

Análise Interpretativa dos Estágios de Degradação de Rolamentos: Uma Abordagem com Algoritmos SHAP e Classificação por Árvore de Decisão

Francisco G. D. da Silva Filho 

Universidade Federal do Ceará (UFC)

gdias27@alu.ufc.br

Joaquim O. F. M. Filho 

Universidade Federal do Ceará (UFC)

joaquim1905@alu.ufc.br

Vandilberto P. Pinto 

Universidade da Integração Internacional da Lusofonia Afro-Brasileira (UNILAB)

vandilberto@unilab.edu.br

Márcio A. B. Amora 

Universidade Federal do Ceará (UFC)

marcio@sobral.ufc.br

Resumo – A degradação de rolamentos é uma questão crítica em sistemas mecânicos, levando à deterioração do desempenho e a possíveis falhas. A classificação dos estágios de degradação em sistemas de rolamentos é essencial para a manutenção eficaz e prevenção de paradas inesperadas. Com base nisso, o estudo apresenta uma metodologia para a detecção dos estágios de degradação em sistemas de rolamentos, empregando múltiplas técnicas como: transformada de Fourier, análise de componentes principais e árvores de decisão. Os resultados demonstram que essa abordagem permite a detecção do atual estágio de degradação com técnicas que comumente exigem um esforço computacional menor em comparação com métodos previamente utilizados na literatura. Também, a análise comparativa revela um desempenho superior em termos de vida útil dos rolamentos após a detecção de falhas nos estágios mais avançados, superando os resultados da literatura. Além disso, a interpretação dos resultados por meio de gráficos SHAP identifica faixas de valores dentro das características cruciais para a classificação de falhas, como: desvio absoluto médio e coeficientes cepstrais de predição linear. A análise de dependência entre essas características fornece percepções valiosas sobre os padrões de falha. A acurácia de classificação obtida foi superior a 90%, demonstrando a utilidade e eficiência do método proposto para o aprimoramento da manutenção preditiva em mancais de rolamento.

Palavras-chave – Detecção de estágio de degradação, sistemas de rolamentos, manutenção preditiva, análise de componentes principais, árvores de decisão, explicações aditivas de Shapley, classificação.

Abstract – The degradation of bearings is a critical issue in mechanical systems, leading to performance deterioration and potential failures. Classifying degradation stages in bearing systems is essential for effective maintenance and the prevention of unexpected downtime. Based on this, the study presents a methodology for detecting degradation stages in bearing systems by employing multiple techniques, such as Fourier transform, principal component analysis, and decision trees. The results demonstrate that this approach enables the detection of the current degradation stage using techniques that generally require lower computational effort compared to methods previously used in the literature. Additionally, the comparative analysis reveals superior performance in terms of bearing lifespan after fault detection in advanced stages, surpassing results from the literature. Moreover, interpreting the results through SHAP plots identifies value ranges within crucial features for fault classification, such as mean absolute deviation and linear predictive cepstral coefficients. The dependency analysis between these features provides valuable insights into failure patterns. The achieved classification accuracy exceeded 90%, demonstrating the effectiveness and efficiency of the proposed method in enhancing predictive maintenance for rolling bearings.

Keywords – Detection of degradation stage, bearing systems, predictive maintenance, principal component analysis, decision trees, shapley additive explanations, classification.

1. INTRODUÇÃO

Os rolamentos são elementos de máquina que desempenham um papel crítico em equipamentos mecânicos, influenciando diretamente a eficiência e a produção em plantas industriais. Devido a esse fator, a manutenção preditiva desses componentes torna-se necessária para garantir uma operação contínua e eficaz. Tradicionalmente, essa manutenção é baseada na observação de

dados de alta frequência coletados por sensores e tem sido uma abordagem importante para prever a necessidade de intervenções [1]. No entanto, com o advento do aprendizado de máquina, várias metodologias foram desenvolvidas para realizar a estimativa da vida útil remanescente e a detecção de falhas nesses componentes [2], [3], [4].

Para o desenvolvimento desses procedimentos, um modelo de aprendizado de máquina é treinado, geralmente de forma supervisionada, para prever ou detectar falhas em rolamentos. No primeiro caso, a técnica visa estimar a curva de degradação do componente para prever quando a substituição é necessária. Em contrapartida, o segundo modelo é treinado usando sinais de vibração contendo amostras rotuladas como saudáveis ou com algum tipo de falha. A primeira metodologia apenas indica quando a manutenção é necessária, mas pode falhar em fornecer outras informações essenciais para a compreensão completa do problema. Por outro lado, a segunda metodologia identifica se o dispositivo está em boas condições ou já possui uma falha, mas não informa precisamente quando a manutenção deve ser realizada, apenas especifica o defeito existente no rolamento. Em contraste com essas metodologias, há uma proposta alternativa que visa classificar estágios de degradação de rolamentos, abrindo possibilidades para o desenvolvimento de estudos voltados a analisar e explicar o problema proposto [5]. Nesse contexto, a classificação dos estágios de degradação dos rolamentos surge como uma solução proposta, identificando quais mancais de rolamento precisam ser substituídos em momentos específicos.

Trabalhos como os de [6] utilizam essa metodologia para garantir a detecção precoce da degradação em rolamentos, expondo um sistema capaz de fornecer mais informações do que outras metodologias previamente apresentadas. Em [6], o autor utiliza os estágios de degradação descritos na literatura para prever os estágios dos rolamentos e quando será necessário substituir a peça sem afetar a produção na indústria farmacêutica. Neste trabalho, é utilizado o *deep learning*, que tem sido uma abordagem popular para modelar o estado de degradação, bem como para classificar esses estágios, devido à sua capacidade de lidar com grandes volumes de dados e complexidade de sinais sem exigir conhecimento específico de domínio [7]. No entanto, essas técnicas exigem poder computacional significativo e, muitas vezes, os resultados obtidos não justificam o investimento necessário, apresentando limitações em termos de precisão, aplicabilidade e interpretabilidade. Diante dessas limitações, é imperativo explorar métodos alternativos que ofereçam uma classificação mais eficiente dos estágios de degradação dos rolamentos sem a necessidade de uma infraestrutura computacional excessiva e que forneçam maior explicabilidade para auxiliar o especialista na compreensão do problema proposto.

Assim, este artigo propõe o uso de árvores de decisão em conjunto com algoritmos SHAP para analisar a influência dos atributos ao classificar os estágios de degradação dos rolamentos. Árvores de decisão são conhecidas por sua interpretabilidade e eficiência computacional, sendo esta última uma característica que as torna uma escolha adequada para esse propósito. Juntamente com o classificador, um algoritmo explicável será utilizado para analisar as classificações e identificar padrões que possam auxiliar pesquisadores em estudos futuros na área.

De forma detalhada, o trabalho propõe:

- Com base no conhecimento de domínio, foi desenvolvido um método para segmentação das classes de degradação dos rolamentos usando sinais de vibração de alta frequência, incorporado em um subespaço latente de baixa dimensionalidade por meio da análise de componentes principais, para automatizar o rotulamento dos dados.
- Um classificador supervisionado baseado em árvore de decisão foi projetado para detectar os estágios de degradação dos rolamentos.
- A técnica SHAP (*SHapley Additive exPlanations*) foi utilizada para a análise interpretativa dos resultados obtidos com árvores de decisão.

O artigo está estruturado da seguinte forma: A Seção I introduz o tópico, focando na análise interpretativa dos estágios de degradação dos rolamentos. A Seção II cobre a fundamentação teórica necessária para a compreensão do estudo. A Seção III descreve a metodologia utilizada. Na Seção IV, são apresentados os resultados e discussões baseados na aplicação dos métodos propostos. Finalmente, a Seção V fornece a conclusão do trabalho, destacando as principais descobertas e implicações do estudo além dos trabalhos futuros.

2. FUNDAMENTAÇÃO TEÓRICA

2.1 Mancal de rolamento

Os mancais de rolamentos têm a função de reduzir a fricção, suportando e guiando um eixo rotativo. Atualmente, há dois tipos principais de mancais em uso: mancais de deslizamento e de rolamento. Para aplicações de menor potência e cargas leves, os mancais de rolamento são uma escolha frequente [6]. Esses mancais de rolamento são compostos por quatro partes principais: pista interna, pista externa, elemento rolante e gaiola, que mantém e guia os elementos rolantes. Eles podem conter diferentes tipos de elementos, como esferas, rolos e rolos cônicos, com o contato entre as partes ocorrendo sempre em nível metal-metal [5].

Quando em bom estado, os rolamentos apresentam frequências características em seu sinal de vibração, sem anomalias indicativas de falhas. A Figura 1 ilustra as frequências que se manifestam desde o rolamento saudável até o estágio 4 de degradação. As frequências características estão localizadas em diferentes zonas de frequência, cada uma associada a diferentes estágios de degradação do rolamento. A Zona A abrange frequências de até 1 kHz, a Zona B está na faixa de aproximadamente 1-5kHz, a Zona C se encontra na faixa de aproximadamente 5-6 kHz e a Zona D cobre frequências acima de 20 kHz [8]. As falhas em mancais de rolamento são comuns na indústria e podem ser causadas por corrosão, trincas, contaminação, amassados e erosão, o que, com o tempo, pode comprometer a operação da máquina [9].

