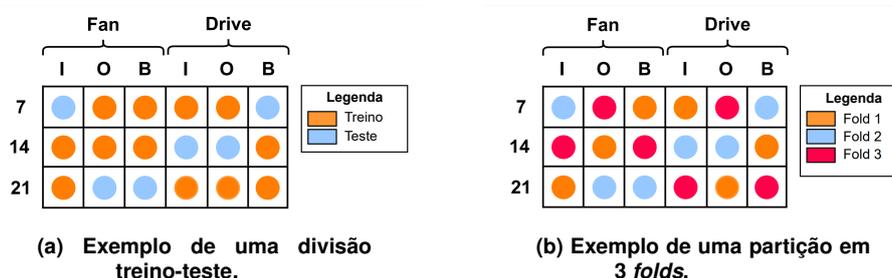


Figura 5. Exemplo da divisão utilizada. [Rosa et al. 2024]

Na Figura 5(a), para cada tipo de falha e localização (que representam uma coluna na tabela) são escolhidos 2 sinais para o treino e 1 para o teste, que possuem tamanhos de falha distintos. Observa-se, também, na Figura 5(b), que a divisão dos *folders* é aleatória, mas garante que não haja vazamento. No processo de otimização de hiperparâmetros foram feitas três combinações com as *folders*. Em cada combinação, treinou-se o modelo com duas *folders* e testou-se na terceira.

Na execução da segunda etapa do CVM-CV, foram feitas $M = 30$ divisões aleatórias (conforme a Figura 5(a)) do conjunto de dados para se avaliar o modelo, variando o parâmetro *random_state* de 1 a 30, enquanto para otimização de hiperparâmetros utilizou-se *random_state* = 42.

2.5. Modelos utilizados

Foram escolhidos os modelos *Logistic Regression* (LR), *Gaussian Naive Bayes* (GNB), *Random Forest* (RF), *Support Vector Machine* (SVM) - descritos em [Maleki et al. 2020] - e *XGBoost* [Chen and Guestrin 2016], buscando-se realizar a otimização de hiperparâmetros para cada um deles. Para os modelos SVM e LR, utilizou-se também uma escalamento dos dados, observando-se um aumento da performance geral. Ao aplicar-se o CVM-CV, fez-se uma busca aleatória na maioria dos parâmetros, de modo a reduzir o viés dos modelos, obtendo-se os resultados apresentados na Tabela 4.

3. Resultados

A métrica escolhida para avaliar os modelos foi a AUROC (*Area Under ROC Curve*), considerando que permite ao usuário escolher um ponto ótimo de operação, fixando uma taxa de verdadeiros positivos. A média e desvio padrão das AUROCs obtidas no conjunto

Tabela 4. Hiperparâmetros otimizados e melhores valores obtidos.

| Modelo | Hiperparâmetros |
|--------|---|
| LR | C = 212.46 |
| SVM | C = 8.31, gamma = 1.88e-3 |
| GNB | var_smoothing = 2.52e-15 |
| RF | ccp_alpha = 1.44e-3, max_depth = 29, min_samples_leaf = 64, min_samples_split = 31, n_estimators = 200 |
| XGB | colsample_bytree = 0.97, eta = 0.38, lambda = 27.53, max_depth = 44, n_estimators = 100, subsample = 0.77 |

de teste dos $M = 30$ modelos estão dadas na Tabela 5, que também apresenta os melhores resultados de [Rosa et al. 2024] ao final.

