

Detecção de anomalias em rolamentos de aerogeradores baseada em modelos não supervisionados

Samuel Bezerra da Gama Neto¹, Gustavo de Novaes Pires Leite^{2*}, Rodrigo Cesar Lira da Silva², Frederico Duarte de Menezes², Alexandre Carlos Araujo da Costa³, Olga de Castro Vilela³, Leonardo José de Petribú Brennand³, Rodrigo Henrique de Lima Farias³, Giovanni Moura de Holanda⁴, Fabrício Cristófani⁴, Edgar Lessa Venancio⁵, Isabela Niedo Marchiori⁵, Jefferson Silva Barbosa⁵

¹Universidade Federal de Pernambuco (UFPE), Brasil

²Instituto Federal de Educação, Ciência e Tecnologia de Pernambuco (IFPE), Brasil

³Centro de Energias Renováveis da UFPE (CER-UFPE), Brasil

⁴Fundação para Inovações Tecnológicas (FITec), Brasil

⁵CPFL Renováveis, Brasil

* Autor para correspondência

samuel.gneto@ufpe.br, gustavonovaes@recife.ifpe.edu.br, rodrigo.lira@paulista.ifpe.edu.br, fredericomenezes@recife.ifpe.edu.br, alexandre.acosta@ufpe.br, olga.vilela@ufpe.br, leonardo.brennand@ufpe.br, rodrigo.hlfarias@ufpe.br, gholanda@fitec.org.br, fabriciocristofani@fitec.org.br, edgar.venancio@cpfl.com.br, imarchiori@cpfl.com.br, jefferson.barbosa@cpfl.com.br

RESUMO

O planejamento da manutenção de parques eólicos é uma tarefa complexa, uma vez que os aerogeradores estão expostos a condições ambientais adversas que podem acelerar o desgaste dos componentes e que dificultam a identificação prévia de falhas. Além disso, a falta de um plano de manutenção adequado pode comprometer a confiabilidade dos aerogeradores, reduzindo a eficiência e diminuindo a margem de lucro dos parques eólicos. Neste contexto, uma estrutura para detecção de anomalias de rolamentos foi aplicada aos dados de vibração de três parques eólicos localizados no Nordeste do Brasil. Foram utilizados dois procedimentos para a rotulação dos dados e extração de atributos em diferentes domínios para o treinamento de doze modelos não supervisionados. Para detecção final da anomalia, uma estratégia de comitês de modelos baseada na votação majoritária foi adotada. A validação da metodologia foi realizada para o caso de ocorrência de falha no rolamento do eixo de alta velocidade do *gearbox* de quatro aerogeradores. Os resultados demonstraram que o método proposto conseguiu identificar corretamente anomalias em dois dos aerogeradores analisados. Os dados foram obtidos no âmbito de um projeto de P&D Aneel, chamado de CPFL.APE, e que está sendo desenvolvido por um consórcio entre UFPE, IFPE, FITEC e CPFL.

Palavras-chaves: Detecção de anomalias; Aerogeradores; Rolamentos; Aprendizagem não supervisionada

ABSTRACT

Maintenance planning for wind farms is a complex task, as wind turbines are exposed to adverse environmental conditions that can accelerate component wear and make it difficult to identify early faults. Additionally, the lack of an adequate maintenance plan can compromise the reliability of the wind turbines, reducing efficiency and decreasing the profit margins of the wind farms. In this context, a framework for bearing anomaly detection was applied to vibration data from three wind farms located in the Brazilian Northeast. Two procedures were used for data labeling, and feature extraction was performed in different domains for the training of twelve unsupervised models. For final anomaly detection, a majority voting-based combination strategy was adopted. The methodology was validated for the case of high-speed shaft bearing failure in the gearboxes of four wind turbines. The results demonstrated that the proposed method correctly identified anomalies in two of the analyzed wind turbines. The data were obtained under an Aneel R&D project called CPFL.APE, that is being developed by a consortium between UFPE, IFPE, FITEC and CPFL.

Keywords: Anomaly detection; Wind turbines; Bearings; Unsupervised learning

1. INTRODUÇÃO

Aerogeradores (AEGs) são dispositivos rotativos cuja geração de energia elétrica está diretamente ligada à disponibilidade e à velocidade do vento. Estes equipamentos são constituídos de sete principais componentes: pás, *hub*, torre, nacele, rolamento principal, *gearbox* e gerador (Olabi *et al.*, 2021). Segundo Santelo *et al.* (2022), um dos principais contribuintes para o tempo de inatividade dos AEGs é o *gearbox*, devido ao seu tamanho e complexidade. As falhas nesse componente são causadas principalmente pela quebra dos dentes das engrenagens e pelo desgaste dos rolamentos.

Em AEGs, os rolamentos podem ser afetados por cargas excessivas devido à dinâmica do fluxo de vento e às condições ambientais, resultando em vários modos de falha, como deformação plástica, fadiga e corrosão (Liu; Zhang, 2020). Neste contexto, visando contornar esses problemas diferentes técnicas para a detecção, diagnóstico e prognóstico de falhas em rolamentos foram publicadas na literatura, como a análise de vibração e modelos de aprendizagem de máquina supervisionada, que necessitam de especialistas ou da intervenção humana.