2.2 Estágios de degradação em mancais de rolamento

Os estágios de degradação dos rolamentos são fundamentais para a manutenção preditiva e a detecção precoce de falhas. Eles são descritos em [10], [11] e [8] da seguinte forma e são apresentados na Figura 1.

- **Estágio 1:** Neste estágio inicial, ocorre um aumento de alta frequência na região de Energia de Pico, que está na faixa ultrassônica. É essencial o uso de sensores projetados especificamente para detectar nessa região. Circuitos especializados filtram apenas esses sinais, e a inspeção física do rolamento pode não revelar defeitos identificáveis. Logo, o estágio 1 é unido ao estágio saudável no presente trabalho.
- **Estágio 2:** Neste estágio, começam a surgir sinais relacionados às frequências naturais de ressonância das partes do rolamento, à medida que os defeitos começam a afetar os componentes do rolamento.
- **Estágio 3:** Aqui, as frequências fundamentais dos defeitos do rolamento estão presentes, juntamente com possíveis harmônicos, dependendo da quantidade e distribuição dos defeitos ao redor das pistas do rolamento. As frequências harmônicas são moduladas pela velocidade do eixo.
- **Estágio 4:** Este é o estágio final antes da falha catastrófica do rolamento. Neste ponto, há várias frequências fundamentais moduladas e harmônicas, indicando uma distribuição dos defeitos ao redor das pistas do rolamento. Devido à degradação avançada, as folgas internas aumentam, permitindo maior vibração do eixo, com aumento nas frequências do eixo relacionadas ao balanceamento ou desalinhamento. Nas fases finais do estágio 4, as frequências fundamentais do rolamento diminuem e são substituídas por ruído aleatório em frequências mais altas.

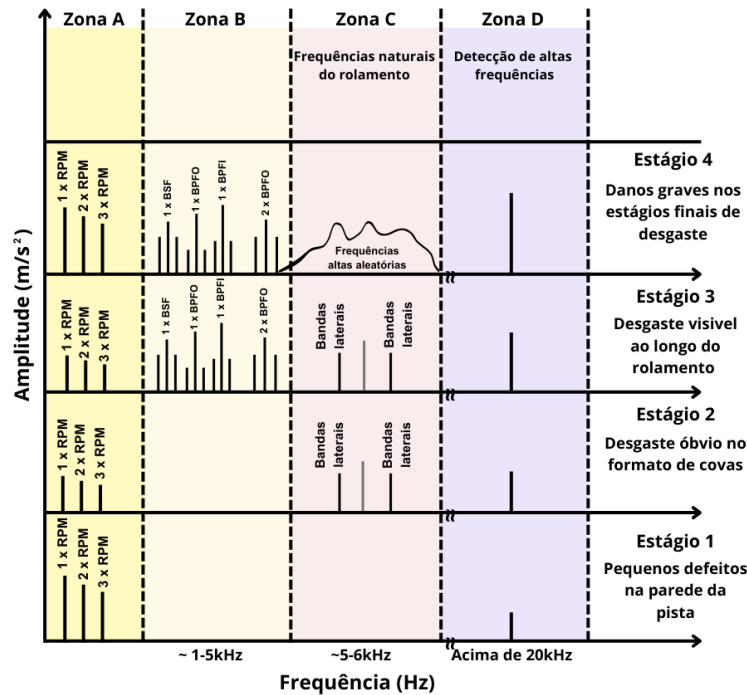


Figura 1: Evolução das frequências do rolamento nos estágios de degradação (Adaptado de [11]).

As frequências que se manifestam durante o processo de degradação segundo [11] são:

- **RPM (Rotações Por Minuto, em inglês *Revolutions Per Minute*):** Refere-se à frequência de rotação do eixo.
- **BSF (Frequência de Rotação da Esfera, em inglês *Ball Spin Frequency*):**

$$BSF = \frac{f_r}{2} \left(\frac{d}{D} \right) \left(1 - \left(\frac{d}{D} \right)^2 \cos^2 \theta \right). \quad (1)$$

- **BPFO (Frequência de Passagem da Esfera pelo Anel Externo, em inglês *Ball Pass Frequency Outer*):**

$$BPFO = \frac{n}{2} f_r \left(1 + \frac{d}{D} \cos \theta \right). \quad (2)$$

- BPFÍ (Frequência de Passagem da Esfera pelo Anel Interno, em inglês *Ball Pass Frequency Inner*):

$$BPFÍ = \frac{n}{2} f_r \left(1 - \frac{d}{D} \cos \theta \right). \quad (3)$$

onde f_r , é a frequência de rotação do eixo, d o diâmetro da esfera, D o diâmetro da pista onde as esferas estão localizadas, θ o ângulo de contato e n o número de esferas no rolamento.

2.3 Árvores de Decisão

As árvores de decisão são uma abordagem eficaz e relativamente simples para representar o conhecimento. Seguindo o princípio de dividir e conquistar, elas dividem o universo de dados em subconjuntos sucessivos até que cada um pertença a uma única classe ou até que uma classe seja predominante, eliminando a necessidade de mais divisões [12]. Os resultados desses subconjuntos, obtidos pela construção da árvore de decisão, são organizados de forma compacta e usados para classificar novos exemplos. Para entender melhor o funcionamento das árvores de decisão, considere um universo de amostras representado por dois atributos, X_1 e X_2 , Figura 2-a).

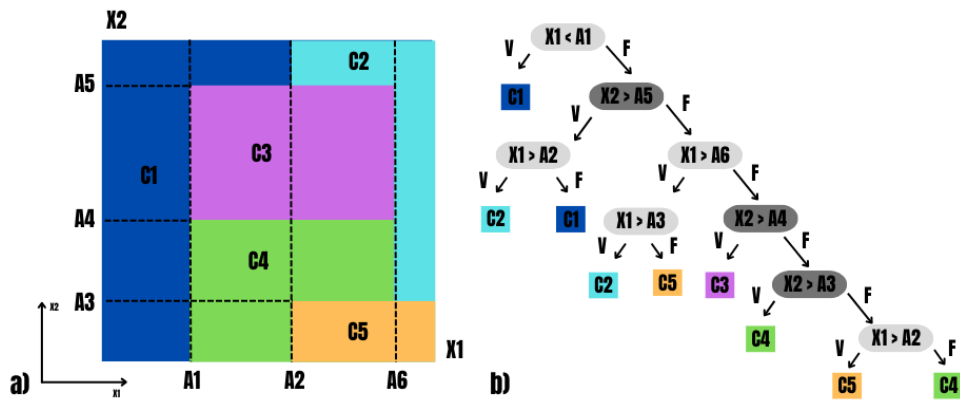


Figura 2: Plano de rótulos (Fonte: Adaptado de [12]).

Neste plano, as classes A_1 , A_2 , A_3 , A_4 , A_5 e A_6 representam os valores que X_1 e X_2 podem assumir, e as fronteiras (linhas tracejadas) entre elas marcam os limites entre as regiões, como mostrado no exemplo. A árvore de decisão que representa esse conjunto de classes está na 2-b).

As árvores de decisão são amplamente utilizadas em diversas situações devido à sua simplicidade e eficácia. Elas são aplicadas, por exemplo, em diagnósticos médicos [13], análise de crédito em conjunto com detecção de fraudes [14], previsão de demanda [15], recrutamento de funcionários e tomada de decisões empresariais [16].

2.4 SHAP

O SHAP é uma técnica utilizada para explicar previsões individuais em um contexto de modelo tipo caixa-preta. As previsões baseiam-se no valor de *Shapley*, que representa a contribuição média de uma característica em todas as combinações possíveis de características para o resultado da previsão. A fórmula que descreve o SHAP é a seguinte [17]:

$$\phi_i(f) = \sum_{S \subseteq F \setminus i} \frac{|S|!(|F| - |S| - 1)!}{|F|!} [f(S \cup i) - f(S)], \quad (4)$$

onde: $\phi_i(f)$ representa a contribuição da i -ésima característica para a previsão do modelo f , F é o conjunto de todas as características, S é um subconjunto de características que não inclui a i -ésima característica, $|S|$ é a quantidade de características no subconjunto S , $|F|$ é o número total de características, $f(S)$ é a previsão do modelo quando o conjunto de características S está presente, $f(S \cup i)$ é a previsão do modelo quando a i -ésima característica é adicionada ao conjunto de características S .

Para aplicar o SHAP, primeiramente ajusta-se um modelo de aprendizado de máquina aos dados e, em seguida, utiliza-se a biblioteca SHAP para gerar as explicações necessárias para compreender as decisões do modelo ao nível local e global. Isso permite que os usuários determinem quais características são mais importantes para as previsões do modelo e como essas características influenciam os resultados. A importância do SHAP reside na sua capacidade de fornecer explicações precisas e confiáveis para as previsões do modelo, ajudando os usuários a entender como cada característica afeta a saída do modelo [17]. No entanto, há desafios, como a necessidade de lidar em muitas situações com modelos complexos e de grande escala, a interpretação das explicações em contextos específicos e a avaliação da qualidade das explicações geradas.