Tabela 5. Média e desvio padrão percentual das ROC AUCs obtidas para os modelos que detectam falha em *inner*, *outer* ou *ball*.

| Features | Modelo | Tipo de falha | | |
|----------------------|----------|-------------------|--------------------|--------------------|
| | | Inner | Outer | Ball |
| Abburi - Tempo | SVM | 54.49±16.93 | 69.19±16.84 | 53.93±13.32 |
| | XGB | 57.59±20.00 | 66.07±18.33 | 67.79±14.28 |
| | LR | 45.73±20.1 | 64.69±25.14 | 46.58±17.36 |
| | GNB | 35.09±13.82 | 90.4±9.53 | 67.96±22.29 |
| | RF | 53.48±11.94 | 56.34±13.85 | 46.88±8.64 |
| Abburi - Freq. | SVM | 68.05±9.44 | 64.4±14.7 | 56.92±8.15 |
| | XGB | 76.13±7.81 | 66.43±15.12 | 55.84±14.14 |
| | LR | 75.65±12.09 | 54.31±23.52 | 40.31±10.89 |
| | GNB | 36.61±20.77 | 52.73±21.02 | 65.39±8.02 |
| | RF | 78.49±7.59 | 69.65±16.46 | 58.16±9.78 |
| A (Tempo) | SVM | 73.98±12.89 | 49.48±13.96 | 66.06±12.13 |
| | XGB | 72.96±12.77 | 61.52±16.78 | 64.79±12.21 |
| | LR | 62.74±20.6 | 72.08±25.06 | 54.93±10.25 |
| | GNB | 67.03±14.98 | 65.34±26.12 | 59.78±10.78 |
| | RF | 76.26±9.16 | 56.61±17.24 | 67.16±13.34 |
| B (Freq.) | SVM | 86.7±11.21 | 70.61±13.33 | 54.47±8.62 |
| | XGB | 96.98±4.09 | 87.15±11.89 | 52.75±10.15 |
| | LR | 96.89±4.09 | 71.84±19.4 | 44.95±11.93 |
| | GNB | 97.02±3.92 | 92.77±11.29 | 52.75±4.33 |
| | RF | 97.19±3.78 | 91.71±12.54 | 43.75±6.71 |
| C (Tempo + Freq.) | SVM | 84.97±6.73 | 65.18±16.67 | 60.66±8.43 |
| | XGB | 96.63±4.0 | 88.01±10.85 | 56.81±9.88 |
| | LR | 92.56±8.73 | 78.71±16.78 | 48.05±11.0 |
| | GNB | 86.29±18.4 | 90.11±14.2 | 55.58±3.83 |
| | RF | 98.08±2.63 | 84.04±12.82 | 57.35±9.83 |
| Espectrograma | RESNET18 | 89.3±5.1 | 89.2±0.5 | 94.6±7.2 |

Observa-se que houve um alto desempenho tanto para os modelos de *inner* e *outer* com os conjuntos de *features* B (Freq.) e C (Tempo + Freq.), indicando um ganho em agregar as frequências de falha em conjunto com as PFRs no treinamento dos modelos. No entanto, os resultados obtidos para os modelos de *ball* foram muito inferiores, corroborando a análise feita em [Smith and Randall 2015], em que é discutido como as amostras de *ball* do CWRU são difíceis de classificar por possuírem poucos elementos associados ao fenômeno físico da falha em elemento rolante. Para os modelos de *ball*, os melhores resultados de nosso trabalho foram com as *features* do domínio do tempo de

[Abburi et al. 2023].

3.1. Análise da correlação

Com o objetivo de aumentar a explicabilidade do modelo e verificar a relação entre as features propostas e os rótulos, calculou-se a correlação entre eles. Foram consideradas as correlações para os conjuntos Abburi - Tempo, A (Tempo) e B (Freq.), excluindo-se o grupo Abburi - Freq., composto por 1024 bins da FFT de cada sinal, devido à baixa interpretabilidade das bins individualmente. A Tabela 6 apresenta os cinco maiores valores absolutos da correlação entre amostras com falha em trilha interna e amostras sem essa falha.

