



UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO TECNOLÓGICO
DEPARTAMENTO DE ENGENHARIA DE PRODUÇÃO E SISTEMAS
CURSO DE GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO MECÂNICA

Pedro Oliveira Onório

**Previsão da vida útil remanescente de máquinas industriais utilizando
algoritmos de aprendizado de máquina e inteligência artificial explicável**

Florianópolis

2026

Pedro Oliveira Onório

**Previsão da vida útil remanescente de máquinas industriais utilizando
algoritmos de aprendizado de máquina e inteligência artificial explicável**

Trabalho de Conclusão de Curso submetido ao curso de Engenharia de Produção Mecânica do Centro Tecnológico da Universidade Federal de Santa Catarina como requisito parcial para a obtenção do título de Bacharel em Engenharia Mecânica com habilitação em Engenharia de Produção.

Orientador: Prof. Enzo Morosini Frazzon

Florianópolis

2026

Onorio, Pedro Oliveira

Previsão da vida útil remanescente de máquinas industriais utilizando algoritmos de aprendizado de máquina e inteligência artificial explicável / Pedro Oliveira Onorio ; orientador, Enzo Morosini Frazzon, 2026.
87 p.

Trabalho de Conclusão de Curso (graduação) - Universidade Federal de Santa Catarina, Centro Tecnológico, Graduação em Engenharia de Produção Mecânica, Florianópolis, 2026.

Inclui referências.

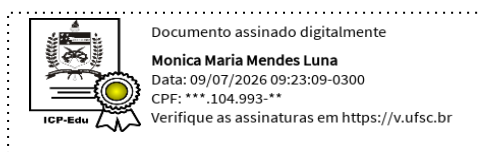
1. Engenharia de Produção Mecânica. 2. Manutenção Preditiva. 3. Vida Útil Remanescente. 4. Aprendizado de Máquina. 5. Interpretabilidade. I. Frazzon, Enzo Morosini. II. Universidade Federal de Santa Catarina. Graduação em Engenharia de Produção Mecânica. III. Título.

Pedro Oliveira Onório

Previsão da vida útil remanescente de máquinas industriais utilizando algoritmos de aprendizado de máquina e inteligência artificial explicável

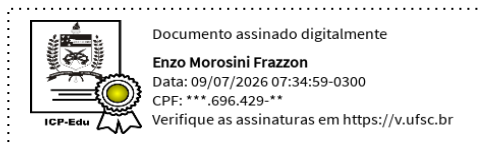
Este Trabalho de Conclusão de Curso foi julgado adequado para obtenção do título de Bacharel em Engenharia de Produção Mecânica e aprovado em sua forma final pelo Curso Engenharia de Produção Mecânica.

Florianópolis, 03 de julho de 2026.



Coordenação do Curso

Banca examinadora



Prof. Enzo Morosini Frazzon, Dr.

Orientador

Prof.(a) Analúcia Schiaffino Morales, Dr.(a)
Universidade Federal de Santa Catarina

Prof. Lynceo Falavigna Braghirolli, Dr.
Universidade Federal de Santa Catarina

Florianópolis, 2026

AGRADECIMENTOS

Agradeço, primeiramente, aos meus pais e irmão, por todo o apoio, incentivo e confiança ao longo da minha trajetória acadêmica. Este trabalho é resultado não apenas do esforço dedicado durante a graduação, mas também de todos os valores, ensinamentos e oportunidades que me proporcionaram ao longo da vida.

Agradeço também às minhas avós e aos meus tios, que sempre estiveram presentes, oferecendo apoio, incentivo e palavras de encorajamento nos momentos em que mais precisei. O carinho e a confiança de vocês foram fundamentais durante esta caminhada.

À minha namorada, expresso minha gratidão pelo companheirismo, pela compreensão e pelo apoio constante ao longo desta etapa. Obrigado por estar ao meu lado nos momentos de desafio, celebrando as conquistas e tornando o percurso mais leve e significativo.

Ao meu amigo e companheiro de trabalho Lúcio, agradeço pela amizade, pelas conversas, pelas trocas de conhecimento e pelo apoio ao longo dos últimos anos. Sua presença tornou a jornada profissional e acadêmica muito mais enriquecedora.

Ao meu orientador, Professor Enzo, agradeço pela orientação, pela confiança e pelos ensinamentos compartilhados ao longo desta trajetória. Sou igualmente grato pela oportunidade de integrar o ProLogIS, ambiente que teve papel fundamental na minha formação acadêmica e profissional.

Agradeço especialmente aos colegas do laboratório, que estiveram presentes durante grande parte da minha graduação. As discussões, os projetos desenvolvidos em conjunto, as trocas de conhecimento e os momentos compartilhados ao longo desses anos contribuíram significativamente para meu crescimento.

Por fim, agradeço a todos os professores, colegas e demais pessoas que, de alguma forma, contribuíram para minha formação e para a realização deste trabalho. Cada apoio recebido ao longo dessa trajetória teve papel importante para que este momento se tornasse possível.

RESUMO

A crescente adoção de estratégias de manutenção preditiva na Indústria 4.0 tem impulsionado o uso de modelos de aprendizado de máquina para estimar a Vida Útil Remanescente (*Remaining Useful Life* - RUL) de equipamentos industriais, com o objetivo de reduzir falhas não planejadas e otimizar a tomada de decisão. Nesse contexto, este trabalho tem como objetivo avaliar a eficácia de modelos de aprendizado de máquina para previsão da RUL a partir de dados temporais de sensores, integrando técnicas de interpretabilidade para análise das previsões. A metodologia adotada baseia-se em uma abordagem experimental aplicada ao conjunto de dados público “PHM *Data Challenge* 2018”, envolvendo etapas de análise exploratória, pré-processamento dos dados, desenvolvimento de modelos e análise de resultados. Foi proposta uma arquitetura em dois estágios, composta por um classificador baseado em XGBoost para detecção do início da degradação e por modelos recorrentes (LSTM e GRU) para estimativa da RUL. Os resultados demonstram que o modelo LSTM apresentou desempenho competitivo em termos de erro absoluto. No entanto, a análise revelou limitações nos extremos do horizonte de RUL, associadas à distribuição dos dados de treinamento. A aplicação do método SHAP permitiu interpretar as previsões, evidenciando que o modelo baseia suas estimativas principalmente em variáveis relacionadas ao tempo de operação, em detrimento de sensores diretamente associados ao mecanismo físico da falha. Esses achados indicam que, na ausência de sinais claros de degradação progressiva nos dados, o modelo tende a explorar correlações mais estáveis, ainda que menos representativas do fenômeno físico. Conclui-se que a integração entre modelagem preditiva e técnicas de interpretabilidade contribui não apenas para a estimativa da RUL, mas também para a compreensão do comportamento dos modelos e suas limitações, oferecendo subsídios relevantes para aplicações em manutenção preditiva.

Palavras-chave: Manutenção Preditiva; Vida Útil Remanescente; RUL; Aprendizado de Máquina; Interpretabilidade.

ABSTRACT

The increasing adoption of predictive maintenance strategies within the Industry 4.0 paradigm has driven the use of machine learning models to estimate the Remaining Useful Life (RUL) of industrial equipment, aiming to reduce unplanned failures and improve decision-making processes. In this context, this study aims to evaluate the effectiveness of machine learning models for RUL prediction based on temporal sensor data, integrating interpretability techniques to analyze model predictions. The methodology follows an experimental approach applied to the public PHM Data Challenge 2018 dataset, encompassing exploratory data analysis, data preprocessing, model development, and results evaluation. A two-stage architecture is proposed, consisting of an XGBoost-based classifier to detect the onset of degradation and recurrent neural networks (LSTM and GRU) to estimate RUL. The results show that the LSTM model achieved competitive performance in terms of absolute error. However, the analysis revealed performance limitations at both ends of the RUL horizon, mainly associated with the distribution of training data. The application of the SHAP method enabled the interpretation of model predictions, indicating that the model relies predominantly on time-related variables rather than sensors directly linked to the physical failure mechanism. These findings suggest that, in the absence of clear progressive degradation signals, the model tends to exploit more stable correlations, even if they are less representative of the underlying physical process. It is concluded that integrating predictive modeling with interpretability techniques not only supports RUL estimation but also enhances the understanding of model behavior and limitations, providing valuable insights for predictive maintenance applications.

Keywords: Predictive Maintenance; Remaining Useful Life; RUL; Machine Learning; Interpretability.

LISTA DE FIGURAS

Figura 1 - Etapas de desenvolvimento.....	38
Figura 2 - Processos de coleta e preparação dos dados.....	39
Figura 3 - Processos de desenvolvimento e treinamento dos modelos	40
Figura 4 - <i>Pipeline</i> de previsão da RUL.....	42
Figura 5 - Equipamento e processo de ataque por feixe de íons.....	48
Figura 6 - Limiar normal e de degradação.....	53
Figura 7 - TTF_F1 ao longo da sessão	54
Figura 8 - Sensor FLOWCOOLPRESSURE em função do TTF	55
Figura 9 - Top 25 atributos por ganho no XGBoost.....	58
Figura 10 - Resultados do teste do XGBoost.....	59
Figura 11 - Arquitetura dos modelos.....	63
Figura 12 - Previsão aleatória (LSTM)	67
Figura 13 - Previsão aleatória (GRU).....	67
Figura 14 - Gráfico de enxame de abelhas	70
Figura 15 - Perfil de atenção temporal.....	72
Figura 16 - Importância SHAP faixa distante	72
Figura 17 - Importância SHAP faixa média.....	73
Figura 18 - Importância SHAP faixa próxima.....	73

LISTA DE TABELAS

Tabela 1 - Trabalhos correlatos.....	34
Tabela 2 - Variáveis de cada sensor.....	49
Tabela 3 - Estatísticas descritivas	51
Tabela 4 - Definição de limiares na literatura.....	52
Tabela 5 - Distribuição de registros por zona	53
Tabela 6 - Análise de correlação entre os sensores.....	56
Tabela 7 - Resultado das etapas de filtragem	61
Tabela 8 - Comparação de desempenho com a literatura	64
Tabela 9 - Resultados por faixa de RUL	65
Tabela 10 - Desempenho por tipo de intervalo	67
Tabela 11 - Contribuição dos sensores.....	69

SUMÁRIO

1	INTRODUÇÃO	12
1.1	PROBLEMA DE PESQUISA.....	14
1.2	JUSTIFICATIVA	14
1.3	OBJETIVOS.....	15
1.3.1	Objetivo geral	15
1.3.2	Objetivos específicos	15
1.4	DELIMITAÇÃO DA PESQUISA	16
1.5	ESTRUTURA DO TRABALHO	17
2	FUNDAMENTAÇÃO TEÓRICA.....	18
2.1	MANUTENÇÃO PREDITIVA (PDM).....	18
2.2	PREVISÃO DA VIDA ÚTIL REMANESCENTE (RUL).....	20
2.3	APRENDIZADO DE MÁQUINA APLICADO À RUL	22
2.3.1	Fundamentos de Aprendizado de Máquina.....	22
2.3.2	Paradigmas de Aprendizado de Máquina.....	23
2.3.3	Previsão de RUL baseada em dados	24
2.3.4	Abordagens baseadas em <i>Deep Learning</i>	25
2.3.5	Abordagens baseadas em Árvores e métodos de <i>ensemble</i>	27
2.3.6	Long Short-Term Memory (LSTM).....	29
2.3.7	eXtreme Gradient Boosting (XGBoost)	30
2.4	INTERPRETABILIDADE EM APRENDIZADO DE MÁQUINA (EXPLAINABLE AI - XAI).....	31
2.5	TRABALHOS CORRELATOS.....	33
3	ABORDAGEM METODOLÓGICA.....	37
3.1	ENQUADRAMENTO METODOLÓGICO	37
3.2	PROCEDIMENTOS METODOLÓGICOS	37
3.2.1	Coleta e preparação dos dados	38
3.2.2	Desenvolvimento e treinamento dos modelos de ML	39
3.2.3	Previsão e análise da RUL	42
3.2.4	Aplicação do SHAP	44
3.2.5	Ferramentas utilizadas	45
4	RESULTADOS E DISCUSSÕES.....	47
4.1	COLETA E PREPARAÇÃO DOS DADOS	47

4.1.1	Banco de dados “PHM Data Challenge 2018”	47
4.1.2	Análise exploratória e preparação dos dados	51
4.1.2.1	<i>Visão Geral da Sessão.....</i>	51
4.1.2.2	<i>Estrutura Temporal.....</i>	52
4.1.2.3	<i>Modo Operacional e Correlação entre Sensores.....</i>	54
4.2	DESENVOLVIMENTO E TREINAMENTO DOS MODELOS DE ML	57
4.2.1	Classificador XGBoost	57
4.2.1.1	<i>Engenharia de Atributos (Feature Engineering)</i>	57
4.2.1.2	<i>Estratégia de Divisão, Limiares e Resultados</i>	59
4.2.2	Preparação das Sequências para o LSTM	60
4.2.2.1	<i>Processo de Filtragem</i>	60
4.2.2.2	<i>Limite CLF e definição da variável-alvo de RUL</i>	62
4.2.2.3	<i>Arquitetura e Configuração de Treinamento</i>	62
4.3	PREVISÃO E ANÁLISE DA RUL	63
4.3.1	Resultados e Análise	64
4.4	ANÁLISE SHAP.....	68
4.4.1	Resultados da Análise SHAP	68
5	CONCLUSÃO E TRABALHOS FUTUROS	75
	REFERÊNCIAS.....	77

1 INTRODUÇÃO

A indústria moderna, imersa no paradigma da Indústria 4.0, enfrenta o desafio constante de maximizar a disponibilidade de seus ativos, melhorar o desempenho e a eficiência (Achouch *et al.*, 2022) e reduzir custos operacionais não planejados (Alaoui, Meddaoui e Hain, 2025). Falhas inesperadas em máquinas industriais representam um dos maiores gargalos da manufatura, resultando em paradas de linha, custos elevados com manutenção corretiva, podendo levar o ecossistema da cadeia de suprimentos a mau funcionamento (Mallioris, Aivazidou e Bechtsis, 2024) e potenciais riscos à segurança (Zonta *et al.*, 2020). Neste cenário, as estratégias de manutenção evoluíram significativamente, transitando de abordagens reativas (corretivas) e preventivas (baseadas em tempo fixo) para metodologias proativas e inteligentes (Daoudi, Smail e Aboussaleh, 2023).

A Manutenção Preditiva (*Predictive maintenance* - PdM), impulsionada pela vasta coleta de dados via sensores da Internet das Coisas (*Internet of Things* - IoT) e pela capacidade analítica do *Big Data*, emergiu como uma estratégia central da Indústria 4.0 (Alaoui, Meddaoui e Hain, 2025). O objetivo da PdM é antecipar falhas antes que elas ocorram, permitindo que as intervenções sejam agendadas no momento ideal, minimizando o impacto na produção, diminuindo o tempo de inatividade, reduzindo custos operacionais e aumentando a produtividade (Cummins *et al.*, 2024; Mallioris, Aivazidou e Bechtsis, 2024). Para alcançar essa antecipação, modelos de Inteligência Artificial (IA), principalmente Aprendizado de Máquina (*Machine Learning* - ML) e Aprendizado Profundo (*Deep Learning* - DL) têm sido amplamente aplicados para analisar dados temporais complexos de sensores (vibração, temperatura, acústica) e identificar padrões de degradação (Zonta *et al.*, 2020).

Dentro do escopo da PdM, a previsão da Vida Útil Remanescente (*Remaining Useful Life* - RUL) é uma das tarefas mais críticas e desafiadoras (Kobayashi e Alam, 2024). Modelos de ML que preveem a RUL com acurácia resultam em tempo de inatividade reduzido, custos menores e melhoria na qualidade da produção (Syed *et al.*, 2025). Contudo, modelos de ML são frequentemente tratados como "caixas-pretas" (*black-boxes*), uma vez que suas previsões são baseadas na identificação de padrões implícitos nos dados, dificultando a compreensão direta dos fatores que

levam a uma determinada estimativa (Kobayashi e Alam, 2024; Mansourvar, Jahangoshai Rezaee e Eshkevari, 2025).

Na prática industrial, essa limitação representa uma barreira relevante para a adoção desses modelos (Cummins *et al.*, 2024). A tomada de decisões de alto impacto, como a interrupção de uma linha de produção, exige não apenas previsões acuradas, mas também justificativas compreensíveis. Quando uma estimativa de RUL baixa não é acompanhada por uma explicação clara das variáveis envolvidas, os gestores de manutenção tendem a hesitar em utilizá-la como base para ação (Alaoui, Meddaoui e Hain, 2025).

A literatura recente demonstra um foco intenso no aumento da acurácia dos modelos preditivos de RUL, frequentemente priorizando desempenho preditivo em detrimento da interpretabilidade (Alaoui, Meddaoui e Hain, 2025). Esse cenário reforça o desafio prático associado à opacidade dos modelos, uma vez que a adoção de sistemas de PdM depende fundamentalmente da confiança do usuário final (Kobayashi e Alam, 2024). A mera previsão de uma RUL é insuficiente para a tomada de decisão se a causa raiz da degradação não for compreendida (Cummins *et al.*, 2024). Existe, portanto, uma demanda contínua por abordagens que permitam correlacionar os resultados da previsão com os vetores-chave de degradação (Zhang e Wang, 2025), bem como melhorar a confiabilidade e a comunicação das estimativas de incerteza (Lefrouni e Taibi, 2025).

Nesse cenário, métodos de Inteligência Artificial Explicável (*Explainable Artificial Intelligence – XAI*), como o SHAP (*SHapley Additive exPlanations*), surgem como ferramentas fundamentais para a interpretação dos resultados. Esses métodos permitem decompor a previsão do modelo em contribuições individuais das variáveis de entrada, fornecendo explicações locais para cada instância analisada. Dessa forma, é possível identificar quais sensores ou variáveis exerceram maior influência na estimativa da RUL, permitindo validar o comportamento do modelo com base no conhecimento físico do processo (Baptista *et al.*, 2024; Cummins *et al.*, 2024).

Além de aumentar a transparência, a utilização de técnicas como o SHAP possibilita a análise crítica do modelo, permitindo identificar padrões inesperados, inconsistências ou dependências indevidas nos dados. Esse processo contribui não apenas para o aumento da confiança dos usuários, mas também para o aprimoramento contínuo dos modelos, tornando-os mais aderentes à realidade

operacional e mais robustos para aplicação em ambientes industriais (Cummins *et al.*, 2024).

1.1 PROBLEMA DE PESQUISA

Diante da complexidade dos modelos de PdM e da necessidade de transparência para a tomada de decisão gerencial, identifica-se um desafio de pesquisa na aplicação sistemática de métodos de XAI que sejam robustos e acionáveis. Logo, este trabalho busca responder à seguinte questão de pesquisa:

Como modelos de ML podem ser aplicados para prever a RUL de máquinas industriais a partir de dados temporais de sensores, e de que forma técnicas de interpretabilidade, como SHAP, podem contribuir para identificar as variáveis mais relevantes e apoiar decisões de manutenção preditiva?

1.2 JUSTIFICATIVA

A realização deste trabalho justifica-se por três pilares principais: sua relevância econômica e operacional, sua contribuição técnica e acadêmica, e sua aplicação direta para a Engenharia de Produção.

Do ponto de vista econômico e operacional, as falhas em máquinas na indústria moderna levam à perda de produção e à redução da competitividade. Os custos de manutenção representam uma parcela significativa dos custos de fabricação, variando entre 15% e 60% do custo final do produto, podendo atingir até 50% do custo total de produção na indústria pesada (Shukla, Nefti-Meziani e Davis, 2022). Neste contexto, a implementação bem-sucedida de uma PdM baseada em IA acarreta implicações profundas e multidimensionais, conforme apontado por revisões da literatura. Tais implicações englobam benefícios operacionais (como aumento da disponibilidade e eficiência global do equipamento), econômicos (redução de custos e perdas de produção) e de segurança (mitigação proativa de falhas) (Lefrouni e Taibi, 2025).

Sob a perspectiva técnica e acadêmica, a literatura sobre PdM demonstra um esforço contínuo na busca por maior acurácia preditiva, frequentemente empregando modelos de DL (P e A, 2024). No entanto, como apontado na contextualização (Seção **Erro! Fonte de referência não encontrada.**), muitos desses modelos operam como

"caixas-pretas", o que gera desconfiança e dificulta sua adoção prática (Cummins *et al.*, 2024). A pesquisa acadêmica, portanto, apresenta uma demanda contínua por trabalhos que equilibrem acurácia preditiva com interpretabilidade, um desafio central da área de XAI (Ferraro *et al.*, 2023).

Este TCC se justifica ao contribuir para o corpo de conhecimento sobre interpretabilidade na PdM, por meio da integração de um modelo de ML de alto desempenho para previsão da RUL com uma análise do método SHAP. O método não apenas elucida por que uma previsão foi feita, mas também demonstra a contribuição de cada atributo (*feature*) para o resultado da previsão (Zhang e Wang, 2025). Isso permite validar se o modelo está aprendendo padrões físicos reais ou se baseando em correlações enganosas.

Para a Engenharia de Produção, a justificativa reside na transformação de um modelo de ML de uma "caixa-preta" em uma ferramenta de apoio à decisão transparente. Ao identificar quais sensores (e, por consequência, quais componentes físicos) mais contribuem para o desgaste, o engenheiro de produção pode tomar decisões mais assertivas, indo além do "quando" manter (prognóstico) e começando a investigar o "porquê" da falha (diagnóstico), otimizando não só a manutenção, mas potencialmente o próprio processo produtivo.

1.3 OBJETIVOS

1.3.1 Objetivo geral

O objetivo geral deste trabalho é avaliar a eficácia da utilização de modelos de ML para previsão da RUL de máquinas industriais com base em dados temporais, integrando métodos de interpretabilidade para apoio a tomadas de decisão no contexto da manutenção preditiva.

1.3.2 Objetivos específicos

Para atender ao objetivo geral, os seguintes objetivos específicos são traçados:

- Preparar um banco de dados público contendo medições temporais de sensores de máquinas industriais.
- Desenvolver modelos preditivos de RUL utilizando técnicas de ML supervisionado.
- Interpretar as previsões, identificando as variáveis mais relevantes para o processo de degradação.
- Avaliar o desempenho preditivo e a capacidade explicativa dos modelos desenvolvidos, comparando os resultados obtidos com abordagens reportadas na literatura e discutindo suas implicações para manutenção preditiva.

1.4 DELIMITAÇÃO DA PESQUISA

Este trabalho delimita-se à aplicação de modelos de ML supervisionado para a tarefa específica de prognóstico (previsão da RUL), utilizando uma abordagem orientada a dados (*data-driven*). Portanto, não fazem parte do escopo o desenvolvimento de modelos baseados em física (*physics-based*) ou modelos híbridos.

Quanto ao objeto de estudo, a pesquisa restringe-se à análise do banco de dados público “PHM *Data Challenge* 2018”. Consequentemente, os resultados e as variáveis analisadas estão condicionados às características específicas deste conjunto de dados, que contém dados operacionais de equipamentos de ataque por feixe de íons utilizados na indústria de semicondutores. Não serão realizadas coletas de dados primários em ambiente fabril físico (“chão de fábrica”).

No que tange à interpretabilidade, o estudo foca na aplicação do método SHAP como técnica *post-hoc* e agnóstica ao modelo. Outras técnicas de XAI (como LIME ou Counterfactuals) não serão objeto de análise comparativa aprofundada, sendo o SHAP escolhido por sua robustez teórica e consistência matemática.

Por fim, a validação dos modelos propostos é estritamente computacional (via métricas de desempenho e gráficos de interpretabilidade). A implementação do modelo em um sistema de tempo real, a integração com CLPs (Controladores Lógicos Programáveis) ou a análise financeira do retorno sobre o investimento (ROI) da solução proposta não estão contempladas neste trabalho.

Assim, as conclusões deste trabalho devem ser interpretadas no contexto do banco “PHM *Data Challenge* 2018” e do recorte experimental adotado, não como

generalização direta para todos os cenários de previsão de RUL em ambientes industriais.

1.5 ESTRUTURA DO TRABALHO

Este Trabalho de Conclusão de Curso está estruturado em cinco capítulos, visando uma apresentação clara e lógica da pesquisa.

O Capítulo 1, esta Introdução, contextualiza o tema da PdM na Indústria 4.0, apresenta o problema de pesquisa focado na interpretabilidade dos modelos da RUL, justifica a relevância do estudo e define os objetivos da pesquisa.

O Capítulo 2 apresenta a Fundamentação Teórica. Este capítulo contempla uma fundamentação teórica sobre os conceitos-chave necessários para o desenvolvimento do trabalho, abordando PdM, a estimação da RUL, os fundamentos de ML e, com especial detalhe, os métodos de Interpretabilidade de Modelos, focando no SHAP.