2.4.1 Biblioteca SHAP

A biblioteca SHAP [18] é uma ferramenta amplamente utilizada para explicar os resultados de modelos de aprendizado de máquina, oferecendo uma interpretação acessível das contribuições das características para as previsões. Entre as visualizações mais úteis estão os gráficos de força e gráficos de dependência, que auxiliam na compreensão detalhada do comportamento do modelo. Dentre as várias ferramentas disponibilizadas pela biblioteca além de suas possibilidades de análise, o presente trabalho se limita a análise de gráficos de impacto de valores dos atributos no SHAP e dos gráficos de dependência entre atributos.

Os gráficos de força do SHAP são úteis para uma compreensão do funcionamento do algoritmo, onde é possível visualizar a contribuição de cada característica para uma predição individual. Eles mostram como cada característica predis põe a predição para cima ou para baixo em relação ao valor esperado e um exemplo pode ser visto na Figura 3 (Adaptada de [18]), onde atributos como Idade, Sexo, Pressão Arterial e Índice de Massa Corporal, são utilizados para prever a chance do paciente apresentar hipertensão, de modo que cada valor do atributo apresenta uma contribuição na decisão do algoritmo quanto a classe da amostra. É importante destacar que assim como as forças de cada atributo, o valor de *base rate* é determinado no momento do treinamento do classificador.

O *base rate* em SHAP refere-se ao valor médio da predição feita pelo modelo, ou seja, a predição esperada sem qualquer informação adicional sobre os dados específicos. É o ponto de partida que indica o comportamento geral do modelo antes de considerar as contribuições individuais das variáveis.

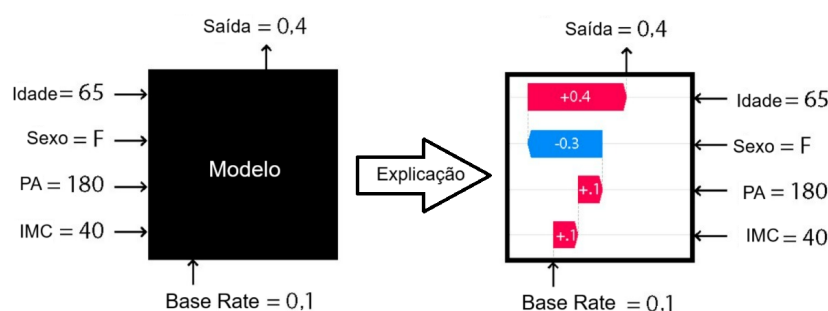


Figura 3: Funcionamento do SHAP.

O gráfico de impactos dos valores SHAP, conhecido como *summary plot*, combina as informações de distribuição de valores SHAP com os valores dos atributos correspondentes em uma única visualização. Ao analisar a Figura 4-a) (Adaptada de [18]) é possível observar como cada ponto no gráfico representa um par de valor de atributo e valor SHAP para uma observação específica no conjunto de dados em cada um dos 10 atributos. O eixo vertical do gráfico lista os diferentes atributos, enquanto o eixo horizontal mostra os valores SHAP, que indicam a magnitude e a direção do impacto de cada atributo na predição.

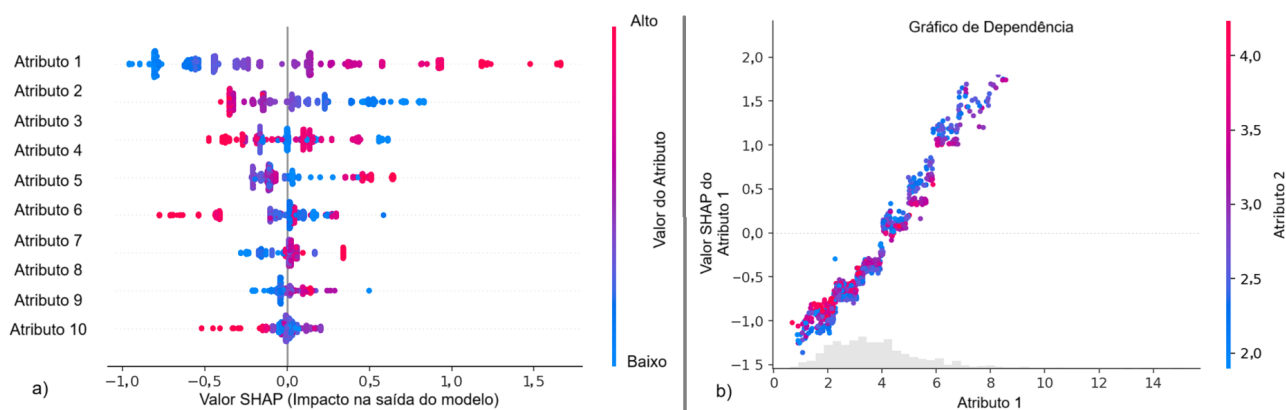


Figura 4: a) Gráfico de impacto de atributos, b) Gráfico de dependência entre atributos.

Os gráficos de dependência fornecem informações sobre a relação entre uma característica específica e a saída do modelo, considerando os valores das outras características constantes. Essa visualização é crucial para entender a sensibilidade do modelo em relação a diferentes variáveis e identificar possíveis interações entre características. Um exemplo sobre este método pode ser observado na Figura 4-b) (Adaptada de [18]), onde é possível observar para dois atributos uma região bidimensional, e como essa região vai estar relacionada a o impacto SHAP de um dos atributos na classificação do modelo. Além de verificar a importância, tal gráfico serve para que o analista consiga propor e analisar relações entre os atributos e a predição.

3 METODOLOGIA

Para realizar a detecção dos estágios de degradação dos rolamentos e sua análise interpretativa é proposto um método de múltiplas etapas observável na Figura 5.

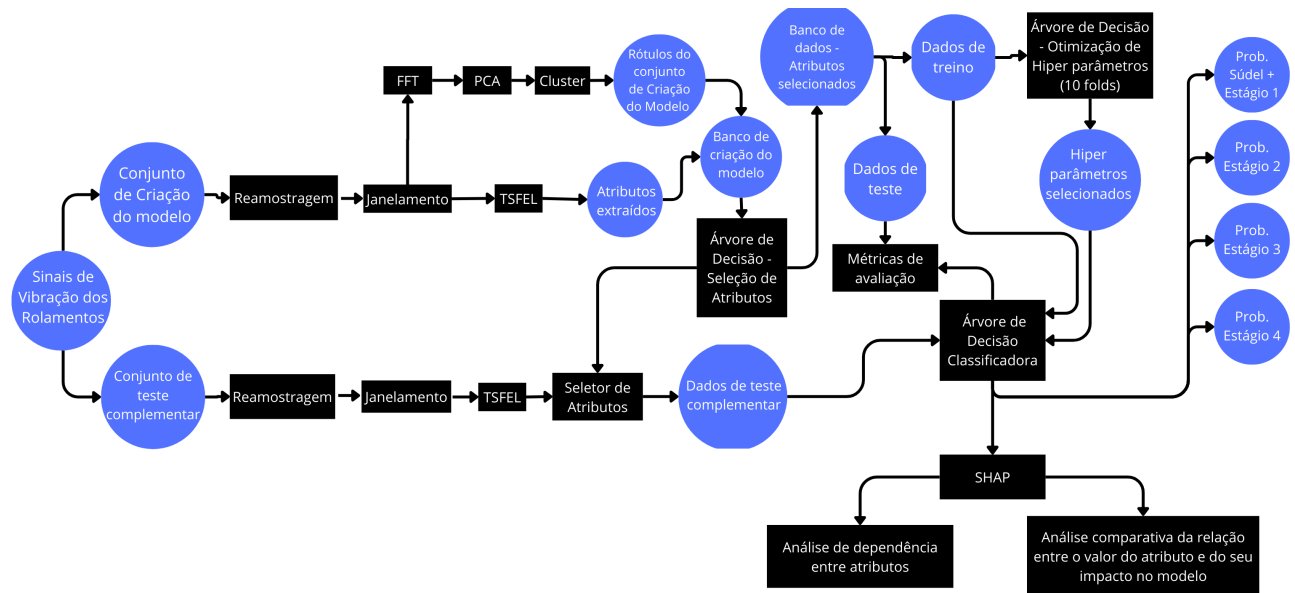


Figura 5: Método proposto para a classificação e análise dos estágios de degradação.

Inicialmente, os sinais de vibração dos rolamentos, obtidos do banco de dados *FEMTO Bearing* de [19]. Nesta base de dados os rolamentos já se mostram previamente separados entre o conjunto de criação do modelo e conjunto de teste complementar, passam por uma redução na taxa de amostragem. Os dados reduzidos são então segmentados em janelas de tamanho uniforme. Para cada janela do conjunto de criação do modelo, aplica-se a transformada rápida de Fourier (FFT, do inglês, *Fast Fourier Transform*), que converte o sinal do domínio do tempo para o domínio da frequência. Em seguida, a Análise de Componentes Principais (PCA, do inglês, *Principal Component Analysis*) é utilizada para reduzir a dimensionalidade dos dados em cada janela transformada. Após esse processamento, os dados são agrupados em *clusters*, que representam diferentes estágios de degradação e permitem a rotulagem das amostras do conjunto.

A biblioteca *Time Series Feature Extraction Library* (TSFEL) é empregada para extrair características tanto dos dados do conjunto de criação do modelo quanto do conjunto de teste complementar. Na etapa seguinte a extração de atributos do conjunto de criação do modelo, os dados rotulados passam por um processo de seleção de atributos mais importantes e em seguida são utilizados para treinar um classificador supervisionado, especificamente uma árvore de decisão, que é otimizada por meio de validação cruzada e busca randômica. O objetivo é prever a probabilidade de um sinal de vibração pertencer a um dos estágios de degradação (Saudável + Estágio 1, Estágio 2, Estágio 3 e Estágio 4).

