

UNIVERSIDADE DO ESTADO DE SANTA CATARINA – UDESC
CENTRO DE CIÊNCIAS TECNOLÓGICAS – CCT
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO APLICADA

GABRIEL VINICIUS BOIN FREITAS

PREVISÃO DE FALHAS EM BOMBAS INDUSTRIAIS DE TRANSPORTE
DUTOVIÁRIO DE PETRÓLEO E DERIVADOS COM REDES NEURAS ARTIFICIAIS
PROFUNDAS E TRANSFERÊNCIA DE APRENDIZADO ENTRE DOMÍNIOS

JOINVILLE

2026

GABRIEL VINICIUS BOIN FREITAS

**PREVISÃO DE FALHAS EM BOMBAS INDUSTRIAIS DE TRANSPORTE
DUTOVIÁRIO DE PETRÓLEO E DERIVADOS COM REDES NEURAIS ARTIFICIAIS
PROFUNDAS E TRANSFERÊNCIA DE APRENDIZADO ENTRE DOMÍNIOS**

Dissertação apresentada ao Programa de Pós-Graduação em Computação Aplicada do Centro de Ciências Tecnológicas da Universidade do Estado de Santa Catarina, como requisito parcial para a obtenção do grau de Mestre em Computação Aplicada.

Orientador: Yuri Kaszubowski Lopes

JOINVILLE

2026

**Ficha catalográfica elaborada pelo programa de geração automática da
Biblioteca Universitária Udesc,
com os dados fornecidos pelo(a) autor(a)**

Freitas, Gabriel Vinicius Boin
PREVISÃO DE FALHAS EM BOMBAS INDUSTRIAIS DE
TRANSPORTE DUTOVIÁRIO DE PETRÓLEO E
DERIVADOS COM REDES NEURAIS ARTIFICIAIS
PROFUNDAS E TRANSFERÊNCIA DE APRENDIZADO
ENTRE DOMÍNIOS / Gabriel Vinicius Boin Freitas. -- 2026.
124 p.

Orientador: Yuri Kaszubowski Lopes
Dissertação (mestrado) -- Universidade do Estado de
Santa Catarina, Centro de Ciências Tecnológicas, Programa
de Pós-Graduação em Computação Aplicada, Joinville, 2026.

1. Vida Útil Remanescente. 2. Manutenção Preditiva. 3.
Engenharia de Confiabilidade. 4. Transferência de
Aprendizado. 5. Bombas de Oleoduto.. I. Lopes, Yuri
Kaszubowski. II. Universidade do Estado de Santa Catarina,
Centro de Ciências Tecnológicas, Programa de
Pós-Graduação em Computação Aplicada. III. Título.

GABRIEL VINICIUS BOIN FREITAS

**PREVISÃO DE FALHAS EM BOMBAS INDUSTRIAIS DE TRANSPORTE
DUTOVIÁRIO DE PETRÓLEO E DERIVADOS COM REDES NEURAIIS ARTIFICIAIS
PROFUNDAS E TRANSFERÊNCIA DE APRENDIZADO ENTRE DOMÍNIOS**

Dissertação apresentada ao Programa de Pós-Graduação em Computação Aplicada do Centro de Ciências Tecnológicas da Universidade do Estado de Santa Catarina, como requisito parcial para a obtenção do grau de Mestre em Computação Aplicada.

Orientador: Yuri Kaszubowski Lopes

BANCA EXAMINADORA:

Dr. Yuri Kaszubowski Lopes (orientador/presidente)
Universidade do Estado de Santa Catarina

Membros:

Dr. Lucas Hermann Negri
Instituto Federal de Mato Grosso do Sul

Dr. Roberto Silvio Ubertino Rosso Junior
Universidade do Estado de Santa Catarina

Joinville, 22 de julho de 2026

AGRADECIMENTOS

Agradeço à minha mulher, Fabiana, por nunca deixar eu desistir, sempre ser minha amiga, me fazer sorrir quando estou triste, me motivar a ser a melhor versão de mim. Também sou grato aos meus finados pais, Luiz Carlos e Maria do Rosário, por sempre terem investido e me estimulado em meus estudos. Meu grande amigo João Paulo Gois foi uma inspiração para eu decidir trilhar o caminho acadêmico. Nada poderia ser feito de forma zelosa, sem a competente, técnica e inteligente ajuda que meu orientador sempre dedicou a mim. Um muito obrigado por me iluminar neste caminho nobre do aprender. Gostaria de agradecer também à toda equipe da UDESC, que sempre foi muito prestativa e solícita em tudo que foi necessário para a conclusão desta trajetória. A qualidade da estrutura da Universidade é notável. Foi de grande valia poder ter contado com bolsa PROMOP fornecida pela UDESC durante todo o Mestrado. É muito satisfatório perceber a boa utilização dos impostos em uma Universidade pública de tanta qualidade.

RESUMO

O autor investiga a estimação da Vida Útil Remanescente, ou RUL (do inglês *Remaining Useful Life*), das bombas do conjunto B-0302, responsável pela pressurização de trecho do Oleoduto Santa Catarina–Paraná (OSPAR), entre o Terminal Aquaviário de São Francisco do Sul (TEFRAN) e a Refinaria Presidente Getúlio Vargas (REPAR). As quatro bombas operam em paralelo e em regime de rodízio. A pesquisa integra Engenharia de Confiabilidade, aprendizado profundo e transferência de aprendizado para avaliar em que condições históricos de falha e sinais de sensores podem apoiar a manutenção preditiva em um sistema dutoviário. Registros de manutenção extraídos do sistema de gestão foram analisados por meio de TTF, MTBF, MTTR e ajustes das distribuições Exponencial e Weibull. A análise revelou comportamentos distintos entre as bombas: parte dos eventos esteve associada a mecanismos elétricos, de controle ou instrumentação pouco compatíveis com o padrão progressivo do domínio de origem e com os sensores disponíveis, enquanto a B-0302D apresentou histórico mais compatível com degradação mecânica progressiva, especialmente relacionada à vedação e ao selo mecânico, além de maior riqueza sensorial. Como domínio de origem, utilizou-se o subconjunto FD001 da base C-MAPSS (do inglês *Commercial Modular Aero-Propulsion System Simulation*), desenvolvida pela NASA, no qual foi reproduzido o AGATT (do inglês *Attention-based Graph Attention Transformer*), modelo que combina convolução, atenção em grafos e *Transformers*. A melhor execução na FD001 alcançou $R^2 = 0,8885$ e RMSE de 13,88 ciclos; em 30 execuções, o RMSE médio foi de $15,99 \pm 1,39$ ciclos, evidenciando variabilidade experimental. A adaptação ao domínio industrial exigiu harmonizar sensores de naturezas distintas. Para isso, foi desenvolvida uma projeção físico-estatística com diversidade condicional, baseada em plausibilidade física, similaridade entre distribuições e penalização da reutilização excessiva de um mesmo sensor. O AGATT pré-treinado foi empregado como extrator de características congelado, sendo treinada apenas uma nova cabeça de regressão. Na B-0302D, a configuração principal obteve $R^2 = 0,467$, RMSE de 10,43 dias, MAE de 8,79 dias, viés de +0,04 dia e *Score* NASA de 45,92. Uma LSTM industrial e um AGATT treinado do zero apresentaram RMSE médios de 16,34 e 18,26 dias, respectivamente, reforçando o benefício da transferência. No limiar de 15 dias, a função de perda conservadora identificou quatro dos cinco casos críticos, sem falsos positivos. Os resultados indicam que a transferência de aprendizado pode apoiar a estimação de RUL quando a falha é progressiva e observável nos sensores, mas possui limitações em eventos abruptos ou pouco instrumentados. A principal contribuição é um fluxo auditável e metodologicamente documentado que integra confiabilidade clássica, *benchmarking* em base pública, harmonização entre domínios, linhas de base e avaliação operacional em um ativo industrial real. Como a validação foi retrospectiva e baseada em três trajetórias da B-0302D, os achados requerem confirmação prospectiva.

Palavras-chave: Vida Útil Remanescente; Manutenção Preditiva; Engenharia de Confiabilidade; Transferência de Aprendizado; C-MAPSS; Bombas de Oleoduto.