Nesse sentido, este estudo propõe o uso de modelos não supervisionados para detecção de anomalias de rolamentos utilizando dados de vibração de AEGs de 2,1 MW de três parques eólicos localizados no Nordeste do Brasil, que são oriundos de um projeto de P&D ANEEL, chamado de CPFL.APE, e que está sendo desenvolvido por um consórcio entre UFPE, IFPE, FITEC e CPFL. Modelos de detecção de anomalias são capazes de detectar pontos de dados que se desviam do padrão normal, cujas principais vantagens são: (i) detecção de *outliers*; (ii) habilidade de ajustar e adaptar para diferentes aplicações (Omar *et al.*, 2013; Boukerche *et al.*, 2020).

Portanto, os principais objetivos deste estudo são:

- i) Identificar os eventos de falhas dos rolamentos dos AEGs disponibilizados pelos parques eólicos;

- ii) Extrair atributos nos domínios do tempo, frequência, teoria da informação e mecânico dos dados brutos de vibração;
- iii) Aplicar modelos não supervisionados para detecção de anomalias;
- iv) Utilizar um comitê de modelos para detecção de anomalias em AEGs.

Este trabalho está organizado da seguinte forma: a Seção 2 apresenta estudos relacionados à detecção de anomalias, a Seção 3 detalha a metodologia adotada, a Seção 4 fornece resultados e discussão. Por fim, a Seção 5 apresenta as conclusões e direções de pesquisas futuras.

2. ESTADO DA ARTE

No contexto da detecção de anomalias, diversas técnicas têm sido aplicadas a AEGs. Um exemplo é o estudo de Lin *et al.* (2020), no qual o modelo de floresta de isolamento (do inglês, *Isolation Forest*) foi utilizado para remover pontos extremos do sistema supervisório. Em seguida, uma rede neural de aprendizagem profunda foi empregada para prever a potência dos AEGs. Os resultados mostraram que o erro quadrático médio foi de 0,114 para os modelos treinados com os dados brutos e de 0,003 para os modelos treinados com os dados filtrados.

Para detecção de falhas em AEGs, Trizoglou *et al.* (2021) implementaram uma estrutura baseada na detecção de anomalias, seleção de atributos e treinamento de modelos com o comportamento livre de falhas. Para a detecção de anomalias foi adotada a floresta de isolamento. Os modelos utilizados para a detecção de falhas foram o *Extreme Gradient Boost* (XGBoost) e a rede neural de memória de curto e longo prazo (LSTM).

Chen *et al.* (2021) propuseram o uso de uma LSTM combinada com um *autoencoder* para detecção de anomalias. Nessa abordagem, o modelo foi treinado com dados do sistema supervisório referentes a períodos livres de falhas. De modo que as previsões de condições anômalas resultassem em um erro de reconstrução maior do que o observado para condições normais. Uma abordagem semelhante foi aplicada por Xiang *et al.* (2022), que implementaram uma técnica de detecção de anomalias em séries temporais combinando uma rede neural convolucional, uma rede neural bidirecional *Gated Recurrent Unit* (GRU) e um mecanismo de atenção.

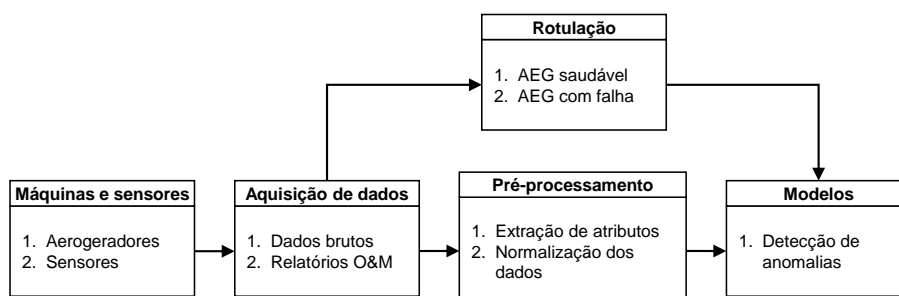
Para classificação dos dados, Leite *et al.* (2023) aplicaram uma metodologia buscando a detecção de anomalias para o rolamento da extremidade sem acionamento do gerador de AEGs utilizando dados de vibração de doze aerogeradores. A metodologia analisou individualmente onze modelos e combinou uma estratégia 5x2 de validação cruzada para garantir a melhor generalização dos modelos. Os resultados alcançaram uma acurácia 96,3% na detecção de falhas.

Munguba *et al.* (2024) utilizaram um comitê de técnicas de detecção de anomalias na sua proposta. Eles empregaram essas técnicas nos rolamentos principais, no rolamento do eixo de alta velocidade da caixa de engrenagens, e nos rolamentos da extremidade com e sem acionamento do gerador. As saídas das combinações foram comparadas pelos valores dos rótulos obtidos através da análise de vibração, alcançando os melhores resultados da métrica de área sob a curva (AUC) de 98% para o caso do rolamento principal.

3. MÉTODO DO ESTUDO

A metodologia aplicada neste estudo, cujo objetivo é avaliar a detecção de anomalias de rolamentos de AEGs, pode ser resumida por meio do diagrama apresentado na Figura 1, no qual se têm os seguintes passos principais: (i) aquisição de dados; (ii) pré-processamento; (iii) rotulação; (iv) desenvolvimento de modelos de detecção de anomalias.

Figura 1 – Metodologia para detecção de anomalias aplicada a dados de vibração do CMS de AEGs.



Fonte: Autores, 2024.