Posteriormente, a biblioteca SHAP é aplicada para a análise de interpretabilidade do modelo, permitindo a avaliação do gráfico de impacto e da dependência das características no processo de decisão. O SHAP fornece explicações detalhadas das previsões do modelo, facilitando uma compreensão mais profunda dos fatores que influenciam a degradação dos mancais e auxiliando no desenvolvimento de estratégias de manutenção preditiva mais eficazes. Por fim, com um modelo treinado e validado por suas métricas, realiza-se, a partir do banco de dados complementar, a análise da evolução dos estágios por meio de suas probabilidades.

3.1 Banco de dados

O sistema proposto utiliza o conjunto de dados **FEMTO Bearing** [19], para treinamento e avaliação. Este conjunto de dados inclui medições de vibração horizontal e vertical de mancais obtidas por meio de experimentos. Para acelerar a degradação, os mancais foram submetidos a condições de estresse além das recomendações do fabricante, resultando em um total de 17 mancais testados sob três diferentes condições de carga e velocidade de rotação [19] que são apresentados a seguir:

- Primeira condição de operação: 1800 rpm e 4000 N;
- Segunda condição de operação: 1650 rpm e 4200 N;
- Terceira condição de operação: 1500 rpm e 5000 N.

O banco de dados está organizado como expresso na Tabela 1 (Adaptada de [19]), e essa organização é levada em consideração na utilização dos dados no presente trabalho.

Tabela 1: Divisão do conjunto de dados.

Conjunto de Dados	Condição 1	Condição 2	Condição 3
Conjunto de criação do modelo	Bearing1_1	Bearing2_1	Bearing3_1
	Bearing1_2	Bearing2_2	Bearing3_2
Conjunto de teste complementar	Bearing1_3	Bearing2_3	Bearing3_3
	Bearing1_4	Bearing2_4	
	Bearing1_5	Bearing2_5	
	Bearing1_6	Bearing2_6	
	Bearing1_7	Bearing2_7	

Os experimentos foram encerrados quando as leituras do acelerômetro atingiram 20 g (gravidade) por segurança, indicando a fase final de degradação do mancal. A variabilidade na vida útil dos mancais, com registros entre 28 minutos e 7 horas. As falhas dos mancais não são especificadas, e é possível que um mancal tenha sofrido múltiplos defeitos simultaneamente [19].

3.2 Reamostragem e janelamento

Primeiramente, utilizando o banco de dados FEMTO Bearing [19], os sinais de vibração de ambos os conjuntos foram submetidos a uma redução de amostragem. Durante os experimentos, as vibrações foram amostradas a cada 10 segundos com uma frequência de 25.600 Hz durante 0,1 segundos. No entanto, essa alta frequência de amostragem pode não ser viável em aplicações industriais, uma vez que as frequências de interesse (além das ultrassônicas) estão abaixo de 6 kHz, conforme mostrado na Figura 1. Pelo Teorema de Nyquist [20], uma frequência de coleta de 25.600 Hz fornece informações até a frequência de 12.800 Hz. Ao observar a Figura 1, é possível observar que a coleta foi inadequada para captar frequências ultrassônicas presentes na zona D, já que elas estão acima de 20kHz. A zona imediatamente abaixo é a zona C, que cobre uma faixa de frequência com limite superior em 6 kHz.

Ao observar a relação entre os estágios e as zonas, é possível ver que a Zona D se altera somente na transição entre a condição saudável e o estágio 1. De modo que esses dados são relevantes somente nessa transição. Logo, para fins de análise a condição de saúde com o estágio 1, os dados em frequência ultrassônica referentes a Zona D podem ser suprimidos. Portanto, o sinal de vibração bruto teve sua taxa de amostragem reduzida pela metade para 12.800 Hz, o que nos dá informações no domínio da frequência até 6.400 Hz. Essa redução preserva as informações importantes ao mesmo tempo que diminui a necessidade computacional. Após isso, os dados foram divididos com base nas coletas de 0,1 segundos, resultando em janelas de 1.280 amostras.

3.3 Rotulagem

Embora os estágios de degradação abordados na Seção 2 sejam definidos na literatura, os dados não estão rotulados em seus estágios de degradação quando disponibilizados. Em [19], o conjunto de criação do modelo é descrito como um conjunto composto por rolamentos que operam desde o funcionamento pleno até a quebra, garantindo assim a passagem por todos os estágios de degradação. A partir dessa informação, é desenvolvido um método para realizar a rotulagem dos referidos dados em quatro grupos, correspondentes aos quatro estágios de degradação do rolamento. As séries temporais dos rolamentos rotulados serão utilizadas para treinar o modelo de classificação, permitindo a identificação automática do nível de degradação em qualquer estágio. O processo de rotulagem é subdividido em outros processos, que estão descritos nas subseções a seguir.

3.3.1 Transformada de Fourier

Após a divisão da série temporal em janelas, cada janela passou por uma FFT para converter o sinal do domínio do tempo para o domínio da frequência. Após a transformação para o domínio da frequência, foram obtidas frequências de até 6.400 Hz, onde se espera, pela literatura [8], detectar os sinais mais evidentes de degradação dos mancais. Essa redução na frequência de amostragem reduz a quantidade de dados, mantendo a faixa de frequência relevante para análise.

3.3.2 Análise dos Componentes Principais

Posteriormente, o uso da PCA é realizado para diminuir a dimensionalidade dos dados de cada janela transformada. O PCA é utilizado como uma técnica de redução de dimensionalidade. Essa técnica é amplamente reconhecida por sua capacidade de capturar a variância dos dados em componentes principais. No entanto, observa-se que os primeiros componentes principais explicam apenas uma pequena parte da variância total. Portanto, decidiu-se manter os primeiros 40 componentes principais de cada sinal tratado no domínio da frequência, correspondendo entre 30% e 60% da variância total [6]. Isso, garante uma boa representação dos dados e diminui consideravelmente a dimensão dos dados.

3.3.3 Cluster

Após transformar os sinais de vibração no espaço de menor dimensionalidade utilizando o PCA, os dados foram agrupados em regiões (*clusters*), usando o método *k*-means, que divide o sinal nos estágios de degradação e possibilita rotular as amostras.

A rotulagem do conjunto de criação do modelo foi realizada utilizando quatro clusters. A escolha do número de regiões foi feita com base no conhecimento sobre os estágios de degradação [5] e no trabalho já realizado em [6]. No entanto, para novas amostras, não é assegurada eficácia da técnica, uma vez que não há garantia de que o rolamento tenha passado por todas as fases do processo de degradação, impossibilitando o uso direto do agrupamento em *clusters*. Os resultados para os *clusters* das três situações de carga aplicada ao rolamento podem ser vistos na Figura 6.

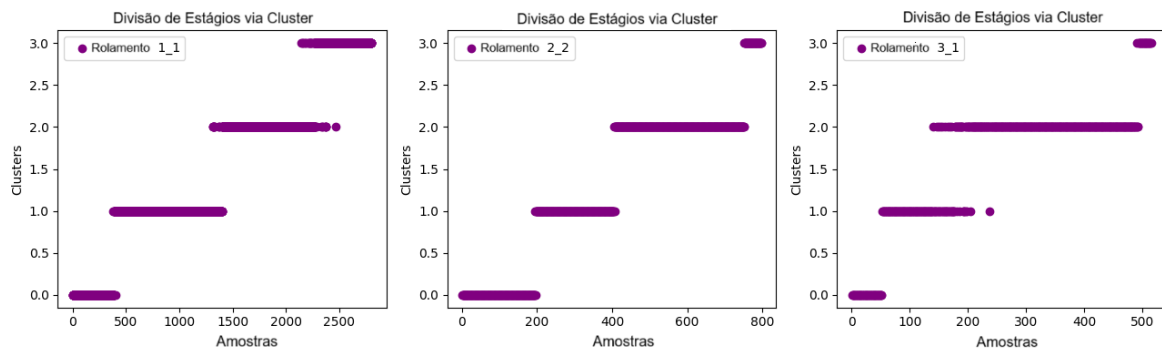


Figura 6: Rotulagem dos estágios usando o *cluster*.

3.4 Extração de características

Diferente do agrupamento via *k-means*, que exige que o rolamento tenha passado por todas as fases do processo, a representação por atributos pode ser realizada em qualquer momento da série temporal. Assim, um classificador será treinado a partir dos rolamentos do conjunto de criação do modelo que operaram até a plena falha, para associar os atributos extraídos aos respectivos estágios de degradação, permitindo, posteriormente, a classificação em qualquer fase da operação. A extração de atributos, realizada com a biblioteca TSFEL [21], foi empregada tanto na construção do modelo quanto na análise de novos dados, que podem ou não ter operado até a falha completa. Além disso, o processo de extração contribui para a otimização de atributos e hiperparâmetros, visando uma classificação mais precisa e uma análise interpretativa focada nos atributos mais relevantes.