ABSTRACT

The author investigates Remaining Useful Life (RUL) estimation for the B-0302 pump set, which pressurizes a section of the Santa Catarina–Paraná Pipeline (OSPAR), between the São Francisco do Sul Waterway Terminal (TEFRAN) and the Presidente Getúlio Vargas Refinery (REPAR). The four pumps operate in parallel and under a rotation scheme. The research integrates Reliability Engineering, deep learning, and transfer learning to assess under which conditions failure histories and sensor signals can support predictive maintenance in an oil pipeline system. Maintenance records extracted from the management system were analyzed using Time to Failure (TTF), Mean Time Between Failures (MTBF), Mean Time To Repair (MTTR), and Exponential and Weibull distribution fitting. The analysis revealed distinct behaviors among the pumps: some events involved electrical, control, or instrumentation mechanisms that were poorly aligned with the progressive source-domain pattern and the available sensors, whereas Pump B-0302D presented a history more consistent with progressive mechanical degradation, particularly involving sealing and mechanical seal issues, as well as a richer sensor set. The FD001 subset of the NASA Commercial Modular Aero-Propulsion System Simulation (C-MAPSS) dataset was used as the source domain. The Attention-based Graph Attention Transformer (AGATT), which combines convolution, graph attention, and Transformer mechanisms, was reproduced on this dataset. Its best execution achieved $R^2 = 0.8885$ and a Root Mean Square Error (RMSE) of 13.88 cycles. Across 30 executions, the mean RMSE was 15.99 ± 1.39 cycles, highlighting experimental variability. Adaptation to the industrial domain required harmonizing sensors with different physical meanings. A physical-statistical projection with conditional diversity was developed using physical plausibility, distributional similarity, and a penalty against excessive reuse of the same sensor. The pretrained AGATT was employed as a frozen feature extractor, and only a new regression head was trained. For Pump B-0302D, the main configuration achieved $R^2 = 0.467$, RMSE of 10.43 days, Mean Absolute Error (MAE) of 8.79 days, bias of +0.04 day, and NASA Score of 45.92. An industrial Long Short-Term Memory (LSTM) network and an AGATT trained from scratch achieved mean RMSE values of 16.34 and 18.26 days, respectively, reinforcing the benefit of transfer learning. At the 15-day threshold, the conservative loss identified four of the five critical cases without false positives. The results indicate that transfer learning can support RUL estimation when degradation is progressive and observable in the available sensors, but its effectiveness is limited for abrupt or poorly instrumented events. The main contribution is an auditable and methodologically documented workflow integrating classical reliability analysis, public-dataset benchmarking, domain harmonization, baseline comparisons, and operational evaluation for a real industrial asset. Since the validation was retrospective and based on three failure trajectories from Pump B-0302D, the findings require prospective confirmation.

Keywords: Remaining Useful Life; Predictive Maintenance; Reliability Engineering; Transfer Learning; C-MAPSS; Pipeline Pumps.

LISTA DE ILUSTRAÇÕES

Figura 1 – Sistema logístico da REPAR	18
Figura 2 – PBS: Diagrama de Venn	40
Figura 3 – PBS: Evolução de publicações por ano	41
Figura 4 – PBS: Tecnologia usada e RMSE FD001	42
Figura 5 – PBS: Publicações por tecnologia	43
Figura 6 – PBS: Publicações <i>Transformer</i> com RMSE entre 9,35 e 10,99 ciclos	43
Figura 7 – PBS: Resumo	44
Figura 8 – Curvas de aderência – Bomba A	52
Figura 9 – Curvas de aderência – Bomba B	53
Figura 10 – Curvas de aderência – Bomba C	55
Figura 11 – Curvas de aderência – Bomba D	57
Figura 12 – FD001: Painel dos 21 sensores alinhados à falha	65
Figura 13 – FD001: Variância por sensor	65
Figura 14 – FD001: Correlação entre os 14 sensores utilizados	66
Figura 15 – FD001: Correlação média absoluta por sensor	67
Figura 16 – Arquitetura didática do AGATT reproduzido	68
Figura 17 – AGATT na C-MAPSS: curvas de perda no treinamento e validação	74
Figura 18 – AGATT na C-MAPSS: RUL real vs. RUL previsto	75
Figura 19 – AGATT na C-MAPSS: distribuição de erros	75
Figura 20 – AGATT na C-MAPSS: distribuição do RMSE em 30 execuções	76
Figura 21 – Transfer Learning: falhas B-0302	80
Figura 22 – Transfer Learning: sensores por bomba	84
Figura 23 – Transfer Learning: correlação Bomba D	86
Figura 24 – Transfer Learning: séries pré-falha Bomba D	88
Figura 25 – Transfer Learning: distribuições s2	97
Figura 26 – Transfer Learning: distribuições s7	97
Figura 27 – Transfer Learning: distribuições s12	97
Figura 28 – Transfer Learning: distribuições s17	98
Figura 29 – Transfer Learning: projeção Bomba D	100
Figura 30 – Transfer Learning: mapa de aderência	102
Figura 31 – Transfer Learning: janela de geração	103
Figura 32 – Transfer Learning: sensibilidade ao RUL	105
Figura 33 – Transfer Learning: resultado NASA	107
Figura 34 – Transfer Learning: <i>loss</i> conservadora	113
Figura 35 – Transfer Learning: comparação operacional	115

LISTA DE TABELAS

Tabela 1 – Síntese didática das famílias de redes discutidas na revisão.	34
Tabela 2 – Comparação consolidada dos 21 modelos avaliados na base FD001 do C-MAPSS.	45
Tabela 3 – Datas de falha e TTF – Bomba A	52
Tabela 4 – Datas de falha e TTF – Bomba B	54
Tabela 5 – Datas de falha e TTF – Bomba C	55
Tabela 6 – Datas de falha e TTF – Bomba D	57
Tabela 7 – Parâmetros ajustados e confiabilidades previstas para 90 e 180 dias.	58
Tabela 8 – Resumo dos subdatasets da C-MAPSS.	61
Tabela 9 – Condições operacionais e sensores disponíveis na base C-MAPSS.	63
Tabela 10 – Fluxo da arquitetura AGATT reproduzida.	71
Tabela 11 – Hiperparâmetros e configuração de treinamento do AGATT na C-MAPSS FD001.	72
Tabela 12 – Métricas da melhor execução do AGATT na C-MAPSS FD001.	73
Tabela 13 – Transfer Learning: resultados por horário	90
Tabela 14 – Transfer Learning: geração do backbone	92
Tabela 15 – Transfer Learning: cabeça de regressão	93
Tabela 16 – Transfer Learning: configuração experimental	94
Tabela 17 – Transfer Learning: agrupamento físico	96
Tabela 18 – Transfer Learning: projeção fold D-1	101
Tabela 19 – Transfer Learning: arquitetura LSTM	109
Tabela 20 – Transfer Learning: triagem LSTM	110
Tabela 21 – Transfer Learning: comparação com linhas de base	111

LISTA DE ABREVIATURAS E SIGLAS

ACM	Associação para Maquinaria Computacional, do inglês <i>Association for Computing Machinery</i>
AGATT	Transformador de Atenção com Grafo Baseado em Atenção, do inglês <i>Attention-based Graph Attention Transformer</i>
AM-TCN	Rede Temporal Convolucional com Mecanismo de Atenção, do inglês <i>Attention Mechanism Temporal Convolutional Network</i>
ATS2S	Modelo Sequência para Sequência com Atenção, do inglês <i>Attention-based Sequence-to-Sequence model</i>
B*-DAE-TF	Modelo combinado de Memória Longo-Curto Bidirecional, Autoencoder de Redução de Ruído e Transformador, do inglês <i>Bidirectional Long Short-Term Memory – Denoising Autoencoder – Transformer</i>
BPR	Razão de Derivação, do inglês <i>Bypass Ratio</i>
CBAM	Módulo de Atenção Convolucional em Bloco, do inglês <i>Convolutional Block Attention Module</i>
C-MAPSS	Simulação Modular Comercial de Sistema de Propulsão Aeronáutica, do inglês <i>Commercial Modular Aero-Propulsion System Simulation</i>
Celesc	Centrais Elétricas de Santa Catarina
CNN	Rede Neural Convolucional, do inglês <i>Convolutional Neural Network</i>
CPU	Unidade Central de Processamento, do inglês <i>Central Processing Unit</i>
CUDA	Arquitetura Unificada de Dispositivos de Cálculo, do inglês <i>Compute Unified Device Architecture</i>
cuDNN	Biblioteca de Redes Neurais Profundas da NVIDIA para CUDA, do inglês <i>CUDA Deep Neural Network Library</i>
D-PSOSEN	Rede Temporal Convolucional com Atenção Dupla e Otimização por Enxame de Partículas, do inglês <i>Dual Attention Temporal Convolutional Network – Particle Swarm Optimization Stacked Ensemble Network</i>
DBN	Rede de Crença Profunda, do inglês <i>Deep Belief Network</i>
DDP-NN	Rede Neural com Paralelismo de Dados Distribuído, do inglês <i>Distributed Data Parallel Neural Network</i>
DL	Aprendizado Profundo, do inglês <i>Deep Learning</i>
DS-GNN	Rede Neural em Grafos de Dupla Escala, do inglês <i>Dual-Scale Graph Neural Network</i>
EPR	Razão de Pressão do Motor, do inglês <i>Engine Pressure Ratio</i>