3.1 AQUISIÇÃO DOS DADOS

De acordo com a Figura 1, o primeiro passo da metodologia consiste no acesso aos dados brutos dos acelerômetros no banco de dados do *Condition Monitoring System* (CMS). Além disso, os relatórios de operação e manutenção (O&M) foram utilizados para identificação de manutenções, substituição de componentes e registros de falhas em AEGs, incluindo os modos e as datas em que essas falhas foram observadas, para auxiliar no processo de rotulagem dos dados.

Os dados fornecidos pelo CMS dos parques eólicos analisados são divididos em:

- i) Dados brutos de vibração: representa a série temporal de aceleração, medida em g , com uma taxa de frequência, fs , e período de coleta, T . Para o sistema CMS analisado, $fs = 20.000 \text{ Hz}$ e $T = 20s$, o que totaliza 400.000 amostras por acelerômetro e por coleta;
- ii) Dados de tendência: esses são dados coletados e extraídos automaticamente pelo CMS determinados pelo RMS (*root mean square*) do sinal de vibração.

3.2 PRÉ-PROCESSAMENTO

Com respeito ao segundo passo no diagrama da Figura 1 (pré-processamento), duas etapas foram empregadas: extração de atributos e normalização dos dados. Deve-se ressaltar que os dados de vibração de componentes mecânicos de carga variável como aqueles componentes presentes nos AEGs são complexos, altamente voláteis e imprevisíveis, principalmente devido a rajadas de vento, mudanças de direção e cisalhamento da velocidade do vento, que produzem assimetrias no rotor (Feng *et al.*, 2015). Para capturar as diferentes dinâmicas das formas de onda da vibração em um ambiente tão complexo, foram extraídos atributos de domínios distintos (domínios do tempo, frequência, teoria da informação e

mecânico) (Caesarendra; Tjahjowidodo, 2017). Tais características são definidas de forma a fornecer informações suficientes aos modelos de aprendizado de máquina para lhes permitir a separação de estados saudáveis dos estados defeituosos dos AEGs. Os atributos extraídos do sinal de vibração utilizados para o treinamento dos modelos foram:

- i) Domínio do tempo: raiz quadrada média, variância, curtose, pico a pico, simetria e fator de crista;
- ii) Domínio da frequência: frequência média, frequência central, frequência RMS e frequência de desvio padrão;
- iii) Domínio da teoria da informação: entropia, divergência Jensen-Shannon, complexidade estatística da permutação, do *power spectrum* e da *wavelet* da função de massa de probabilidade;
- iv) Domínio mecânico: harmônicos e *sidebands* da frequência de *Ball Pass Frequency Outer* (BPFO), *Ball Pass Frequency Inner* (BPFI) e *Ball Spin Frequency* (BSF), e velocidade do eixo.

A segunda etapa do pré-processamento, antes do treinamento dos modelos de detecção de anomalias, consistiu na normalização do conjunto de dados. A normalização é um procedimento para evitar que ocorra influência no desempenho dos modelos de aprendizagem de máquina pela diferença de variabilidade entre os atributos do conjunto de dados (Singh; Singh, 2020). No caso específico, foi aplicada a normalização (Z-score) em relação à média μ e ao desvio padrão σ , conforme a Eq. 1, para cada amostra x_n de cada atributo i do conjunto de dados.

$$x_{i,n}^* = \frac{x_{i,n} - \mu_i}{\sigma_i} \quad (1)$$

3.3 ROTULAÇÃO

O terceiro passo da metodologia, que ocorre simultaneamente ao passo 2, envolve o procedimento de rotulação dos dados. Esse procedimento inclui a classificação dos AEGs em dois grupos, saudáveis e com falhas, com o objetivo de auxiliar na avaliação da performance dos modelos. Nesta etapa, foram utilizadas duas abordagens distintas: registro de falhas e limites de vibração conforme norma.

A rotulação com base no registro de falhas consistiu na análise das informações registradas por componente (rolamento principal, *gearbox* e gerador), concentrando-se principalmente em problemas relacionados a engrenagens e rolamentos.

Na abordagem baseada nos limites de vibração conforme norma, foram examinados os alarmes de falhas obtidos pelos dados de tendência do CMS. Nesse caso, considera-se como alarme de falha qualquer valor de vibração acima de uma das duas zonas de vibrações determinadas pela norma ISO 20816-21, conforme a Tabela 1.

De acordo com a norma ISO 20816-21, a Zona B representa valores de vibrações nas quais os

componentes do AEG podem operar por um longo período. Por sua vez, a Zona C representa valores não adequados para a operação por um longo período. Com esses valores de vibrações é recomendado realizar investigação da razão do aumento da vibração. Por fim, a Zona D é a mais perigosa para a máquina; segundo a norma vigente, esses valores de vibrações vão certamente danificar o AEG e seus componentes.

Tabela 1 – Limites das zonas de vibrações conforme a ISO 20816-21.

Componente do AEG	Medição em aceleração (m/s ²)		Medição em velocidade (mm/s)	
	Zona B/C	Zona C/D	Zona B/C	Zona C/D
Nacele e torre	0,1 Hz a 10 Hz		0,1 Hz a 10 Hz	
	0,3	0,5	60	100
Rotor	0,1 Hz a 10 Hz		0,1 Hz a 1000 Hz	
	0,3	0,5	2,0	3,2
Gearbox	0,1 Hz a 10 Hz		0,1 Hz a 1000 Hz	
	0,3	0,5	2,0	5,6
	0,1 Hz a 10 Hz			
	7,5	12		
Gerador	10 Hz a 5000 Hz		10 Hz a 1000 Hz	
	10	16	6	10

Fonte: Adaptado da ISO 20816-21.