3.5 Otimização de atributos e parâmetros da Árvore de Decisão

O processo de otimização se inicia pela seleção dos atributos considerados mais importantes. O TSFEL é capaz de extrair automaticamente mais de 60 características em domínios estatísticos, temporais e espectrais além de uma variedade de coeficientes e atributos adicionais. Por exemplo, a função LPCC extrai um número específico de coeficientes, que pode variar de acordo com o parâmetro configurado e características derivadas de transformadas e análises avançadas, como energia da *wavelet*, entropia espectral e outros atributos relacionados ao domínio do tempo e da frequência. Logo, dentre a grande variedade de atributos é necessário selecionar quais atributos são relevantes para o classificador atuar de forma efetiva, excluindo atributos que não apresentam nenhuma importância na classificação.

3.5.1 Seleção de atributos

Os atributos extraídos foram unidos aos rótulos gerados. Os dados já rotulados foram classificados de acordo com sua relevância, utilizando a função *feature importance* de uma árvore de decisão. A função *feature importance* pode ser encontrada na biblioteca *scikit-learn* [22], uma das principais bibliotecas de aprendizado de máquina em *Python*. Em seguida, foi realizada uma análise para observar a evolução da acurácia ao adicionar progressivamente as características mais relevantes ao conjunto de treinamento. O objetivo foi determinar o número de características necessárias para alcançar a máxima acurácia ou identificar quando ela se estabilizaria. Os resultados podem ser observados na Figura 7, onde vemos que com 14 atributos a acurácia obtida foi estável comparada aos 50 atributos mais relevantes.

A importância dos atributos extraídos, expressa em valores unitários, pode ser observada na Figura 8. Os atributos selecionados estão descritos conforme [23], [24] e [25]:

- **Frequência Mediana:** Em análise de vibração, a frequência mediana é a frequência abaixo da qual 50% da energia do sinal de vibração é concentrada. Esta métrica é útil para identificar mudanças no padrão de frequência ao longo do tempo, como a presença de falhas incipientes em componentes de máquinas. Quando há um aumento na frequência mediana, pode indicar desgaste progressivo ou a presença de defeitos.

- **LPCC (Coeficiente Cepstral Linear Predictivo):**

LPCC_8: Os Coeficientes Cepstrais Lineares Predictivos são extraídos do espectro de um sinal de vibração, representando suas características em termos de amplitude e frequência. O oitavo coeficiente (LPCC_8) captura uma parte específica do

espectro que pode estar associada a certas características do sinal, como padrões recorrentes que indicam tipos específicos de falhas.

LPCC_2: Similar ao LPCC_8, mas representa uma parte diferente do espectro. Como o segundo coeficiente, ele tende a capturar informações mais gerais ou globais do espectro.

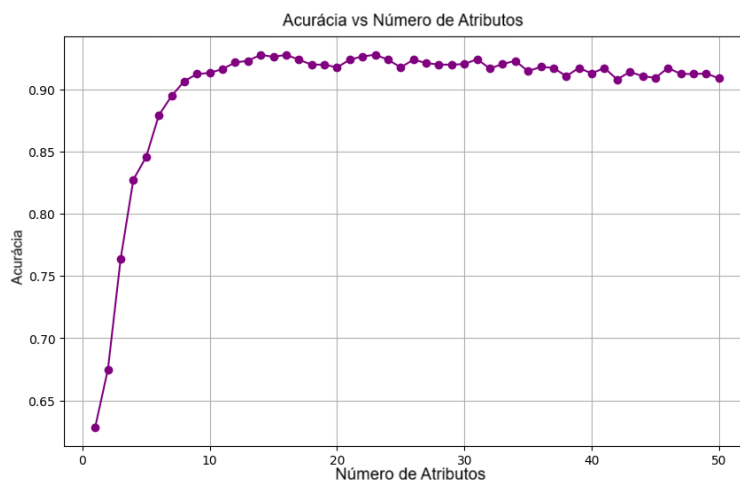


Figura 7: Seleção do número de características necessárias.

- **Roll-off Espectral:** Esta métrica define o ponto de frequência abaixo do qual uma porcentagem específica da energia total do espectro está concentrada. Em análise de vibração, pode ajudar a distinguir entre sinais de vibração normais e anomalias como ruído ou ressonância indesejada.
- **Frequência Fundamental:** Esta é a frequência mais baixa presente em um sinal de vibração e é frequentemente associada ao movimento periódico ou ao ciclo fundamental de um componente mecânico. Mudanças na frequência fundamental podem indicar alterações na condição operacional ou no estado de saúde de uma máquina.
- **Taxa de Cruzamento por Zero:** Mede quantas vezes o sinal de vibração cruza o eixo zero ao longo do tempo. Isso pode estar relacionado à quantidade de oscilação ou pulsação no sistema. Taxas elevadas de cruzamento por zero podem sugerir condições de vibração de alta frequência, como cavitação ou ressonância.
- **Diferença Absoluta Mediana:** Refere-se à mediana das diferenças absolutas entre valores consecutivos de um sinal de vibração. É uma medida robusta da variação central, e pode ser usada para detectar mudanças sutis na amplitude do sinal que podem ser indicativas de desgaste ou danos progressivos.
- **Desvio Absoluto Médio:** Mede a média das distâncias absolutas de cada valor em relação à média do sinal de vibração. Esse desvio dá uma ideia da variabilidade global do sinal, e aumentos nessa métrica podem indicar uma dispersão maior das frequências, potencialmente causada por uma condição de falha.
- **Centroide Espectral:** Representa o “centro de massa” do espectro de frequências do sinal de vibração. Frequentemente relacionado à percepção de “brilho” em sinais acústicos, no contexto de vibração, pode indicar o deslocamento das frequências dominantes no espectro devido a mudanças operacionais ou falhas.
- **Coefficiente Médio da FFT_1:** Refere-se ao valor médio do primeiro coeficiente obtido através da Transformada Rápida de Fourier (FFT) do sinal. Ele captura a componente de frequência mais baixa, que pode estar associada à rotação ou à frequência operacional fundamental da máquina.
- **Energia da Wavelet_5:** Mede a energia do sinal em uma determinada sub-banda de frequência após a decomposição wavelet. A sub-banda correspondente ao índice 5 captura uma faixa específica de frequências que pode estar associada a certas características de falha, como o desgaste de rolamentos.
- **Assimetria Espectral:** Avalia a simetria da distribuição espectral em torno de sua média. Desvios na assimetria podem sugerir a predominância de frequências altas ou baixas, que por sua vez podem estar associadas ao início de falhas ou a condições operacionais fora do normal.
- **Entropia de Wavelet:** É uma medida da complexidade ou da desordem do sinal obtida através da decomposição wavelet. Em análise de vibração, a entropia wavelet pode indicar a presença de irregularidades ou a complexidade de padrões de falha, fornecendo uma visão detalhada da condição de saúde do equipamento.

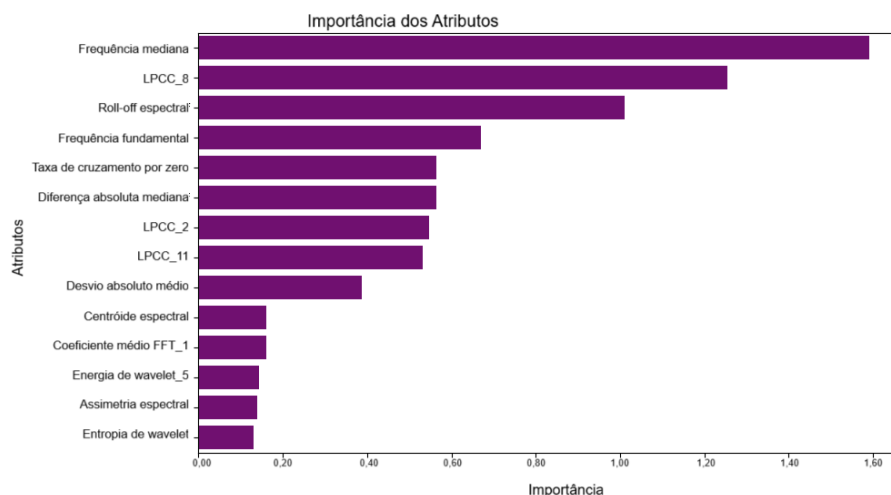


Figura 8: Importância das características selecionadas.

Após a seleção das características, o conjunto de dados de criação do modelo foi reduzido somente para os atributos selecionados. Além disso o conjunto foi dividido entre conjunto de treino e de teste derivados do conjunto de criação do modelo, onde 80% (2410 amostras) são destinadas ao treinamento e 20% (602 amostras) são destinadas ao teste.

3.6 Criação do classificador

Com o processo de busca randômica e validação cruzada são definidos, utilizando os dados de treinamento desenvolvido na etapa anterior, os melhores hiperparâmetros para a árvore de decisão buscando classificar o estágio de degradação de um sinal de vibração pertencente a uma das classes (Saudável + Estágio de Falha 1, Estágio de Falha 2, Estágio de Falha 3 e Estágio de Falha 4). O processo foi realizado em 300 interações de uma validação cruzada com 10 *folds*. Os hiperparâmetros foram selecionados com base no desempenho do melhor modelo alcançado durante a etapa de busca aleatória. Informações sobre os parâmetros e suas descrições são apresentadas na Tabela 2.