F1-score	Métrica de desempenho baseada na média harmônica entre precisão e revocação, do inglês <i>F1-score</i>
FD001	Sub-base 001 da C-MAPSS
FD002	Sub-base 002 da C-MAPSS
FD003	Sub-base 003 da C-MAPSS
FD004	Sub-base 004 da C-MAPSS
FN	Falso Negativo
FP	Falso Positivo
GAT	Rede de Atenção em Grafos, do inglês <i>Graph Attention Network</i>
GNN	Rede Neural em Grafos, do inglês <i>Graph Neural Network</i>
GPU	Unidade de Processamento Gráfico, do inglês <i>Graphics Processing Unit</i>
GRU	Unidade Recorrente com Portas, do inglês <i>Gated Recurrent Unit</i>
IEEE	Instituto de Engenheiros Eletricistas e Eletrônicos, do inglês <i>Institute of Electrical and Electronics Engineers</i>
IP	Parâmetro elétrico registrado no sistema de monitoramento
IA	Parâmetro elétrico registrado no sistema de monitoramento
IQR	Intervalo Interquartil, do inglês <i>Interquartile Range</i>
KS	Kolmogorov–Smirnov
LSTM	Memória de Longo e Curto Prazo, do inglês <i>Long Short-Term Memory</i>
LTS	Suporte de Longo Prazo, do inglês <i>Long-Term Support</i>
MAE	Erro Absoluto Médio, do inglês <i>Mean Absolute Error</i>
MAPE	Erro Percentual Absoluto Médio, do inglês <i>Mean Absolute Percentage Error</i>
M*-DTCN	Rede Temporal Convolutacional com Atenção Multicanal e Dilatação, do inglês <i>Multi-Channel Attention – Dilated Temporal Convolutional Network</i>
MedAE	Erro Absoluto Mediano, do inglês <i>Median Absolute Error</i>
MODBUS	Protocolo de comunicação industrial MODBUS
MSE	Erro Quadrático Médio, do inglês <i>Mean Squared Error</i>
MTBF	Tempo Médio Entre Falhas, do inglês <i>Mean Time Between Failures</i>
MTTR	Tempo Médio para Reparo, do inglês <i>Mean Time To Repair</i>
NASA	Administração Nacional de Aeronáutica e Espaço, do inglês <i>National Aeronautics and Space Administration</i>
OLAPA	Oleoduto Araucária–Paranaguá

OPASC	Oleoduto Paraná–Santa Catarina
OSPAR	Oleoduto Santa Catarina–Paraná
PBS	Pesquisa Bibliográfica Sistemática
PHM	Prognóstico e Gerenciamento da Saúde de Ativos, do inglês <i>Prognostics and Health Management</i>
PID	Controlador proporcional–integral–derivativo
PSO	Otimização por Enxame de Partículas, do inglês <i>Particle Swarm Optimization</i>
REPAR	Refinaria Presidente Getúlio Vargas
ReLU	Unidade Linear Retificada, do inglês <i>Rectified Linear Unit</i>
RMSE	Raiz do Erro Quadrático Médio, do inglês <i>Root Mean Square Error</i>
RNN	Rede Neural Recorrente, do inglês <i>Recurrent Neural Network</i>
RUL	Vida Útil Remanescente, do inglês <i>Remaining Useful Life</i>
SAP	Sistema de Aplicações e Produtos, do alemão <i>Systeme, Anwendungen und Produkte</i>
SVM	Máquina de Vetor de Suporte, do inglês <i>Support Vector Machine</i>
TAP	Comutador de derivações do transformador, do inglês <i>Tap Changer</i>
TCN	Rede Temporal Convolucional, do inglês <i>Temporal Convolutional Network</i>
TEFRAN	Terminal Aquaviário de São Francisco do Sul
TEPAR	Terminal Aquaviário de Paranaguá
TEMHSA	Mecanismo Temporal Aprimorado de Autoatenção Multi-Cabeças, do inglês <i>Temporal Enhanced Multi-Head Self-Attention</i>
TTF	Tempo até a Falha, do inglês <i>Time To Failure</i>
VN	Verdadeiro Negativo
VP	Verdadeiro Positivo
WSL2	Subsistema Windows para Linux 2, do inglês <i>Windows Subsystem for Linux 2</i>

SUMÁRIO

1	INTRODUÇÃO	15
1.1	CONTEXTUALIZAÇÃO	15
1.1.1	Sistema logístico da REPAR	17
1.2	PROBLEMA DE PESQUISA	18
1.2.1	Delimitação do estudo	19
1.2.2	Justificativa	20
1.3	OBJETIVOS	20
1.4	PRINCIPAIS CONTRIBUIÇÕES	21
1.5	ORGANIZAÇÃO DO TRABALHO	23
2	REVISÃO DA LITERATURA	24
2.1	PROGNÓSTICO, SAÚDE DE ATIVOS E VIDA ÚTIL REMANESCENTE	24
2.2	ENGENHARIA DA CONFIABILIDADE	25
2.2.1	Falha, modo de falha e registro de manutenção	25
2.2.2	Confiabilidade, probabilidade de falha e taxa de risco	26
2.2.3	Distribuições Exponencial e Weibull	26
2.2.4	Confiabilidade histórica e prognóstico por sensores	27
2.3	APRENDIZADO DE MÁQUINA PARA PREVISÃO DE RUL	27
2.3.1	Da janela de sensores ao aprendizado do modelo	28
2.3.2	Redes neurais artificiais e aprendizado profundo	29
2.3.3	Redes convolucionais	30
2.3.4	Redes recorrentes, LSTM e GRU	30
2.3.5	Autocodificadores	31
2.3.6	Mecanismos de atenção e Transformers	32
2.3.7	Redes neurais em grafos e atenção em grafos	32
2.3.8	Modelo híbrido AGATT	33
2.3.9	Transferência de aprendizado entre domínios	35
2.4	BASE C-MAPSS COMO REFERÊNCIA EM PROGNÓSTICO	35
2.5	LEVANTAMENTO EXPLORATÓRIO	36
2.6	PESQUISA BIBLIOGRÁFICA SISTEMÁTICA (PBS)	37
2.6.1	Palavras-chave, fraseamento e campo a ser pesquisado	38
2.6.2	Bases de busca de referências, justificativa e período de busca	38
2.6.3	Critérios de seleção, inclusão, exclusão e forma de combinação	38
2.6.4	Execução da PBS	39
2.6.5	Síntese Analítica dos Modelos Seleccionados na PBS	44
2.6.6	Discussão e conclusão sobre a PBS	47