O Capítulo 3 descreve os Procedimentos Metodológicos. Esta seção detalha as etapas de pré-processamento de dados e a metodologia para o desenvolvimento dos modelos preditivos de RUL. Além disso, explica como o método de interpretabilidade será aplicado para analisar as previsões.

O Capítulo 4 apresenta e discute os Resultados. O capítulo é dividido para abordar tanto o desempenho preditivo dos modelos desenvolvidos quanto a análise da capacidade explicativa, identificando as variáveis mais relevantes e discutindo as implicações práticas dos achados para a PdM.

Finalmente, o Capítulo 5 traz a Conclusão, que sintetiza as principais descobertas, verifica o atendimento aos objetivos propostos (geral e específicos), aponta as limitações encontradas durante o estudo e sugere direções para trabalhos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

Nesta seção, são apresentados os conceitos essenciais para o desenvolvimento da pesquisa, fornecendo o embasamento sobre a evolução das estratégias de manutenção, a importância do prognóstico da RUL, a aplicação de ML e a necessidade de interpretabilidade dos modelos.

A discussão dos métodos de ML busca contextualizar a seleção das arquiteturas utilizadas neste estudo. A escolha das abordagens analisadas considera sua aderência às características do problema de previsão da RUL, à natureza temporal dos dados e à sua recorrência na literatura aplicada à manutenção preditiva.

2.1 MANUTENÇÃO PREDITIVA (PDM)

A gestão da manutenção é um pilar para a segurança e eficiência industrial. Estratégias de manutenção ineficazes são apontadas como uma das principais causas de acidentes, sendo que falhas técnicas e de equipamento associadas à manutenção respondem por uma parcela significativa desses incidentes (Aktas e Kagnicioglu, 2023). Além dos riscos à segurança, as falhas de manutenção impactam diretamente a viabilidade operacional e financeira, resultando em alto tempo de inatividade (*downtime*), atrasos, aumento nos custos de fabricação e, consequentemente, perda de receita (Gupta, Kumar e Maiti, 2024; Lefrouni e Taibi, 2025).

Historicamente, as estratégias de manutenção evoluíram. Elas são amplamente classificadas em reativas (não planejadas), que ocorrem após a falha, e proativas (planejadas) (Patil *et al.*, 2022). A manutenção reativa refere-se à estratégia onde as atividades são executadas somente após a ocorrência da falha. A estratégia proativa, por outro lado, envolve medidas de precaução tomadas antecipadamente, seja para reparar, substituir ou realizar qualquer atividade de manutenção, visando prevenir o sistema de falhas súbitas. A abordagem proativa é, por sua vez, classificada em subcategorias, das quais se destacam a manutenção preventiva e a preditiva (Errandonea, Beltrán e Arrizabalaga, 2020).

As abordagens proativas tradicionais, como a manutenção preventiva (baseada em tempo fixo), muitas vezes falham em superar os desafios de competitividade do mercado (Gupta, Kumar e Maiti, 2024). Isso ocorre porque uma

manutenção realizada de forma muito antecipada pode reduzir a vida útil de peças e impor paradas desnecessárias, enquanto a manutenção tardia pode levar a danos em cascata e riscos de segurança (Ramezani *et al.*, 2023).

Neste contexto, surge a PdM como uma estratégia proativa avançada. A PdM é uma abordagem que utiliza o monitoramento de condições operacionais dos equipamentos, por meio de sensores e sistemas industriais, aliado a algoritmos de ML, com a finalidade de otimizar o desempenho e a longevidade dos ativos (Gupta, Kumar e Maiti, 2024; Lefrouni e Taibi, 2025). O objetivo é usar ferramentas de monitoramento em tempo real para observar o desempenho dos ativos, identificar sinais de deterioração e prever seu índice de saúde (Gupta, Kumar e Maiti, 2024).

A PdM está intrinsecamente ligada ao conceito de Prognóstico e Gestão da Saúde (*Prognostics and Health Management* - PHM), um *framework* computacional que visa estimar o estado de saúde dos sistemas e prever sua evolução, incluindo a estimativa da RUL (Umeda *et al.*, 2021). No paradigma da Indústria 4.0, a PdM revoluciona a manutenção, deixando de ser um centro de custo para se tornar uma oportunidade de negócio (Gupta, Kumar e Maiti, 2024).

A aplicação da PdM torna-se particularmente relevante em ambientes industriais de alta complexidade e sensibilidade, como a manufatura de semicondutores. Esse setor é caracterizado por elevados níveis de automação, alta precisão e forte dependência da confiabilidade dos equipamentos, nos quais falhas podem comprometer significativamente a continuidade da produção e gerar impactos financeiros expressivos (Palsaniya, Patil e Thanu, 2025; Phongmoo *et al.*, 2025).

Por esse motivo, estratégias avançadas de manutenção tornam-se fundamentais, uma vez que abordagens tradicionais não são capazes de atender aos requisitos de disponibilidade e estabilidade desses sistemas. Assim, a PdM tem sido amplamente investigada como uma abordagem capaz de prever falhas e estimar o tempo até a falha, contribuindo para o aumento da disponibilidade dos equipamentos e a otimização das operações (Biebl *et al.*, 2020; Jalali *et al.*, 2019).

Dessa forma, a utilização de dados industriais reais, provenientes de sensores, registros operacionais e históricos de manutenção, torna-se essencial para o desenvolvimento e validação de modelos preditivos. Esses dados refletem a complexidade, variabilidade e natureza temporal dos sistemas industriais, permitindo avaliar o desempenho de modelos de previsão de falhas e de RUL em cenários próximos da aplicação prática (Palsaniya, Patil e Thanu, 2025). Além disso, a

interpretabilidade dos modelos é um fator crítico para sua adoção no ambiente industrial, uma vez que a transparência das previsões influencia diretamente a confiança dos especialistas e a efetividade das decisões de manutenção (Jalali *et al.*, 2019).

Sob essa perspectiva, a capacidade de antecipar o comportamento futuro dos sistemas torna-se um dos principais diferenciais da PdM moderna. No âmbito do PHM, o prognóstico não se limita à detecção de falhas iminentes, mas envolve a estimativa da evolução do processo de degradação ao longo do tempo, permitindo inferir quando um sistema atingirá um estado crítico de operação (Jardine, Lin e Banjevic, 2006; Si *et al.*, 2011). Dentre as diferentes formas de representar essa informação, a estimativa da RUL destaca-se como uma das abordagens mais relevantes e amplamente adotadas, sendo discutida em maior detalhe na seção seguinte (Lei *et al.*, 2018; Pecht, 2008).

2.2 PREVISÃO DA VIDA ÚTIL REMANESCENTE (RUL)

A previsão da RUL constitui um dos elementos centrais das estratégias de PdM, sendo amplamente utilizada como suporte à tomada de decisão em ambientes industriais. A RUL pode ser definida como o intervalo de tempo restante até que um componente ou sistema atinja um estado de falha ou um limite de desempenho previamente estabelecido, fornecendo uma medida quantitativa da condição futura dos ativos. Nesse sentido, o prognóstico desempenha um papel fundamental nas arquiteturas modernas de PdM, sendo responsável por estimar a evolução do estado de saúde dos sistemas ao longo do tempo e apoiar o planejamento de intervenções (Bougacha, Varnier e Zerhouni, 2021; Soualhi *et al.*, 2022; Vogl, Weiss e Helu, 2019).

Os avanços recentes em sensores e ML têm viabilizado a construção de modelos capazes de estimar a RUL a partir de dados coletados continuamente durante a operação dos sistemas (Khan *et al.*, 2022). Sob essa perspectiva, abordagens *data-driven* utilizam técnicas de ML para extrair padrões a partir de dados brutos de sensores, parâmetros operacionais e registros históricos, permitindo não apenas a detecção e o diagnóstico de falhas, mas também a antecipação de sua ocorrência com maior precisão (Darwish, 2024; Gupta, Kumar e Maiti, 2024). Dessa forma, a RUL destaca-se como uma das principais saídas dos modelos de prognóstico, por representar de forma quantitativa o processo de degradação e apoiar

decisões mais informadas em contextos industriais (Apeiranthitis *et al.*, 2026; Huang *et al.*, 2024).

A partir desse cenário, a estimativa da RUL assume um papel central no suporte à tomada de decisão em manutenção, uma vez que permite a definição de estratégias mais eficientes para a gestão de ativos. A partir de previsões acuradas, torna-se possível programar intervenções de manutenção de forma proativa, mitigando falhas em desenvolvimento antes que evoluam para eventos críticos e reduzindo a ocorrência de paradas não planejadas. Além disso, a informação sobre a RUL contribui diretamente para o planejamento logístico, incluindo o gerenciamento de estoques e a garantia da disponibilidade de peças de reposição no momento adequado, evitando interrupções na operação de sistemas industriais (Florian, Sgarbossa e Zennaro, 2021; Goby, Brandt e Neumann, 2023; Irfan *et al.*, 2017; Nguyen e Medjaher, 2019). Assim, a previsão da RUL fornece subsídios essenciais para a redução de incertezas e para a tomada de decisões mais assertivas ao longo do ciclo de vida dos ativos (Darwish, 2024).

Do ponto de vista da modelagem, a estimativa da RUL está intrinsicamente associada à análise de dados temporais, uma vez que o processo de degradação de sistemas ocorre de forma progressiva ao longo do tempo e depende do histórico operacional dos ativos. Dessa forma, a modelagem da RUL requer a capacidade de capturar dependências temporais e padrões dinâmicos presentes em séries de dados sensoriais, frequentemente caracterizadas por alta dimensionalidade, ruído e variabilidade operacional (Lei *et al.*, 2018; Zhang, Yang e Wang, 2019). Além disso, a relação entre as variáveis monitoradas e o processo de degradação é tipicamente não linear e influenciada por múltiplos fatores, o que torna a tarefa de previsão ainda mais desafiadora (Si *et al.*, 2011; Wang, Zhao e Addepalli, 2020).

Nesse cenário, diferentes abordagens de ML têm sido exploradas para a estimativa da RUL, destacando-se, por um lado, modelos capazes de capturar dependências temporais em dados sequenciais, como as Redes Neurais Recorrentes (*Recurrent Neural Networks* - RNNs), e, por outro, métodos baseados em árvores que lidam de forma eficaz com relações não lineares e dados heterogêneos (Dalzochio *et al.*, 2020; Lefrouni e Taibi, 2025; Zhang, Yang e Wang, 2019). Essas abordagens são fundamentadas nos conceitos de ML, discutidos na seção seguinte, e posteriormente exploradas no contexto da previsão de RUL.

2.3 APRENDIZADO DE MÁQUINA APLICADO À RUL

A aplicação de técnicas de ML tem se consolidado como uma das principais abordagens para a previsão da RUL em sistemas industriais. Diferentemente de métodos baseados exclusivamente em modelos físicos, as abordagens *data-driven* permitem extrair padrões complexos diretamente a partir de grandes volumes de dados operacionais, viabilizando o prognóstico mesmo em cenários onde o comportamento de degradação não é completamente conhecido (Darwish, 2024; Gupta, Kumar e Maiti, 2024). Nesta seção, são apresentados os fundamentos do ML, seus principais paradigmas e as estratégias utilizadas na modelagem da RUL, com ênfase em técnicas amplamente empregadas na literatura, incluindo modelos baseados em árvores e métodos de DL.

2.3.1 Fundamentos de Aprendizado de Máquina

O ML pode ser definido como um conjunto de métodos computacionais que permitem a sistemas aprenderem padrões a partir de dados e melhorarem seu desempenho em tarefas específicas com base na experiência, sem a necessidade de programação explícita de regras (Murphy, 2022). Essa mudança de paradigma, de abordagens baseadas em regras para abordagens *data-driven*, representa uma transformação fundamental no desenvolvimento de sistemas computacionais (Kumar, Jairam e Ambedkar, 2023). Em sua essência, o ML utiliza modelos estatísticos e probabilísticos para representar incertezas, identificar padrões em dados e realizar previsões ou apoiar processos de tomada de decisão com base em representações aprendidas (Murphy, 2022; Pandey, 2023).

Os sistemas de ML distinguem-se de abordagens algorítmicas tradicionais por serem intrinsecamente orientados a dados, construindo modelos a partir de exemplos em vez de dependerem de regras codificadas manualmente (Cabrera *et al.*, 2026; Kumar, Jairam e Ambedkar, 2023). Além disso, abordagens modernas enfatizam a modelagem probabilística, permitindo não apenas realizar previsões, mas também quantificar a incerteza associada a essas previsões, o que é essencial em contextos industriais e de tomada de decisão (Murphy, 2022). Na prática, o desenvolvimento de soluções de ML envolve uma metodologia completa que inclui etapas de pré-processamento de dados, engenharia de atributos, seleção de modelos, validação e avaliação, sendo a capacidade de generalização, ou seja, o desempenho em dados

não vistos, um dos pilares centrais dessa área (Alzubi, Nayyar e Kumar, 2018; Mohri, Rostamizadeh e Talwalkar, 2012).

2.3.2 Paradigmas de Aprendizado de Máquina

De forma geral, o ML pode ser categorizado em três paradigmas principais: aprendizado supervisionado, não supervisionado e por reforço, cada um caracterizado por diferentes tipos de dados disponíveis, objetivos de aprendizado e formas de interação com o ambiente. Esses paradigmas representam abordagens distintas para a extração de conhecimento a partir de dados, dando origem a diferentes famílias de algoritmos e aplicações, com níveis variados de complexidade, interpretabilidade e dependência de dados (Alzubi, Nayyar e Kumar, 2018; Gupta *et al.*, 2022; Pandey, 2023).

O aprendizado supervisionado (*supervised learning*) é o paradigma mais amplamente utilizado, sendo caracterizado pela disponibilidade de dados rotulados compostos por pares de entrada e saída (Gupta *et al.*, 2022; Pandey, 2023). Nesse contexto, o objetivo do modelo é aprender uma função de mapeamento que permita prever corretamente os rótulos de novos dados não observados durante o treinamento. Esse processo de aprendizado é guiado pelos próprios rótulos, que funcionam como referência para a correção dos erros do modelo ao longo do treinamento (Bhushan, 2025; Gupta *et al.*, 2022). As tarefas supervisionadas são geralmente classificadas em problemas de classificação, quando os rótulos são categóricos, e regressão, quando os valores a serem previstos são contínuos, sendo este último especialmente relevante para a estimativa da RUL (Bhushan, 2025; Gupta *et al.*, 2022). Diversos algoritmos se enquadram nesse paradigma, incluindo árvores de decisão, máquinas de vetor de suporte, redes neurais e métodos de *ensemble*, os quais diferem em capacidade de representação, custo computacional e robustez a ruído (Pandey, 2023).

Por outro lado, o aprendizado não supervisionado (*unsupervised learning*) trata de cenários em que não há dados rotulados disponíveis, sendo o objetivo principal identificar estruturas, padrões ou relações intrínsecas nos dados (Alzubi, Nayyar e Kumar, 2018; Pandey, 2023). Nesse paradigma, o modelo busca organizar ou representar os dados de maneira significativa sem a orientação explícita de um sinal de saída. Entre as tarefas mais comuns estão o agrupamento (*clustering*), a redução de dimensionalidade e a estimação de densidade, frequentemente utilizadas

para análise exploratória, compressão de dados e extração de características (Pandey, 2023; Swain, 2025). Embora menos diretamente aplicado à previsão de RUL, o aprendizado não supervisionado pode desempenhar um papel relevante em etapas preliminares, como detecção de anomalias, identificação de estados operacionais e construção de representações mais informativas para modelos supervisionados (Bhushan, 2025; Pandey, 2023).

Já o aprendizado por reforço (*Reinforcement Learning*) constitui um paradigma distinto, no qual um agente aprende a tomar decisões sequenciais por meio da interação com um ambiente, recebendo sinais de recompensa ou penalidade em resposta às suas ações (Alzubi, Nayyar e Kumar, 2018; Murphy, 2022). Diferentemente do aprendizado supervisionado, não há rótulos explícitos indicando a ação correta, sendo o aprendizado conduzido por tentativa e erro, com o objetivo de maximizar uma recompensa acumulada ao longo do tempo (Gupta *et al.*, 2022; Mohri, Rostamizadeh e Talwalkar, 2012). Esse paradigma é particularmente adequado para problemas de controle e otimização sequencial, como robótica e sistemas autônomos (Kumar, Jairam e Ambedkar, 2023; Pandey, 2023). No contexto da PdM, embora menos utilizado diretamente na previsão de RUL, o aprendizado por reforço pode ser aplicado em estratégias de decisão relacionadas à otimização de políticas de manutenção e operação de sistemas (Engbers e Freitag, 2025).

De maneira geral, observa-se que, enquanto o aprendizado supervisionado é mais adequado para tarefas preditivas com objetivos bem definidos, como a estimativa da RUL, os paradigmas não supervisionados e por reforço desempenham papéis complementares, contribuindo para a compreensão dos dados e para a tomada de decisão em níveis mais estratégicos dentro do contexto de sistemas industriais complexos.

2.3.3 Previsão de RUL baseada em dados

No contexto da previsão da RUL, predominam abordagens baseadas em aprendizado supervisionado, uma vez que a tarefa consiste essencialmente em estimar o tempo restante até a falha com base em dados históricos e operacionais previamente rotulados, conforme apresentado anteriormente. Esses modelos exploram dados provenientes de múltiplas fontes, como sensores embarcados, parâmetros operacionais e registros de manutenção, permitindo capturar padrões

associados ao processo de degradação dos sistemas ao longo do tempo (Khan *et al.*, 2022).

Nesse cenário, a abordagem *data-driven* fundamenta-se na aplicação de técnicas avançadas para extrair informações relevantes a partir de dados brutos, apoiando tarefas de detecção, diagnóstico e, sobretudo, prognóstico de falhas (Gupta, Kumar e Maiti, 2024). Modelos de ML têm se consolidado como ferramentas robustas para lidar com grandes volumes de dados heterogêneos, sendo capazes de aprender relações complexas, frequentemente não lineares, e integrar diferentes tipos de variáveis simultaneamente, característica essencial em ambientes industriais complexos (Darwish, 2024; Li e Li, 2025; Subashree *et al.*, 2026).

Entre as diversas abordagens aplicadas à previsão de RUL, destacam-se modelos baseados em DL, que têm recebido crescente atenção na literatura, e métodos de ML tradicionais, especialmente aqueles baseados em árvores e frequentemente combinados por técnicas de *ensemble* (Dalzochio *et al.*, 2020; Lefrouni e Taibi, 2025; Zhang, Yang e Wang, 2019). Essas abordagens diferem em termos de capacidade de representação, custo computacional e interpretabilidade, sendo frequentemente exploradas de forma complementar (Bhushan, 2025; Pandey, 2023; Swain, 2025).

2.3.4 Abordagens baseadas em *Deep Learning*

O DL pode ser definido como uma subárea do ML que se baseia no uso de Redes Neurais Artificiais (*Artificial Neural Networks*) com múltiplas camadas de processamento, capazes de aprender representações de dados em diferentes níveis de abstração (LeCun, Bengio e Hinton, 2015). As Redes Neurais (*Neural Networks*), por sua vez, consistem em modelos formados por unidades interconectadas organizadas em camadas, que transformam os dados de entrada por meio de combinações lineares e funções não lineares, permitindo capturar relações complexas entre variáveis (Goodfellow, Bengio e Courville, 2017). Já o termo “profundo” refere-se à profundidade dessas arquiteturas, isto é, ao número de camadas intermediárias responsáveis por transformar progressivamente as representações de entrada em descrições mais complexas e informativas (Goodfellow, Bengio e Courville, 2017). Com base nisso, o DL distingue-se por sua capacidade de aprender automaticamente padrões hierárquicos dos dados, nas quais níveis mais elevados capturam

características cada vez mais abstratas, aspecto central para o desempenho desses modelos em problemas complexos (Bengio, 2009; LeCun, Bengio e Hinton, 2015).

O DL tem recebido grande atenção no campo de PHM devido à sua capacidade de modelar relações complexas em dados de alta dimensionalidade (Ramezani *et al.*, 2023). Arquiteturas como Redes Neurais Convolucionais (*Convolutional Neural Networks* - CNNs), RNNs e Redes Neurais Profundas (*Deep Neural Network* - DNN) são particularmente eficazes no processamento de séries temporais multivariadas, frequentemente encontradas em sistemas industriais (Tama *et al.*, 2023; Wu, J.-Y. *et al.*, 2021; Zhang, Yang e Wang, 2019).

Uma das principais vantagens dessas abordagens é a capacidade de realizar extração automática de características, reduzindo a necessidade de engenharia manual de atributos e frequentemente resultando em melhor desempenho preditivo (Darwish, 2024; Gupta, Kumar e Maiti, 2024). As arquiteturas profundas destacam-se especialmente em tarefas de prognóstico devido à sua habilidade de aprender representações hierárquicas diretamente dos dados sensoriais brutos (Serradilla *et al.*, 2022; Zhang *et al.*, 2019).

Dentre essas arquiteturas, as RNNs destacam-se por sua capacidade de modelar dados sequenciais, sendo especialmente adequadas para o tratamento de séries temporais. Diferentemente de redes neurais tradicionais, as RNNs apresentam conexões recorrentes entre seus neurônios, permitindo que informações de etapas anteriores sejam incorporadas no processamento de entradas atuais. Essa característica possibilita a modelagem de dependências temporais e dinâmicas ao longo do tempo, aspecto fundamental para a análise de processos de degradação em sistemas industriais (Goodfellow, Bengio e Courville, 2017; Shi, Shi e Li, 2017). Durante o treinamento, os parâmetros dessas redes são ajustados iterativamente por meio de algoritmos de retropropagação, permitindo que o modelo aprenda padrões sequenciais complexos a partir dos dados (Magnusson, Olsson e Tran, 2023; Naumov, 2017). Entretanto, as RNNs ainda possuem dificuldades quando se é necessário lidar com longos intervalos de tempo (Sahoo *et al.*, 2019).

Nesse contexto, Nguyen e Medjaher (2019) argumentam que modelos baseados em *Long Short-Term Memory* (LSTM), um tipo de RNN, são mais adequados do que abordagens tradicionais quando o mecanismo de degradação é complexo e difícil de modelar explicitamente, especialmente em sistemas com múltiplos sensores e dependências temporais. Adicionalmente, Hong *et al.* (2020)

demonstram que a combinação de diferentes arquiteturas de DL, como CNN, RNN, LSTM e BiLSTM, pode resultar em modelos híbridos mais robustos e com maior acurácia preditiva.

Apesar do desempenho superior, modelos de DL apresentam limitações importantes, como a necessidade de grandes volumes de dados rotulados, elevado custo computacional e baixa interpretabilidade, sendo frequentemente considerados “caixas-pretas” (Bourokba, Hamdi, El e Njah, 2024; Gupta, Kumar e Maiti, 2024). Esses fatores representam desafios relevantes para aplicações industriais, onde transparência e confiabilidade são essenciais (Alaoui, Meddaoui e Hain, 2025). Tais limitações motivam a investigação de abordagens alternativas ou complementares, como os métodos de *ensemble* (Biggio e Kastanis, 2020; El-Brawany *et al.*, 2023; Fink *et al.*, 2020).

2.3.5 Abordagens baseadas em Árvores e métodos de *ensemble*

Modelos baseados em árvores de decisão (*decision trees*) constituem uma das abordagens mais tradicionais, interpretáveis e amplamente utilizadas no ML, sendo aplicados tanto em tarefas de classificação quanto de regressão (Breiman *et al.*, 1984; Hastie, Tibshirani e Friedman, 2009). Esses modelos podem ser descritos como estruturas hierárquicas baseadas em regras, nas quais o espaço de atributos é particionado recursivamente em regiões cada vez mais homogêneas por meio de testes sucessivos sobre as variáveis de entrada. Cada nó interno representa uma condição sobre um atributo, enquanto os nós folha correspondem às predições do modelo (Quinlan, 1986; Sutton, 2005).

Do ponto de vista teórico, árvores de decisão são modelos não paramétricos, não dependendo de suposições prévias sobre a distribuição dos dados, o que permite lidar de forma eficaz com diferentes tipos de variáveis, incluindo dados contínuos e categóricos, além de apresentar relativa robustez a outliers e dados faltantes. Adicionalmente, esses modelos são capazes de capturar relações não lineares complexas entre variáveis sem a necessidade de transformações explícitas, característica particularmente relevante em cenários industriais com múltiplas variáveis interdependentes (Hastie, Tibshirani e Friedman, 2009; Sutton, 2005).