Deste modo, os AEGs que registraram o maior número de alarmes na zona crítica de vibração foram classificados com falha. Para cada AEG com falha, foram selecionados dois aerogeradores vizinhos saudáveis, desde que não apresentassem um número significativo de alarmes que os classificasse também com falha.

3.4 DESCRIÇÃO DOS MODELOS

No tocante ao quarto passo no diagrama da Figura 1 (desenvolvimento de modelos de detecção de anomalias), foram implementados, doze modelos de aprendizado não supervisionado para detecção de anomalias e consequentes falhas nos aerogeradores: *Angle-base Outlier Detection*, *Clustering-Based Local Outlier*, *Connectivity-Based Local Outlier*, *Isolation Forest*, *Histogram-based Outlier Detection*, *K-Nearest Neighbors Detector*, *Local Outlier Factor*, *One-class SVM detector*, *Principal Component Analysis*, *Minimum Covariance Determinant*, and *Stochastic Outlier Selection*.

Os modelos de detecção de anomalias são algoritmos utilizados para identificar padrões no conjunto de dados que não correspondem ao comportamento normal, ou seja, para detectar pontos de dados que são considerados como *outliers* (Agrawal; Agrawal, 2015). Nesse sentido, os modelos foram implementados utilizando a biblioteca PyCaret (Ali, 2020), uma ferramenta *low-code* em Python que reúne diversos modelos de aprendizado de máquina para uso em diferentes aplicações, como aprendizado

supervisionado (classificação e regressão), aprendizado não supervisionado (*clustering* e detecção de anomalias) e séries temporais (previsão).

Para a detecção final da anomalia, adotou-se um comitê de modelos baseados em votação. Cada modelo foi treinado separadamente e suas previsões foram combinadas para determinar a decisão final. A estratégia de combinação adotada foi a votação majoritária, na qual um ponto no conjunto de dados foi considerado anômalo somente quando pelo menos metade dos modelos do comitê, ou seja, 6, indicassem anomalias.

4. ANÁLISE E DISCUSSÃO DOS RESULTADOS

Nesta seção, são apresentados os resultados da metodologia proposta para detecção de anomalias em rolamentos de AEGs. Inicialmente, serão detalhados os resultados da seleção dos AEGs categorizados como saudáveis e com falha, obtidos por meio do procedimento de rotulação, juntamente com os atributos extraídos durante o pré-processamento. Em seguida, serão apresentados os resultados da combinação dos modelos de detecção para cada AEG.

Os resultados foram obtidos pela aplicação da metodologia a uma base de dados extraída de três parques eólicos situados no Brasil, todos com AEGs do mesmo fabricante e com capacidade de 2,1 MW.

4.1 PRÉ-PROCESSAMENTO E ROTULAÇÃO

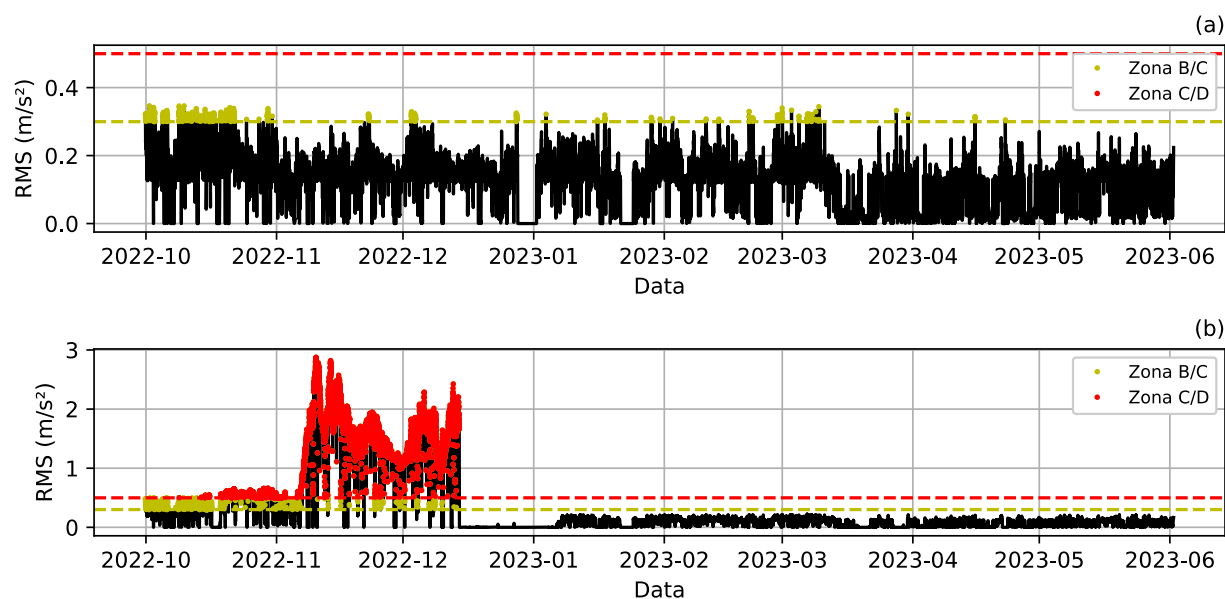
Como mencionado, os passos seguintes após a aquisição dos dados foram o pré-processamento e a rotulação dos AEGs. Para a rotulação dos dados, foram considerados os registros de falhas e os alarmes baseados nos limites de vibração de cada componente, conforme a norma ISO 20816-21.

