



Análise Fatorial e Random Forest para Monitoramento de Vibrações em Turbinas Eólicas

Marina Geremia¹ | marina.geremia@ifsc.edu.br
Raiza Silveira Medina² | raiza.sm1991@aluno.ifsc.edu.br
João Victor Rochinski Vieira³ | joao.vrv2000@aluno.ifsc.edu.br
Eduardo Raimundo Rodycz⁴ | eduardo.r2003@aluno.ifsc.edu.br
Gabriel Zapelini Barcellos⁵ | gabriel.zb07@aluno.ifsc.edu.br
Glauco Cardozo⁶ | glauco.cardozo@ifsc.edu.br
Bruno Panerai Velloso⁷ | bruno.velloso@ifsc.edu.br
Sergio Luciano Avila⁸ | sergio.avila@ifsc.edu.br

RESUMO

Este trabalho apresenta uma proposta para o monitoramento de turbinas eólicas por meio da combinação de Análise Fatorial (FA) e Random Forest Regression (RF). O estudo parte do desafio de lidar com grandes volumes de dados de sensores de vibração instalados no drivetrain, sistema responsável pela transmissão de potência da turbina. A metodologia permitiu reduzir redundâncias e selecionar variáveis mais representativas, que foram usadas em modelos de Random Forest capazes de alcançar coeficientes de determinação (R^2) superiores a 0,90 em regiões de maior atividade vibracional. Os resultados indicam o potencial da integração entre estatística e aprendizado de máquina para a construção de modelos mais interpretáveis e robustos, fortalecendo estratégias de manutenção em sistemas de energia renovável.

Palavras-chave: turbinas eólicas; análise fatorial; random forest; vibrações; manutenção.

1 INTRODUÇÃO

As turbinas eólicas são peças centrais na transição energética e na busca por fontes renováveis de baixo impacto ambiental (ZHANG et al., 2024). No entanto, a operação desses equipamentos envolve desafios técnicos importantes. O drivetrain, conjunto formado por eixos, rolamentos e caixas de engrenagens, é uma das partes mais suscetíveis a falhas, estando associado a elevados custos de manutenção e períodos de inatividade (JARDINE; LIN; BANJEVIC, 2006).

Uma das formas mais eficientes de antecipar falhas é a análise de vibração, capaz de detectar irregularidades ainda em estágios iniciais (LEI et al., 2013; WIDODO; YANG, 2007). Nesse contexto, surge a necessidade de métodos que ajudem a selecionar as variáveis mais relevantes em meio ao grande volume de dados coletados por sensores.

Este trabalho propõe a integração de Análise Fatorial (HAIR et al., 2009; FABRIGAR; WEGENER, 2012) e Random Forest Regression (BREIMAN, 2001) para reduzir a complexidade dos dados e, ao mesmo tempo, indicar quais variáveis e sensores mais contribuem para explicar o comportamento vibracional das turbinas.



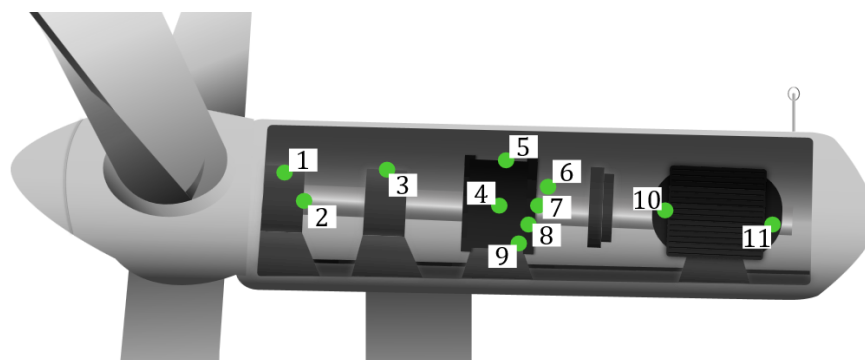
1.1 Metodologia

Os dados foram obtidos de sensores de vibração instalados em diferentes pontos do drivetrain, incluindo eixo de baixa velocidade, eixo intermediário, caixa planetária e gerador. O objetivo foi identificar métricas e posições de sensores com maior poder explicativo.

Na primeira etapa, aplicou-se a Análise Fatorial, técnica utilizada para reduzir multicolinearidade e identificar fatores latentes (HAIR et al., 2009; JOLLIFFE; CADIMA, 2016). A adequação dos dados foi verificada pelos testes de Kaiser-Meyer-Olkin (KMO) e de esfericidade de Bartlett (COSTELLO; OSBORNE, 2005). A quantidade de fatores foi definida pelo critério do autovalor maior que 1 e pela análise do scree plot, seguida da rotação Varimax, que facilita a interpretação das variáveis em cada fator (FABRIGAR; WEGENER, 2012).

Na sequência, as variáveis selecionadas foram usadas como entradas em um modelo de Random Forest Regression, algoritmo reconhecido pela robustez em lidar com relações não lineares (BREIMAN, 2001). O desempenho foi avaliado por meio do coeficiente de determinação (R^2), métrica consolidada para aferir a qualidade de ajuste de modelos (HAIR et al., 2009).