Apesar dessas vantagens, árvores de decisão individuais apresentam limitações importantes, especialmente relacionadas à alta variância e à sensibilidade a pequenas variações nos dados de treinamento, o que pode comprometer sua

capacidade de generalização (Breiman, 1996; Jurek *et al.*, 2014). Essa instabilidade inerente frequentemente resulta em ajuste excessivo (*overfitting*), mas torna esses modelos candidatos ideais para compor métodos de *ensemble*, nos quais múltiplos modelos são combinados para produzir previsões mais robustas e estáveis (Jurek *et al.*, 2014).

Sob essa perspectiva, o aprendizado por *ensemble* pode ser entendido como uma abordagem que visa construir modelos preditivos mais fortes a partir da combinação de múltiplos modelos base, frequentemente denominados aprendizes fracos (*weak learners*) (Hastie, Tibshirani e Friedman, 2009; Sutton, 2005). Essa estratégia tem sido amplamente empregada na previsão da RUL devido à sua capacidade de melhorar a acurácia e a estabilidade das previsões, especialmente em cenários caracterizados por dados sensoriais de alta dimensionalidade e elevados níveis de ruído (Allal *et al.*, 2026; Serenko *et al.*, 2024; Zhang, Yang e Wang, 2019). Além disso, modelos de *ensemble* são capazes de capturar diferentes fontes de incerteza, incluindo tanto a incerteza epistêmica, associada à variabilidade do modelo, quanto a incerteza aleatória inerente aos dados (Allal *et al.*, 2026).

Diferentes estratégias de *ensemble* têm sido propostas na literatura, destacando-se principalmente os métodos de *Bagging*, *Stacking* e *Boosting*, que diferem na forma como os modelos de base são construídos e combinados (Jurek *et al.*, 2014; Serenko *et al.*, 2024). O *Bagging* (*Bootstrap Aggregating*) baseia-se na geração de múltiplos subconjuntos de dados por amostragem com reposição, sobre os quais modelos independentes são treinados e posteriormente combinados por meio de agregação, geralmente por média ou votação, com o objetivo de reduzir a variância do modelo final (Murphy, 2022; Sutton, 2005). Essa abordagem é particularmente eficaz para estabilizar aprendizes instáveis, como árvores de decisão profundas, e apresenta a vantagem de ser facilmente paralelizável (Dietterich, 2000; Jurek *et al.*, 2014).

O *Stacking* (*Stacked Generalization*), por sua vez, adota uma arquitetura em múltiplos níveis na qual diferentes modelos base (nível 0) são treinados e suas previsões são utilizadas como entrada para um meta-modelo (nível 1), que aprende a combinar essas saídas para minimizar o erro global de generalização (Breiman, 1996; Wolpert, 1992). Esta técnica permite deduzir e corrigir os vieses dos modelos de nível 0 ao projetar suas predições em um segundo espaço de busca, explorando a complementaridade entre diferentes algoritmos (Wolpert, 1992). Para garantir a

eficácia dessa integração e evitar o vazamento de dados, o treinamento do nível 1 é realizado com previsões "fora da amostra", obtidas obrigatoriamente por meio de processos de validação cruzada (Breiman, 1996; Wolpert, 1992).

Em contraste, o *Boosting* constitui uma abordagem sequencial, na qual modelos são treinados iterativamente, com cada novo modelo focando nos erros cometidos pelos anteriores (Jurek *et al.*, 2014; Sutton, 2005). Ao atribuir maior peso a instâncias difíceis, o *Boosting* direciona o processo de aprendizado para regiões mais complexas do espaço de atributos, permitindo a construção de modelos altamente expressivos a partir da combinação de aprendizes fracos (Murphy, 2022; Sutton, 2005). Essa estratégia tem se mostrado particularmente eficaz em problemas de prognóstico, sendo amplamente aplicada na previsão de RUL (Serenko *et al.*, 2024; Singh e Bobde, 2025b).

Dentre as diversas abordagens apresentadas, arquiteturas baseadas em LSTM, no contexto de DL, e modelos baseados em *Boosting*, como o XGBoost, destacam-se na literatura de PHM pela sua eficácia na previsão da RUL, sendo, portanto, selecionados como foco de análise neste trabalho.

2.3.6 Long Short-Term Memory (LSTM)

No campo do DL, as redes LSTM surgem como uma evolução das RNNs tradicionais, superando as dificuldades destas em lidar com longos intervalos de tempo (Akram e El, 2016; Sahoo *et al.*, 2019). A principal vantagem da LSTM é a sua capacidade de capturar dependências e tendências de longo prazo em séries temporais complexas e voláteis (Kotsompolis *et al.*, 2026). Seu funcionamento depende de portões (*gates*) que possibilitam um maior controle de quais informações serão ou não utilizadas pelo modelo, são eles: Portão de Esquecimento (*Forget Gate*), Portão de Entrada (*Input Gate*) e Portão de Saída (*Output Gate*) (Sahoo *et al.*, 2019). De acordo com Kotsompolis *et al.* (2026) e Yuvaraju e Divya (2020), o funcionamento dos portões ocorre da seguinte forma:

- Portão de Esquecimento (*Forget Gate*): utiliza uma função sigmoide para selecionar quais informações obsoletas serão descartadas do estado da célula.

$$f_t = \sigma (W_f \cdot [h_{t-1}, X_t] + b_f) \quad (1)$$

- Portão de Entrada (*Input Gate*): seleciona as novas informações a serem armazenadas, combinando funções sigmoide e tangente hiperbólica para atualizar o estado da célula com novos valores candidatos.

$$I_t = \sigma (W_i \cdot [h_{t-1}, X_t] + b_i) \quad (2)$$

$$C_t = \tanh (W_c \cdot [h_{t-1}, X_t] + b_c) \quad (3)$$

- Portão de Saída (*Output Gate*): regula qual parte do estado da célula influenciará o estado oculto h_t , determinando a saída do modelo para o passo de tempo atual.

$$O_t = \sigma (W_o \cdot [h_{t-1}, X_t] + b_o) \quad (4)$$

$$C_t = f_t * C_{t-1} + I_t * C'_t \quad (5)$$

$$h_t = O_t * \tanh (C_t) \quad (6)$$

Ao permitir que informações relevantes sejam transportadas através de muitos passos de tempo, enquanto sinais irrelevantes são descartados, a LSTM demonstra desempenho superior na modelagem de dinâmicas mutáveis e condições operacionais variáveis (Kotsompolis *et al.*, 2026). Por esses motivos é tipicamente utilizada para previsão de RUL no campo do PHM (Apeiranthitis *et al.*, 2026).

2.3.7 eXtreme Gradient Boosting (XGBoost)

O XGBoost é uma implementação avançada do algoritmo de *Gradient Boosting* baseada em árvores de decisão, projetada para correção sequencial de erros (Ok e Emmanuel, 2025; Rajagukguk *et al.*, 2025). Diferente de métodos como o *Random Forest*, que treina árvores de forma independente, o XGBoost adiciona sucessivas árvores ao modelo focando nos erros residuais das iterações anteriores, o que proporciona uma acurácia preditiva superior (Karakılıç, Hataş e Pacal, 2025). Um diferencial crítico desta arquitetura é a inclusão de um termo de regularização $\Omega(f)$ na função objetivo L , apresentada a seguir, para penalizar a complexidade do modelo (Karakılıç, Hataş e Pacal, 2025; Kotsompolis *et al.*, 2026).

$$L = \sum_{i=1}^N L(\hat{y}_i, y_i) + \sum_{k=1}^K \Omega(f_k) \quad (7)$$

$$\Omega(f) = \gamma T + 0.5 \lambda \|\omega\|^2 \quad (8)$$

Onde $\hat{y}_i$ representa o valor previsto, y_i denota o valor real, N corresponde ao número de observações e K indica o total de árvores empregadas. A diferenciação do algoritmo reside na aplicação do termo de regularização $\Omega(f)$, no qual γ e λ atuam como coeficientes de punição. Especificamente, o parâmetro γ controla o ganho mínimo de redução de perda necessário para realizar a divisão de um nó interno, sendo que valores mais elevados de γ resultam em árvores mais simples e generalistas. Adicionalmente, T define o número de ramos (ou folhas) e ω representa o peso atribuído a cada folha (Bentéjac, Csörgő e Martínez-Muñoz, 2019; Kotsompolis *et al.*, 2026). O algoritmo opera através de um método ganancioso (*greedy*) para otimizar esta estrutura, ajustando o número de folhas e os pesos de cada nó por meio dos referidos coeficientes de punição (Karakılıç, Hataş e Pacal, 2025; Kotsompolis *et al.*, 2026).

2.4 INTERPRETABILIDADE EM APRENDIZADO DE MÁQUINA (*EXPLAINABLE AI* - XAI)

A natureza "caixa-preta" de algoritmos complexos de ML e DL constitui uma barreira significativa para sua adoção em domínios críticos (Bourokba, Hamdi, El e Njah, 2024). Embora esses modelos apresentem um sucesso notável na PdM baseada em dados, a opacidade inerente a suas arquiteturas impede que operadores e técnicos compreendam o que ocorre no sistema, o motivo de determinado evento e a forma adequada de reação (Ferraro *et al.*, 2023; Kundu e Hoque, 2023). Esta falta de transparência gera desconfiança, uma vez que as equipes de manutenção enfrentam dificuldades para identificar as características que fundamentam uma previsão (Gupta, Kumar e Maiti, 2024).

Essa opacidade torna-se um problema crítico na previsão da RUL, aplicada frequentemente em setores de alta responsabilidade, como o aeroespacial e o ferroviário, nos quais decisões incorretas resultam em perdas financeiras severas ou riscos à vida humana. Uma vez que o objetivo final da RUL é o suporte à decisão, o usuário final necessita de clareza sobre os momentos em que pode confiar na estimativa do modelo (Gawde *et al.*, 2024; Li *et al.*, 2024). Nesse cenário,

compreender as razões por trás de uma previsão possui uma importância equivalente à própria acurácia do modelo em diversas aplicações (Lundberg e Lee, 2017).

A XAI surge como a solução para este impasse, pois integra técnicas de interpretabilidade aos modelos de IA para a detecção e previsão de falhas (Gawde *et al.*, 2024). A XAI refere-se a métodos que produzem modelos explicáveis e elucidam o processo de decisão de um algoritmo, o que permite a compreensão dos resultados por seres humanos (Bourokba, Hamdi, El e Njah, 2024). Tais técnicas oferecem *insights* sobre os fatores que contribuem para a previsão, o que viabiliza uma análise de causa raiz precisa e intervenções de manutenção direcionadas (Gawde *et al.*, 2024; Singh e Bobde, 2025b). Assim, o especialista de domínio pode complementar as informações do modelo com seu próprio conhecimento técnico, o que resulta em uma estratégia de manutenção mais confiável (Gawde *et al.*, 2024).

No entanto, existe um *trade-off* conhecido entre a acurácia de um modelo e sua interpretabilidade: de modo geral, os modelos de maior desempenho técnico são os menos interpretáveis (Kamwa, Samantaray e Joos, 2013). Para mitigar este conflito, a maioria das estruturas de XAI utiliza métodos "agnósticos ao modelo" (*model-agnostic*). Estas técnicas não buscam a criação de modelos inerentemente transparentes, que poderiam ter sua precisão reduzida, mas sim a explicação de modelos "caixa-preta" já existentes por meio de uma análise externa (Ahmad *et al.*, 2022; Bennetot *et al.*, 2022).

Dentre os métodos agnósticos, o SHAP destaca-se como uma técnica de análise *post-hoc*, conceito que designa o estudo realizado após o treinamento do modelo para extrair conhecimento sem a necessidade de modificar sua arquitetura interna (Barredo Arrieta *et al.*, 2020). Por constituir uma estrutura abrangente e robusta, o SHAP permite o entendimento do sistema em dois níveis complementares: a explicação global, que identifica o impacto geral de cada atributo sobre o comportamento do modelo em todo o conjunto de dados, e a explicação local, que justifica os motivos de uma previsão específica para uma única unidade ou falha isolada (Ferraro *et al.*, 2023; Gawde *et al.*, 2024; Linardatos, Papastefanopoulos e Kotsiantis, 2020).

O funcionamento do método baseia-se na teoria dos jogos cooperativos e atribui a cada atributo um valor de importância por meio do cálculo dos "Valores de Shapley" (Gawde *et al.*, 2024; Singh e Bobde, 2025b). Tais valores representam a contribuição marginal média de cada característica para a previsão sob a análise de

todas as combinações possíveis de variáveis (Ensarioglu, 2025; Gawde *et al.*, 2024). Na prática, o SHAP quantifica o impacto de cada sensor no resultado final e facilita a identificação dos fatores de maior influência no risco de falha previsto (Ensarioglu, 2025; Lundberg e Lee, 2017). Além disso, a ferramenta possui versatilidade para aplicação em diversos modelos, desde redes neurais profundas até algoritmos baseados em árvores, embora a interpretação dos resultados exija cautela devido à complexidade das interações entre as variáveis (Gawde *et al.*, 2024).

A preferência pelo SHAP em relação a outras técnicas, como o LIME (*Local Interpretable Model-agnostic Explanations*), deve-se à sua superioridade em estabilidade e consistência na geração de explicações (Kundu e Hoque, 2023). Enquanto o LIME restringe-se ao nível local e pode apresentar perda de fidelidade ao simplificar excessivamente o modelo "caixa-preta", o SHAP captura interações complexas e garante maior transparência para a previsão da RUL (Ensarioglu, 2025; Gawde *et al.*, 2024; Singh e Bobde, 2025b).

O uso de explicações *post-hoc* via SHAP revela se as recomendações automáticas derivam de sinais físicos coerentes ou de correlações espúrias ou enganosas (Danach *et al.*, 2026; Gawde *et al.*, 2024). O diagnóstico de falhas torna-se, portanto, um processo rastreável, no qual cada agendamento de manutenção vincula-se a padrões específicos de sensores. Essa rastreabilidade constitui uma trilha de auditoria valiosa para a conformidade regulatória e para a segurança em domínios de alta criticidade, como a aviação e a energia, o que apoia a análise de causa raiz e permite intervenções direcionadas (Danach *et al.*, 2026; Ensarioglu, 2025; Gawde *et al.*, 2024).

2.5 TRABALHOS CORRELATOS

A literatura sobre PdM e previsão da RUL apresenta uma evolução rápida, impulsionada pela Indústria 4.0 e pela necessidade de maior confiabilidade em sistemas complexos (Achouch *et al.*, 2022; Zonta *et al.*, 2020). Esta seção apresenta uma análise de estudos recentes que empregam abordagens baseadas em ML e DL para previsão da RUL extraídos da base de dados Scopus.

Com o objetivo de sintetizar o panorama atual das pesquisas em RUL associadas a métodos de explicabilidade, a Tabela 1 apresenta uma comparação

entre os estudos discutidos, destacando os modelos utilizados, as bases de dados empregadas e as principais vantagens e limitações identificadas na literatura.

Tabela 1 - Trabalhos correlatos

Autor	Modelo	Base de Dados	Técnica XAI	Vantagens	Limitações
Ma <i>et al.</i> (2020)	XGBoost (<i>Ensemble</i>)	NASA C-MAPSS	SHAP	Utiliza SHAP para interpretar a influência de diferentes sensores na previsão da RUL.	N/A
Hong <i>et al.</i> (2020)	1D CNN, LSTM e Bi-LSTM (<i>Deep Learning</i>)	NASA C-MAPSS (FD001)	SHAP	Alcança alta precisão combinando extração de características espaciais e temporais com redução de dimensionalidade.	A explicação via SHAP não é absoluta e depende da precisão do modelo subjacente, podendo ser afetada por ruído nos dados dos sensores.
Mwangi <i>et al.</i> (2025)	<i>Attention-based</i> GRU / LSTM / <i>Ensemble</i> , DL	MAN B&W 6S46MC-C7	SHAP	Apresenta um <i>framework</i> modular sensível à carga que integra detecção de anomalias, agrupamento de falhas e prognose da RUL com interpretabilidade física.	A sobreposição de sinal característico da falha em cargas elevadas dificulta o diagnóstico preciso e a implementação real exige alta capacidade computacional.
Nguyen e Medjaher (2019)	<i>Long Short-Term Memory</i> (LSTM) (<i>Deep Learning</i>)	NASA C-MAPSS	N/A	Propõe um <i>framework</i> dinâmico completo, desde a prognose baseada em dados até a tomada de decisão de manutenção, fornecendo probabilidades de falha em diferentes janelas de tempo.	O modelo considera apenas manutenção perfeita e foca em falhas de componentes específicos sem integrar o planejamento de produção completo.
Serenko <i>et al.</i> (2024)	RNN com LSTM (<i>Deep Learning</i>)	NASA C-MAPSS	N/A	Melhora a confiabilidade das previsões ao estimar a própria incerteza do modelo através de <i>Deep Ensembling</i> .	Modelos RNN tradicionais podem sofrer com o desaparecimento de gradientes e limitações em funções de ativação comuns.
Huang <i>et al.</i> (2024)	<i>Cross-Validation Accuracy Weighted Probability Ensemble</i> (CAWPE) (<i>Ensemble</i>)	BackBlaze SMART data (ST4000DM000, ST8000DM002, etc.)	N/A	Estende o ciclo de detecção de falhas e melhora a precisão ao transformar o problema em classificação multi-classe baseada em RUL. Utiliza uma camada <i>softmax</i> para aumentar a interpretabilidade dos resultados	Elevado custo computacional decorrente do processamento de múltiplos atributos, a sensibilidade ao <i>overfitting</i> em arquiteturas de <i>Ensemble</i> complexas e a ambiguidade nas informações temporais ao utilizar diferenciações de ordem superior
Darwish (2024)	ATCN-LSTM (<i>Deep Learning</i>)	2018 PHM Data Challenge Dataset (subconjunto 01_M02)	N/A	O modelo integra TCN, LSTM e mecanismos de auto-atenção para capturar dependências de longo alcance e padrões temporais, superando modelos convencionais.	Necessidade de desenvolver modelos individuais para cada equipamento, devido às variações nos estados operacionais e de consumo das máquinas

Singh <i>et al.</i> (2018)	<i>Random Forest (Ensemble)</i> e <i>Dynamic Time Warping (DTW)</i>	2018 PHM Data Challenge Dataset	N/A	Propõe uma metodologia orientada a dados para estimativa simultânea da RUL para múltiplas falhas usando correspondência de curvas via DTW.	As previsões da RUL não seguem a realidade de forma consistente quando o sinal de falha é fraco ou está distante do tempo de falha real.
Zhao, Zhu e Zhao (2022)	<i>Transformer</i> e <i>Random Forest (Deep Learning)</i>	2018 PHM Data Challenge Dataset (subconjunto 03_M01_DC)	N/A	Integra o módulo <i>Transformer</i> com <i>Random Forest</i> para identificar estados de saúde e prever RUL com alta eficiência computacional e precisão.	O custo computacional do <i>Transformer</i> aumenta rapidamente com o aumento de parâmetros, dificultando tarefas de alta dimensão ou sequências longas.
TV <i>et al.</i> (2018)	<i>Deep Recurrent Neural Network (RNN/LSTM)</i> (<i>Deep Learning</i>)	2018 PHM Data Challenge Dataset	N/A	Utiliza um conjunto de modelos LSTM profundos para estimar simultaneamente a RUL para três diferentes modos de falha.	Extrema dificuldade em modelar a grande variância nos valores da RUL e ausência de conhecimento sobre o início exato da degradação nos dados de treino.
Huang <i>et al.</i> (2018)	LSTM e <i>Random Forest Classifier</i> (<i>Deep Learning</i>)	2018 PHM Data Challenge Dataset	N/A	Incorpora uma etapa de detecção precoce do modo de degradação via <i>Random Forest</i> antes da aplicação do LSTM, lidando melhor com falhas abruptas.	O erro absoluto no início do monitoramento é maior devido à limitação de dados históricos disponíveis nesse estágio inicial.
Wu, S. <i>et al.</i> (2021)	GRU, GRU-FCs e <i>Random Forest (Ensemble)</i> e <i>Deep Learning</i>	2018 PHM Data Challenge Dataset (subconjunto 01_M02)	N/A	Modelos baseados em GRU com triagem de amostras superaram abordagens LSTM em precisão e eficiência para a previsão da RUL.	A introdução de camadas complementares (FCs) melhorou a precisão para determinados modos de falha, porém reduziu o desempenho em outros.

Fonte – Autor (2026)

A partir da análise dos estudos revisados, observa-se que modelos de DL frequentemente alcançam elevados níveis de acurácia na previsão da RUL, especialmente em cenários caracterizados por grande volume de dados e dependências temporais complexas (Ensarioglu, 2025; Wu, J.-Y. *et al.*, 2021). Entretanto, modelos baseados em *ensemble*, como *Random Forest* e XGBoost, mantêm competitividade em termos de desempenho preditivo em diferentes bases de dados amplamente utilizadas na literatura, como NASA C-MAPSS e PHM Challenge (Huang *et al.*, 2024; Ma *et al.*, 2020; Serenko *et al.*, 2024; Singh e Bobde, 2025a). Além disso, são capazes de identificar com maior precisão o início da etapa de degradação da máquina, dificuldade enfrentada por TV *et al.* (2018).

Conforme destacado por Biggio e Kastanis (2020), não existe uma abordagem universal aplicável a todos os cenários de PHM. A escolha do modelo depende

fundamentalmente da natureza dos dados e do problema analisado. Portanto, a seleção final do modelo deve ser conduzida por meio de avaliação comparativa sob métricas consistentes.

Diante desse panorama, o presente trabalho propõe uma união entre dois paradigmas amplamente utilizados na literatura para previsão da RUL: os modelos de regressão, LSTM e GRU, representando abordagens baseadas em DL para séries temporais, e o modelo de classificação, o XGBoost, representando métodos baseados em *ensemble*. O modelo será avaliado utilizando uma base de dados e métricas de desempenho experimental. Adicionalmente, será aplicada o SHAP, uma técnica de XAI, ao modelo com o objetivo de analisar a contribuição das variáveis de entrada na previsão da RUL e investigar o nível de interpretabilidade obtido.

3 ABORDAGEM METODOLÓGICA

Nessa seção será apresentada a metodologia de pesquisa, que será dividida em duas partes: (i) o enquadramento metodológico e (ii) os procedimentos metodológicos.

3.1 ENQUADRAMENTO METODOLÓGICO

O presente trabalho enquadra-se, quanto à sua abordagem, como uma pesquisa quantitativa. Esta classificação se justifica pela natureza dos dados e dos objetivos, sendo possível quantificar numericamente a relação entre as variáveis (sinais dos sensores) e o evento de interesse (a RUL). A pesquisa desenvolve modelos preditivos descritos em linguagem matemática e computacional e utiliza técnicas analíticas e estatísticas para calcular valores numéricos (ex: métricas de erro como RMSE) para avaliar o sistema estudado (Gil, 2002; Miguel *et al.*, 2010). Quanto à sua natureza, é uma pesquisa aplicada, pois tem como objetivo gerar conhecimentos para a aplicação prática, visando especificamente apoiar e melhorar as decisões de PdM na indústria (Gil, 2002).

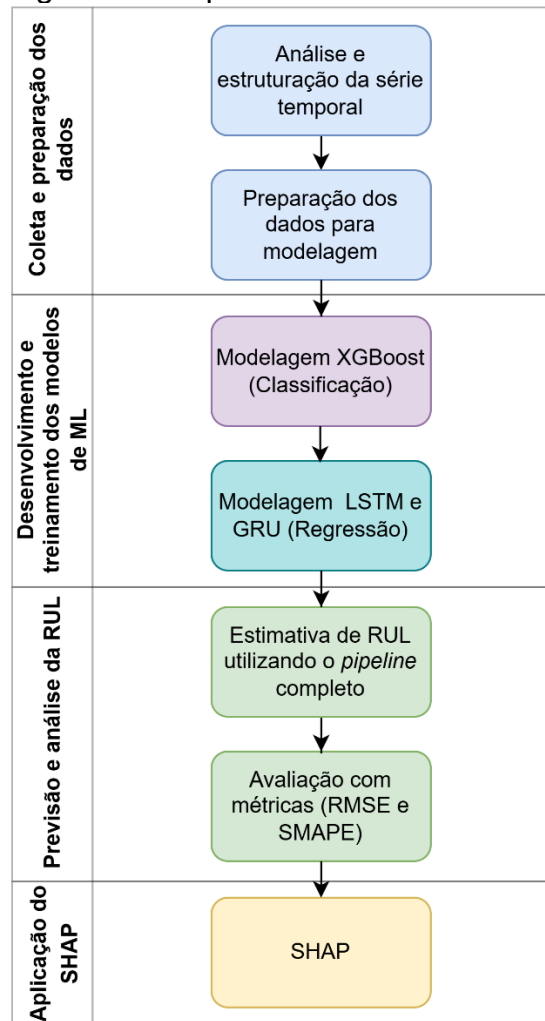
Em relação aos objetivos, o trabalho é caracterizado como uma pesquisa exploratória. Embora utilize técnicas de modelagem estabelecidas, o objetivo principal é o aprimoramento de ideias e a descoberta de intuições sobre a interpretabilidade desses modelos. A análise de exemplos (via SHAP) para estimular a compreensão de quais variáveis impulsionam a degradação alinha-se com a definição de Gil (2002). A fonte de informação é caracterizada como uma fonte secundária, o banco de dados público “PHM *Data Challenge* 2018” (Gil, 2002). A principal técnica aplicada sobre esta fonte documental é a Modelagem e Simulação, que busca imitar e prever o comportamento de um sistema real, o processo de degradação da máquina, à medida que este evolui no tempo (Miguel *et al.*, 2010).