Tabela 2: Melhores Hiperparâmetros encontrados e suas descrições

Parâmetro	Valor	Comentário
'splitter'	'best'	Escolhe a melhor divisão possível com base no critério de impureza definido pelo parâmetro 'criterion', que, neste caso, é 'entropy'. O parâmetro 'splitter' mede a desordem dos dados em cada nó e seleciona a divisão que proporciona a melhor separação entre as classes, tornando a árvore mais eficiente. A opção 'best' mostrou-se superior em comparação à opção 'random' que não busca a divisão ótima, mas seleciona aleatoriamente uma entre as melhores opções disponíveis.
'min_samples_split'	2	Define o número mínimo de amostras necessárias para dividir um nó. O valor 2 significa que um nó precisa de pelo menos duas amostras para ser considerado para divisão.
'min_samples_leaf'	1	Indica o número mínimo de amostras que devem estar presentes em um nó folha. Com o valor 1, cada nó folha deve conter pelo menos uma amostra.
'max_depth'	None	Define a profundidade máxima da árvore. None significa que os nós serão expandidos até que todas as amostras sejam perfeitamente classificadas ou até que todas as folhas contenham menos do que min_samples_split amostras.
'criterion'	'entropy'	Define a função usada para medir a qualidade de uma divisão. 'Entropy' refere-se à entropia da informação, que é uma medida da incerteza ou impureza da divisão. O algoritmo da árvore de decisão tentará minimizar a entropia nas divisões feitas.

Por fim, os dados de treinamento e teste tratados anteriormente bem como os hiperparâmetros selecionados são utilizados para criar e treinar a árvore de decisão que irá ser responsável pela classificação dos estágios de degradação do rolamento. Após ser treinada, a árvore é disponibilizada para análise dos dados do conjunto de teste complementar. A representação visual da árvore de decisão, com seus nós e regras de decisão estão disponíveis na página do GitHub [26].

3.7 Classificação de novos dados

Para que seja realizada a classificação de novos dados, o modelo desenvolvimento precisa receber informações para que os dados sejam padronizados de acordo com o modelo de entrada. Tais informações que estão relacionadas a coleta são: a frequência, o tamanho das janelas e o intervalo entre as coletas, informações importantes e cruciais no pre-processamento dos dados. Após essa padronização os atributos são extraídos e os dados são enviados ao modelo que retorna a probabilidade da amostra pertencer a cada um dos estágios. O processo de funcionamento está descrito na Figura 9.

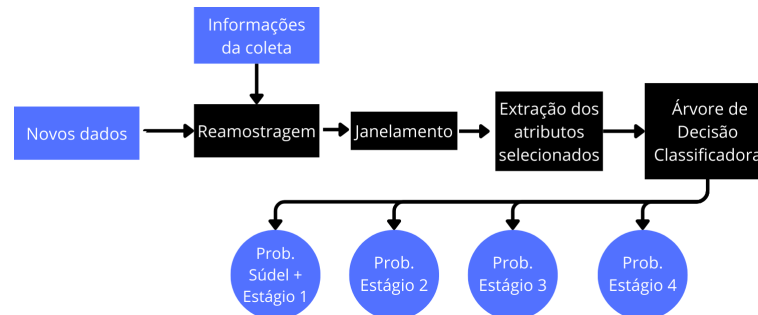


Figura 9: Método proposto - Novos dados.

Os dados dos acelerômetros vertical e horizontal são carregados e processados para prever os estágios de degradação. Inicialmente, são obtidas informações sobre a frequência de coleta, o tamanho das janelas e o intervalo entre coletas. Em seguida, os dados passam por reamostragem e janelamento, permitindo a extração de atributos relevantes. Esses atributos são então combinados para representar melhor o comportamento do sistema. Com os dados preparados, um modelo de árvore de decisão previamente treinado é carregado para calcular as probabilidades de cada estágio de degradação. Por fim, os resultados são exibidos, fornecendo uma análise detalhada do estado do sistema com base nos sensores analisados.

3.8 Análise interpretativa

Por fim, a biblioteca SHAP é empregada para análise de interpretabilidade do modelo, avaliando a importância e a dependência das características no processo de decisão. O SHAP permite extrair informações importantes ao fornecer uma explicação detalhada das previsões do modelo. Essas informações são cruciais para expandir os estudos na área, oferecendo uma compreensão mais profunda dos fatores que influenciam a degradação dos mancais e possibilitando o desenvolvimento de estratégias de manutenção preditiva mais eficazes.

3.9 Métricas de avaliação

Neste trabalho foram utilizadas as seguintes métricas de avaliação de desempenho do classificador [12]:

- **Acurácia (ACC)**

A acurácia é considerada uma das métricas mais simples e importantes. Ela avalia o percentual de classificações corretas, que pode ser obtido pela razão entre o número de previsões corretas e o total de amostras, e pode ser descrita através da seguinte equação 5:

$$ACC = \frac{V_p + V_n}{V_p + F_p + V_n + F_n}, \quad (5)$$

onde V_p pode ser denominado de verdadeiro positivo, V_n verdadeiro negativo, F_p falso positivo e F_n falso negativo. Embora a acurácia seja amplamente utilizada, ela pode ser insuficiente em cenários onde os dados estão desbalanceados, já que, neste caso, um resultado geral das previsões, necessariamente, não expressa o bom desempenho do classificador.

- **Precisão (P)**

A precisão é definida como a proporção de verdadeiros positivos entre todas as instâncias preditas como positivas, conforme a equação 6:

$$P = \frac{V_p}{V_p + F_p}. \quad (6)$$

- **Sensibilidade (S)**

A sensibilidade é uma métrica que reflete a capacidade do modelo em identificar corretamente as instâncias positivas. Ela é calculada como a proporção de verdadeiros positivos entre todas as instâncias que deveriam ter sido preditas como positivas, conforme a equação 7:

$$S = \frac{V_p}{V_p + F_n}. \quad (7)$$

• F1-Score (F1)

O F1-Score é uma métrica que combina a precisão e a sensibilidade em uma única medida. Ele é especialmente útil em cenários de classificação desbalanceada, pois fornece uma visão mais completa do desempenho do modelo. O F1-Score é calculado como a média harmônica da precisão e da sensibilidade, conforme a equação 8:

$$F1 = 2 \cdot \frac{P \cdot S}{P + S}. \quad (8)$$

4. RESULTADOS

Nesta seção são detalhados e discutidos os resultados obtidos. As informações presentes estão divididas segundo a metodologia apresentada na seção 3.

4.1 Classificação

Com o modelo treinado e empregando os 20% do banco de dados restante é realizado um teste final, ao qual conseguiu uma acurácia de 92,7%, resultado superior ao melhor resultado alcançado por [6] de 89%. Além da acurácia, os dados de precisão e sensibilidade podem ser observados na Tabela 3

Classe	Precisão (P)	Sensibilidade (S)	F1-Score (F1)
Saúde + Estágio 1	0,96	0,91	0,94
Estágio 2	0,92	0,90	0,91
Estágio 3	0,92	0,94	0,93
Estágio 4	0,94	0,96	0,95
Acurácia	0,927 (1507 amostras)		

Tabela 3: Precisão, Sensibilidade e F1-Score para cada classe.

Analisando os dados de precisão e sensibilidade, é possível observar que no geral, as métricas tem valores próximo e uma análise mais consistente das classificações errôneas toma como base a matriz de confusão no conjunto de teste pode ser visualizada na Figura 10.

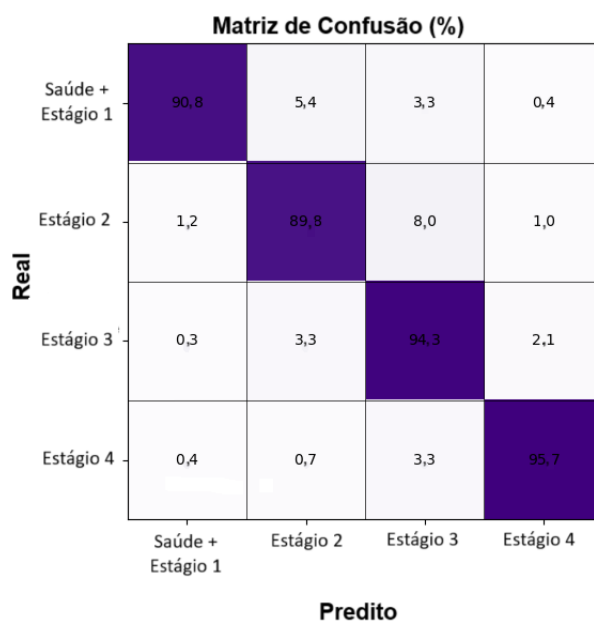


Figura 10: Matriz de Confusão.

A matriz de confusão apresenta, em suas linhas, as classes reais e, em suas colunas, as classes preditas. Dessa forma, cada amostra é contabilizada em uma posição $n \times m$, onde n corresponde à linha da classe real e m à coluna da classe predita. Quando $n = m$, ou seja, a amostra é contabilizada na diagonal principal, significa que a amostra foi classificada corretamente.

No contexto da análise do processo evolutivo do presente trabalho, classificações incorretas podem ser aceitáveis, desde que ocorram para classes adjacentes. Como observado, a maior parte das amostras classificadas incorretamente na Figura 10 foram alocadas nas classes vizinhas, o que pode ser explicado pelos momentos de transição entre os estágios, onde as fronteiras entre as classes não estão claramente definidas.