Tabela 6. Features com maior correlação aos labels de falha em trilha interna.

| Feature | Correlação com rótulos de <i>inner</i> |
|-------------------------------|--|
| envelope/PFR_BPFI_3/3000-6000 | 0.7362 |
| envelope/PFR_BPFI_4/3000-6000 | 0.6896 |
| envelope/PFR_BPFI_2/3000-6000 | 0.6837 |
| envelope/BPFI_1X/global | 0.5923 |
| envelope/PFR_BPFI_1/3000-6000 | 0.5811 |

Pela tabela, é possível identificar que há alta correlação das PFRs calculadas via extração do espectro do envelope entre 3kHz - 6kHz, indicando que essa é uma faixa adequada para identificação de falhas no CWRU. Além disso, observa-se que as métricas conferem com os rótulos, ou seja, PFRs aplicadas às frequências de falha de trilha interna (BPFI) são as mais correlatas com sinais que possuem esse tipo de falha.

Repetindo-se o experimento para trilha externa, obteve-se um resultado semelhante, com as maiores correlações associadas aos harmônicos da BPFO, alternando entre o envelope sem filtragem e filtrado na banda entre 3kHz - 6kHz. Na Tabela 7 são dados os resultados discutidos.

Tabela 7. Features com maior correlação aos labels de falha em trilha externa.

| Feature | Correlação com rótulos de <i>outer</i> |
|----------------------------|--|
| envelope/BPFO_2X/3000-6000 | 0.7115 |
| envelope/BPFO_2X/global | 0.6858 |
| envelope/BPFO_1X/global | 0.6471 |
| envelope/BPFO_1X/3000-6000 | 0.6400 |
| envelope/BPFO_3X/3000-6000 | 0.5996 |

Por fim, calculando as correlações para os rótulos de ball, obtiveram-se os resultados da Tabela 8.

As maiores correlações encontradas para *ball* foram com *features* extraídas do domínio do tempo e a PFR calculada para o primeiro harmônico da BPFO (sem relação com falha no elemento rolante). Observa-se que os valores encontrados foram baixos, o que explica o desempenho obtido nos modelos, indicando que os conjuntos de *features* utilizados não são capazes de representar adequadamente a falha em elemento rolante para os sinais do CWRU. Isso evidencia a necessidade de um melhor *feature engineering*, buscando-se explorar outros elementos associados à falha nessa localidade do rolamento.

Tabela 8. Features com maior correlação aos labels de falha em elemento rolante.

| Feature | Correlação com rótulos de <i>ball</i> |
|-----------------------------------|---------------------------------------|
| mean | 0.2529 |
| envelope/PFR_BPFO_1/3000-6000 | 0.2324 |
| envelope/skewness/global | 0.2290 |
| envelope/crest_factor_plus/global | 0.2232 |
| envelope/crest_factor/global | 0.2231 |

Com os resultados obtidos para os modelos de *inner* e *outer*, compreende-se que a alta correlação entre as features propostas neste artigo e os rótulos do conjunto de dados conferem maior explicabilidade ao modelo, pelo fato de que se relacionam diretamente com os fenômenos físicos.

4. Conclusão

Neste artigo, avaliou-se o desempenho de diferentes conjuntos de *features* com base na metodologia de [Rosa et al. 2024] em métodos tradicionais de aprendizado de máquina. Foram propostas *features* relacionadas ao fenômeno físico de falhas em rolamentos, as *Peak-To-Floor Ratios* (PFRs), visando melhorar desempenho e explicabilidade. Com abordagem multirrótulo, treinaram-se três modelos para identificar falhas específicas, atingindo AUROCs médias de 98,08%, 92,77% e 67,96% para *inner*, *outer* e *ball*, respectivamente. Testaram-se cinco conjuntos de *features* distintos, sendo três deles já discutidos na literatura, identificando ganho com o grupo B (Freq.) - que inclui as *features* desenvolvidas (PFRs) - nos modelos que detectam falha em trilha interna e externa.