A Figura 2 apresenta um exemplo da determinação dos alarmes pela aplicação dos limites de vibração para o sensor do rolamento do eixo de alta velocidade do *gearbox*, considerando as Zonas B/C e C/D. Assim, os aerogeradores identificados com falhas foram aqueles que apresentaram o maior número de alarmes na soma das duas zonas. Lembrado que, para cada AEG com falha, foram selecionados dois vizinhos espacialmente mais próximos, desde que estes não apresentassem um número significativo de alarmes. Para a rotulação, devido à disponibilidade dos dados, foram utilizados os dados de tendência entre os períodos de setembro de 2022 a junho de 2023.

A Tabela 2 apresenta os resultados da identificação dos AEGs com falhas, juntamente com suas descrições, e dos AEGs saudáveis. Nesta etapa, foram selecionados 22 aerogeradores dos três parques eólicos, dos quais 8 apresentavam falhas e 14 estavam em condições saudáveis.

Na etapa de pré-processamento, foram extraídos atributos dos sinais de vibração de cada sensor em diferentes domínios para a detecção de anomalia: tempo, frequência, teoria da informação e mecânico. Contudo, os dados brutos de vibração foram disponibilizados apenas em amostras espaçadas entre os períodos de 01-2022 a 12-2023, devido à grande quantidade de informações geradas pelos sensores e o número de AEGs. A aplicação da extração de atributos resultou no conjunto de dados com 9.180 instâncias e 1.081 colunas.

Figura 2 – Aplicação dos limites de vibração para obtenção dos alarmes de falha. (a) AEG saudável; (b) AEG com falha.



Fonte: Autores, 2024.

Tabela 2 – Resultado do procedimento de rotulação.

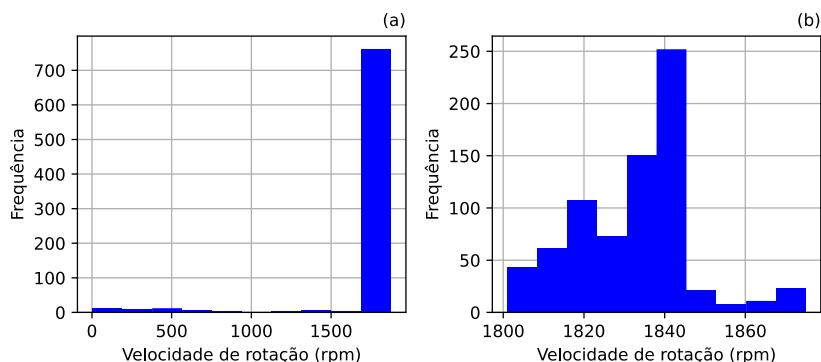
Parque	AEG	Falha/Descrição	Fonte	AEG saudáveis
A	5	Rolamento NDE do gerador		9, 10
	14	Rolamento do planetário		12, 15
B	8	Rolamento LSS do <i>gearbox</i> Rolamento IMS do <i>gearbox</i>	Dados de tendência	9, 10
	22	Rolamento HSS do <i>gearbox</i> Rolamento NDE do gerador		21, 23
C	9	Deslocamento das engrenagens planetárias da <i>gearbox</i>	Log de alarmes	10, 12
	13	Rolamento LSS do <i>gearbox</i> Rolamento IMS do <i>gearbox</i> Rolamento HSS do <i>gearbox</i>	Dados de tendência	2, 4
	14	Danos na pista dos rolamentos HSS da <i>gearbox</i> e trinca encontrada no alojamento do rolamento IMS	Log de alarmes	2, 4
	26	Rolamento HSS do <i>gearbox</i>	Dados de tendência	24, 27

Fonte: Autores, 2024.

Neste estudo, os dados do sensor radial do rolamento do eixo de alta velocidade do *gearbox* foram utilizados para detectar situações anômalas que poderiam indicar falhas, visto que quatro AEGs

listados na Tabela 2 compartilharam o mesmo problema. Os dados originais foram filtrados para selecionar apenas instâncias com velocidades iguais ou superiores a 1.800 rpm, considerando assim apenas o conjunto de dados dos instantes em que os AEGs estavam em funcionamento. A Figura 3 apresenta o histograma da velocidade de rotação para os dois conjuntos de dados.

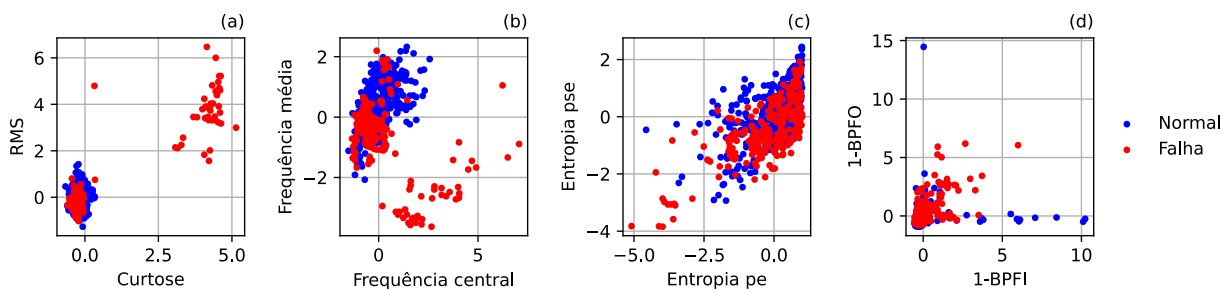
Figura 3 – Histograma da velocidade de rotação em rpm. (a) Conjunto de dados original; (b) Conjunto de dados filtrado.