2.6.7	Riscos, limitações e formas de validação da PBS	48
3	ENGENHARIA DE CONFIABILIDADE NA B-0302	49
3.1	FONTE DOS DADOS E PROCESSAMENTO DOS REGISTROS	49
3.2	INDICADORES DE CONFIABILIDADE	50
3.3	RESULTADOS POR BOMBA	51
3.3.1	Bomba A	51
3.3.2	Bomba B	53
3.3.3	Bomba C	55
3.3.4	Bomba D	56
3.4	PREVISÕES DE CONFIABILIDADE EM 90 E 180 DIAS	58
3.5	SÍNTESE DO CAPÍTULO	59
4	REPRODUÇÃO DE BENCHMARKING NA C-MAPSS	60
4.1	BASE C-MAPSS	60
4.2	POR QUE UTILIZAR A C-MAPSS NESTA DISSERTAÇÃO	61
4.3	ESTRUTURA DOS DADOS	62
4.4	CONSTRUÇÃO DO ALVO RUL	63
4.5	JANELAMENTO TEMPORAL	64
4.6	NORMALIZAÇÃO DOS SENSORES	64
4.7	ANÁLISE EXPLORATÓRIA DOS SENSORES	64
4.8	ARQUITETURA AGATT	67
4.8.1	Entrada do modelo	67
4.8.2	Camada convolucional <i>depthwise</i>	69
4.8.3	Canal espacial	69
4.8.4	Canal temporal	70
4.8.5	Concatenação espaço-temporal	70
4.8.6	Codificação posicional	70
4.8.7	Blocos Transformer	71
4.9	RESUMO DA ARQUITETURA REPRODUZIDA	71
4.10	HIPERPARÂMETROS E DECISÕES DE TREINAMENTO	72
4.11	MÉTRICAS DE AVALIAÇÃO	73
4.12	RESULTADOS DA REPRODUÇÃO	73
4.13	ANÁLISE DE REPRODUTIBILIDADE	76
4.14	COMPARAÇÃO COM O ARTIGO ORIGINAL	77
4.15	INTERPRETAÇÃO DOS RESULTADOS	77
4.16	SÍNTESE DO CAPÍTULO	78
5	TRANSFER LEARNING PARA B-0302	79
5.1	CARACTERIZAÇÃO DO DOMÍNIO INDUSTRIAL DE DESTINO	79

5.1.1	Histórico de falhas e construção das séries	80
5.1.2	Protocolo de validação por série de falha e pontos ocultados	81
5.1.3	Síntese comparativa das bombas A, B, C e D	82
5.1.4	Sensores disponíveis por bomba	83
5.1.5	Análise exploratória dos sensores industriais da Bomba D	84
5.1.6	Normalização e suavização temporal sem uso de observações futuras . .	87
5.1.7	Influência do horário de coleta dos sensores	89
5.1.8	Da reprodução com janela 30 ao <i>backbone</i> com janela 35	91
5.2	PIPELINE GERAL DE TRANSFERÊNCIA DE APRENDIZADO	92
5.2.1	Configuração experimental principal	94
5.3	PROJEÇÃO DOS SENSORES INDUSTRIAIS PARA O ESPAÇO FD001 .	95
5.3.1	Agrupamento físico dos sensores e alternativas de projeção	95
5.3.2	Comparação visual das distribuições na escolha do de-para	96
5.3.3	Distância composta e diversidade condicional	98
5.4	SENSIBILIDADE À JANELA DE GERAÇÃO	102
5.5	SENSIBILIDADE AO INTERVALO DE RUL OCULTADO	104
5.6	RESULTADO PRINCIPAL DA TRANSFERÊNCIA DE APRENDIZADO .	105
5.7	LINHAS DE BASE E ABLAÇÃO DA TRANSFERÊNCIA	108
5.7.1	LSTM treinada exclusivamente com dados industriais	108
5.7.2	AGATT treinado do zero no domínio industrial	110
5.8	LIMITAÇÕES DA SELEÇÃO EXPERIMENTAL	112
5.9	CALIBRAÇÃO OPERACIONAL CONSERVADORA	112
5.10	COMPARAÇÃO ENTRE DESEMPENHO GLOBAL E DESEMPENHO OPERACIONAL	114
5.11	SÍNTESE DO CAPÍTULO	116
6	CONCLUSÃO	117
6.1	TRABALHOS FUTUROS	118
6.2	CONSIDERAÇÕES FINAIS	118
	REFERÊNCIAS	119

1 INTRODUÇÃO

Este capítulo apresenta o contexto, o problema de pesquisa, a justificativa, os objetivos, as delimitações e as contribuições desta dissertação. O trabalho investiga a estimação da Vida Útil Remanescente, ou RUL (do inglês *Remaining Useful Life*), em bombas industriais de transporte dutoviário de petróleo e derivados, tendo como estudo de caso o conjunto B-0302, localizado no Terminal Aquaviário de São Francisco do Sul-SC (TEFRAN).

O autor descreve três perspectivas. A primeira é a Engenharia de Confiabilidade, empregada para caracterizar de forma exploratória o histórico de falhas das bombas A, B, C e D. A segunda é o prognóstico orientado por dados, representado pela reprodução do modelo AGATT (do inglês *Attention-based Graph Attention Transformer*) na base pública C-MAPSS (do inglês *Commercial Modular Aero-Propulsion System Simulation*), subconjunto FD001. A terceira é a transferência de aprendizado entre domínios, na qual representações aprendidas em motores turbofan simulados são adaptadas ao domínio real das bombas B-0302 por meio de harmonização sensorial, projeção físico-estatística e treinamento de uma nova cabeça de regressão sobre um *backbone* congelado.

O foco não está em afirmar que um único modelo possa prever indistintamente todos os modos de falha do conjunto. O autor investiga em quais condições essa transferência é tecnicamente defensável, quais limitações surgem quando há poucas trajetórias industriais e como o desempenho global de regressão se relaciona com a utilidade operacional de alarmes próximos da falha.

1.1 CONTEXTUALIZAÇÃO

A confiabilidade de equipamentos industriais é um fator crítico para a continuidade operacional de sistemas logísticos de grande porte. Na indústria de petróleo e derivados, falhas em bombas, compressores, válvulas e demais ativos rotativos podem provocar perdas econômicas, restrições operacionais, riscos ambientais e impactos à segurança. Nesse contexto, a manutenção preditiva busca identificar sinais de degradação antes da ocorrência de uma falha funcional, permitindo planejar inspeções e intervenções com maior antecedência.

A Transpetro S.A., subsidiária logística da Petrobras, opera uma malha nacional de dutos e terminais destinada ao transporte e ao armazenamento de petróleo, derivados e biocombustíveis. Entre os sistemas operados pela empresa, destaca-se a rota que conecta o Terminal Aquaviário de São Francisco do Sul, em Santa Catarina, à Refinaria Presidente Getúlio Vargas (REPAR), em Araucária, Paraná. Essa rota possui relevância logística para o abastecimento regional e depende da disponibilidade de ativos críticos de bombeamento.

O conjunto B-0302, localizado no TEFRAN, fornece a pressurização necessária ao transporte de petróleo bruto pelo Oleoduto Santa Catarina–Paraná (OSPAR) em direção à REPAR. Uma indisponibilidade inesperada nesse conjunto pode reduzir temporariamente a capacidade de movimentação, exigir alternativas operacionais mais onerosas e aumentar a necessidade de

manutenção corretiva emergencial.

O conjunto é composto pelas Bombas B-0302A, B-0302B, B-0302C e B-0302D. Duas ou três unidades podem operar simultaneamente, enquanto outra permanece em reserva. O rodízio busca distribuir o desgaste e equilibrar horas de funcionamento. As bombas possuem horímetros; contudo, a análise histórica adotou dias de calendário porque não foi possível obter, para todo o período, uma série de horas por unidade contínua e sincronizada com o sensoramento e as falhas.

A operação também é influenciada pela tarifação de energia elétrica. Em Santa Catarina, o horário de ponta definido pela Celesc ocorre das 18h30 às 21h30 (Celesc, 2024). Quando a programação operacional permite, busca-se evitar esse intervalo por corresponder ao período de maior custo energético. Essa característica produz uma rotina diária que pode influenciar o estado operacional capturado pelos sensores.

O horário de extração dos dados sensoriais foi tratado como uma variável metodológica porque cada série utiliza uma observação por sensor e por dia, e diferentes horários podem capturar operação, reserva ou transição de regime. Foram avaliadas 00:00, 06:00, 10:00, 12:00 e 15:00. A extração às 00:00 foi adotada para a B-0302D por apresentar o melhor desempenho retrospectivo, sem pressupor que a bomba estivesse operando nesse horário em todas as datas.

A manutenção corretiva, executada após a falha, tende a produzir maior pressão operacional e menor flexibilidade de planejamento. Em sentido oposto, intervenções preventivas excessivamente antecipadas podem gerar indisponibilidades desnecessárias, substituição prematura de componentes e uso ineficiente de recursos. Um desafio central da gestão de ativos consiste, portanto, em determinar quando intervir, equilibrando risco, custo e disponibilidade.