Figura 1 - Posicionamento dos sensores no drivetrain da turbina eólica



Fonte: Elaborado pelos autores (2025)

2. Resultados

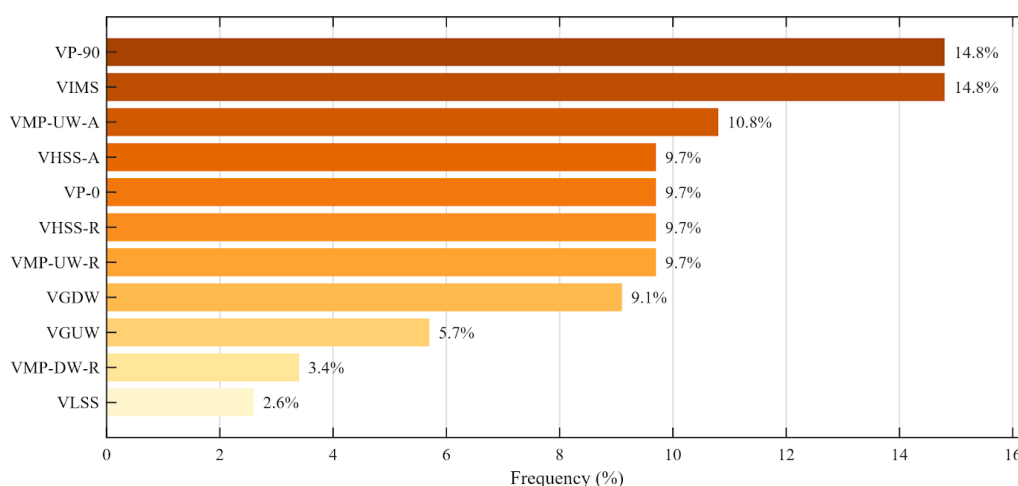
A análise fatorial identificou quatro fatores principais que explicam a maior parte do comportamento vibracional, destacando como variáveis representativas o Defect Factor (DEF), o Crest Factor (FC), o valor de pico (Peak_g) e o RMS, métricas reconhecidas em estudos anteriores pela relevância no diagnóstico de falhas (AMRUTHNATH; GUPTA, 2019; WIDODO; YANG, 2007). Do ponto de vista da localização, sensores instalados no eixo intermediário, na caixa planetária e no gerador



surgiram como os mais relevantes, confirmando a importância dessas regiões no acoplamento dinâmico do drivetrain (JARDINE; LIN; BANJEVIC, 2006).

O Random Forest apresentou valores de R^2 superiores a 0,90 em sensores de áreas com maior atividade vibracional, como eixos e engrenagens. Nos rolamentos principais, porém, situados em regiões de menor sensibilidade modal, o desempenho foi inferior, com dificuldade de generalização em alguns casos. Esse contraste reforça que previsões robustas exigem considerar informações de sensores distribuídos por toda a estrutura, e não apenas aqueles próximos ao ponto de interesse (BREIMAN, 2001; LEI et al., 2013).

Figura 2 – Frequência de contribuição dos sensores para a variável-alvo em diferentes turbinas



Fonte: Elaborado pelos autores (2025)

3. Conclusões

O estudo demonstrou que a integração entre Análise Fatorial e Random Forest é eficaz para lidar com grandes volumes de dados de sensores em turbinas eólicas, permitindo selecionar variáveis-chave, reduzir redundâncias e ampliar a interpretabilidade dos modelos (AMRUTHNATH; GUPTA, 2019; HAIR et al., 2009). As variáveis derivadas mais recorrentes e os sensores localizados em regiões de maior acoplamento estrutural mostraram-se decisivos para o monitoramento, o que reforça a visão do drivetrain como um sistema dinâmico integrado em que falhas locais podem gerar efeitos globais (JARDINE; LIN; BANJEVIC, 2006; WIDODO; YANG, 2007).

Apesar dos avanços, o trabalho apresenta limitações: utilizou apenas variáveis derivadas e foi baseado em turbinas de uma mesma planta. Como perspectivas futuras, recomenda-se explorar sinais brutos de vibração, validar os resultados em modelos

diferentes e investigar a aplicação de métodos mais avançados de inteligência artificial, como redes neurais (ZHANG et al., 2024).

REFERÊNCIAS

AMRUTHNATH, N.; GUPTA, T. Factor analysis in fault diagnostics using random forest. *Industrial Engineering & Management*, v. 8, n. 2, p. 1–7, 2019.

BREIMAN, L. Random forests. *Machine Learning*, v. 45, n. 1, p. 5–32, 2001.

COSTELLO, A. B.; OSBORNE, J. W. Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis. *Practical Assessment, Research, and Evaluation*, v. 10, n. 7, p. 1–9, 2005.

FABRIGAR, L. R.; WEGENER, D. T. *Exploratory factor analysis*. New York: Oxford University Press, 2012.

HAIR, J. F.; BLACK, W. C.; BABIN, B. J.; ANDERSON, R. E. *Análise multivariada de dados*. 6. ed. Porto Alegre: Bookman, 2009.

JOLLIFFE, I. T.; CADIMA, J. Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A*, v. 374, n. 2065, 2016.

JARDINE, A. K. S.; LIN, D.; BANJEVIC, D. A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing*, v. 20, n. 7, p. 1483–1510, 2006.

LEI, Y.; LIN, J.; HE, Z.; ZUO, M. J. A review on empirical mode decomposition in fault diagnosis of rotating machinery. *Mechanical Systems and Signal Processing*, v. 35, n. 1–2, p. 108–126, 2013. doi:10.1016/j.ymssp.2012.09.015.

TSE, P. W.; ZHANG, J. Y.; WANG, X. J. Blind source separation and blind equalization algorithms for mechanical signal separation and identification. *Journal of Vibration and Control*, v. 12, n. 4, p. 395–423, 2006. doi:10.1177/1077546306063248.

WIDODO, A.; YANG, B. S. Support vector machine in machine condition monitoring and fault diagnosis. *Mechanical Systems and Signal Processing*, v. 21, n. 6, p. 2560–2574, 2007.