3.2 PROCEDIMENTOS METODOLÓGICOS

A pesquisa consistiu nas seguintes etapas de desenvolvimento conforme apresentado pela Figura 1: (i) Coleta e preparação dos dados; (ii) Desenvolvimento e

treinamento do modelo de ML; (iii) Previsão e análise da RUL; e, finalmente (iv) Aplicação do SHAP para melhor explicabilidade do modelo.

Figura 1 - Etapas de desenvolvimento



Fonte – Autor (2026)

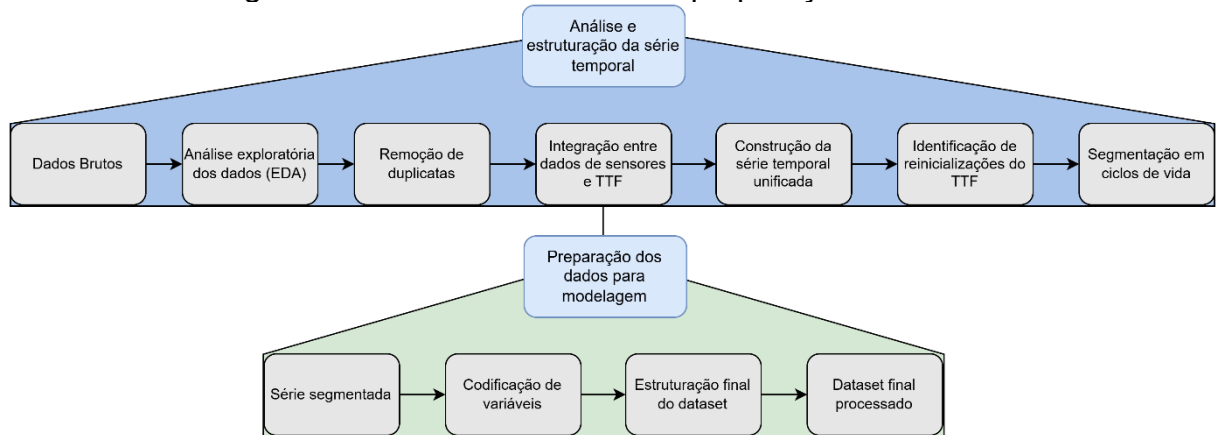
3.2.1 Coleta e preparação dos dados

A etapa de coleta e preparação dos dados tem como objetivo garantir a consistência e a adequação das informações para o treinamento dos modelos. Inicialmente, foi realizada uma Análise Exploratória dos Dados (*Exploratory Data Analysis* - EDA), contemplando a investigação da estrutura temporal, a identificação de padrões operacionais e a análise de correlação entre os sensores, com o objetivo de identificar redundâncias, relações de dependência e possíveis implicações para as etapas posteriores de modelagem.

Na sequência, foram realizados o tratamento de registros duplicados e a integração entre os dados de sensores e as informações de tempo até a falha (*Time-*

To-Failure - TTF), resultando em um conjunto de dados unificado indexado temporalmente. O processo completo de coleta e preparação dos dados é ilustrado na Figura 2.

Figura 2 - Processos de coleta e preparação dos dados



Fonte – Autor (2026)

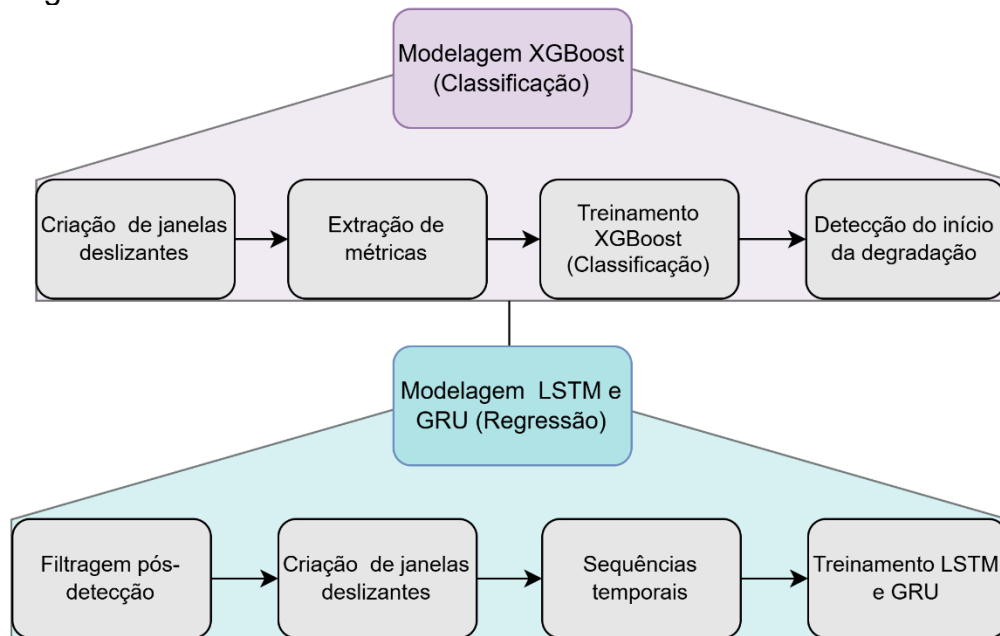
Para lidar com a possibilidade de múltiplos ciclos de degradação ocorrerem dentro de uma mesma execução operacional, foi desenvolvido um procedimento de segmentação baseado na identificação de reinicializações do TTF ao longo do tempo. A partir desse processo, foi criado um atributo representando ciclos de vida independentes, utilizado como referência para todas as etapas subsequentes.

No tratamento das variáveis, foi aplicada codificação apropriada para atributos categóricos relevantes ao processo. Ao final dessa etapa, obteve-se uma série temporal estruturada, limpa e segmentada em ciclos de vida, que serviu como base para a construção das amostras utilizadas nos modelos de ML, realizada posteriormente na etapa de desenvolvimento dos modelos. Essa estruturação garante a coerência temporal dos dados e preserva a integridade dos ciclos de degradação para as etapas de modelagem.

3.2.2 Desenvolvimento e treinamento dos modelos de ML

Com os dados preparados, o desenvolvimento dos modelos foi estruturado em uma arquitetura de dois estágios, composta por um modelo de classificação e um modelo de regressão, cada um responsável por uma etapa específica do problema de prognóstico. A Figura 3 ilustra as etapas do desenvolvimento.

Figura 3 - Processos de desenvolvimento e treinamento dos modelos



Fonte – Autor (2026)

Inicialmente, para viabilizar o uso de modelos de ML, as séries temporais foram organizadas por meio da técnica de janelas deslizantes (*sliding windows*). Nessa abordagem, cada amostra corresponde a um conjunto consecutivo de observações ao longo do tempo, com comprimento fixo, no qual cada linha representa um instante temporal e cada coluna um atributo de entrada. O valor alvo associado a cada janela é definido a partir do último instante da sequência, correspondendo ao valor de RUL naquele ponto. O deslocamento sucessivo da janela ao longo da série permite a geração de múltiplas amostras sobrepostas, preservando a continuidade temporal dos dados e ampliando o conjunto de treinamento. Essa abordagem é particularmente adequada para modelos capazes de capturar dependências temporais, como RNNs.

No primeiro estágio, foi utilizado um modelo baseado em XGBoost para a classificação do estado operacional do equipamento. Para isso, as janelas construídas a partir das séries temporais passaram por um processo de engenharia de atributos (*feature engineering*). Nessa etapa, os dados sequenciais foram transformados em vetores de características por meio da extração de métricas estatísticas e temporais representativas do comportamento recente dos sinais, resultando em um formato adequado para modelos baseados em árvores de decisão.

Adicionalmente, foram aplicadas etapas específicas de preparação dos dados voltadas ao treinamento do classificador, incluindo o tratamento do desbalanceamento

entre as classes de operação normal e degradação. Esse desbalanceamento é inerente ao problema, uma vez que eventos de falha são significativamente menos frequentes do que períodos de operação regular. Consequentemente, estratégias de balanceamento foram empregadas com o objetivo de melhorar a capacidade do modelo em identificar corretamente o início da degradação.

O segundo estágio foi dedicado ao desenvolvimento do modelo de regressão para estimativa da RUL, utilizando RNN, especificamente o LSTM. Diferentemente do modelo de classificação, esse modelo opera diretamente sobre sequências temporais de dados de sensores, preservando a dependência temporal entre as observações. As entradas são compostas por sequências de comprimento fixo, construídas a partir de janelas deslizantes, nas quais cada amostra representa um trecho contínuo do histórico do equipamento.

Para a organização das sequências temporais utilizadas, foram definidos segmentos operacionais a partir dos identificadores de execução disponíveis na sessão. Cada unidade de execução originou inicialmente um segmento e, quando foram identificados aumentos abruptos nos contadores de TTF dentro da mesma execução, foram criadas subdivisões adicionais. Essa etapa teve como objetivo delimitar trechos temporalmente coerentes para a geração das janelas deslizantes, evitando que uma mesma sequência atravessasse reinicializações locais dos contadores de TTF.

Para o treinamento desse modelo, foram consideradas apenas as sequências correspondentes ao período posterior à detecção do início da degradação pelo classificador, direcionando o aprendizado para padrões associados à proximidade da falha. Essa estratégia permite concentrar o modelo na dinâmica de degradação do sistema, reduzindo a influência de períodos prolongados de operação estável.

Além do modelo LSTM, foi implementada uma arquitetura baseada em GRU, com o objetivo de realizar uma análise comparativa. Essa escolha foi motivada pelo trabalho de Wu *et al.* (2021), que reportam desempenho superior do GRU em relação ao LSTM no problema de estimativa de RUL. Entretanto, os autores utilizam como entrada um atributo adicional baseado no resíduo de pressão, o qual não foi incorporado neste estudo. Dessa forma, o experimento conduzido permite avaliar se a superioridade do GRU se mantém na ausência dessa informação adicional.

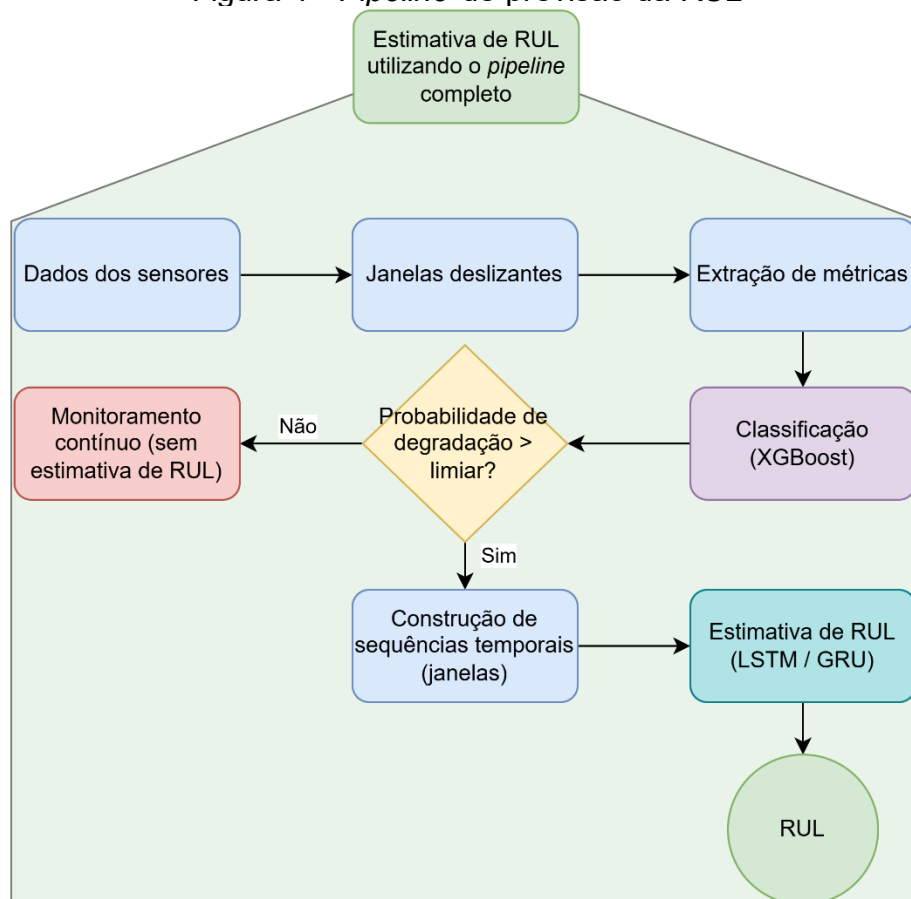
A divisão dos dados em conjuntos de treinamento, validação e teste foi realizada de forma temporal, respeitando os limites dos segmentos identificados e

evitando vazamento de informação entre os conjuntos. Essa divisão foi efetuada antes da construção das janelas deslizantes, garantindo que não houvesse sobreposição temporal entre os dados utilizados em cada etapa. Por fim, os dados de entrada foram normalizados com base exclusivamente no conjunto de treinamento, assegurando consistência tanto no processo de aprendizado quanto na aplicação dos modelos.

3.2.3 Previsão e análise da RUL

A etapa de previsão consiste na aplicação sequencial do *pipeline* desenvolvido, ilustrado na Figura 4, simulando o funcionamento do sistema em um cenário real de monitoramento. Nesse cenário, o modelo de classificação previamente treinado é aplicado continuamente sobre os dados mais recentes de sensores, organizados em janelas temporais, avaliando, para cada janela, a probabilidade de o sistema se encontrar em estado de degradação. O classificador não atua apenas como um pré-processamento, mas como um mecanismo de controle do sistema, determinando o momento a partir do qual a estimativa de RUL se torna relevante e definindo a região operacional na qual o modelo de regressão é aplicado.

Figura 4 - *Pipeline* de previsão da RUL



Fonte – Autor (2026)

A partir dessas estimativas, é aplicado um critério de decisão baseado em um limiar, responsável por definir o estado operacional do equipamento. A classificação resultante atua como um gatilho para o restante do método. Caso o equipamento seja classificado como operando em condição normal, o sistema permanece em modo de monitoramento contínuo, sem gerar estimativas de RUL. Por outro lado, ao identificar o início da degradação, o modelo de regressão é ativado.

Uma vez ativado, o modelo de regressão recebe como entrada sequências temporais recentes dos dados de sensores, estruturadas em janelas deslizantes de comprimento fixo. Essas janelas são construídas a partir do instante em que a degradação é detectada pelo classificador, de forma consistente com o processo de treinamento do modelo, que foi realizado exclusivamente com dados pertencentes à fase de degradação. Essa estratégia garante que o modelo opere sobre padrões representativos do processo de falha, preservando a dependência temporal entre as observações. A saída do modelo corresponde à estimativa da RUL associada ao instante mais recente da sequência analisada.

A adoção desse mecanismo de ativação condicional permite restringir a geração de previsões a períodos em que há evidência consistente de degradação, reduzindo a ocorrência de estimativas irrelevantes durante a operação normal e aumentando a confiabilidade do sistema.

Por fim, para a avaliação do desempenho do sistema, são empregadas métricas apropriadas para problemas de regressão em séries temporais. Entre elas, destacam-se a raiz do erro quadrático médio (*Root Mean Squared Error* - RMSE), que quantifica a magnitude média dos erros de previsão, e o erro percentual absoluto médio simétrico (*Symmetric Mean Absolute Percentage Error* - SMAPE), que permite avaliar a precisão relativa das estimativas, especialmente em diferentes escalas de RUL. As formulações matemáticas são apresentadas a seguir:

$$RMSE = \sqrt{\frac{1}{n} \sum_{n=1}^N (\hat{y}_n - y_n)^2} \quad (9)$$

$$SMAPE = \frac{100\%}{n} \sum_{n=1}^N \frac{2|\hat{y}_n - y_n|}{(\hat{y}_n + y_n)} \quad (10)$$

Onde N é o tamanho da amostra, $\hat{y}_n$ é a RUL prevista e y_n a RUL real.

Já para a avaliação do desempenho do modelo de classificação, foram empregadas métricas amplamente utilizadas em problemas de classificação binária, incluindo matriz de confusão, precisão (*precision*), *recall* e a área sob a curva ROC (*Area Under the Receiver Operating Characteristic Curve* – AUC-ROC). A matriz de confusão permite analisar a distribuição das previsões corretas e incorretas entre as classes, evidenciando a ocorrência de falsos positivos e falsos negativos. A métrica de precisão quantifica a proporção de instâncias corretamente classificadas como degradação entre todas as previsões positivas, enquanto o *recall* mede a capacidade do modelo de identificar efetivamente os eventos de degradação. A AUC-ROC, por sua vez, avalia o desempenho do classificador ao longo de diferentes limiares de decisão, sendo particularmente relevante em cenários com desbalanceamento de classes, como o presente estudo. Em conjunto, essas métricas fornecem uma visão abrangente do desempenho do modelo, permitindo avaliar tanto sua capacidade discriminativa quanto sua adequação ao objetivo de detecção precoce da degradação.

Os resultados obtidos a partir dessas métricas são analisados de forma comparativa, permitindo avaliar o desempenho do método proposto e sua adequação ao problema de previsão de RUL.

3.2.4 Aplicação do SHAP

Para a aplicação do SHAP ao modelo de regressão, foi utilizado o método *GradientExplainer*, escolhido por sua compatibilidade direta com modelos implementados em *TensorFlow/Keras* (biblioteca utilizada para o desenvolvimento do modelo de ML) e por sua eficiência computacional em cenários de alta dimensionalidade. Por ser uma abordagem agnóstica a modelos, esse método utiliza gradientes para estimar as contribuições das variáveis, preservando a estrutura temporal dos dados de entrada e retornando valores SHAP com a mesma dimensionalidade das sequências utilizadas pelo modelo.

A definição do conjunto de referência (*background*) é um passo fundamental na aplicação do SHAP, pois estabelece o comportamento esperado do modelo na ausência de informações específicas. O conjunto de referência (*background*) foi composto por 100 amostras selecionadas aleatoriamente do conjunto de treinamento normalizado. A escolha desse tamanho busca equilibrar a estabilidade das estimativas dos valores SHAP e o custo computacional associado ao processo.

Os valores SHAP foram calculados para um subconjunto de 500 amostras do conjunto de teste, selecionadas de forma aleatória. Esse conjunto foi definido de forma a abranger diferentes estágios do ciclo de degradação, permitindo uma análise abrangente do comportamento do modelo ao longo do tempo.

Como os valores SHAP são originalmente expressos em unidades normalizadas, foi realizada uma etapa adicional de conversão para unidades de tempo, permitindo sua interpretação direta no contexto da RUL. Essa conversão foi baseada no valor máximo de RUL observado no conjunto de treinamento e no intervalo de amostragem do banco de dados, possibilitando associar as contribuições das variáveis a variações reais na estimativa de tempo até a falha.

A consistência dos valores obtidos foi verificada por meio da comparação entre a soma das contribuições SHAP e as predições do modelo, assegurando que a decomposição da saída fosse coerente com os princípios teóricos da abordagem. A partir dos valores calculados, foram realizadas diferentes análises com o objetivo de compreender o comportamento do modelo. Inicialmente, foi avaliada a importância global das variáveis, considerando a contribuição média absoluta de cada sensor ao longo de todas as amostras e instantes temporais. Em seguida, foi analisado o perfil de atenção temporal do modelo, permitindo identificar quais regiões da janela temporal exercem maior influência nas predições.

Adicionalmente, foi construída uma representação bidimensional relacionando sensores e instantes temporais, possibilitando identificar padrões específicos de dependência entre variáveis e momentos do histórico do equipamento. Por fim, as amostras foram agrupadas de acordo com diferentes faixas de RUL, permitindo investigar variações na importância das variáveis ao longo do processo de degradação.

Essas análises fornecem uma visão abrangente do funcionamento interno do modelo, permitindo avaliar se as predições estão fundamentadas em variáveis fisicamente relevantes ao problema, especialmente aquelas associadas ao sistema de resfriamento *FlowCool*, diretamente relacionado às falhas consideradas neste estudo.

3.2.5 Ferramentas utilizadas

Todo o desenvolvimento computacional deste trabalho foi realizado em linguagem Python (versão 3.13.7), utilizando bibliotecas amplamente adotadas na

área de ciência de dados e ML. As bibliotecas pandas e NumPy foram empregadas para manipulação, organização e operações vetorizadas sobre os dados, enquanto Matplotlib e Seaborn foram utilizadas para a construção de visualizações. A biblioteca SciPy foi utilizada para cálculos estatísticos auxiliares, e o scikit-learn foi empregado nas etapas de pré-processamento, divisão dos dados e cálculo das métricas de avaliação.

O modelo de classificação foi desenvolvido com o algoritmo XGBoost, escolhido por sua eficiência computacional e bom desempenho em dados tabulares com desbalanceamento de classes. Para o desenvolvimento dos modelos de regressão baseados em RNNs, foi utilizada a biblioteca TensorFlow, por meio da API Keras, permitindo a implementação das arquiteturas LSTM e GRU. A análise de interpretabilidade foi conduzida com a biblioteca SHAP, enquanto a biblioteca PyArrow foi utilizada para leitura e escrita eficiente dos dados em formato Parquet ao longo do pipeline.

4 RESULTADOS E DISCUSSÕES

Esta seção apresenta os resultados obtidos ao longo das diferentes etapas do *pipeline* desenvolvido para estimativa da RUL em um contexto de PdM. Os resultados são discutidos em conjunto com as decisões metodológicas que os motivaram, incluindo experimentos que não produziram melhorias e cuja documentação é relevante para compreender os limites do problema.

Inicialmente, são apresentadas as etapas de coleta e preparação dos dados, abrangendo a caracterização do conjunto de dados e a análise exploratória. Em seguida, são discutidos o desenvolvimento e o treinamento dos modelos de ML empregados no *pipeline* proposto. Por fim, são apresentados os resultados relacionados ao desempenho preditivo e à interpretabilidade dos modelos, com foco na análise das variáveis mais relevantes para o processo de degradação.

4.1 COLETA E PREPARAÇÃO DOS DADOS

No início do método, as etapas de coleta e preparação dos dados são fundamentais para estruturar adequadamente o problema de prognóstico. Nesta subseção, apresenta-se o conjunto de dados utilizado, destacando suas principais características, variáveis e contexto operacional. Em seguida, é conduzida uma análise exploratória detalhada, com o objetivo de identificar padrões relevantes, compreender a estrutura temporal dos registros e subsidiar as decisões de pré-processamento adotadas. Essas etapas garantem a consistência dos dados e sua adequação às fases subsequentes de modelagem.