Nesse cenário, o RUL representa uma estimativa do tempo ou do número de ciclos restantes até que um ativo atinja uma condição de falha ou degradação crítica. Uma estimativa de RUL não substitui o diagnóstico de engenharia, mas pode complementar a tomada de decisão ao indicar tendências de redução da vida remanescente, apoiar a priorização de inspeções e contribuir para o planejamento de intervenções.

Tradicionalmente, a Engenharia de Confiabilidade utiliza históricos de falhas e tempos até a falha para estimar indicadores como Tempo até a Falha, ou TTF (do inglês *Time To Failure*), Tempo Médio Entre Falhas, ou MTBF (do inglês *Mean Time Between Failures*), Tempo Médio para Reparo, ou MTTR (do inglês *Mean Time To Repair*), taxa de falha e confiabilidade. Distribuições estatísticas, como Exponencial e Weibull, são frequentemente utilizadas nessas análises (ReliaSoft Corporation, 2023; Nelson, 1982). No presente trabalho, esses métodos fornecem uma caracterização exploratória do comportamento histórico das bombas, pois o número de intervalos disponíveis por unidade é reduzido.

O avanço do sensoramento industrial permite complementar essa leitura histórica com modelos capazes de explorar séries temporais multivariadas de pressão, temperatura, vibração, corrente elétrica e outras variáveis operacionais. Redes recorrentes, redes LSTM (do inglês *Long Short-Term Memory*), redes GRU (do inglês *Gated Recurrent Unit*), redes convolucionais, auto-

encodificadores e, mais recentemente, arquiteturas baseadas em atenção, grafos e *Transformers* são empregadas para aprender padrões de degradação diretamente dos dados (Liu, Wen e Wang, 2025; Wu, Zhang e Li, 2024).

Uma das bases públicas mais utilizadas nesse campo é a C-MAPSS, desenvolvida pela NASA (do inglês *National Aeronautics and Space Administration*). A base simula a degradação de motores turbofan e fornece trajetórias completas de operação até a falha, permitindo treinar e comparar modelos de RUL em ambiente padronizado (Saxena *et al.*, 2008b). Embora a C-MAPSS represente motores aeronáuticos, e não bombas centrífugas, sua estrutura multivariada e suas trajetórias *run-to-failure* tornam-na uma referência adequada para estudar métodos de prognóstico.

Nesta dissertação, o subconjunto FD001 da C-MAPSS é utilizado como domínio de origem para a reprodução e a avaliação do AGATT (Zeng, Wu e Cheng, 2023). A arquitetura combina mecanismos de atenção, redes de atenção em grafos e blocos *Transformer*, buscando modelar simultaneamente dependências temporais e relações entre sensores. Essa característica motivou sua escolha para um problema industrial em que a degradação pode se manifestar por alterações conjuntas em variáveis hidráulicas, elétricas, térmicas e mecânicas.

A aplicação do AGATT às bombas B-0302 não pode ser feita por correspondência direta. O FD001 representa degradação progressiva, enquanto o histórico industrial também contém eventos elétricos, de comando e de instrumentação. Existem técnicas específicas para prever falhas elétricas com sensoriamento apropriado (Nandi, Toliyat e Li, 2005); a limitação deste estudo decorre da compatibilidade entre o domínio de origem, os sinais disponíveis e o mecanismo de falha, não de uma impossibilidade geral de prognóstico elétrico. Por isso, a transferência foi tratada como adaptação entre domínios. O procedimento final utiliza o AGATT pré-treinado como extrator de características congelado e treina apenas uma nova cabeça de regressão no domínio industrial, após a projeção físico-estatística dos canais das bombas para as 14 dimensões esperadas pelo modelo de origem.

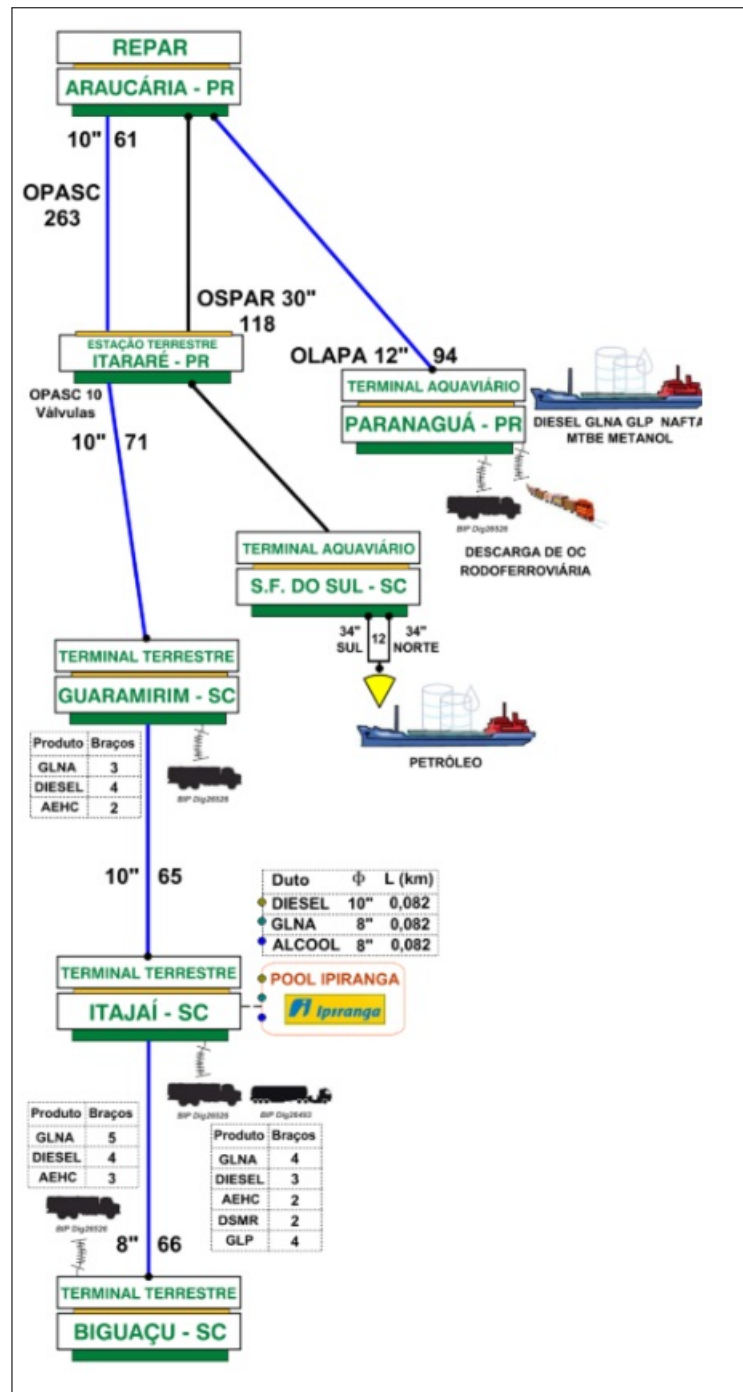
1.1.1 Sistema logístico da REPAR

A Figura 1 apresenta o sistema logístico associado à REPAR e a posição dos principais dutos e terminais.

No arranjo da Figura 1, o OSPAR liga o TEFTRAN à região de Itararé, de onde o petróleo segue para a refinaria. Nesse sistema, o conjunto B-0302 fornece energia hidráulica para vencer as restrições de elevação e perda de carga do percurso, contribuindo para a continuidade do escoamento.

A relevância do conjunto decorre, portanto, de sua função sistêmica. O monitoramento de sua condição não busca apenas reduzir o tempo de reparo de uma unidade, mas também preservar flexibilidade operacional em um sistema de bombeamento redundante. É nesse contexto que a presente pesquisa combina confiabilidade, prognóstico de RUL e transferência de aprendizado.

Figura 1 – Sistema logístico da REPAR, com os oleodutos OPASC, OSPAR e OLAPA, os terminais TEFran e TEPAR, a estação de Itararé e os terminais terrestres de Guaramirim, Itajaí e Biguaçu.



Fonte: Cortesia da Transpetro S.A.

1.2 PROBLEMA DE PESQUISA

O problema investigado consiste em estimar retrospectivamente o RUL das bombas do conjunto B-0302 a partir de históricos de falha e séries temporais de sensores, utilizando conhecimento previamente aprendido em uma base pública de degradação.