Fonte: Autores, 2024.

Para o treinamento dos modelos, o conjunto de dados filtrado foi normalizado pelo método Z-score. A aplicação das operações anteriores gerou uma base de dados resultante de 752 instâncias e 206 colunas. A Figura 4 mostra o gráfico de dispersão de alguns atributos extraídos em diferentes domínios, usados na detecção de anomalias.

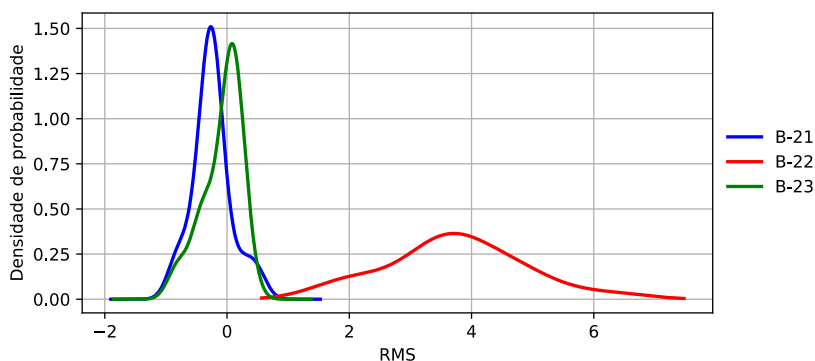
Figura 4 – Gráfico de dispersão. (a) Tempo; (b) Frequência; (c) Teoria da informação; (d) Mecânico.



Fonte: Autores, 2024.

A Figura 5 mostra a comparação da densidade de probabilidade do atributo RMS dos AEGs B-21, B-22 e B-23 pela estimativa de densidade kernel. Observa-se que os AEGs saudáveis, B-21 e B-23, exibem densidades de probabilidade similares e compactas. Em contraste, o AEG B-22 apresenta valores mais dispersos.

Figura 5 – Densidade de probabilidade do atributo RMS.



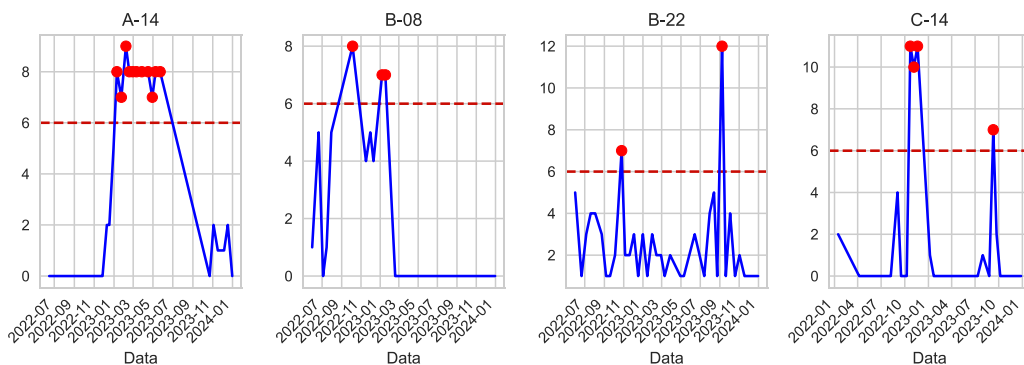
Fonte: Autores, 2024.

4.2 DETECÇÃO DE ANOMALIAS

A seguir é apresentado o estudo de caso da aplicação da metodologia nos dados brutos de vibração obtidos do CMS. Para a operação de detecção de anomalia foram utilizados doze modelos de aprendizagem não supervisionada implementadas pela biblioteca PyCaret, com os hiperparâmetros padrões indicados na biblioteca.

Para aumentar o desempenho na detecção de anomalias e reduzir a incidência de falsos alarmes, adotou-se uma estratégia de combinação das saídas dos modelos por meio de votação majoritária. Assim, o alarme é acionado apenas se mais de seis modelos classificarem o ponto como anômalo. A Figura 6 apresenta os AEGs considerados com falha da Tabela 2 que apresentaram alarmes de anomalias.

Figura 6 – Alarmes de anomalias para os aerogeradores com falha.