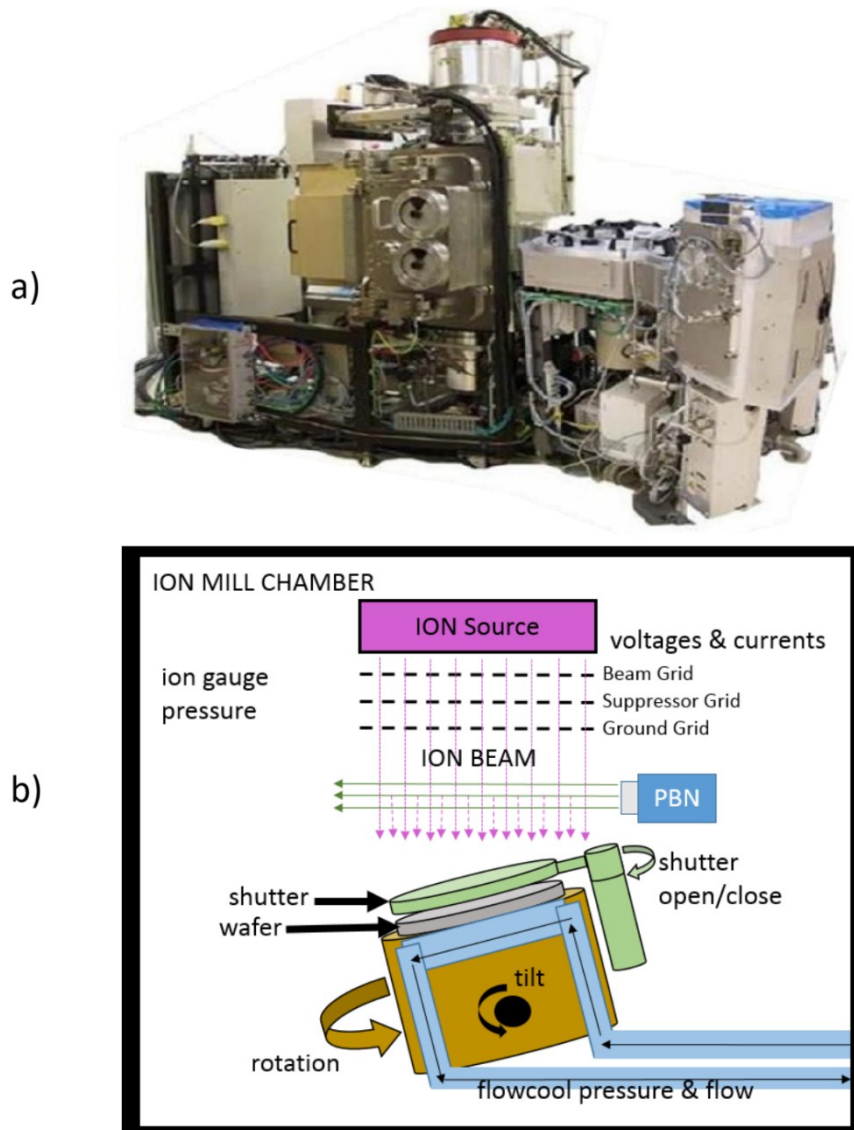
4.1.1 Banco de dados “PHM Data Challenge 2018”

O conjunto de dados utilizado neste trabalho é proveniente do “PHM *Data Challenge* 2018”, uma competição organizada pela *Prognostics and Health Management Society* com o objetivo de estimular o desenvolvimento de métodos para diagnóstico de falhas e previsão de RUL em sistemas industriais.

O problema proposto baseia-se em um equipamento de *ion mill etching* (ataque por feixe de íons), mostrado na Figura 5a, utilizado na fabricação de semicondutores. Nesse processo, íons são acelerados por um campo elétrico e direcionados à superfície do *wafer* (lâmina de material semicondutor), promovendo a

remoção controlada de material. O sistema opera em diferentes etapas de processamento (*recipes*), nas quais parâmetros como tensão, corrente, ângulo de inclinação e velocidade de rotação são ajustados. Durante a operação, o *wafer* é resfriado por um sistema denominado *FlowCool*, responsável por manter condições térmicas adequadas ao processo. A Figura 5b ilustra o processo.

Figura 5 - Equipamento e processo de ataque por feixe de íons



Fonte – PHM DATA CHALLENGE (2018)

Os dados disponibilizados consistem em séries temporais de sensores coletadas ao longo da operação de múltiplos equipamentos, com amostragem em intervalos regulares de tempo. Cada registro contém 24 variáveis mostradas na Tabela 2, incluindo informações de contexto operacional e medições de sensores. Dentre essas variáveis, 16 correspondem diretamente a sensores do equipamento,

responsáveis por monitorar grandezas físicas como pressão, corrente, tensão, vazão de gases e parâmetros de movimento mecânico.

Tabela 2 - Variáveis de cada sensor

ID	Nome do Parâmetro	Tipo	Descrição
S1	<i>time</i>	Numérico	tempo
S2	<i>Tool</i>	Categórico	ID da ferramenta
S3	<i>stage</i>	Categórico	etapa de processamento do <i>wafer</i>
S4	<i>Lot</i>	Categórico	ID do <i>wafer</i>
S5	<i>runnum</i>	Numérico	número de vezes que a ferramenta foi operada
S6	<i>recipe</i>	Categórico	descreve as configurações da ferramenta usadas para processar o <i>wafer</i>
S7	<i>recipe_step</i>	Categórico	etapa de processo de uma receita
S8	IONGAUGE PRESSURE	Numérico (Sensor)	leitura de pressão para a câmara de processo principal sob vácuo
S9	ETCHBEAMVOLTAGE	Numérico	potencial de voltagem aplicado à placa de feixe do conjunto da grade
S10	ETCHBEAMCURRENT	Numérico	corrente de íons impactando a grade de feixe, determinando a quantidade de íons acelerados através do conjunto da grade para o <i>wafer</i>
S11	ETCHSUPPRESSORVOLTAGE	Numérico	potencial de voltagem aplicado à placa supressora do conjunto da grade
S12	ETCHSUPPRESSORCURRENT	Numérico (Sensor)	corrente de íons impactando a placa da grade supressora
S13	FLOWCOOLFLOWRATE	Numérico	taxa de fluxo de hélio através do circuito de resfriamento (<i>flowcool</i>), controlado pelo controlador de fluxo de massa
S14	FLOWCOOLPRESSURE	Numérico (Sensor)	pressão de hélio resultante no circuito de resfriamento (<i>flowcool</i>)
S15	ETCHGASCHANNEL1READBACK	Numérico	taxa de fluxo de argônio para o conjunto da fonte na câmara de vácuo
S16	ETCHPBGASREADBACK	Numérico	taxa de fluxo de argônio para o conjunto PBN na câmara
S17	FIXTURETILTANGLE	Numérico	configuração do ângulo de inclinação do suporte do <i>wafer</i>
S18	ROTATIONSPEED	Numérico	configuração da velocidade de rotação do <i>wafer</i>
S19	ACTUALROTATIONANGLE	Numérico (Sensor)	medida do ângulo de rotação do <i>wafer</i>
S20	FIXTURESHUTTERPOSITION	Numérico	configuração de abrir/fechar o obturador para blindagem do <i>wafer</i>
S21	ETCHSOURCEUSAGE	Numérico	contador de uso para o consumível do conjunto da grade
S22	ETCHAUXSOURCE TIMER	Numérico	contador de uso para o consumível das blindagens da câmara
S23	ETCHAUX2SOURCE TIMER	Numérico	contador de uso para o consumível das blindagens da câmara
S24	ACTUALSTEPDURATION	Numérico (Sensor)	duração de tempo medida para uma etapa específica

Fonte – PHM DATA CHALLENGE (2018)

As variáveis do banco de dados podem ser agrupadas em três categorias principais. A primeira corresponde a variáveis de identificação e contexto operacional, como o tempo, a ID da ferramenta, a etapa de processamento, o ID do *wafer*, o número de vezes que a ferramenta foi operada, bem como os parâmetros de configuração representados por *recipe* e *recipe_step*.

A segunda categoria é composta pelas variáveis de sensores, que representam medições físicas do sistema durante a operação. Entre elas destacam-

se leituras de pressão da câmara de vácuo, parâmetros elétricos associados ao feixe de íons, variáveis relacionadas ao sistema *FlowCool* além de medições de fluxo de gases, ângulo de inclinação do *wafer*, velocidade de rotação e posição do obturador (FIXTURESHUTTERPOSITION), que indica se o *wafer* está exposto ao feixe de íons.

A terceira categoria compreende variáveis de uso e temporização do equipamento, como contadores de utilização de componentes e a duração real de cada etapa do processo.

Além dos dados de sensores, o banco de dados inclui arquivos auxiliares contendo os instantes de ocorrência de falhas e informações sobre o TTF. As falhas consideradas estão associadas ao sistema *FlowCool* e são classificadas em três tipos:

- I. *FlowCool Pressure Dropped Below Limit* (queda de pressão abaixo do limite operacional – Falha F1);
- II. *FlowCool Pressure Too High Check FlowCool Pump* (pressão excessiva - Falha F2);
- III. *FlowCool Leak* (vazamento - Falha F3).

O instante de falha registrado corresponde ao momento em que o equipamento é interrompido para manutenção, não sendo necessariamente o início do processo de degradação.

A partir dessas informações, o objetivo do problema consiste em, a cada instante de tempo, estimar o tempo restante até a próxima falha com base apenas nos dados históricos disponíveis até aquele momento. Esse cenário caracteriza um problema típico de PdM baseado em séries temporais, no qual padrões de degradação devem ser inferidos a partir do comportamento dos sensores ao longo do tempo.

Dentre os modos de falha disponíveis no conjunto de dados, este trabalho concentra-se especificamente na Falha F1, associada à queda de pressão do sistema de resfriamento *FlowCool* (*FlowCool Pressure Dropped Below Limit*). A escolha dessa falha foi motivada pela maior disponibilidade e consistência dos ciclos de degradação observados no banco de dados, fator importante para o treinamento e avaliação dos modelos de ML. Além disso, a análise foi restrita à sessão identificada como “01_M02”, correspondente aos dados coletados da máquina M02 durante a sessão operacional 01, selecionada tanto pela quantidade representativa de ciclos de falha e registros operacionais quanto pelo fato de ser utilizada em estudos anteriores, como Darwish

(2024) e Wu *et al.* (2021), permitindo uma comparação mais consistente entre os resultados obtidos.

4.1.2 Análise exploratória e preparação dos dados

Antes do desenvolvimento dos modelos, realizou-se uma EDA detalhada da sessão 01_M02, com o objetivo de compreender a estrutura temporal dos dados, caracterizar os ciclos da Falha F1 e identificar padrões relevantes para as decisões de pré-processamento. A análise foi dividida em 3 partes principais: Visão Geral da Sessão, Estrutura Temporal e Modo Operacional e Correlação entre Sensores

4.1.2.1 Visão Geral da Sessão

A análise inicial evidenciou que a sessão abrange aproximadamente 422 dias de operação ($3,28 \times 10^6$ a $3,98 \times 10^7$ segundos), com 1.677 *runnums* (número de vezes que a ferramenta foi operada) distintos e 5.103.087 registros após remoção de duplicatas temporais. O intervalo de amostragem nominal é de 4 segundos, com desvios pontuais, foram observados *gaps* de até 16.464 s dentro do mesmo *runnum*, refletindo pausas não documentadas na coleta. Entre *runnums* consecutivos, o tempo de não atividade (*idle*) tem mediana de 1.454 s e máximo de 511.624 s (aproximadamente 6 dias), indicando que a sessão inclui interrupções prolongadas de produção. As estatísticas estão apresentadas na Tabela 3.

Tabela 3 - Estatísticas descritivas

Indicador	Valor
Duração total da sessão	422.5 dias (10,141 h)
Número de <i>runnums</i>	1,677
Registros (após deduplicação)	5,103,087
Intervalo de amostragem mediano (<i>intra-runnum</i>)	4 s
Gap máximo <i>intra-runnum</i>	16,464 s (4.6 h)
Gap <i>inter-runnum</i> — mediana	1,454 s (~24 min)
Gap <i>inter-runnum</i> — média	9,236 s (~2.6 h)
Gap <i>inter-runnum</i> — máximo	511,624 s (~5.9 dias)

Fonte – Autor (2026)

A ausência de valores é bem baixa, apenas 7 registros com FIXTURESHUTTERPOSITION nulo (0,0001%) e 7.455 duplicatas temporais (0,146%). O TTF_F1 (*Time-to-Failure* da Falha F1) está definido para 90,5% dos

registros, sendo que os 9,5% restantes correspondem a períodos após o último ciclo documentado, onde não há falha futura dentro da janela de observação.

4.1.2.2 Estrutura Temporal

A variável TTF_F1 é o eixo organizador de toda a estrutura: ela define as zonas operacionais para o classificador, o horizonte de previsão para o regressor e os ciclos que estruturam a divisão treino/validação/teste. A análise exploratória desta variável foi, portanto, o passo mais crítico da EDA.

A modelagem da degradação exige a conversão do TTF_F1 contínuo em rótulos discretos para o classificador XGBoost. Dois limiares devem ser definidos: o limite superior da zona normal e o limite superior da zona de degradação. A região entre eles é excluída do treinamento, pois corresponde a uma transição gradual em que rótulos binários seriam ambíguos. A escolha desses valores foi orientada pela literatura do “PHM *Data Challenge* 2018”, onde os autores listados na Tabela 4 divergem consideravelmente nas definições.

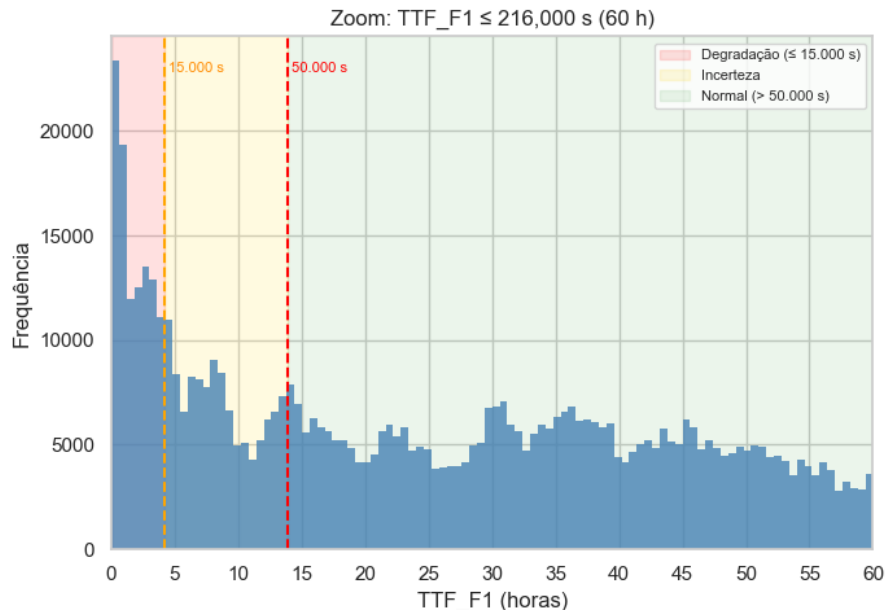
Tabela 4 - Definição de limiares na literatura

Referência	Limiar normal	Limiar degradação	Unidade
Huang <i>et al.</i> (2018)	> 50.000	< 5.000	segundos
Wu, S. <i>et al.</i> (2021)	> 6.250 amostras (~25.000 s)	≤ 500 amostras (~2.000 s)	amostras × 4 s
Darwish (2024)	> 5.000 amostras (~20.000 s)	≤ 500 amostras (~2.000 s)	amostras × 4 s
Zhao, Zhu e Zhao (2022)	> 5.500	< 5.000	segundos

Fonte – Autor (2026)

Conforme mostra a Figura 6, o limiar de 50.000 s para a classe normal foi adotado seguindo Huang *et al.* (2024) e serve também como TTF_MAX para o LSTM: define o horizonte máximo de RUL modelável pelo regressor. Já o limiar de 15.000 s para a classe de degradação foi definido empiricamente a partir dos resultados obtidos nas etapas posteriores de modelagem, apresentando desempenho superior ao limiar de 5.000 s utilizado por Huang *et al.* (2024). A adoção desse valor amplia a região classificada como degradação e reduz a zona de incerteza excluída do treinamento.

Figura 6 - Limiar normal e de degradação



Fonte – Autor (2026)

A distribuição resultante revela uma assimetria marcante entre as zonas, conforme mostra a Tabela 5, onde apenas 2,3% dos registros pertencem à zona de degradação. Esse desbalanceamento severo exige estratégias explícitas no treinamento do classificador, discutidas na Seção 4.2.1.

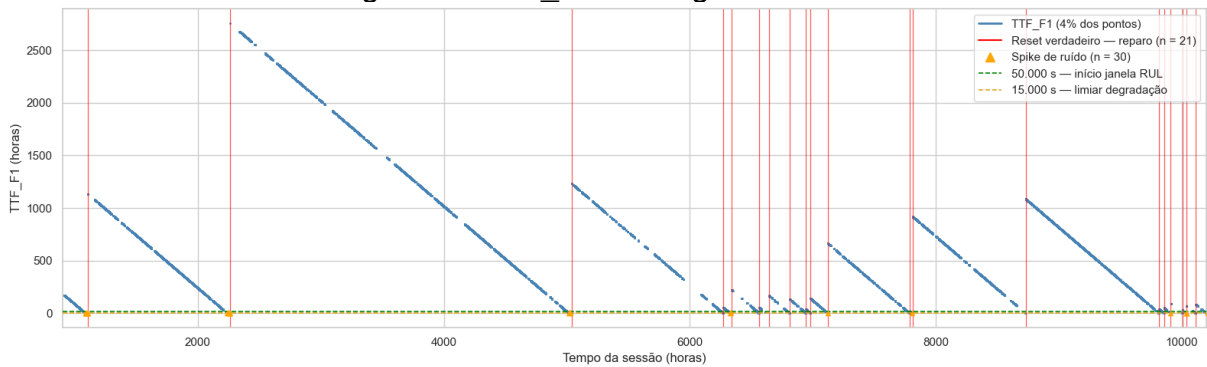
Tabela 5 - Distribuição de registros por zona

Zona	Critério	Registros	Proporção
Normal	$TTF_F1 > 50.000$ s	4.398.089	95,2%
Incerteza	$15.000 \text{ s} < TTF \leq 50.000 \text{ s}$	115.481	2,5%
Degradação	$TTF_F1 \leq 15.000$ s	104.180	2,3%

Fonte – Autor (2026)

Já a etapa de identificação dos ciclos de degradação a partir da série TTF_F1 exigiu uma análise cuidadosa. A Figura 7 mostra que o TTF_F1 decresce monotonamente dentro de cada ciclo (à medida que a falha se aproxima) e aumenta abruptamente quando a máquina é reparada, reiniciando a contagem. No entanto, pequenas flutuações positivas também ocorrem dentro dos ciclos, decorrentes de ruído de medição ou recalibrações pontuais, tornando necessário distinguir reinicializações reais do TTF de variações associadas ao ruído.

Figura 7 - TTF_F1 ao longo da sessão



Fonte – Autor (2026)

Analisando todos os aumentos positivos do TTF_F1, observou-se um salto natural na distribuição: 30 eventos classificados como ruído atingiram no máximo 34.306 s, enquanto os 21 eventos de reparo verdadeiro partiram de 53.718 s. Resultando em 22 ciclos até a falha (*run-to-failure*) identificados na sessão, cada ciclo corresponde a um período de operação contínua desde um reparo do *FlowCool* até a próxima Falha F1. Como unidade de controle, foi atribuído um identificador único (*“cycle_id”*) que delimita os limites de cada ciclo e será útil na divisão treino/validação/teste dos modelos subsequentes.

Um aspecto estrutural relevante é que um ciclo de Falha F1 não corresponde a um único *runnum*: ele abrange desde o reparo até a próxima falha, cruzando dezenas a centenas de *wafer runs*. Isso tem implicação direta na divisão treino/validação/teste, respeitar os limites de ciclo é condição necessária para evitar perda de informações.

Na zona de degradação ($TTF_F1 \leq 15.000$ s), todos os 22 ciclos apresentam volume de dados abundante, em média 4.961 amostras brutas por ciclo. Esse número, entretanto, não representa diretamente o volume disponível para o LSTM, pois os filtros subsequentes, a extração do último segmento contínuo e o ponto de aviso do classificador reduzem substancialmente cada ciclo antes que o regressor seja treinado.

4.1.2.3 Modo Operacional e Correlação entre Sensores

O campo FIXTURESHUTTERPOSITION (FSP) registra o estado do obturador da máquina e determina o modo operacional associado a cada registro. Como se trata de uma variável categórica, foi aplicada codificação *one-hot encoding*, transformando

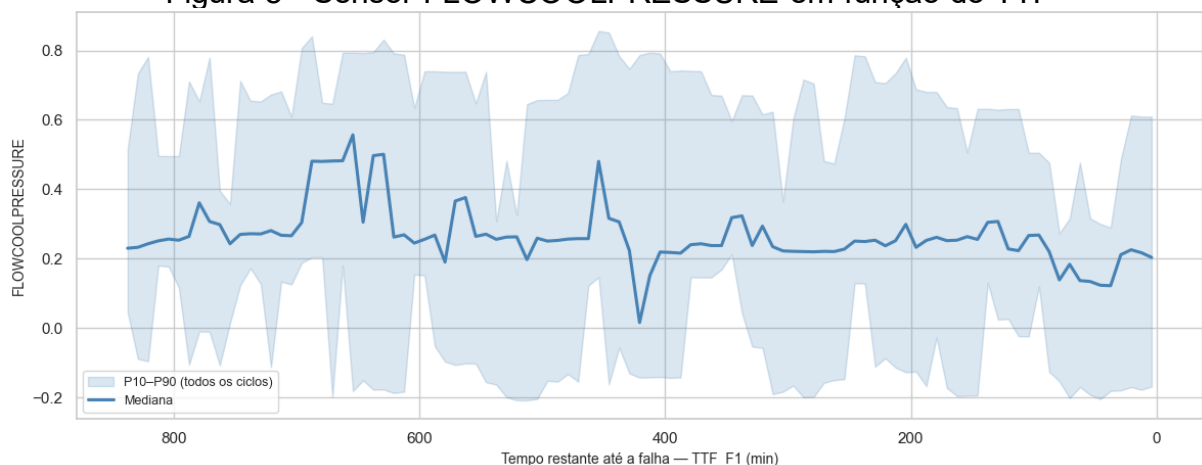
os diferentes estados observados em cinco colunas binárias, nomeadas conforme o padrão SHUTTER_x, correspondentes aos valores 0, 1, 2, 3 e 255 identificados no conjunto de dados.

A posição 1 corresponde ao processamento ativo de *wafers*, definição adotada por Wu *et al.* (2021) e Darwish (2024) como condição representativa de operação efetiva do equipamento. As demais posições representam estados operacionais distintos e apresentam padrões de sensores significativamente diferentes, indicando mudanças no regime de funcionamento da máquina.

A distribuição da sessão 01_M02 confirma a predominância do estado operacional ativo: a posição 1 representa 80,25% dos registros (4.101.355 amostras), enquanto os demais estados somam menos de 20% do conjunto de dados. Essa distinção entre modos operacionais mostrou-se relevante para as etapas posteriores de modelagem, especialmente no pipeline de regressão da RUL, no qual foram considerados exclusivamente registros associados ao processamento ativo.

Dado que a Falha F1 é definida pela queda de pressão do sistema *FlowCool*, torna-se relevante examinar diretamente o comportamento do sensor FLOWCOOLPRESSURE em função do TTF. A Figura 8 apresenta essa relação considerando todos os ciclos de operação disponíveis. Observa-se que a mediana do sinal permanece relativamente estável ao longo de praticamente todo o horizonte temporal, sem evidenciar uma trajetória de degradação progressiva. A faixa entre os percentis P10 e P90 indica elevada dispersão entre os ciclos, revelando alta variabilidade do sinal mesmo sob condições operacionais semelhantes.

Figura 8 - Sensor FLOWCOOLPRESSURE em função do TTF



Fonte – Autor (2026)

Mesmo em instantes próximos à falha, não se observa um padrão consistente que permita distinguir de forma robusta o estado crítico do sistema com base exclusivamente nesse sensor. Esse comportamento indica que, apesar de diretamente relacionado ao mecanismo de falha, o sensor FLOWCOOLPRESSURE apresenta capacidade limitada de fornecer um sinal preditivo estável ao longo do tempo. Como consequência, modelos que dependem desse tipo de informação tendem a ter dificuldade em aprender padrões robustos para a estimativa da RUL.

Posteriormente foi feita uma análise de correlação entre os 16 sensores, a Tabela 6 mostra os resultados. Foram identificando 6 pares com correlação $\geq 0,90$, agrupados em dois padrões distintos. O primeiro e mais crítico é o cluster formado pelos três odômetros (ETCHSOURCEUSAGE, ETCHAUXSOURCETIMER e ETCHAUX2SOURCETIMER) com correlações de 0,9998 a 1,0000 entre si. Do ponto de vista informacional, os três representam um único sinal: o tempo acumulado de operação da fonte iônica, expresso em três variáveis redundantes. O segundo padrão envolve pares de grandezas elétricas acopladas fisicamente: ETCHBEAMVOLTAGE - ETCHSUPPRESSORVOLTAGE e ETCHBEAMCURRENT - ETCHSUPPRESSORCURRENT.

Tabela 6 - Análise de correlação entre os sensores

Par	Correlação	Interpretação
ETCHAUXSOURCETIMER $\leftrightarrow$ ETCHAUX2SOURCETIMER	1	Virtualmente idênticos
ETCHSOURCEUSAGE $\leftrightarrow$ ETCHAUXSOURCETIMER	0,9998	Grupo colinear
ETCHSOURCEUSAGE $\leftrightarrow$ ETCHAUX2SOURCETIMER	0,9998	Grupo colinear
ETCHBEAMCURRENT $\leftrightarrow$ ETCHSUPPRESSORCURRENT	0,9816	Par elétrico acoplado
ETCHBEAMVOLTAGE $\leftrightarrow$ ETCHSUPPRESSORVOLTAGE	0,9756	Par elétrico acoplado
ETCHBEAMVOLTAGE $\leftrightarrow$ ETCHBEAMCURRENT	0,9047	Dependência operacional

Fonte – Autor (2026)

A redundância dos sensores com natureza temporal tem implicação direta nos resultados da análise SHAP (Seção 4.4.1): quando os três dominam a importância global do modelo, isso representa na prática a dominância de um único atributo temporal. No XGBoost, o corte baseado no ganho (*gain*) (Seção 4.2.1) controlou indiretamente esse efeito ao selecionar apenas os atributos mais informativos de cada grupo.