Após isso, os rolamentos que compõe o conjunto de teste complementar foram aplicados, e os resultados para o rolamento 1_5 podem ser vistos na Figura 11. É possível a partir de uma análise visual observar como o rolamento se alterna entre os estágios conforme ocorre a degradação e como o aumento de defeitos acelera o processo de degradação diminuindo o tempo dos estágios seguintes.

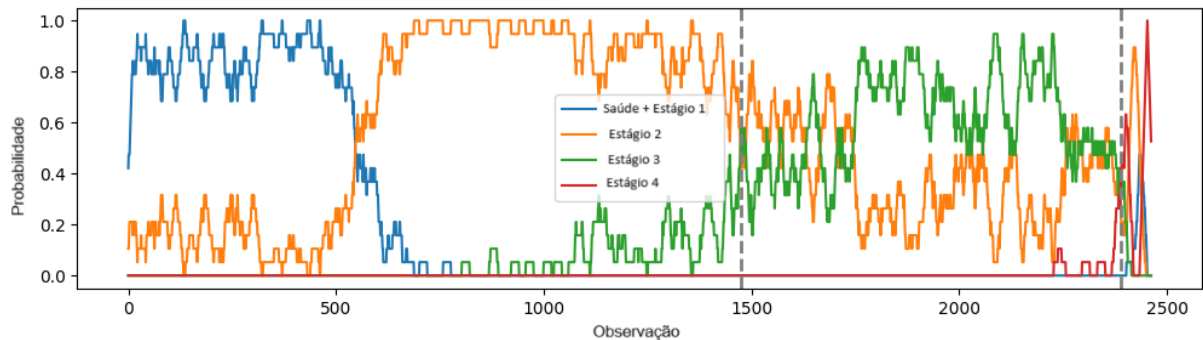


Figura 11: Classificação - Rolamento 1_5.

4.2 Análise através do SHAP

Dada a vasta quantidade de informações que podem ser extraídas usando o método SHAP para cada uma das quatro classes possíveis, o estudo focou na análise de padrões e relações nas classificações do algoritmo no estágio 4. Esta análise pode oferecer percepções valiosas sobre como as falhas se manifestam e, mais especificamente, os tipos de falhas que podem surgir. Observa-se antecipadamente que, embora o número de atributos tenha sido reduzido na etapa de seleção de características, ainda há uma quantidade considerável de atributos e análises potenciais a serem realizadas. No entanto, esta seção se limita às principais análises observadas e conduzidas pelo autor. Isso não implica que os demais atributos e suas correlações sejam irrelevantes para o estudo, mas sim que o autor optou por focar em um conjunto reduzido de variáveis, com o objetivo de simplificar a análise e destacar os aspectos mais pertinentes.

Inicialmente, o gráfico que correlaciona os valores de uma característica com os valores SHAP da respectiva amostra foi extraído, conforme mostrado na Figura 12.

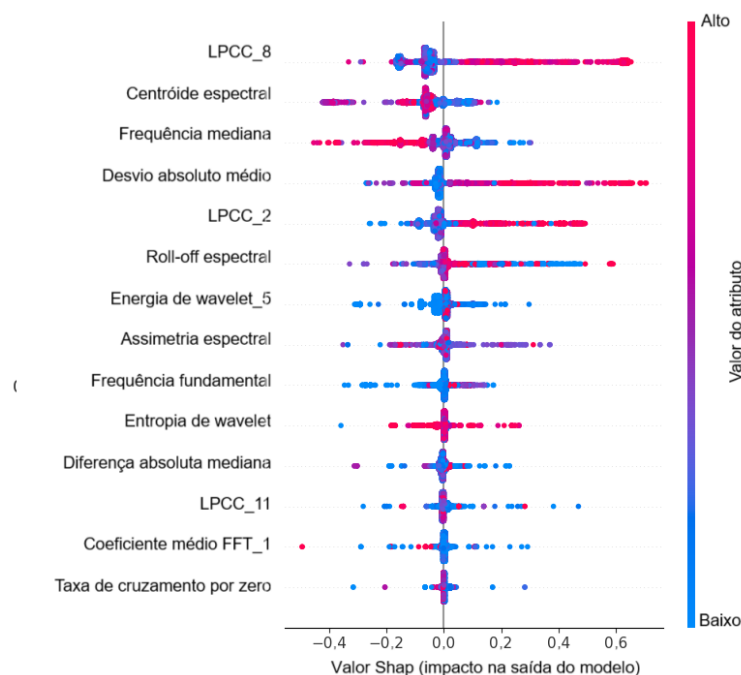


Figura 12: Gráfico de Impacto SHAP - Estágio 4.

Observa-se que, para as características desvio absoluto médio, LPPC_8 e LPPC_2.1, os valores mais altos estão diretamente relacionados aos maiores impactos positivos na classificação do estágio 4. Isso sugere que essas informações estão intimamente ligadas a eventos característicos desse estágio. Os coeficientes LPCC capturam informações sobre o espectro de potência de um sinal, sendo úteis na análise de sinais de vibração para identificar padrões e frequências dominantes. Por exemplo, na detecção de falhas em máquinas como motores industriais, os LPCCs podem ser usados para extrair características distintas de padrões de vibração associados a diferentes tipos de falhas mecânicas.

Por outro lado, o desvio médio absoluto é uma métrica estatística crucial para avaliar a dispersão dos pontos de dados em relação à média do conjunto. Ela fornece uma indicação direta da variabilidade dos dados, independentemente dos *outliers*, sendo uma ferramenta valiosa para entender a distribuição e consistência dos dados. A Frequência Mediana pode ser interpretada como a frequência abaixo da qual metade da energia vibracional está concentrada. Valores mais altos dessa característica têm um impacto negativo na classificação do estágio 4, indicando sua forte influência na classificação de outras classes.

Outra análise que pode ser realizada é sobre como a dependência entre duas características pode afetar a decisão do algoritmo, seja de forma positiva ou negativa. Neste estudo, essa análise é limitada à identificação de regiões com maior impacto na decisão do algoritmo e à sugestão de possíveis relações, abrindo espaço para estudos futuros.

A primeira relação de dependência é analisada entre as características LPCC_8 e o centroide espectral, e pode ser observada na Figura 13-a).

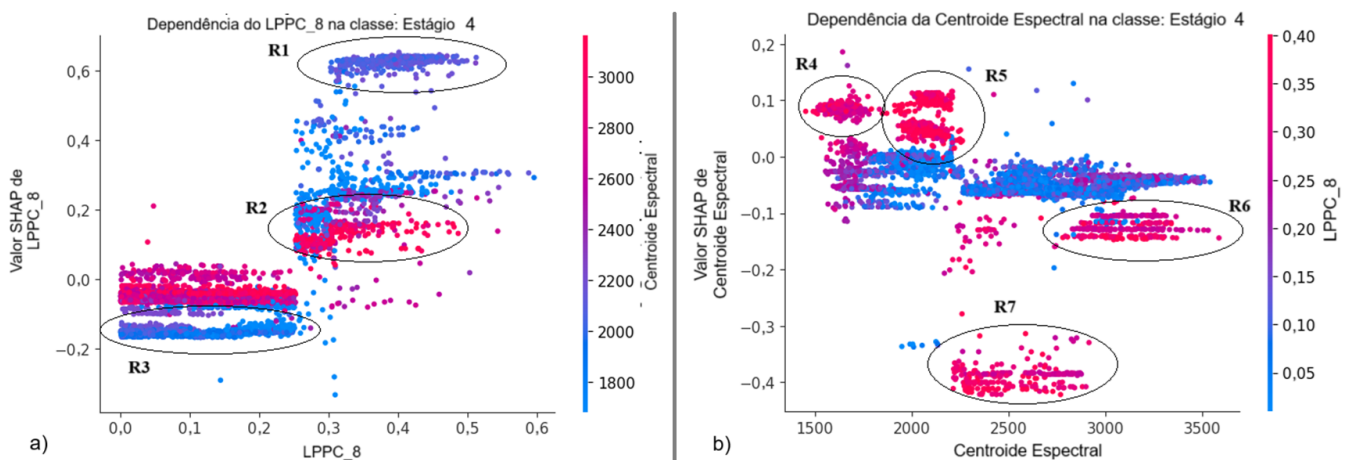


Figura 13: Estágio 4 a) Dependência entre LPCC_8 e centroide espectral, b) Dependência entre Centroide Espectral e LPCC_8.

A relação entre o coeficiente LPCC_8 e o centroide espectral é significativa, pois ambos fornecem informações cruciais sobre as características espectrais de um sinal de vibração. Enquanto o LPCC_8 captura aspectos específicos do espectro de frequência relacionados à forma da onda do sinal, o centroide espectral fornece uma medida agregada da distribuição de energia espectral. Essas duas medidas complementares permitem uma compreensão abrangente das propriedades espectrais.

Como visto, o gráfico apresenta em seus eixos a faixa de valores de LPCC_8 e os valores SHAP correspondentes para cada uma das amostras. Os valores do centroide espectral, por sua vez, têm sua faixa de valores destacada usando um padrão de cores, assim é possível observar e destacar algumas regiões. A Região R1 denota valores de LPCC_8 entre 0,3 e 0,5 para valores de centroide espectral entre 2000 e 2400. Tal região influencia fortemente os valores SHAP para as amostras contidas nela. A Região R2 também se destaca, que, embora tenha os mesmos valores de LPCC_8, por ter uma faixa maior de centroide espectral (acima de 2800), apresenta valores SHAP próximos de zero. A Região R3, por outro lado, apresenta valores de centroide espectral na mesma faixa que R1, mas apresenta valores de LPCC_8 entre 0 e 0,3, o que indica uma forte influência negativa no SHAP, sugerindo a não-pertinência à classe.