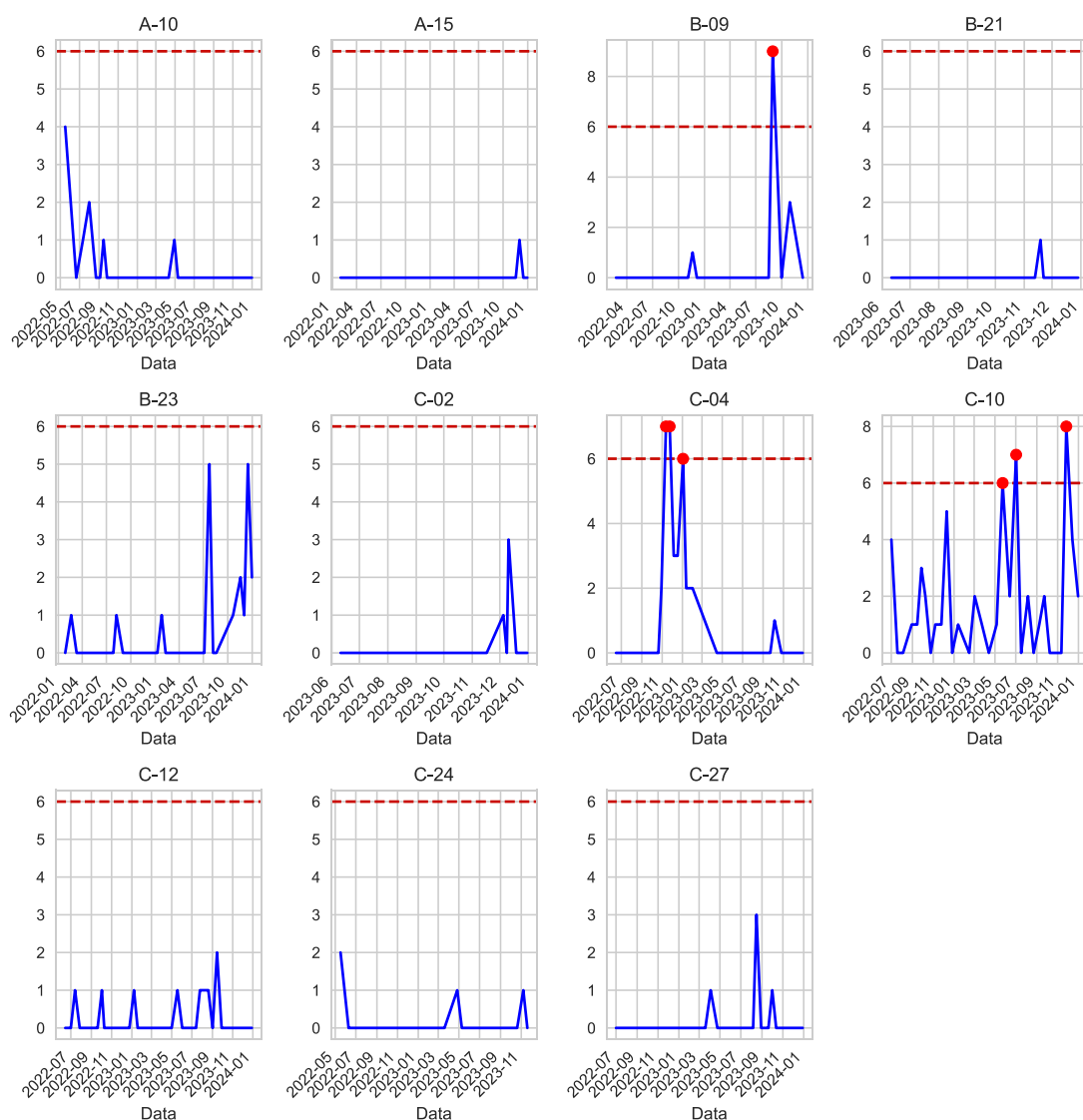
Fonte: Autores, 2024.

Para falhas relacionadas ao rolamento do eixo de alta velocidade do *gearbox*, conforme mostrado na Tabela 2, apenas os AEGs B-22, C-13, C-14 e C-26 apresentaram esse tipo de falha em algum momento. No entanto, pela Figura 6, observa-se que os AEGs A-14, B-8, B-22 e C-14 foram os que tiveram anomalias indicadas por seis ou mais modelos em seus dados. Conforme listado na Tabela 2, nos AEGs A-14 e B-8 foram identificadas falhas relacionadas a outros componentes do *gearbox*. Desse modo, o resultado da Figura 2 levanta a hipótese de haver atravessamento de frequências entre as vibrações dos componentes, resultando na detecção de anomalias pelo sensor considerado. Por outro lado, os AEGs B-22 e C-14 apresentaram alarmes durante o período de rotulação de falha do rolamento HSS do *gearbox*.

Além disso, para os mesmos AEGs foram identificadas anomalias fora dos limites dos dados de tendência analisados, logo não há como confirmar se houve alarmes para esses AEGs ou se estes podem ser considerados falsos positivos.

Para validação da metodologia, a Figura 7 apresenta o resultado da detecção de anomalias para os AEGs considerados saudáveis da Tabela 2.

Figura 7 – Alarmes de anomalias para os aerogeradores saudáveis.



Fonte: Autores, 2024.

De acordo com a Figura 7, somente os AEGs B-9, C-4 e C-10 apresentaram indicações de anomalias. Contudo, os alarmes dos AEGs B-9 e C-10 ocorreram fora do período de análise dos dados de tendência. Para o AEG C-4, no entanto, são necessárias mais investigações sobre a classificação como saudável estabelecida na Tabela 2.

Para os demais AEGs classificados como saudáveis que apresentaram pelo menos uma indicação de anomalia, conforme mostrado na Figura 7, podem ser feitas as seguintes análises:

- i) Devido à parametrização, alguns modelos podem identificar eventos de anomalias para as condições normais de operação, resultando em falsos positivos que não necessariamente indicam problemas reais;
- ii) A variabilidade nas condições ambientais e operacionais pode ter influenciado na indicação de eventos de anomalias;
- iii) Especificamente para os AEGs A-15, B-21, B-23, C-2 e C-27, foram identificadas anomalias após o período de análise dos dados de tendência;

5. CONCLUSÕES

Neste estudo foi aplicado uma abordagem para detecção de anomalias para o rolamento do eixo de alta velocidade do *gearbox* de AEGs, utilizando dados de três parques eólicos em operação no Nordeste do Brasil.