Os resultados da EDA indicam que variáveis diretamente associadas ao mecanismo de falha, como o sensor FLOWCOOLPRESSURE, não apresentam padrões claros de degradação progressiva ao longo do tempo. Esse comportamento

sugere possíveis limitações para modelos baseados em aprendizado supervisionado, que dependem da existência de sinais consistentes para capturar a dinâmica de falha, aspecto investigado nas seções subsequentes.

4.2 DESENVOLVIMENTO E TREINAMENTO DOS MODELOS DE ML

No estágio seguinte, são desenvolvidos e treinados os modelos de ML responsáveis pela detecção de degradação e pela estimativa da RUL. Esta seção descreve, inicialmente, a construção do classificador baseado em XGBoost, incluindo as etapas de engenharia de atributos, definição de limiares e avaliação de desempenho. Em seguida, são apresentadas as estratégias de preparação das sequências temporais utilizadas como entrada para os modelos recorrentes, com destaque para os processos de filtragem e definição da variável-alvo de RUL. Por fim, são detalhadas a arquitetura e as configurações de treinamento dos modelos baseados em RNNs, consolidando o processo de modelagem proposto.

4.2.1 Classificador XGBoost

Com os dados preparados e os ciclos de falha identificados, o primeiro estágio da etapa é o treinamento de um classificador XGBoost capaz de detectar o início da degradação da máquina. A zona de incerteza ($15.000 \text{ s} < \text{TTF_F1} \leq 50.000 \text{ s}$) foi excluída do treinamento por corresponder a uma transição gradual onde rótulos binários seriam contraditórios, o classificador recebe apenas exemplos claramente normais ($\text{TTF_F1} > 50.000 \text{ s}$) ou claramente em degradação ($\text{TTF_F1} \leq 15.000 \text{ s}$).

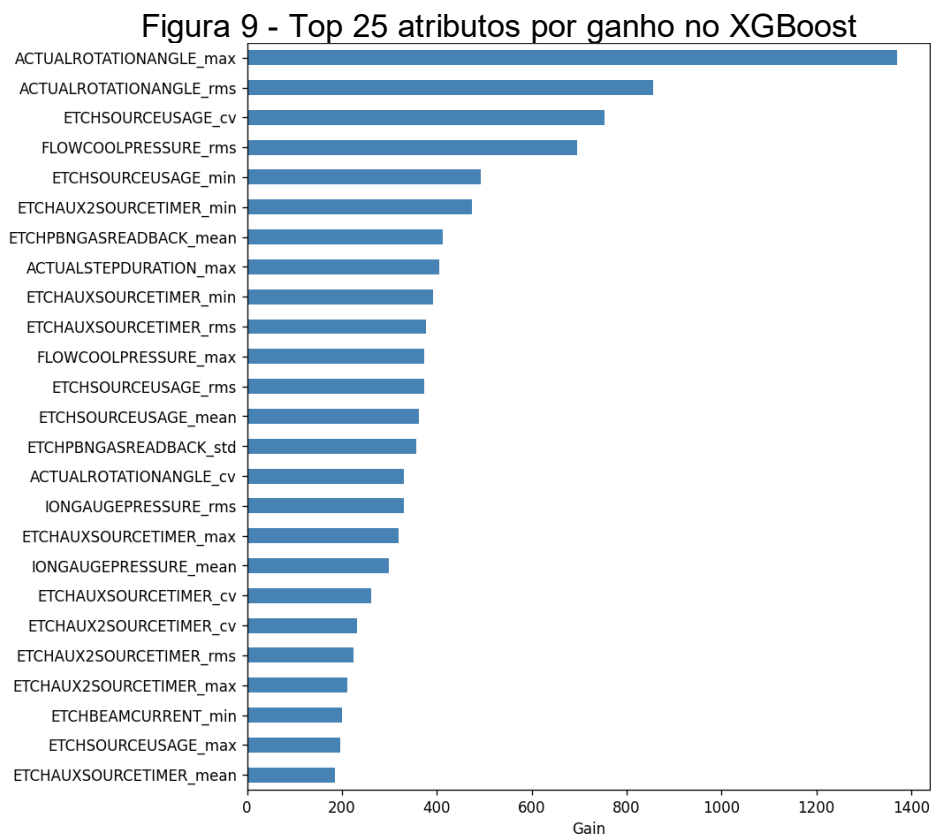
4.2.1.1 Engenharia de Atributos (*Feature Engineering*)

Para cada janela de 500 registros (~33 minutos), foram calculadas 10 estatísticas por sensor: média, desvio padrão, RMS, assimetria (*skewness*), curtose, inclinação linear (*slope*), mínimo, máximo, coeficiente de variação e autocorrelação *lag-1*, que mede a correlação do sinal com sua observação imediatamente anterior, totalizando 165 atributos iniciais.

A inclusão de curtose e autocorrelação *lag-1*, realizada durante a revisão do *pipeline*, merece discussão. A curtose captura a presença de caudas pesadas na distribuição do sinal, característica de eventos anômalos esporádicos que precede

falhas em sistemas mecânicos. A autocorrelação *lag-1* mede a persistência temporal do sinal: em operação normal, sensores tendem a variar de forma suave e correlacionada, na aproximação de uma falha, essa regularidade decresce. Ambas as estatísticas contribuíram para a melhora do AUC-ROC do classificador.

Após o cálculo dos atributos, aplicou-se um corte de acordo com a importância do atributo: atributos abaixo do percentil 25 de ganho foram removidas e o modelo foi retreinado, resultando em 47–49 atributos finais, a Figura 9 mostra os 25 atributos com maior ganho. Esse processo reduz o risco de ajuste excessivo por eliminação de atributos de baixa relevância e reduz o custo de inferência.



Fonte – Autor (2026)

O ranking de importância revela que os dois atributos de maior ganho são ACTUALROTATIONANGLE_max e ACTUALROTATIONANGLE_rms, seguidas por múltiplas estatísticas dos três odômetros (ETCHSOURCEUSAGE, ETCHAUXSOURCETIMER e ETCHAUX2SOURCETIMER), que juntos ocupam 13 das 24 posições no top 25. O sensor diretamente ligado à Falha F1, FLOWCOOLPRESSURE, aparece apenas nas posições 4 e 11 (rms e max). Esse padrão indica que o classificador detecta a degradação primariamente por meio de

indicadores indiretos de desgaste acumulado, o tempo de operação registrado pelos odômetros e as variações no padrão de rotação do *fixture*, e não pelo sinal direto de pressão do *FlowCool*.

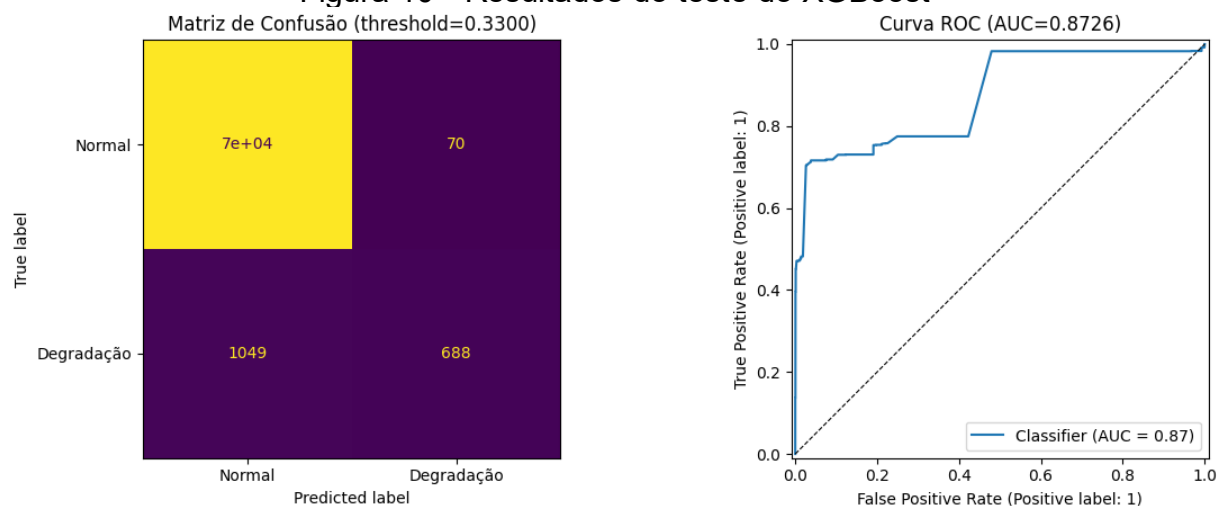
4.2.1.2 Estratégia de Divisão, Limiares e Resultados

A escolha da estratégia de divisão revelou-se determinante para a qualidade do classificador. A abordagem inicial, divisão temporal por *cycle_id* é a abordagem mais intuitiva, porém, com apenas 22 ciclos disponíveis, essa estratégia apresenta alta variância: avaliando o mesmo modelo treinado em 10 sorteios aleatórios distintos de ciclos de teste, o AUC-ROC oscilou entre 0,52 e 0,90. Por isso, fica difícil garantir uma divisão equilibrada sem escolher a semente com base no resultado do teste.

A solução adotada foi a divisão por *runnum*: os *runnums* foram distribuídos aleatoriamente entre treino (65%), validação (15%) e teste (20%), de modo que ciclos vizinhos no tempo apareçam em splits diferentes. As janelas deslizantes continuam sendo calculadas por *cycle_id* completo, sem cruzar limites de ciclo, preservando a integridade temporal local. Para lidar com o desbalanceamento extremo entre as classes, os exemplos normais foram subamostrados na razão 10:1 em relação aos exemplos de degradação durante o treinamento.

Essa estratégia, porém, introduz uma limitação: alguns ciclos ficam com 0% ou 100% de amostras de degradação no conjunto de validação, conforme os *runnums* próximos à falha foram alocados. Como mostra a Figura 10, o modelo atingiu AUC-ROC de 0,8726 no teste, evidenciando boa capacidade discriminativa.

Figura 10 - Resultados do teste do XGBoost



Fonte – Autor (2026)

O classificador atingiu *recall* de 39,6% e Precisão de 90,8% para degradação no teste, com base na matriz de confusão mostrada na Figura 10. O critério de seleção foi maximizar o *recall* sem baixar muito a precisão, ou seja, o objetivo era capturar o maior número possível de segmentos de degradação para alimentar o LSTM, aceitando falsos alarmes dentro de um limite tolerável. A precisão elevada de 90,8% não é um objetivo deliberado, mas uma consequência do desbalanceamento extremo da classe positiva (1.737 amostras de degradação contra 69.627 normais). O *recall* de 39,6% implica que aproximadamente 60% dos eventos de degradação não são detectados, uma limitação estrutural discutida nas próximas seções.

4.2.2 Preparação das Sequências para o LSTM

Com o classificador XGBoost treinado, o segundo estágio do método requer a transformação dos dados brutos em sequências adequadas para o treinamento do LSTM. Esse processo não é trivial: os dados contêm saltos temporais, períodos de ociosidade e trechos sem dados de degradação que, se incluídos diretamente, degradariam o aprendizado do modelo de RUL. A preparação envolve uma cadeia de filtros sucessivos, cada um com justificativa na literatura ou nas características específicas da sessão 01_M02, seguida pela definição do objetivo e da normalização. O resultado é um conjunto de sequências contínuas e fisicamente coerentes, prontas para o treinamento do LSTM.

4.2.2.1 Processo de Filtragem

O processo de preparação dos dados para o LSTM aplica quatro filtros a cada segmentos operacionais, motivados pela literatura e pelas características específicas do banco de dados. Dos 1.703 segmentos operacionais originalmente formados para a organização das séries temporais, 88 apresentaram observações válidas da Falha F1 dentro do horizonte de RUL adotado ($TTF\ F1 \leq 15.000\ s$).

O primeiro filtro aplicado consistiu em manter apenas os registros com *shutter* ativo ($SHUTTER_1 = 1$), configuração utilizada por Wu *et al.* (2021), Darwish (2024) e TV *et al.* (2018). Esse filtro remove períodos em que a máquina não está processando *wafers* ativamente, estados ociosos onde os sensores não refletem o desgaste operacional. O filtro eliminou 28,2% dos registros da zona de degradação.

Em seguida, foram analisados se, dentro de cada ciclo, havia saltos superiores a 1 hora na série temporal filtrada, indicando pausas de manutenção preventiva ou troca de turno (eventos que não representam progressão da degradação). Nesses casos, apenas o segmento final (o mais próximo da falha) é retido, garantindo que a trajetória de degradação seja contínua e fisicamente coerente.

Para o terceiro filtro, o XGBoost é aplicado aos segmentos remanescentes para detectar o ponto de aviso. Segmentos com menos de 500 amostras FSP=1 são descartados antes mesmo da aplicação do modelo, volume insuficiente para calcular sequer uma janela de atributos. Para os demais, a probabilidade de degradação é suavizada por média móvel de 20 previsões e o LSTM recebe apenas os dados a partir do primeiro instante em que essa probabilidade suavizada supera o limite do classificador (CLF) de 0,10. Essa abordagem foca o modelo de RUL na zona onde a degradação já está confirmada, evitando que o LSTM receba dados de operação normal onde o valor de RUL é alto e pouco discriminativo.

Por último, segmentos com menos de 90 amostras FSP=1 após o ponto de aviso são descartados, o mínimo necessário para pelo menos duas janelas deslizantes completas de 45 amostras.

O resultado do processo de filtragem é sintetizado na Tabela 7 abaixo:

Tabela 7 - Resultado das etapas de filtragem

Etapas	Descartados	Restantes
Segmentos com dados da Falha F1 disponíveis	—	88
Muito curto para extrair atributos (< 500 amostras FSP=1)	13	75
XGBoost nunca atingiu limite de 0,10	26	49
Curto após ponto de aviso (< 90 amostras FSP=1)	1	48
Segmentos válidos para o LSTM	—	48

Fonte – Autor (2026)

Os 13 segmentos descartados por volume insuficiente nunca chegaram a ser avaliados pelo XGBoost trata-se de uma limitação de dados, não do classificador. Dos 75 segmentos efetivamente avaliados, o XGBoost detectou o ponto de aviso em 49 (65,3%). Nos 26 segmentos restantes, a probabilidade suavizada não superou o limiar de 0,10 ao longo dos registros disponíveis, impedindo a definição de um ponto de aviso e, conseqüentemente, a geração de uma sequência para a etapa de regressão. Esses casos foram considerados falsos negativos do classificador, pois representavam segmentos com degradação real que não foram encaminhados ao

LSTM. Por fim, após a aplicação do quarto filtro, permaneceram 48 segmentos válidos para o modelo LSTM. Sendo 80% treino (38 segmentos), 8% para validação (4 segmentos) e 12% para teste (6 segmentos)

4.2.2.2 *Limite CLF e definição da variável-alvo de RUL*

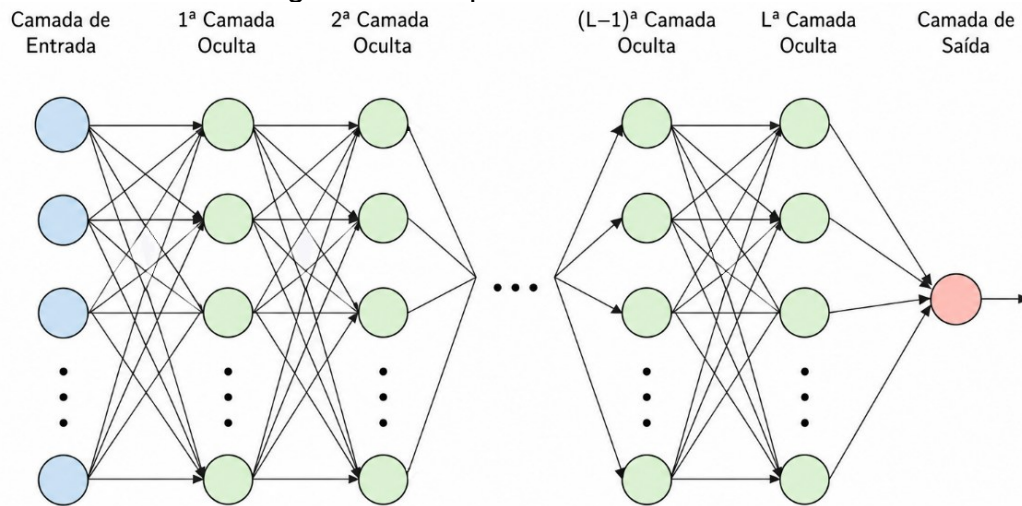
A escolha do limite CLF foi investigada empiricamente. Com limite CLF = 0,20, o número de segmentos válidos para o LSTM caiu para menos de 40, produzindo RMSE piores. Com limite CLF = 0,10, foram obtidos 48 segmentos válidos, um ganho significativo dado o volume total disponível. Essa relação revela um *trade-off*, um limite mais alto produz avisos mais confiáveis, mas descarta segmentos onde a probabilidade nunca alcança esse patamar. Com apenas 88 segmentos disponíveis na sessão 01_M02, o volume de treino supera a confiabilidade do aviso, mais segmentos com avisos menos precisos resultam em melhor generalização do LSTM do que menos segmentos com avisos mais tardios e seguros.

Já a variável-alvo de RUL não foi definida em segundos de tempo real, mas como contagem regressiva de amostras FSP=1 restantes a partir do ponto de aviso. Essa escolha elimina a dependência do intervalo de amostragem uniforme, o banco de dados PHM 2018 não garante regularidade temporal estrita, e converter amostras para segundos via um fator fixo (4 s/amostra) introduziria erros acumulados. Para o treinamento dos modelos recorrentes, essa contagem foi normalizada pelo maior valor observado no conjunto de treino, igual a 3.040 amostras. Essa transformação teve como objetivo padronizar a escala da variável-alvo e melhorar a estabilidade numérica do processo de otimização. A normalização foi ajustada exclusivamente no conjunto de treino, e o mesmo fator foi aplicado aos conjuntos de validação e teste sem reajuste, prevenindo vazamento de informação.

4.2.2.3 *Arquitetura e Configuração de Treinamento*

Foram implementadas e avaliadas duas arquiteturas recorrentes, conforme descrito na Seção 3.2.2, ambas podem ser representadas pela Figura 11 com pequenas diferenças.

Figura 11 - Arquitetura dos modelos



Fonte – Autor (2026)

Tanto o LSTM quanto o GRU contêm um tensor de entrada no formato de 45 unidades de tempo por 21 atributos, seguido por 2 camadas ocultas com 64 unidades no caso do LSTM e 3 camadas ocultas no caso do GRU. Por fim, uma camada de saída de 1 unidade. O tamanho da janela foi definido em 45 unidades de tempo, buscando equilibrar a captura de dependências temporais relevantes e o custo computacional do treinamento.

A escolha da arquitetura LSTM segue diretamente Wu *et al.* (2021), que adotam a mesma configuração como referência para o problema. O GRU foi incluído para replicar a comparação presente no artigo de referência. O número de unidades (64) foi mantido fixo. Com apenas 38 segmentos de treino, aumentar a capacidade do modelo agravaria o ajuste excessivo sem fornecer dados suficientes para explorar a capacidade adicional.

O treinamento utilizou o otimizador RMSprop com taxa de aprendizado de 0,001 e função de perda baseado no erro quadrático médio (*Mean Squared Error – MSE*). A janela deslizante de 45 amostras e passo 1 gerou aproximadamente 45.000 sequências de treino.

4.3 PREVISÃO E ANÁLISE DA RUL

Com os modelos treinados, o estágio seguinte do método consiste na avaliação do desempenho das arquiteturas propostas para a estimativa da RUL. São analisadas as duas abordagens mencionadas anteriormente, LSTM e GRU. A análise

dos resultados é organizada em quatro níveis de detalhe: métricas globais no conjunto de teste, desempenho segmentado por faixa de RUL, inspeção visual das previsões por ciclo e impacto dos ajustes de pós-processamento. Esse escalonamento permite evidenciar padrões de comportamento que não são capturados por métricas agregadas isoladamente.

4.3.1 Resultados e Análise

A avaliação inicial do desempenho dos modelos foi realizada por meio de métricas globais de erro no conjunto de teste, permitindo uma comparação direta entre as arquiteturas LSTM e GRU. A Tabela 8 apresenta os valores de RMSE e SMAPE obtidos por cada modelo e comparações com a literatura, fornecendo uma visão geral da acurácia das previsões de RUL.

A comparação com a literatura deve ser interpretada com cautela, pois os trabalhos relacionados diferem quanto ao recorte do banco de dados, à definição dos limiares de degradação, à forma de construção da variável-alvo e ao conjunto de atributos utilizado como entrada. Conforme discutido na Tabela 4, há diferenças relevantes até mesmo na definição das zonas normal e de degradação entre Huang et al. (2018), Wu et al. (2021), Darwish (2024) e Zhao, Zhu e Zhao (2022). Dessa forma, as métricas apresentadas na Tabela 8 servem como referência comparativa, mas não constituem uma comparação controlada sob as mesmas condições experimentais.

Tabela 8 - Comparação de desempenho com a literatura

Modelo	Referência	RMSE (s)	SMAPE
DW-LSTM	Wu <i>et al.</i> (2021)	1849	23,56%
DW-GRU	Wu <i>et al.</i> (2021)	686	14,13%
LSTM	Huang <i>et al.</i> (2024)	1877	13,90%
LSTM (este trabalho)	Autor (2026)	1.677	51,56%
GRU (este trabalho)	Autor (2026)	1.970	56,61%

Fonte – Autor (2026)

O modelo LSTM obteve RMSE de 1.677 s e SMAPE de 51,56% no conjunto de teste. Já o GRU apresentou um RMSE de 1.970 s e SMAPE de 56,61%. O resultado superior do LSTM para este banco de dados específico diverge dos resultados encontrados por Wu *et al.* (2021), nos quais a arquitetura DW-GRU

(*Degradation Warning* GRU), baseada em um mecanismo prévio de detecção do início da degradação para ativação da previsão de RUL, superou o DW-LSTM.

O RMSE de 1.677 s supera o de Wu *et al.* (2021) DW-LSTM (1.849 s) e o Huang *et al.* (2024) LSTM (1.877 s), demonstrando que a implementação é competitiva em erro absoluto. O SMAPE global de 51,56%, porém, é substancialmente maior que os reportados na literatura (13,9–23,56%), indicando que, embora o modelo apresente bom desempenho em termos absolutos, há limitações na precisão relativa das estimativas ao longo de diferentes escalas de RUL.

Essa aparente inconsistência entre RMSE e SMAPE sugere que o erro do modelo não é uniforme ao longo do horizonte de previsão. Para investigar esse comportamento, o conjunto de teste foi segmentado em três faixas de RUL real, permitindo analisar o desempenho do modelo em cenários com distintos níveis de proximidade da falha. Os resultados dessa análise segmentada são apresentados na Tabela 9.

O limite superior da faixa distante (6.000 s) corresponde aproximadamente a 50% do valor máximo de RUL, separando a faixa em que o modelo opera com sinal claro de urgência daquela em que o horizonte é longo demais para predições precisas. O limite inferior de 2.000 s corresponde ao limite de degradação adotado por Wu *et al.* (2021) (500 amostras $\times$ 4 s), delimitando a faixa crítica próxima à falha. A Figura 12 (LSTM) e a Figura 13 (GRU) ilustram exemplos representativos das predições em cada faixa.