Na Figura 13-b), por outro lado, apresenta em seus eixos a faixa de valores de centroide espectral e os valores SHAP correspondentes para cada uma das amostras. Os valores de LPCC_8 tem sua faixa de valores destacada usando um padrão de cores. Neste caso, podemos identificar que as regiões R4, R5 e R6, que, embora apresentem valores de centroide espectral distribuídos por quase toda a faixa e apresenta LPCC_8 acima de 0,25, têm pouco impacto na classificação, com valores SHAP entre -0,1 e 0,1. Já a região R7, que possui os mesmos valores de LPCC_8 das regiões R4, R5 e R6, mas concentrados em uma faixa específica do centroide espectral (cerca de 2500), apresenta impacto elevado em módulo, porém negativo, indicando que os dados não pertencem a esse estágio, sendo um forte indicativo de que os dados associados a região R7 pertencem a outro estágio de degradação.

5. CONCLUSÃO

No sistema proposto, observou-se a eficácia das técnicas FFT, PCA e k-means na rotulagem, bem como das árvores de decisão na detecção e classificação dos estágios de falhas em mancais. Essas técnicas contribuíram significativamente para a melhoria da manutenção preditiva, mostrando-se computacionalmente eficientes e apresentando acurácia superior em comparação com trabalhos anteriores.

Com base no conhecimento de domínio, foi desenvolvido um método para segmentação das classes de degradação, utilizando sinais de vibração de alta frequência. Esses sinais após serem transformados para o domínio da frequência utilizando a FFT, foram projetados em um subespaço latente de baixa dimensionalidade por meio da PCA, automatizando o rotulamento dos dados e o possibilitando o treinamento do classificador a partir de um banco de dados com rótulos. Além disso, um classificador supervisionado baseado em árvore de decisão foi implementado para detectar os estágios de degradação dos rolamentos, demonstrando alto desempenho na classificação dos dados. A análise interpretativa dos resultados foi realizada por meio da técnica SHAP, que destacou as relações quantitativas entre o valores de características e o impacto daquela amostra na classificação dos estágios de falha.

Para o estágio 4, no desvio absoluto médio, valores mais altos do atributo impactam positivamente a classificação do estágio enquanto valores mais baixos não apresentam nenhum impacto, no LPCC 8 e LPCC 2, valores mais altos apresentam um alto impacto na classificação enquanto valores menores apresentam em menor escala um impacto negativo. Tal análise pode ser expandida para cada uma das classes de atributos. Essa interpretação ofereceu informações valiosas sobre os padrões de falha possibilitando a abertura de estudos do comportamento do rolamento durante o processo de degradação.

Os resultados indicam que o método proposto melhorou a classificação dos estágios de degradação dos mancais. Essa capacidade de previsão de evolução dos estágios fornece uma janela temporal crucial para a substituição preventiva das peças, aumentando a confiabilidade do sistema. Além disso, a relação entre LPCC 8 e o centroide espectral revelou padrões de falha relevantes, que podem ser investigados em estudos futuros.

Por fim, para trabalhos futuros, propõe-se o desenvolvimento de um software de aplicação prática baseada no classificador além da integração de um sistema de predição da vida útil remanescente, utilizando os múltiplos estágios de degradação como dados adicionais. Essa predição não apenas fornecerá informações quantitativas sobre a vida útil restante dos rolamentos, mas também permitirá identificar padrões e correlações que aprimorem a manutenção preditiva e corretiva.

REFERÊNCIAS

- [1] Y. Ran, X. Zhou, P. Lin, Y. Wen and R. Deng. “A Survey of Predictive Maintenance: Systems, Purposes and Approaches”, 2019.
- [2] L. Nie, L. Zhang, S. Xu *et al.*. “Remaining useful life prediction for rolling bearings based on similarity feature fusion and convolutional neural network”. *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, vol. 44, pp. 328, 2022.
- [3] Z. Chen, X. Tu, Y. Hu and F. Li. “Real-Time Bearing Remaining Useful Life Estimation Based on the Frozen Convolutional and Activated Memory Neural Network”. *IEEE Access*, vol. 7, pp. 96583–96593, 2019.
- [4] C. Wang, W. Jiang, X. Yang and S. Zhang. “RUL Prediction of Rolling Bearings Based on a DCAE and CNN”. *Applied Sciences*, vol. 11, pp. 11516, 2021.
- [5] I. STI Vibration Monitoring. “Rolling Element Bearings: Field Application Note”, 2012.
- [6] D. Juodelyte, V. Cheplygina, T. Graversen and P. Bonnet. “Predicting Bearings’ Degradation Stages for Predictive Maintenance in the Pharmaceutical Industry”, 2022.
- [7] L. Zhang, J. J. Lin, B. Liu, Z. Zhang, X. Yan and M. Wei. “A Review on Deep Learning Applications in Prognostics and Health Management”. *IEEE Access*, vol. 7, pp. 162415–162438, 2019.
- [8] C. Scheffer and P. Girdhar. *Practical Machinery Vibration Analysis and Predictive Maintenance*. Elsevier, 2004.
- [9] R. Zambrano López. “Classificação Automática de Defeitos em Máquinas Rotativas”. Ph.D. thesis, Universidade Federal do Rio de Janeiro, Escola Politécnica, 2014. Graduação em Engenharia Eletrônica e de Computação.
- [10] T. Simi and N. Resmi. “A Robust Data-Driven Predictive Maintenance Framework for Industrial Machinery using Explainable Machine Learning Techniques”. In *2023 9th International Conference on Smart Computing and Communications (ICSCC)*, pp. 138–143. IEEE, 2023.
- [11] E. Iunusova, M. K. Gonzalez, K. Szpika and A. Archenti. “Early fault diagnosis in rolling element bearings: comparative analysis of a knowledge-based and a data-driven approach”. *Journal of Intelligent Manufacturing*, vol. 35, no. 5, pp. 2327–2347, 2024.
- [12] K. Faceli, A. C. Lorena, J. Gama and A. C. P. d. L. Ferreira de Carvalho. *Inteligência Artificial: Uma Abordagem de Aprendizado de Máquina*. LTC, 2011.
- [13] C. M. Maia, J. C. M. Gomes and L. D. Chagas. “Estudo sobre o uso de árvores de decisão na área da saúde”. *Anais do Encontro de Computação do Oeste Potiguar ECOP/UFERSA (ISSN 2526-7574)*, vol. 1, no. 1, 2017.
- [14] R. L. Assis. “Detecção de fraudes em cartões de crédito utilizando métodos de baseados em árvores de decisão”. B.S. thesis, Universidade Tecnológica Federal do Paraná, 2023.

- [15] A. P. B. Sobral. “Previsao de carga horária-Uma nova abordagem por Arvore de Decisao”. Ph.D. thesis, PhD thesis, Pontificia Universidade Católica do Rio de Janeiro, 2003.
- [16] L. E. d. Abreu. “People analytics: uso de arvores de decisão na retenção de talentos”. 2022.
- [17] S. M. Lundberg and S. I. Lee. “A unified approach to interpreting model predictions”. In *Advances in Neural Information Processing Systems*, volume 30, pp. 4765–4774, 2017.
- [18] “SHAP Documentation”. Online, <https://shap.readthedocs.io/>, 2024. Accessed: 2024-08-12.
- [19] P. Nectoux, R. Gouriveau, K. Medjaher, E. Ramasso, B. Chebel-Morello, N. Zerhouni and C. Varnier. “PRONOSTIA: An Experimental Platform for Bearings Accelerated Degradation Tests”. In *IEEE International Conference on Prognostics and Health Management (PHM)*, pp. 1–8, 2012.
- [20] C. E. Shannon. *The Mathematical Theory of Communication*. University of Illinois Press, 1949.
- [21] M. Barandas, D. Folgado, L. Fernandes, S. Santos, M. Abreu, P. Bota, H. Liu, T. Schultz and H. Gamboa. “TSFEL: Time Series Feature Extraction Library”. *SoftwareX*, vol. 11, pp. 100456, 2020.
- [22] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, R. Layton, M. Astrahan, S. Hoyer, W. McKinney, P. Vincent, J. VanderPlas, A. Stéfan *et al.*. “Scikit-learn: Machine Learning in Python”, 2011.
- [23] J. K. Sinha. *Vibration Analysis, Instruments, and Signal Processing*. CRC Press, Boca Raton, FL, 2014.
- [24] M. Misiti, Y. Misiti, G. Oppenheim and J.-M. Poggi. *Wavelet Analysis and Applications*. Springer, Berlin, Germany, 2007.
- [25] T. F. Quatieri. *Discrete-Time Speech Signal Processing: Principles and Practice*. Prentice Hall, Upper Saddle River, NJ, 2001.
- [26] F. S. Filho. “Árvore de Decisão”. Online, <https://github.com/Gdias2701/Arv2DT>, 2024. Acessado em: 2 abr. 2025.