Os resultados obtidos para *inner* e *outer* ultrapassaram os de [Rosa et al. 2024], trabalho que aplicou o mesmo split em modelos de *deep learning* com segmentos de 11500 amostras. Com isso, atingiu-se um melhor desempenho com menor custo computacional, considerando que foram utilizados segmentos de 2048 amostras e representações menos complexas dos dados. Além disso, através de análise de correlação, identificou-se que as PFRs aplicadas à BPFI foram as mais correlatas aos rótulos de *inner*, corroborando que utilizar a proeminência dos harmônicos de falha no espectro contribui para o aprendizado do modelo na detecção de falha, ao mesmo tempo que conferem maior explicabilidade.

Para o modelo de *ball*, com resultado negativo, observou-se que o grupo de *features* proposto não possui forte correlação com os dados, dificultando a classificação. A baixa correlação deve-se à dificuldade de diagnóstico das falhas em elementos rolantes do CWRU, que poderia ser mitigada com um melhor *feature engineering*. Deseja-se, em trabalhos futuros, encontrar representações mais adequadas para esses sinais, de modo a atingir desempenhos semelhantes às abordagens de *deep learning*, porém com um menor custo computacional e maior explicabilidade.

Referências

Abhuri, H., Chaudhary, T., Ilyas, S. H. W., Manne, L., Mittal, D., Bowen, E., Williams, D., Snaidauf, D., and Veeramani, B. (2023). A closer look at bearing fault classification approaches. *Annual Conference of the PHM Society*, 15(1).

- Ahmed, H. and Nandi, A. K. (2020). *Condition Monitoring with Vibration Signals: Compressive Sampling and Learning Algorithms for Rotating Machines*, chapter 6, pages 117–128. John Wiley & Sons, Inc, Hoboken, NJ, USA.
- Barreto, L., Rosa, R., Silva, D., and Braga, D. (2023). Fault detection for rotating machinery based on vibration data using machine learning. In *Anais do XX Encontro Nacional de Inteligência Artificial e Computacional*, pages 242–256, Porto Alegre, RS, Brasil. SBC.
- Chen, T. and Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, pages 785–794, New York, NY, USA. Association for Computing Machinery.
- González, M., Díaz, V. G., Pérez, B. L., G-Bustelo, B. C. P., and Anzola, J. P. (2023). Bearing fault diagnosis with envelope analysis and machine learning approaches using cwru dataset. *IEEE Access*, 11:57796–57805.
- Hendriks, J., Dumond, P., and Knox, D. (2022). Towards better benchmarking using the cwru bearing fault dataset. *Mechanical Systems and Signal Processing*, 169:108732.
- Juodelyte, D., Cheplygina, V., Graversen, T., and Bonnet, P. (2022). Predicting bearings degradation stages for predictive maintenance in the pharmaceutical industry. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, KDD '22. ACM.
- Kapoor, S. and Narayanan, A. (2023). Leakage and the reproducibility crisis in machine-learning-based science. *Patterns*, 4(9):100804.
- Maleki, F., Ovens, K., Najafian, K., Forghani, B., Reinhold, C., and Forghani, R. (2020). Overview of Machine Learning Part 1. *Neuroimaging Clinics of North America*, 30(4):e17–e32.
- Neupane, D. and Seok, J. (2020). Bearing fault detection and diagnosis using case western reserve university dataset with deep learning approaches: A review. *IEEE Access*, 8:93155–93178.
- Randall, R. B. (2021). *Vibration-based Condition Monitoring*, chapter 1, pages 1–24. John Wiley Sons, Ltd.
- Rosa, R. K., Braga, D., and Silva, D. (2024). Benchmarking deep learning models for bearing fault diagnosis using the cwru dataset: A multi-label approach. arXiv:2407.14625 [eess].
- Smith, W. A. and Randall, R. B. (2015). Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study. *Mechanical Systems and Signal Processing*, 64–65:100–131.
- Tsamardinos, I., Rakhshani, A., and Lagani, V. (2015). Performance-Estimation Properties of Cross-Validation-Based Protocols with Simultaneous Hyper-Parameter Optimization. *International Journal on Artificial Intelligence Tools*, 24(05):1540023.