A aplicação de modelos profundos a esse domínio apresenta desafios específicos. Pri-

meiro, trajetórias completas até a falha são raras em ambiente industrial, pois a própria manutenção procura impedir que o ativo opere até a falha funcional. Segundo, os eventos registrados possuem naturezas distintas: falhas mecânicas progressivas podem deixar assinaturas graduais via sensores, enquanto ocorrências elétricas, de comando, de comunicação ou de instrumentação podem ser mais abruptas. Terceiro, os sensores industriais não correspondem diretamente às variáveis da C-MAPSS. Quarto, o rodízio e a redundância do conjunto fazem com que o tempo de calendário não represente necessariamente o tempo contínuo de exposição operacional (existem os horímetros nas bombas, mas houve dificuldade técnica para correlacioná-los com os dados dos sensores).

Diante dessas condições, a questão central é formulada da seguinte forma:

Em que medida representações aprendidas por um modelo AGATT no subconjunto C-MAPSS FD001 podem, após harmonização físico-estatística dos sensores e transferência por extração de características, apoiar a estimação retrospectiva do RUL e a emissão de alarmes operacionais nas bombas do conjunto B-0302?

A formulação “em que medida” é importante porque o estudo não parte da premissa de que a transferência funcionará igualmente para todas as bombas ou modos de falha. A investigação procura identificar condições de aplicabilidade, comparar o modelo transferido com linhas de base treinadas apenas no domínio industrial e separar duas dimensões de desempenho: a qualidade global da regressão de RUL e a capacidade de reconhecer situações críticas próximas da falha.

1.2.1 Delimitação do estudo

A pesquisa foi delimitada ao conjunto B-0302 e aos eventos disponíveis no histórico analisado. A Engenharia de Confiabilidade foi utilizada para caracterização exploratória, e não para estabelecer inferências populacionais amplas, devido ao pequeno número de intervalos entre falhas por bomba.

O experimento de origem foi conduzido no FD001 da C-MAPSS. O Capítulo 4 apresenta a reprodução de *benchmarking* com janela de 30 ciclos. Para a transferência industrial, foi utilizada uma instância distinta da mesma família AGATT, treinada no FD001 com janela efetiva de 35 ciclos e selecionada por seu comportamento no domínio de destino entre as alternativas avaliadas.

No domínio industrial, as quatro bombas foram submetidas a uma triagem comparativa. O aprofundamento concentrou-se na Bomba D porque ela reuniu maior riqueza sensorial, histórico de falhas mais compatível com degradação progressiva e melhor desempenho nos experimentos preliminares. A configuração final utilizou 20 canais industriais de entrada, projetados para as 14 dimensões ativas do FD001.

A avaliação principal da Bomba D foi retrospectiva. Os 30 pontos de teste foram derivados de três trajetórias reais de falha, com 10 pontos ocultados por trajetória no intervalo de 5 a 60

dias. Esses pontos não correspondem a 30 falhas independentes. As métricas devem, portanto, ser interpretadas como evidência de aderência ao histórico estudado, sujeita a dependência intrassérie e a viés de seleção experimental. A validação definitiva da abordagem requer novas trajetórias e avaliação prospectiva com o *pipeline* previamente congelado.

1.2.2 Justificativa

O conjunto B-0302 é formado por ativos de elevado valor e importância operacional. Uma falha não planejada pode afetar a capacidade logística, exigir rearranjos de operação, antecipar intervenções corretivas e aumentar a pressão sobre equipes e sobressalentes. Em um sistema redundante, o risco não depende somente da condição de uma bomba, mas da disponibilidade simultânea das demais unidades.

A estimação de RUL pode contribuir para reduzir esse risco ao apoiar três tipos de decisão: priorização de inspeções, planejamento de intervenções e gestão de alarmes. Uma tendência de redução da vida remanescente pode motivar verificações adicionais de vibração, temperaturas, pressões, mancais e vedação. Em horizontes de dias ou semanas, a informação pode auxiliar na programação de parada, mobilização de equipe e disponibilidade de peças.

A relevância científica está na transferência entre domínios estruturalmente distintos. A maior parte dos modelos recentes de RUL é avaliada em bases públicas, com trajetórias mais controladas do que aquelas encontradas em sistemas industriais reais. Avaliar o reaproveitamento de representações aprendidas na C-MAPSS exige enfrentar diferenças de sensores, escalas, regimes de operação e mecanismos de falha.

A relevância metodológica está na construção de um fluxo auditável de adaptação. Em vez de reutilizar automaticamente os pesos de uma rede, o trabalho organiza sensores por famílias físicas, compara distribuições, aplica uma penalização por reutilização, recalcula a projeção dentro de cada rodada de validação e mantém a série externa fora das estatísticas de adaptação. O processo procura reduzir o descompasso entre domínios sem ocultar que a equivalência física é aproximada.

A relevância aplicada está na avaliação conjunta de erro de regressão e alarmes operacionais. Um modelo com menor erro global de regressão não é necessariamente aquele que oferece maior proteção no horizonte crítico. Por isso, a dissertação avalia tanto a perda NASA, voltada ao desempenho global, quanto uma função de perda conservadora, destinada a penalizar com maior intensidade previsões tardias próximas da falha.

1.3 OBJETIVOS

O objetivo geral desta dissertação é avaliar a viabilidade de adaptar, por transferência de aprendizado, uma arquitetura profunda baseada em atenção, grafos e *Transformers* para estimar o RUL das bombas do conjunto B-0302, integrando séries temporais de sensores, históricos de falha e elementos de Engenharia de Confiabilidade.

Para alcançar esse objetivo geral, foram definidos os seguintes objetivos específicos:

- realizar uma Pesquisa Bibliográfica Sistemática (PBS) sobre técnicas recentes de previsão de RUL, com ênfase em modelos baseados em *Transformers* e em aplicações sobre a C-MAPSS;
- caracterizar de forma exploratória o histórico de falhas das bombas B-0302 por meio de TTF, MTBF, MTTR e ajustes das distribuições Exponencial e Weibull;
- reproduzir e avaliar o AGATT no subconjunto C-MAPSS FD001, examinando tanto o melhor desempenho obtido quanto a variabilidade entre execuções;
- documentar o pré-processamento dos domínios de origem e destino, incluindo seleção de sensores, normalização, janelamento, construção do RUL, tratamento de valores inválidos, suavização e definição dos pontos ocultados;
- desenvolver uma estratégia de harmonização físico-estatística que permita projetar os canais industriais para o espaço de 14 sensores ativos esperado pelo AGATT;
- adaptar o AGATT ao domínio B-0302 por extração de características com *backbone* congelado e comparar seu desempenho com uma LSTM industrial e com um AGATT da mesma família treinado do zero, utilizando métricas globais e indicadores de alarme; e
- analisar as condições de aplicabilidade, as limitações e as implicações operacionais da abordagem, propondo diretrizes para validação prospectiva e expansão a outros ativos.

1.4 PRINCIPAIS CONTRIBUIÇÕES

A primeira contribuição é a integração entre Engenharia de Confiabilidade e prognóstico orientado a sensores em um estudo de caso industrial real. A análise de TTF, MTBF, MTTR e distribuições de confiabilidade fornece contexto histórico para interpretar por que algumas bombas e modos de falha são mais previsíveis do que outros. Essa integração evita tratar as métricas do modelo como resultados desvinculados da natureza física e operacional dos eventos.

A segunda contribuição é a reprodução estatisticamente documentada do AGATT no FD001. A melhor execução do experimento de *benchmarking* com janela 30 apresentou Raiz do Erro Quadrático Médio, ou RMSE (do inglês *Root Mean Square Error*), de 13,88 ciclos, Erro Absoluto Médio, ou MAE (do inglês *Mean Absolute Error*), de 10,19 ciclos, $R^2 = 0,8885$ e Score NASA¹ de 275,56. Em 30 execuções, o RMSE médio foi de $15,99 \pm 1,39$ ciclos, evidenciando

¹ A perda NASA (*NASA loss*) é uma adaptação diferenciável da função de pontuação assimétrica do desafio PHM08, empregada como função objetivo durante o treinamento do modelo. Ela penaliza exponencialmente os erros de estimação da vida útil remanescente, atribuindo maior penalização às superestimações, nas quais o modelo prevê uma RUL superior à real. O *NASA Score*, por sua vez, é utilizado como métrica de avaliação após o treinamento e corresponde à soma das penalizações calculadas para as observações do conjunto de teste. Assim, embora ambos utilizem o mesmo princípio de penalização assimétrica, a perda NASA orienta a otimização dos

que a avaliação de um modelo profundo deve considerar não apenas o melhor caso, mas também sua variabilidade experimental.