Tabela 9 - Resultados por faixa de RUL

Modelo	Faixa	RMSE (s)	SMAPE
LSTM	Distante (> 6.000 s)	2.866	54,27%
	Médio (2.000–6.000 s)	977	20,78%
	Próximo (< 2.000 s)	1.962	97,63%
GRU	Distante (> 6.000 s)	3.667	73,88%
	Médio (2.000–6.000 s)	1.026	20,30%
	Próximo (< 2.000 s)	2.239	106,28%

Fonte – Autor (2026)

Faixa distante (RUL > 6.000 s): O modelo apresenta predições aproximadamente constantes, subestimando sistematicamente o tempo restante e resultando no maior RMSE entre as faixas. Esse comportamento está associado à distribuição do conjunto de treino, na qual valores elevados de RUL são pouco

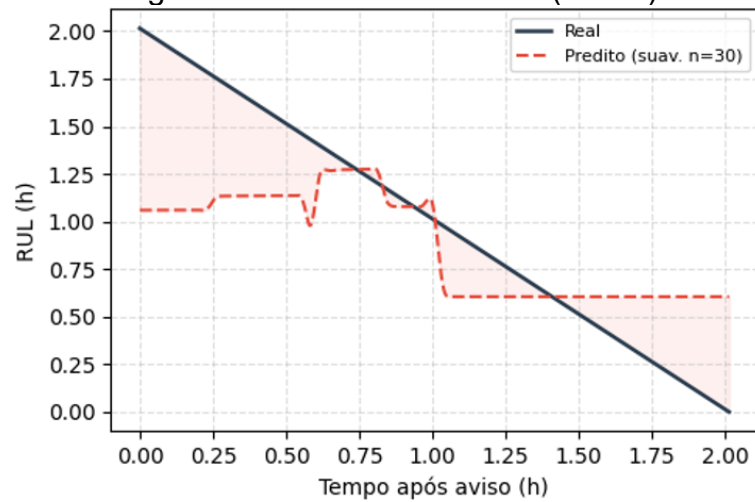
representados. Como consequência, o modelo tende a convergir para estimativas médias quando não há sinal claro de degradação.

Faixa média (2.000 s – 6.000 s): Nesta faixa, observa-se o melhor desempenho do modelo, com SMAPE de 20,78% para o LSTM e 20,30% para o GRU, valor próximo ao reportado por Wu *et al.* (2021) (23,56%). Isso sugere que o modelo opera de forma mais adequada em um regime intermediário de degradação, onde há sinal informativo suficiente e distribuição mais equilibrada de exemplos.

Faixa próxima ($RUL < 2.000$ s): O SMAPE de 97,63% evidencia a presença de um piso de previsão, o LSTM estabiliza suas estimativas em torno de 2.160 s, enquanto o GRU converge para aproximadamente 2.700 s, sem acompanhar a queda da RUL real até zero. Esse comportamento é consistente com resultados reportados por Huang *et al.* (2024) e Wu *et al.* (2021), que também observam limites inferiores semelhantes (~ 2.500 s). A causa está na escassez de exemplos nessa região, já que cada ciclo contribui com poucas amostras próximas à falha.

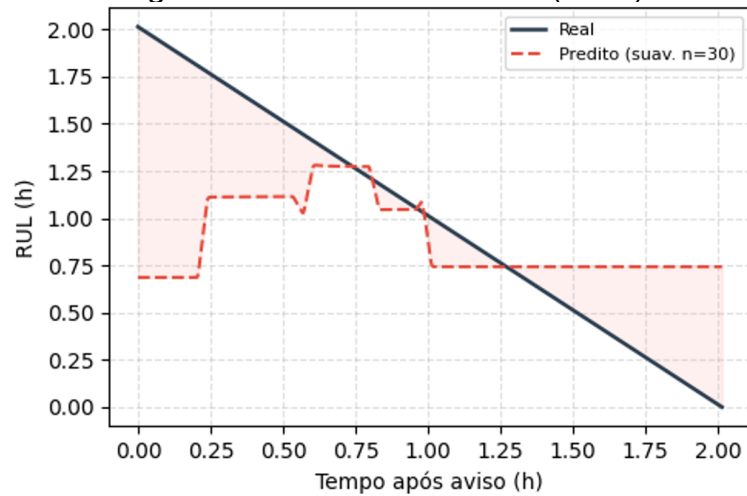
A inspeção visual das predições por ciclo revela padrões que complementam a análise quantitativa. Nas Figura 12 e Figura 13 é possível visualizar que o modelo não acompanha a trajetória real da RUL, manifestando esse comportamento de formas distintas conforme o ciclo. Nos segmentos com RUL inicial longo, as predições oscilam em um intervalo relativamente fixo em vez de decrescer acompanhando a RUL real. Em alguns casos, essas oscilações não apresentam correspondência clara com padrões dos sensores, indicando possível extrapolação fora da distribuição de treino. Em segmentos mais curtos, que iniciam diretamente na faixa crítica, observa-se o comportamento oposto: as predições permanecem sistematicamente acima do valor real. Em ambos os casos, a limitação decorre da baixa diversidade de trajetórias presentes no conjunto de treino.

Figura 12 - Previsão aleatória (LSTM)



Fonte – Autor (2026)

Figura 13 - Previsão aleatória (GRU)



Fonte – Autor (2026)

A análise por faixa permite ainda identificar uma janela ótima de previsão, definida como o intervalo contínuo de RUL com menor SMAPE agregado, com largura mínima de 40 minutos. Os resultados são apresentados na Tabela 10.

Tabela 10 - Desempenho por tipo de intervalo

Modelo	Intervalo	RMSE (s)	SMAPE
LSTM	Global	1.677	51,60%
	Janela ótima (41-83 min)	647	16,50%
GRU	Global	1.970	56,60%
	Janela ótima (41-83 min)	603	13,50%

Fonte – Autor (2026)

Na janela ótima (41–83 minutos), o desempenho melhora significativamente. O SMAPE do LSTM reduz-se de 51,6% para 16,50%, tornando-se competitivo com os valores reportados por Wu *et al.* (2021) (23,6%) e Huang *et al.* (2024) (13,90%). Para o GRU, o SMAPE cai de 56,6% para 13,50%, superando o resultado de Wu *et al.* (2021) (14,13%). Esses resultados indicam que a discrepância observada no SMAPE global não decorre de limitações intrínsecas dos modelos, mas sim do escopo da avaliação. O sistema apresenta maior confiabilidade em um intervalo específico antes da falha, potencialmente relevante para o planejamento de manutenção, cuja aplicabilidade depende das restrições operacionais do ambiente considerado.

4.4 ANÁLISE SHAP

Avaliar o desempenho preditivo do modelo é necessário, mas não suficiente para compreender sua utilidade prática. Um modelo com RMSE competitivo pode estar aprendendo padrões distintos: ora capturando efetivamente a degradação física do sistema, ora explorando correlações indiretas associadas ao tempo de operação ou a características específicas do conjunto de dados. Nesse contexto, esta seção investiga o comportamento interno do modelo final por meio da análise SHAP, com o objetivo de identificar quais atributos e quais instantes temporais exercem maior influência nas previsões de RUL.

4.4.1 Resultados da Análise SHAP

Para essa análise, foi utilizado o *GradientExplainer* da biblioteca SHAP, aplicado ao modelo final. Foram calculados valores SHAP para 500 amostras do conjunto de teste, quantidade considerada suficiente para estimativas estáveis de importância global, mantendo viabilidade computacional. Cada valor SHAP representa, em segundos de RUL, a contribuição de uma variável específica em um determinado instante da janela temporal (45 passos) para a previsão da amostra. A Tabela 11 apresenta os 15 sensores com maior contribuição média absoluta.

Tabela 11 - Contribuição dos sensores

Rank	Sensor	SHAP médio (s)
1	ETCHSOURCEUSAGE	17,3
2	ACTUALSTEPDURATION	12,4
3	ETCHAUXSOURCETIMER	12,2
4	ETCHSUPPRESSORVOLTAGE	10,4
5	ETCHBEAMCURRENT	8,7
6	FIXTURETILTANGLE	8,1
7	ACTUALROTATIONANGLE	8,0
8	ETCHSUPPRESSORCURRENT	7,2
9	ETCHBEAMVOLTAGE	6,0
10	ETCHAUX2SOURCETIMER	4,0
11	FLOWCOOLFLOWRATE	1,6
12	ETCHGASCHANNEL1READBACK	1,5
13	FLOWCOOLPRESSURE	1,0
14	ETCHPBNGASREADBACK	0,0
15	IONGAUGEPRESSURE	0,0

Fonte – Autor (2026)

Os resultados indicam que os três atributos de maior influência, ETCHSOURCEUSAGE, ACTUALSTEPDURATION e ETCHAUXSOURCETIMER, são variáveis relacionadas ao tempo de operação. O primeiro e o terceiro correspondem a contadores acumulativos de uso de componentes do sistema de processamento, enquanto o sensor ACTUALSTEPDURATION representa a duração efetiva de uma etapa de processo. Em conjunto, essas variáveis refletem principalmente o histórico de utilização da ferramenta, e não necessariamente o estado físico instantâneo do sistema.

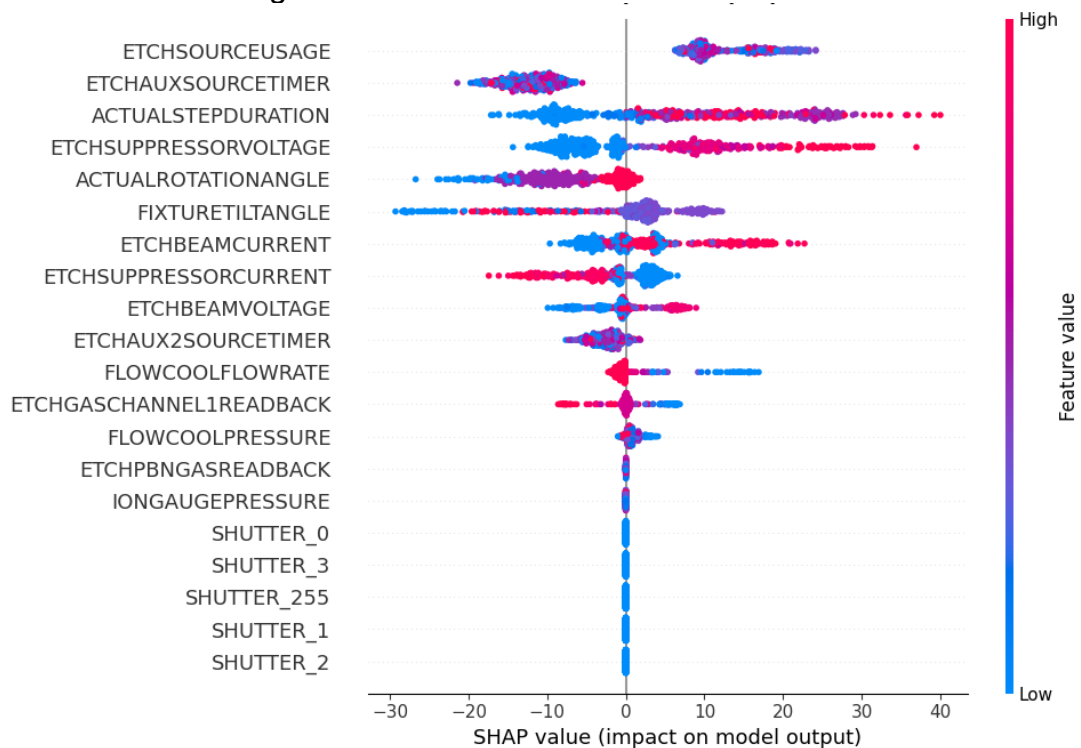
Os sensores classificados entre as posições 4 e 9, como ETCHSUPPRESSORVOLTAGE, ETCHBEAMCURRENT, sensor ETCHSUPPRESSORCURRENT e ETCHBEAMVOLTAGE, estão associados ao comportamento do feixe iônico e às condições de operação da câmara. Embora representem grandezas físicas relevantes, esses parâmetros também tendem a evoluir ao longo do tempo de uso, refletindo indiretamente o desgaste progressivo do sistema.

Em contraste, o sensor FLOWCOOLPRESSURE, diretamente relacionado à Falha F1, aparece apenas na 13ª posição, com contribuição média reduzida. FLOWCOOLFLOWRATE, que também atua no circuito de resfriamento, apresenta importância ligeiramente superior, mas ainda limitada. Em conjunto, ambos

representam uma fração pequena da importância total, aproximadamente 2,6%. Esse resultado indica que o modelo atribui baixa relevância aos sinais diretamente associados ao mecanismo físico da falha, mesmo sendo esses os responsáveis pelo evento crítico analisado.

Embora o ranking global forneça uma visão agregada da importância das variáveis, ele não captura a direção e a variabilidade das contribuições individuais. Para isso, utiliza-se o gráfico de enxame de abelhas (*beeswarm plot*) mostrado na Figura 14, que permite analisar simultaneamente a magnitude, a direção e a distribuição dos valores SHAP. No gráfico, cada ponto representa uma amostra do conjunto de teste, posicionada horizontalmente conforme sua contribuição SHAP (valores positivos aumentam a estimativa de RUL, enquanto negativos a reduzem) e colorida de acordo com o valor da variável (azul para valores baixos e vermelho para altos).

Figura 14 - Gráfico de enxame de abelhas



Fonte – Autor (2026)

Um padrão relevante observado envolve os atributos ETCHSOURCEUSAGE e ETCHAUXSOURCETIMER, ambos contadores acumulativos de uso. Apesar de representarem grandezas similares, suas contribuições apresentam direções opostas: enquanto valores elevados de ETCHSOURCEUSAGE estão associados a

contribuições positivas para a RUL, valores baixos de ETCHAUXSOURCETIMER tendem a reduzir significativamente a estimativa. Essa assimetria constitui um indício de que o modelo pode estar capturando relações específicas ao histórico operacional do conjunto de dados analisado, em vez de padrões físicos generalizáveis.

O sensor ACTUALSTEPDURATION também apresenta contribuições predominantemente positivas para valores elevados, com alta dispersão, sugerindo associação entre etapas mais longas de processamento e maior RUL estimada. Já variáveis como FIXTURETILTANGLE exibem comportamento assimétrico, com valores baixos associados a reduções significativas na estimativa de RUL, indicando possível correlação com estados operacionais próximos à falha.

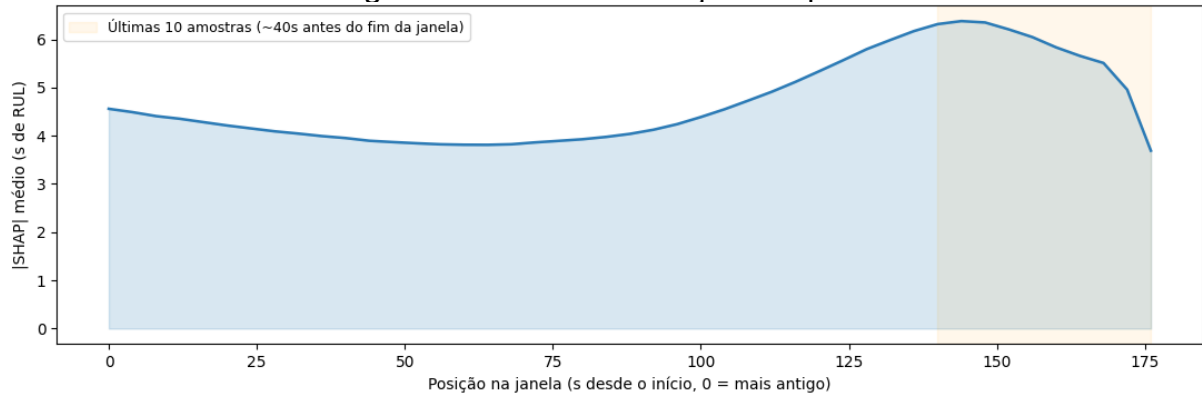
As variáveis referentes ao sensor ACTUALROTATIONANGLE e ETCHSUPPRESSORVOLTAGE, por sua vez, embora apresentem importância média moderada, exibem pontos isolados com contribuições SHAP superiores a 30 s. Esse comportamento evidencia a ocorrência de impactos individuais elevados em determinadas amostras, mesmo quando a contribuição média global é limitada. Tal efeito não é capturado por métodos baseados exclusivamente em importância média, como o ganho do XGBoost.

Por outro lado, os sensores FLOWCOOLPRESSURE e FLOWCOOLFLOWRATE apresentam distribuições fortemente concentradas em torno de zero, independentemente de seus valores. Mesmo em situações em que a pressão do sistema de resfriamento se encontra em níveis críticos, suas contribuições para a previsão permanecem reduzidas. Isso evidencia baixa sensibilidade do modelo a esses sinais, sugerindo dificuldade em capturar diretamente o mecanismo físico associado à Falha F1. Esse resultado é consistente com a EDA, na qual se observou que o comportamento do sensor FLOWCOOLPRESSURE não apresenta um padrão temporal claro de degradação ao longo dos segmentos, sendo caracterizado por elevada variabilidade e baixa capacidade discriminativa. Consequentemente, o modelo tende a priorizar variáveis com comportamento mais estável e informativo ao longo do tempo.

Ao analisar a atenção temporal (SHAP médio por passo temporal da janela) mostrado na Figura 15, é possível ver que a contribuição média dos instantes ao longo da janela é relativamente uniforme na maior parte da sequência, aumentando gradualmente até um pico próximo aos instantes finais, seguido de uma queda abrupta. Esse comportamento sugere que o modelo prioriza padrões de médio prazo

ao longo da janela, em vez de reagir predominantemente aos valores mais recentes dos sensores. Esse resultado reforça a evidência de que o modelo está mais orientado a capturar tendências associadas ao histórico de operação do que variações abruptas próximas à falha.

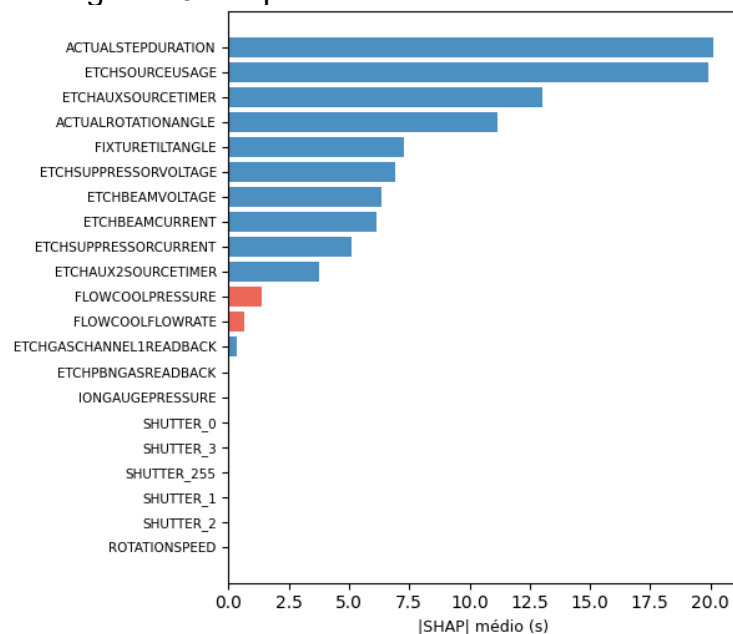
Figura 15 - Perfil de atenção temporal



Fonte – Autor (2026)

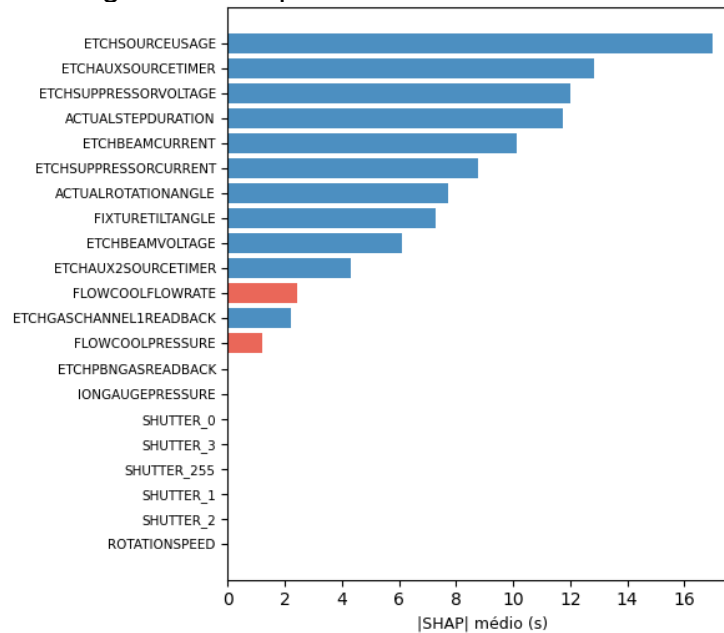
Já a análise segmentada por faixas de RUL (Figura 16-Figura 18) indica que esse comportamento se mantém ao longo de diferentes horizontes de previsão. Mesmo em situações próximas à falha, a contribuição conjunta dos sensores do sistema *FlowCool* permanece reduzida. Isso sugere que o modelo não incorpora de forma consistente o estado crítico do sistema de resfriamento, o que pode impactar sua capacidade de generalização em cenários nos quais o padrão de degradação difere do observado no conjunto de treinamento.

Figura 16 - Importância SHAP faixa distante



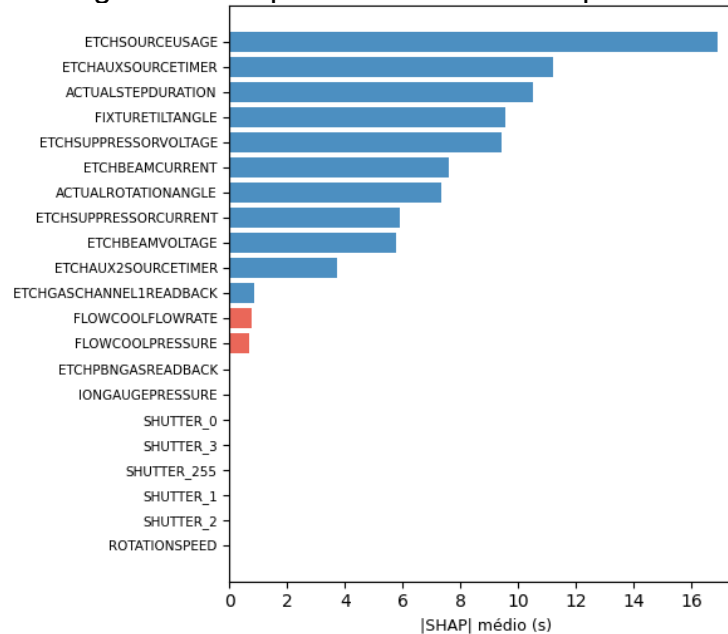
Fonte – Autor (2026)

Figura 17 - Importância SHAP faixa média



Fonte – Autor (2026)

Figura 18 - Importância SHAP faixa próxima



Fonte – Autor (2026)

Por fim, os resultados da análise SHAP apontam para uma limitação estrutural relacionada ao conjunto de dados utilizado. Com aproximadamente 38 segmentos de treino, a quantidade de exemplos disponíveis pode ser insuficiente para que o modelo aprenda padrões de degradação mais complexos ou menos frequentes. Além disso, conforme indicado na análise exploratória, o comportamento de variáveis diretamente associadas ao mecanismo de falha, como o sensor FLOWCOOLPRESSURE, é

caracterizado por elevada variabilidade e ausência de um padrão temporal consistente ao longo dos segmentos, o que limita sua capacidade preditiva ao longo da maior parte do horizonte de operação.

Diante desse cenário, o modelo tende a atribuir maior relevância a variáveis que apresentam variação mais estável e informativa ao longo do tempo, como os contadores acumulativos de uso. Esse comportamento indica uma estratégia de previsão predominantemente baseada no histórico de operação, análoga a abordagens de manutenção preventiva baseadas em tempo, limitando sua aplicação em cenários que exigem monitoramento baseado em condição.

Essa análise evidencia o papel fundamental das técnicas de XAI não apenas na interpretação das predições, mas também na identificação de limitações inerentes aos dados e ao processo de aprendizado, fornecendo subsídios para o aprimoramento de modelos e para o direcionamento de estratégias mais eficazes de coleta e enriquecimento de dados em aplicações de prognóstico de RUL.

5 CONCLUSÃO E TRABALHOS FUTUROS

A presente pesquisa teve como objetivo avaliar a eficácia de uma abordagem de prognóstico de RUL baseada em ML, aplicada a dados reais de sensores industriais no contexto de PdM. Para isso, foi proposta uma arquitetura em dois estágios, composta por um modelo de classificação para detecção do início da degradação e um modelo de regressão baseado em RNN para estimativa da RUL.

Os resultados indicaram que a abordagem proposta apresentou desempenho razoável em termos de erro absoluto, com o modelo LSTM superando o GRU no conjunto avaliado. No entanto, essa diferença deve ser interpretada com cautela, pois não constitui evidência conclusiva de que o modelo tenha aprendido de forma robusta a dinâmica física de degradação do equipamento. A análise por faixas de RUL revelou que o desempenho não foi uniforme ao longo do horizonte de previsão, com limitações tanto em regiões distantes da falha quanto nos instantes mais próximos ao evento crítico. Além disso, a concentração dos dados em determinadas faixas de RUL pode ter favorecido previsões próximas a valores médios, de modo que parte do desempenho observado pode estar associada à distribuição da variável-alvo, e não exclusivamente à identificação de padrões temporais de degradação.