Na aplicação do método, foram utilizados dados de vibração entre os anos de 2022 e 2023 de 22 AEGs, para o treinamento de 12 modelos de detecção de anomalias. O alarme de detecção final foi considerado pela técnica de combinação majoritária. Acerca dos resultados da metodologia aplicada, as principais considerações são:

- i) Os métodos de detecção de anomalias conseguiram reconhecer variações do comportamento padrão nos AEGs B-22 e C-14, os quais estão registrando dados de falhas no rolamento HSS do *gearbox*;
- ii) Foram registradas anomalias em aerogeradores considerados como saudáveis ou com falhas registradas em outros componentes. Essa observação pode estar correlacionada com os hiperparâmetros específicos dos modelos utilizados para análise, que foram aplicados conforme configuração padrão, ou então à possibilidade real de que os AEGs estejam exibindo um comportamento anômalo.

Em síntese, os limites de vibração estabelecidos por norma são indicativos da possibilidade de falha nos componentes dos AEGs. No entanto, os modelos de aprendizado de máquina para detecção de anomalias representam uma alternativa para identificar padrões de falhas, permitindo que estas sejam identificadas antes mesmo de alcançarem níveis severos de vibração.

Dessa forma, os próximos passos em relação à metodologia para trabalhos futuros são:

- i) Expansão da base de dados, agregando mais amostras e ocorrências de falhas dos AEGs, de modo que ajudem a distinguir entre anomalias reais e variações normais no funcionamento dos AEGs;
- ii) Aplicação de técnica de otimização dos hiperparâmetros individuais de cada modelo, para diminuir falsos positivos e aumentar a detecção de casos de anomalias;
- iii) Expandir a abordagem de detecção de anomalias para outros componentes do *drivetrain* dos AEGs;

- iv) Adotar outras abordagens da combinação dos modelos para detecção de anomalias.

6. AGRADECIMENTOS

Os autores reconhecem o financiamento e apoio das empresas do grupo CPFL Energia, sob supervisão da ANEEL - Agência Nacional de Energia Elétrica, número do projeto P&D PD-00063-3090.

REFERÊNCIAS

- AGRAWAL, S.; AGRAWAL, J. Survey on anomaly detection using data mining techniques. **Procedia Computer Science**, v. 60, p. 708-713, 2015.
- ALI, M. PyCaret: An open source, low-code machine learning library in Python. **PyCaret version**, v. 2, 2020.
- BOUKERCHE, A., ZHENG, L.; ALFANDI, O. Outlier detection: Methods, models, and classification. **ACM Computing Surveys (CSUR)**, v. 53, n. 3, p. 1-37, 2020.
- CAESARENDRA, W.; TJAHJOWIDODO, T. A review of feature extraction methods in vibration-based condition monitoring and its application for degradation trend estimation of low-speed slew bearing. **Machines**, v. 5, n. 4, p. 21, 2017.
- CHEN, H. *et al.* Anomaly detection and critical SCADA parameters identification for wind turbines based on LSTM-AE neural network. **Renewable Energy**, v. 172, p. 829-840, 2021.
- FENG, Z.; CHEN, X.; LIANG, M. Iterative generalized synchrosqueezing transform for fault diagnosis of wind turbine planetary gearbox under nonstationary conditions. **Mechanical Systems and Signal Processing**, v. 52, p. 360-375, 2015.
- ISO 20816-1. Mechanical Vibration—Measurement and Evaluation of Machine Vibration—Part 1: General Guidelines; **ISO**: Geneva, Switzerland, 2016.
- LEITE, G. N. P. *et al.* A robust fleet-based anomaly detection framework applied to wind turbine vibration data. **Engineering Applications of Artificial Intelligence**, 2023. v. 126.
- LIN, Z.; LIU, X.; COLLU, M. Wind power prediction based on high-frequency SCADA data along with isolation forest and deep learning neural networks. **International Journal of Electrical Power & Energy Systems**, v. 118, p. 105835, 2020.
- LIU, Z.; ZHANG, L. A review of failure modes, condition monitoring and fault diagnosis methods for large-scale wind turbine bearings. **Measurement**, v. 149, p. 107002, 2020.
- MUNGUBA, C. F. L. *et al.* Ensemble learning framework for fleet-based anomaly detection using wind turbine drivetrain components vibration data. **Engineering Applications of Artificial Intelligence**, 2024. v. 133.
- OLABI, A. G. *et al.* A review on failure modes of wind turbine components. **Energies**, v. 14, n. 17, p. 5241, 2021.
- OMAR, S.; NGADI, A.; JEBUR, H. H. Machine learning techniques for anomaly detection: an overview. **International Journal of Computer Applications**, v. 79, n. 2, 2013.
- SANTELO, T. N.; OLIVEIRA, C. M. R.; MACIEL, C. D.; MONTEIRO, J. R. B. A. Wind turbine failures review and trends. **Journal of Control, Automation and Electrical Systems**, p. 1-17, 2022.
- SINGH, D.; SINGH, B. Investigating the impact of data normalization on classification performance. **Applied Soft Computing**, v. 97, p. 105524, 2020.

TRIZOGLU, P.; LIU, X.; LIN, Z. Fault detection by an ensemble framework of Extreme Gradient Boosting (XGBoost) in the operation of offshore wind turbines. **Renewable Energy**, v. 179, p. 945-962, 2021.

XIANG, L.; YANG, X.; HU, A.; SU, H.; WANG, P. Condition monitoring and anomaly detection of wind turbine based on cascaded and bidirectional deep learning networks. **Applied Energy**, v. 305, p. 117925, 2022.