A terceira contribuição é a estratégia de harmonização entre domínios. Para a Bomba D, 20 canais industriais de entrada foram convertidos em uma representação de 14 dimensões compatível com o AGATT. A projeção combina famílias físicas, distância de Wasserstein², estatística de Kolmogorov–Smirnov³, diferenças de média e desvio-padrão e penalização de diversidade. A matriz é recalculada em cada *fold* utilizando apenas as séries de treinamento, reduzindo o risco de vazamento de informação⁴.

A quarta contribuição é a demonstração de que o benefício observado não decorreu apenas da arquitetura AGATT. A configuração transferida utilizou o *backbone* pré-treinado e congelado e treinou uma cabeça de regressão com 1.921 parâmetros. Seu resultado principal na Bomba D foi $R^2 = 0,467$, RMSE de 10,43 dias, MAE de 8,79 dias, viés de +0,04 dia e Score NASA de 45,92. Em contraste, a LSTM treinada apenas com os dados industriais apresentou R^2 médio negativo e RMSE médio de 16,34 dias, enquanto o AGATT inicializado do zero também apresentou R^2 médio negativo e RMSE médio de 18,26 dias. Essa comparação sustenta, no conjunto estudado, a utilidade das representações aprendidas no domínio de origem.

A quinta contribuição é a separação entre desempenho global e objetivo operacional. A perda NASA produziu a melhor regressão global. A função *horizon_conservative*, por sua vez, identificou quatro dos cinco pontos com RUL real de até 15 dias, sem falsos positivos, resultando em precisão de 1,00, revocação, ou *recall*, de 0,80 e F1-score de 0,889. Esse resultado mostra que a função de perda pode ser calibrada conforme a prioridade operacional, embora a avaliação seja retrospectiva e derivada de somente três trajetórias.

A sexta contribuição é a explicitação das limitações experimentais. O trabalho diferencia o *benchmarking* de janela 30 do *backbone* de janela 35, distingue 20 canais de entrada de 16 sensores utilizados na análise de correlação e de 14 dimensões projetadas, e registra que as escolhas de horário, janela, suavização, horizonte de RUL e função de perda foram realizadas sobre um conjunto industrial reduzido. Essa transparência é parte da contribuição metodológica, pois permite reproduzir, auditar e contestar as decisões tomadas.

parâmetros do modelo, enquanto o *NASA Score* mensura seu desempenho final acumulado. Em ambos os casos, valores menores indicam melhor desempenho.

² A distância de Wasserstein é uma medida de dissimilaridade entre duas distribuições de probabilidade. Intuitivamente, ela representa o menor custo necessário para transformar uma distribuição na outra, considerando simultaneamente a quantidade de massa probabilística deslocada e a distância desse deslocamento. Valores menores indicam distribuições mais semelhantes.

³ A estatística de Kolmogorov–Smirnov corresponde à maior diferença absoluta observada entre as funções de distribuição acumulada de duas amostras, ou entre uma amostra e uma distribuição teórica. Valores menores indicam maior semelhança entre as distribuições, enquanto valores elevados evidenciam diferenças em algum ponto de suas distribuições acumuladas.

⁴ O vazamento de informação entre os conjuntos de treino e teste ocorre quando dados, estatísticas ou características do conjunto de teste influenciam, direta ou indiretamente, o treinamento do modelo. Esse problema pode acontecer, por exemplo, quando a normalização é ajustada com todos os dados antes da separação, quando observações relacionadas aparecem nos dois conjuntos ou quando o conjunto de teste participa da seleção de hiperparâmetros. Como consequência, as métricas obtidas tornam-se excessivamente otimistas e deixam de representar adequadamente a capacidade de generalização do modelo para dados realmente não observados.

Parte dos resultados da PBS e da reprodução experimental do AGATT originou o artigo *A Statistical Analysis of the State-of-the-Art Transformer-Based Remaining Useful Life Models on C-MAPSS FD001*, de autoria de Yuri Kaszubowski Lopes e Gabriel V. B. Freitas, aceito como *short paper* no *DATA 2026 – International Conference on Data Science, Technology and Applications*, em Porto, Portugal. A transferência de aprendizado para as bombas B-0302 constitui a ampliação aplicada desenvolvida nesta dissertação.

1.5 ORGANIZAÇÃO DO TRABALHO

Esta dissertação está estruturada em seis capítulos:

- **Capítulo 1 – Introdução:** apresenta o contexto industrial, o problema de pesquisa, a delimitação, a justificativa, os objetivos e as contribuições;
- **Capítulo 2 – Revisão da Literatura:** reúne os fundamentos de Engenharia de Confiabilidade, prognóstico, aprendizado profundo, mecanismos de atenção, grafos, *Transformers*, AGATT, transferência de aprendizado e PBS;
- **Capítulo 3 – Engenharia de Confiabilidade na B-0302:** caracteriza exploratoriamente os históricos das bombas por meio de TTF, MTBF, MTTR, curvas de confiabilidade e ajustes Exponencial e Weibull;
- **Capítulo 4 – Reprodução de *benchmarking* na C-MAPSS:** descreve o FD001, o pré-processamento, a reprodução do AGATT com janela 30 e a variabilidade observada em múltiplas execuções;
- **Capítulo 5 – *Transfer learning* para B-0302:** apresenta a caracterização industrial, o horário de coleta, o *backbone* de janela 35, a projeção físico-estatística, as análises de sensibilidade, o resultado da Bomba D, as linhas de base sem transferência, os alarmes operacionais e as limitações do protocolo;
- **Capítulo 6 – Conclusão:** retoma os objetivos, sintetiza os achados, discute contribuições e ameaças à validade e propõe trabalhos futuros.

2 REVISÃO DA LITERATURA

Este capítulo apresenta os conceitos necessários para acompanhar a investigação, partindo do problema físico de manutenção até chegar à arquitetura de aprendizado profundo empregada nos experimentos. A exposição procura responder, em sequência, a quatro questões: o que significa estimar a vida remanescente de um ativo; como a Engenharia da Confiabilidade descreve o histórico de falhas; como modelos de aprendizado de máquina extraem padrões de séries temporais; e por que o AGATT e a transferência de aprendizado são adequados ao problema estudado.

As equações são usadas para formalizar relações que serão retomadas nos capítulos experimentais, mas não constituem o ponto de partida da explicação. Antes de cada formulação, apresenta-se a ideia física ou computacional envolvida; depois, discute-se o significado dos termos e sua relação com as bombas B-0302. Essa organização evita tratar o modelo como uma sequência de símbolos desconectada da aplicação.

2.1 PROGNÓSTICO, SAÚDE DE ATIVOS E VIDA ÚTIL REMANESCENTE

O Prognóstico e Gerenciamento da Saúde de Ativos, ou PHM (do inglês *Prognostics and Health Management*), reúne atividades de monitoramento, diagnóstico e prognóstico. Embora esses termos sejam relacionados, eles respondem a perguntas diferentes. O monitoramento verifica se o comportamento atual se afasta do esperado; o diagnóstico procura identificar a natureza e a localização da anomalia; e o prognóstico estima como essa condição pode evoluir e quanto tempo resta até um limite de falha ou desempenho inaceitável (Heng *et al.*, 2009; Liu, Wen e Wang, 2025).

Em uma bomba industrial, por exemplo, uma elevação de vibração pode ser detectada pelo monitoramento. A análise conjunta de frequência, temperatura de mancal e histórico de manutenção pode indicar desalinhamento, desgaste ou defeito em rolamento, constituindo o diagnóstico. O prognóstico exige um passo adicional: estimar se a tendência observada é compatível com degradação progressiva e, em caso afirmativo, quantificar o horizonte provável até a condição crítica.

A Vida Útil Remanescente, ou RUL (do inglês *Remaining Useful Life*), representa esse horizonte (Heng *et al.*, 2009; Saxena *et al.*, 2008a). Sua unidade depende da forma como a exposição do equipamento é registrada. Em uma base de motores, pode ser expressa em ciclos; em um equipamento com horímetro sincronizado, em horas de operação; e, quando a série é organizada por datas, em dias de calendário. A unidade não é apenas uma convenção: ela define o significado operacional da previsão.