A análise de interpretabilidade baseada em SHAP permitiu aprofundar a compreensão do comportamento do modelo, evidenciando que as previsões são predominantemente guiadas por variáveis associadas ao tempo acumulado de operação, como contadores de uso, em detrimento de sensores diretamente relacionados ao mecanismo físico da falha. Esse achado foi consistente com a análise exploratória dos dados, que indicou ausência de um padrão claro de degradação progressiva no sensor crítico de pressão do sistema *FlowCool*. Como consequência, o modelo tende a adotar uma estratégia implícita de previsão baseada no histórico de operação, em vez de responder diretamente ao estado físico instantâneo do sistema. Além disso, a análise permitiu compreender os fatores que influenciam as decisões do modelo.

Os resultados evidenciam uma limitação estrutural importante: a capacidade do modelo de aprender padrões fisicamente significativos está diretamente condicionada à qualidade e à natureza dos dados disponíveis. Em cenários onde o sinal de degradação é abrupto ou pouco representado nos dados de treinamento,

modelos baseados em aprendizado supervisionado tendem a explorar correlações mais estáveis, ainda que menos representativas do fenômeno físico de interesse.

Nesse contexto, o uso de técnicas de XAI mostrou-se fundamental não apenas para a interpretação das predições, mas também como ferramenta diagnóstica do modelo. A análise SHAP permitiu identificar limitações que não seriam evidentes apenas por métricas de desempenho, contribuindo para uma avaliação mais completa da confiabilidade e da aplicabilidade do sistema em contextos industriais.

Dessa forma, os resultados deste trabalho demonstram que a combinação entre modelagem preditiva e técnicas de interpretabilidade não apenas viabiliza a estimativa da RUL, mas também fornece subsídios consistentes para a tomada de decisão em PdM, atendendo ao objetivo proposto de integrar desempenho e explicabilidade em um mesmo *pipeline*.

Como trabalhos futuros, destacam-se algumas direções relevantes. Primeiramente, a ampliação do conjunto de dados, especialmente com maior número de ciclos de falha, pode permitir ao modelo capturar padrões mais robustos de degradação física. Adicionalmente, a incorporação de novas variáveis derivadas ou sensores mais diretamente relacionados ao mecanismo de falha pode melhorar a capacidade preditiva baseada em condição. Outras possibilidades consistem na exploração de arquiteturas híbridas que integrem conhecimento físico ao modelo de ML, reduzindo a dependência exclusiva de padrões estatísticos, e na otimização dos parâmetros do modelo.

Do ponto de vista metodológico, futuras investigações podem explorar estratégias alternativas de modelagem da RUL, como abordagens probabilísticas ou modelos capazes de estimar incerteza, aumentando a confiabilidade das previsões. Por fim, a integração sistemática de técnicas de interpretabilidade ao ciclo de desenvolvimento de modelos configura-se como um passo fundamental para viabilizar a adoção prática de soluções de PdM baseadas em IA.

REFERÊNCIAS

- ACHOUCH, M. *et al.* On Predictive Maintenance in Industry 4.0: Overview, Models, and Challenges. **Applied Sciences**, v. 12, n. 16, p. 8081, 12 ago. 2022.
- AHMAD, K. *et al.* Developing future human-centered smart cities: Critical analysis of smart city security, Data management, and Ethical challenges. **Computer Science Review**, v. 43, p. 100452, fev. 2022.
- AKRAM, M.; EL, C. Sequence to Sequence Weather Forecasting with Long Short-Term Memory Recurrent Neural Networks. **International Journal of Computer Applications**, v. 143, n. 11, p. 7–11, 17 jun. 2016.
- AKTAS, E.; KAGNICIOGLU, C. H. Factors affecting safety behaviors of aircraft maintenance technicians: A study on Civil Aviation Industry in Turkey. **Safety Science**, v. 164, p. 106146, ago. 2023.
- ALAOUI, A. H.; MEDDAOUI, A.; HAIN, M. **AI-Driven Predictive Maintenance for Industry 4.0: A Systematic Review of Models, Methods, and Challenges**, 30 set. 2025.
- ALLAL, Z. *et al.* Machine learning for fuel cell remaining useful life prediction: A review. **Energy Conversion and Management: X**, v. 30, p. 101597, maio 2026.
- ALZUBI, J.; NAYYAR, A.; KUMAR, A. Machine Learning from Theory to Algorithms: An Overview. **Journal of Physics: Conference Series**, v. 1142, p. 012012, nov. 2018.
- APEIRANTHITIS, S. *et al.* From Corrective and Preventive Maintenance to Prognostics. **International Journal of Prognostics and Health Management**, v. 17, n. 1, 5 abr. 2026.
- BAPTISTA, M. L. *et al.* Using Explainable Artificial Intelligence to Interpret Remaining Useful Life Estimation with Gated Recurrent Unit. **Annual Conference of the PHM Society**, v. 16, n. 1, 5 nov. 2024.
- BARREDO ARRIETA, A. *et al.* Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. **Information Fusion**, v. 58, p. 82–115, jun. 2020.

BENGIO, Y. Learning Deep Architectures for AI. **Foundations and Trends® in Machine Learning**, v. 2, n. 1, p. 1–127, 2009.

BENNETOT, A. *et al.* A Practical Guide on Explainable Ai Techniques Applied on Biomedical Use Case Applications. **SSRN Electronic Journal**, nov. 2022.

BENTÉJAC, C.; CSÖRGŐ, A.; MARTÍNEZ-MUÑOZ, G. **A Comparative Analysis of XGBoost**, mar. 2019.

BHUSHAN, R. An Overview of Machine Learning Algorithms, Tasks, and Applications in Brief. **Journal of Systems Engineering and Electronics**, v. 35, p. <https://jseepublisher.com/volume-35>, abr. 2025.

BIEBL, F. *et al.* A conceptual model to enable prescriptive maintenance for etching equipment in semiconductor manufacturing. **Procedia CIRP**, v. 88, p. 64–69, 2020.

BIGGIO, L.; KASTANIS, I. Prognostics and Health Management of Industrial Assets: Current Progress and Road Ahead. **Frontiers in Artificial Intelligence**, v. 3, 9 nov. 2020.

BOUGACHA, O.; VARNIER, C.; ZERHOUNI, N. Review of Post-Prognostics Decision-Making in Prognostics and Health Management. **International Journal of Prognostics and Health Management**, v. 11, n. 2, 24 mar. 2021.

BOUROKBA, A.; HAMDİ, R. EL; NJAH, M. **A Shapley based XAI approach for a turbofan RUL estimation** 2024 21st International Multi-Conference on Systems, Signals & Devices (SSD). **Anais...IEEE**, 22 abr. 2024

BREIMAN, L. *et al.* **Classification And Regression Trees**. [s.l.] Routledge, 1984.

_____. Stacked regressions. **Machine Learning**, v. 24, n. 1, p. 49–64, jul. 1996.

CABRERA, C. *et al.* Machine Learning Systems: A Survey from a Data-Oriented Perspective. **ACM Computing Surveys**, v. 58, n. 5, p. 1–38, 30 abr. 2026.

CUMMINS, L. *et al.* Explainable Predictive Maintenance: A Survey of Current Methods, Challenges and Opportunities. **IEEE Access**, v. 12, p. 57574–57602, 2024.

DALZIOCHIO, J. *et al.* Machine learning and reasoning for predictive maintenance in Industry 4.0: Current status and challenges. **Computers in Industry**, v. 123, p. 103298, dez. 2020.

DANACH, K. *et al.* An explainable data-driven optimization framework for industrial predictive maintenance scheduling. **Results in Engineering**, v. 29, p. 109022, mar. 2026.

DAOUDI, N.; SMAIL, Z.; ABOUSSALEH, M. Machine Learning Based Predictive Maintenance: Review, Challenges and Workflow. *Em: [s.l: s.n.]*. p. 71–88.

DARWISH, A. A Data-driven Deep Learning Approach for Remaining Useful Life in the ion mill etching Process. **Sustainable Machine Intelligence Journal**, v. 8, 9 jun. 2024.

DIETTERICH, T. G. Ensemble Methods in Machine Learning. *Em: [s.l: s.n.]*. p. 1–15.

EL-BRAWANY, M. A. *et al.* Artificial intelligence-based data-driven prognostics in industry: A survey. **Computers & Industrial Engineering**, v. 184, p. 109605, out. 2023.

ENGBERS, H.; FREITAG, M. Automated algorithm selection for Predictive Maintenance: Advances and challenges. **Journal of Manufacturing Systems**, v. 82, p. 964–975, out. 2025.

ENSARIOGLU, K. Hierarchical RUL Prediction for Turbofan Engines Based on Health Stage Classification and Change Point-Guided Data Augmentation. **IEEE Access**, v. 13, p. 134241–134260, 2025.

ERRANDONEA, I.; BELTRÁN, S.; ARRIZABALAGA, S. Digital Twin for maintenance: A literature review. **Computers in Industry**, v. 123, p. 103316, dez. 2020.

FERRARO, A. *et al.* Evaluating eXplainable artificial intelligence tools for hard disk drive predictive maintenance. **Artificial Intelligence Review**, v. 56, n. 7, p. 7279–7314, 13 jul. 2023.

FINK, O. *et al.* Potential, challenges and future directions for deep learning in prognostics and health management applications. **Engineering Applications of Artificial Intelligence**, v. 92, p. 103678, jun. 2020.

FLORIAN, E.; SGARBOSSA, F.; ZENNARO, I. Machine learning-based predictive maintenance: A cost-oriented model for implementation. **International Journal of Production Economics**, v. 236, p. 108114, jun. 2021.

GAWDE, S. *et al.* Explainable Predictive Maintenance of Rotating Machines Using LIME, SHAP, PDP, ICE. **IEEE Access**, v. 12, p. 29345–29361, 2024.

GIL, A. C. **Como elaborar projetos de pesquisa**. 4. ed. São Paulo: Atlas, 2002.

GOBY, N.; BRANDT, T.; NEUMANN, D. Deep reinforcement learning with combinatorial actions spaces: An application to prescriptive maintenance. **Computers & Industrial Engineering**, v. 179, p. 109165, maio 2023.

GOODFELLOW, IAN.; BENGIO, YOSHUA.; COURVILLE, AARON. **Deep learning**. [s.l.] The MIT Press, 2017.

GUPTA, S.; KUMAR, A.; MAITI, J. A critical review on system architecture, techniques, trends and challenges in intelligent predictive maintenance. **Safety Science**, v. 177, p. 106590, set. 2024.

GUPTA, V. *et al.* **An Overview of Supervised Machine Learning Algorithm** 2022 11th International Conference on System Modeling & Advancement in Research Trends (SMART). **Anais...IEEE**, 16 dez. 2022

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The Elements of Statistical Learning**. New York, NY: Springer New York, 2009.

HONG, C. W. *et al.* Remaining Useful Life Prognosis for Turbofan Engine Using Explainable Deep Neural Networks with Dimensionality Reduction. **Sensors**, v. 20, n. 22, p. 6626, 19 nov. 2020.

HUANG, C. *et al.* Prognostics and health management for predictive maintenance: A review. **Journal of Manufacturing Systems**, v. 75, p. 78–101, ago. 2024.

HUANG, W. *et al.* Remaining Useful Life Estimation for Systems with Abrupt Failures. **Annual Conference of the PHM Society**, v. 10, n. 1, 24 set. 2018.

IRFAN, M. *et al.* Condition monitoring of induction motors via instantaneous power analysis. **Journal of Intelligent Manufacturing**, v. 28, n. 6, p. 1259–1267, 11 ago. 2017.

JALALI, A. *et al.* **Predicting Time-to-Failure of Plasma Etching Equipment using Machine Learning** 2019 IEEE International Conference on Prognostics and Health Management (ICPHM). **Anais...IEEE**, jun. 2019

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, out. 2006.

JUREK, A. *et al.* A survey of commonly used ensemble-based classification techniques. **The Knowledge Engineering Review**, v. 29, n. 5, p. 551–581, 3 nov. 2014.

KAMWA, I.; SAMANTARAY, S.; JOOS, G. **On the accuracy versus transparency trade-off of data-mining models for fast-response PMU-based catastrophe predictors** 2013 IEEE Power & Energy Society General Meeting. **Anais...IEEE**, 2013

KARAKILIÇ, M.; HATAŞ, H.; PACAL, İ. **Open-Circuit Fault Detection in T-Type MLI Using XGBoost: A Machine Learning-Based Approach** 2025 7th International Congress on Human-Computer Interaction, Optimization and Robotic Applications (ICHORA). **Anais...IEEE**, 23 maio 2025

KHAN, T. *et al.* **An Explainable Regression Framework for Predicting Remaining Useful Life of Machines** 2022 27th International Conference on Automation and Computing (ICAC). **Anais...IEEE**, 1 set. 2022

KOBAYASHI, K.; ALAM, S. B. Explainable, interpretable, and trustworthy AI for an intelligent digital twin: A case study on remaining useful life. **Engineering Applications of Artificial Intelligence**, v. 129, p. 107620, mar. 2024.

KOTSOMPOLIS, G. *et al.* Smart Forecasting of Carbon Prices Using Machine Learning and Neural Networks: When ARIMA Meets XGBoost and LSTM. **Journal of Forecasting**, v. 45, n. 1, p. 47–60, 16 jan. 2026.

KUMAR, K. J.; JAIRAM, K.; AMBEDKAR, CH. An In-Depth Study Of Machine Learning In Artificial Intelligence. **Educational Administration: Theory and Practice**, 2023.

KUNDU, R. K.; HOQUE, K. A. Explainable Predictive Maintenance is Not Enough: Quantifying Trust in Remaining Useful Life Estimation. **Annual Conference of the PHM Society**, v. 15, n. 1, 26 out. 2023.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. **Nature**, v. 521, n. 7553, p. 436–444, 28 maio 2015.

LEFROUNI, K.; TAIBI, S. Artificial Intelligence Techniques for Industrial Predictive Maintenance: A Systematic Review of Recent Advances. **Journal Européen des Systèmes Automatisés**, v. 58, n. 4, p. 653–668, 30 abr. 2025.

LEI, Y. *et al.* Machinery health prognostics: A systematic review from data acquisition to RUL prediction. **Mechanical Systems and Signal Processing**, v. 104, p. 799–834, maio 2018.

LI, H. *et al.* A review on physics-informed data-driven remaining useful life prediction: Challenges and opportunities. **Mechanical Systems and Signal Processing**, v. 209, p. 111120, mar. 2024.

LI, W.; LI, T. Comparison of deep learning models for predictive maintenance in industrial manufacturing systems using sensor data. **Scientific Reports**, v. 15, n. 1, p. 23545, 2 jul. 2025.

LINARDATOS, P.; PAPASTEFANOPOULOS, V.; KOTSIANTIS, S. Explainable AI: A Review of Machine Learning Interpretability Methods. **Entropy**, v. 23, n. 1, p. 18, 25 dez. 2020.

LUNDBERG, S. M.; LEE, S.-I. **A unified approach to interpreting model predictions** Proceedings of the 31st International Conference on Neural Information Processing Systems. **Anais...: NIPS'17**. Red Hook, NY, USA: Curran Associates Inc., 2017

MA, Z. *et al.* **An interpretability research of the Xgboost algorithm in remaining useful life prediction** 2020 International Conference on Big Data & Artificial Intelligence & Software Engineering (ICBASE). **Anais...IEEE**, out. 2020

MAGNUSSON, L. V.; OLSSON, J. R.; TRAN, C. T. T. Recurrent Neural Networks for Oil Well Event Prediction. **IEEE Intelligent Systems**, v. 38, n. 2, p. 73–80, mar. 2023.

MALLIORIS, P.; AIVAZIDOU, E.; BECHTSIS, D. Predictive maintenance in Industry 4.0: A systematic multi-sector mapping. **CIRP Journal of Manufacturing Science and Technology**, v. 50, p. 80–103, jun. 2024.

MANSOURVAR, Z.; JAHANGOSHAI REZAEE, M.; ESHKEVARI, M. An explainable approach for prediction of remaining useful life in turbofan condition monitoring. **Neural Computing and Applications**, v. 37, n. 17, p. 10621–10645, 26 jun. 2025.

MIGUEL, P. A. C. *et al.* **Metodologia de pesquisa em engenharia de produção e gestão de operações** Elsevier, , 2010.

MOHRI, MEHRYAR.; ROSTAMIZADEH, AFSHIN.; TALWALKAR, AMEET. **Foundations of machine learning**. [s.l.] MIT Press, 2012.

MURPHY, K. P. . **Probabilistic machine learning : an introduction**. [s.l.] The MIT Press, 2022.

MWANGI, M. G. W. *et al.* Hybrid AI-driven condition monitoring and RUL forecasting for multi-fault diagnosis in two-stroke marine diesel engines. **Journal of Ocean Engineering and Science**, dez. 2025.

NAUMOV, M. Feedforward and Recurrent Neural Networks Backward Propagation and Hessian in Matrix Form. **ArXiv**, v. abs/1709.06080, 2017.

NGUYEN, K. T. P.; MEDJAHHER, K. A new dynamic predictive maintenance framework using deep learning for failure prognostics. **Reliability Engineering & System Safety**, v. 188, p. 251–262, ago. 2019.

OK, E.; EMMANUEL, M. Understanding the Gradient Boosting Algorithm in XGBoost. mar. 2025.

P, Y. G.; A, M. **Enhanced Predictive Maintenance through Explainable Multimodal Deep Learning and Human-in-the-Loop Systems** 2024 International Conference on Emerging Research in Computational Science (ICERCS). **Anais...IEEE**, 12 dez. 2024

PALSANIYA, R. C.; PATIL, R.; THANU, R. **Predictive Maintenance Using Machine Learning in Semiconductor Manufacturing** 2025 IEEE 34th International Symposium on Industrial Electronics (ISIE). **Anais...IEEE**, 20 jun. 2025

PANDEY, MRS. A. Machine Learning. **International Journal for Research in Applied Science and Engineering Technology**, v. 11, n. 8, p. 864–869, 31 ago. 2023.

PATIL, A. *et al.* Maintenance strategy selection: a comprehensive review of current paradigms and solution approaches. **International Journal of Quality & Reliability Management**, v. 39, n. 3, p. 675–703, 22 fev. 2022.

PECHT, MICHAEL. **Prognostics and health management of electronics**. [s.l.] Wiley, 2008.

PHONGMOO, S. *et al.* Predictive maintenance in semiconductor manufacturing: Comparative analysis of machine learning models for downtime reduction. **Computers & Industrial Engineering**, v. 205, p. 111211, jul. 2025.

QUINLAN, J. R. Induction of decision trees. **Machine Learning**, v. 1, n. 1, p. 81–106, mar. 1986.

RAJAGUKGUK, I. M. *et al.* Comparative Analysis of XGBoost, Random Forest, and Logistic Regression for Classifying Jakarta's Air Pollution Index (ISPU). **Procedia Computer Science**, v. 269, p. 108–120, 2025.

RAMEZANI, S. B. *et al.* Scalability, Explainability and Performance of Data-Driven Algorithms in Predicting the Remaining Useful Life: A Comprehensive Review. **IEEE Access**, v. 11, p. 41741–41769, 2023.

SAHOO, B. B. *et al.* Long short-term memory (LSTM) recurrent neural network for low-flow hydrological time series forecasting. **Acta Geophysica**, v. 67, n. 5, p. 1471–1481, 20 out. 2019.

SERENKO, I. A. *et al.* Room for Uncertainty in Remaining Useful Life Estimation for Turbofan Jet Engines. **Nelineinaya Dinamika**, v. 20, n. 5, p. 933–943, 2024.

SERRADILLA, O. *et al.* Deep learning models for predictive maintenance: a survey, comparison, challenges and prospects. **Applied Intelligence**, v. 52, n. 10, p. 10934–10964, 18 ago. 2022.

SHI, Z.; SHI, M.; LI, C. **The prediction of character based on recurrent neural network language model** 2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS). **Anais...IEEE**, maio 2017

SHUKLA, K.; NEFTI-MEZIANI, S.; DAVIS, S. A heuristic approach on predictive maintenance techniques: Limitations and scope. **Advances in Mechanical Engineering**, v. 14, n. 6, 14 jun. 2022.

SI, X.-S. *et al.* Remaining useful life estimation – A review on the statistical data driven approaches. **European Journal of Operational Research**, v. 213, n. 1, p. 1–14, ago. 2011.

SINGH, A.; BOBDE, S. **A Review on Predicting Failure in Industrial Machines: Its Methods, Challenges and Future Direction** 2025 1st International Conference on AIML-Applications for Engineering & Technology (ICAET). **Anais...IEEE**, 16 jan. 2025a

_____. **Explaining Condition-Based Maintenance for Metro Trains Using Ensemble ML with XAI and Hybrid Deep Learning Models** 2025 Global Conference in Emerging Technology (GINOTECH). **Anais...IEEE**, 9 maio 2025b

SINGH, K. *et al.* Concurrent Estimation of Remaining Useful Life for Multiple Faults in an Ion Etch Mill. **Annual Conference of the PHM Society**, v. 10, n. 1, 24 set. 2018.

SOUALHI, A. *et al.* PHM SURVEY: Implementation of Prognostic Methods for Monitoring Industrial Systems. **Energies**, v. 15, n. 19, p. 6909, 21 set. 2022.

SUBASHREE, S. *et al.* Adaptive machine learning models for predictive maintenance in industrial internet of things (IIoT) systems. **Scientific Reports**, v. 16, n. 1, p. 12451, 7 mar. 2026.

SUTTON, C. D. Classification and Regression Trees, Bagging, and Boosting. *Em*: [s.l: s.n.]. p. 303–329.

SWAIN, P. **The Artificial Intelligence and Machine Learning Blueprint: Foundations, Frameworks, and Real-World Applications**. [s.l.] Deep Science Publishing, 2025.

SYED, M. A. BIN *et al.* A systematic review of time series algorithms and analytics in predictive maintenance. **Decision Analytics Journal**, v. 15, p. 100573, jun. 2025.

TAMA, B. A. *et al.* Recent advances in the application of deep learning for fault diagnosis of rotating machinery using vibration signals. **Artificial Intelligence Review**, v. 56, n. 5, p. 4667–4709, 9 maio 2023.

TV, V. *et al.* Recurrent Neural Networks for Online Remaining Useful Life Estimation in Ion Mill Etching System. **Annual Conference of the PHM Society**, v. 10, n. 1, 22 set. 2018.

UMEDA, S. *et al.* Planned Maintenance Schedule Update Method for Predictive Maintenance of Semiconductor Plasma Etcher. **IEEE Transactions on Semiconductor Manufacturing**, v. 34, n. 3, p. 296–300, ago. 2021.

VOGL, G. W.; WEISS, B. A.; HELU, M. A review of diagnostic and prognostic capabilities and best practices for manufacturing. **Journal of Intelligent Manufacturing**, v. 30, n. 1, p. 79–95, 9 jan. 2019.

WANG, Y.; ZHAO, Y.; ADDEPALLI, S. Remaining Useful Life Prediction using Deep Learning Approaches: A Review. **Procedia Manufacturing**, v. 49, p. 81–88, 2020.

WOLPERT, D. H. Stacked generalization. **Neural Networks**, v. 5, n. 2, p. 241–259, jan. 1992.

WU, J.-Y. *et al.* Degradation-Aware Remaining Useful Life Prediction With LSTM Autoencoder. **IEEE Transactions on Instrumentation and Measurement**, v. 70, p. 1–10, 2021.

WU, S. *et al.* Remaining useful life prediction for ion etching machine cooling system using deep recurrent neural network-based approaches. **Control Engineering Practice**, v. 109, p. 104748, abr. 2021.

YUVARAJU, M.; DIVYA, M. **Residential Load Forecasting by Recurrent Neural Network on LSTM Model** 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS). **Anais...IEEE**, maio 2020

ZHANG, L. *et al.* A Review on Deep Learning Applications in Prognostics and Health Management. **IEEE Access**, v. 7, p. 162415–162438, 2019.

ZHANG, T.; WANG, H. Explainable remaining useful life uncertainty prediction method for rolling bearing. **Engineering Applications of Artificial Intelligence**, v. 161, p. 112114, dez. 2025.

ZHANG, W.; YANG, D.; WANG, H. Data-Driven Methods for Predictive Maintenance of Industrial Equipment: A Survey. **IEEE Systems Journal**, v. 13, n. 3, p. 2213–2227, set. 2019.

ZHAO, L.; ZHU, Y.; ZHAO, T. Deep Learning-Based Remaining Useful Life Prediction Method with Transformer Module and Random Forest. **Mathematics**, v. 10, n. 16, p. 2921, 13 ago. 2022.

ZONTA, T. *et al.* Predictive maintenance in the Industry 4.0: A systematic literature review. **Computers & Industrial Engineering**, v. 150, p. 106889, dez. 2020.