Considere uma observação realizada 20 dias antes de uma falha conhecida. Retrospectivamente, o RUL real desse ponto é 20 dias. Se um modelo estimar 18 dias, o erro é pequeno e conservador, pois antecipa levemente a proximidade da falha. Se estimar 35 dias, a previsão pode atrasar uma inspeção necessária. Por isso, além das métricas médias de regressão, esta

dissertação avalia o comportamento de alarmes próximos à falha.

A previsão de RUL pode ser construída por três famílias gerais de abordagem: métodos baseados em modelos físicos, métodos estatísticos e métodos orientados a dados (Heng *et al.*, 2009; Wang, Zhao e Addepalli, 2020; Liu, Wen e Wang, 2025). Os primeiros descrevem mecanismos de degradação por equações do sistema; os estatísticos usam históricos de falha para estimar probabilidades e tendências populacionais; e os orientados a dados aprendem relações entre sensores e vida remanescente. O presente trabalho combina as duas últimas perspectivas: a confiabilidade caracteriza o histórico das quatro bombas, enquanto o aprendizado profundo explora as séries multivariadas de sensores.

2.2 ENGENHARIA DA CONFIABILIDADE

A Engenharia da Confiabilidade estuda a capacidade de um item cumprir sua função durante um intervalo e sob condições especificadas (O'Connor e Kleyner, 2012; ISO, 2016). O termo “função” é central. Uma bomba pode continuar girando e, ainda assim, não cumprir adequadamente sua função se não fornecer a pressão, a vazão ou a integridade de vedação exigidas pelo sistema.

No contexto desta dissertação, a confiabilidade não é usada para afirmar que uma distribuição estatística explica completamente o comportamento das bombas. Ela fornece uma leitura histórica: quantos eventos ocorreram, como os intervalos variaram, quanto tempo os reparos demandaram e se há indícios exploratórios de taxa de falha constante, crescente ou decrescente.

2.2.1 Falha, modo de falha e registro de manutenção

Falha é a perda da capacidade de desempenhar uma função requerida. Modo de falha descreve como essa perda se manifesta, por exemplo, vazamento em selo, vibração elevada, falha de partida ou atuação indevida de proteção. Causa de falha corresponde ao mecanismo que levou ao evento, como desgaste, contaminação, problema de alimentação ou defeito de componente. A distinção é importante porque registros de manutenção nem sempre separam claramente sintoma, modo, causa e ação executada (ISO, 2016; IEC, 2018).

Uma ordem de manutenção também não equivale automaticamente a uma falha. Ela pode registrar inspeção, ajuste preventivo, substituição programada, atividade administrativa ou correção de uma perda funcional, distinção coerente com a organização de dados de manutenção proposta pela ISO 14224 (ISO, 2016). Por essa razão, os registros SAP foram interpretados em conjunto com a descrição técnica do evento, evitando contar indiscriminadamente toda intervenção como falha.

Os indicadores temporais mais utilizados neste trabalho são TTF, MTBF e MTTR (O'Connor e Kleyner, 2012; Nelson, 1982). O Tempo até a Falha, do inglês *Time to Failure*, ou TTF, mede o intervalo associado a uma trajetória até determinado evento. O Tempo Médio Entre Falhas, do inglês *Mean Time Between Failures*, ou MTBF, resume os intervalos entre falhas de

um item reparável. O Tempo Médio para Reparo do inglês *Mean Time To Repair*, ou MTTR, resume a duração necessária para restaurar a função. Esses indicadores descrevem dimensões diferentes: frequência de ocorrência e duração da indisponibilidade não devem ser confundidas.

No conjunto B-0302, os intervalos são expressos em dias de calendário porque não foi possível obter, para todo o período, uma série contínua de horímetro sincronizada com os sensores e as falhas. Assim, um intervalo de 100 dias não significa necessariamente 2.400 horas de funcionamento. A bomba pode ter operado somente em parte desse período devido ao rodízio. Essa limitação é discutida novamente nos Capítulos 3 e 5.

2.2.2 Confiabilidade, probabilidade de falha e taxa de risco

A função de confiabilidade responde a uma pergunta simples: qual é a probabilidade de o item ultrapassar determinado horizonte sem falhar? Se T representa o tempo até a falha, a confiabilidade é definida pela Equação 1 (NIST/SEMATECH, 2012c):

$$R(t) = \Pr(T > t) = 1 - F(t). \quad (1)$$

Na Equação 1, $R(t)$ é a probabilidade de sobrevivência além de t ; $\Pr(\cdot)$ denota probabilidade; e $F(t) = \Pr(T \leq t)$ é a probabilidade acumulada de falhar até esse horizonte. Se $R(90) = 0,80$, por exemplo, o modelo indica 80% de probabilidade de sobrevivência além de 90 unidades de tempo, nas condições e hipóteses adotadas.

A taxa de risco, ou *hazard*, responde a outra pergunta: entre os itens que chegaram funcionando ao instante t , qual é a propensão instantânea de falhar naquele ponto? Ela é dada pela Equação 2 (NIST/SEMATECH, 2012b):

$$h(t) = \frac{f(t)}{R(t)}. \quad (2)$$

Na Equação 2, $h(t)$ é a taxa instantânea de falha e $f(t)$ é a densidade do tempo até a falha. A taxa de risco não é a probabilidade acumulada de falhar. Uma analogia útil é distinguir a inclinação da trajetória do valor já acumulado: a função $F(t)$ informa quanto da probabilidade de falha já se acumulou, enquanto $h(t)$ descreve a intensidade instantânea entre os sobreviventes.

2.2.3 Distribuições Exponencial e Weibull

A distribuição Exponencial supõe taxa de falha constante. Essa hipótese é adequada como referência quando o risco não apresenta tendência evidente de aumento ou redução com o tempo. Sua função de confiabilidade é apresentada na Equação 3 (NIST/SEMATECH, 2012a):

$$R(t) = \exp(-\lambda t). \quad (3)$$

Na Equação 3, $\lambda > 0$ é a taxa constante de falha, t é o horizonte e $\exp(\cdot)$ é a função exponencial natural. Uma taxa constante não significa que a probabilidade acumulada permaneça constante; significa que a intensidade instantânea de falha não depende da idade no modelo.

A distribuição Weibull é mais flexível porque permite representar taxas decrescentes, constantes ou crescentes. Sua função de confiabilidade é dada pela Equação 4 (adaptada de Weibull (1951)):

$$R(t) = \exp \left[- \left(\frac{t}{\eta} \right)^\beta \right]. \quad (4)$$

Na Equação 4, $\eta > 0$ é o parâmetro de escala, na mesma unidade de t , e $\beta > 0$ é o parâmetro de forma. Quando $\beta < 1$, a taxa de falha decresce; quando $\beta = 1$, a Weibull se reduz à Exponencial; e quando $\beta > 1$, a taxa aumenta com o tempo. Em termos físicos, um $\beta > 1$ pode ser compatível com desgaste, mas não prova sozinho a existência de um mecanismo de degradação. Com poucos eventos, como no histórico de cada bomba, a interpretação deve permanecer exploratória e ser confrontada com os registros de manutenção.

2.2.4 Confiabilidade histórica e prognóstico por sensores

A confiabilidade e o prognóstico por sensores observam o mesmo ativo em escalas diferentes. A primeira resume o comportamento de uma população ou de um histórico de eventos; o segundo procura estimar o estado de uma trajetória específica a partir de medições atuais. Uma curva Weibull pode indicar aumento histórico da taxa de falha, mas não informa diretamente qual bomba está mais próxima de falhar hoje. Inversamente, um modelo de RUL pode estimar a proximidade de um evento em uma série, mas não substitui a análise da frequência histórica e do impacto sistêmico das falhas.

Essa complementaridade justifica a estrutura da dissertação. O Capítulo 3 usa TTF, MTBF, MTTR e distribuições para caracterizar as quatro unidades. Os Capítulos 4 e 5 tratam do prognóstico orientado a sensores e avaliam se padrões aprendidos em trajetórias de degradação podem ser reutilizados no domínio industrial.

2.3 APRENDIZADO DE MÁQUINA PARA PREVISÃO DE RUL

Aprendizado de máquina reúne métodos que ajustam seus parâmetros a partir de exemplos, em vez de depender exclusivamente de regras programadas manualmente (Mitchell, 1997). Em uma tarefa supervisionada de RUL, cada exemplo contém uma entrada e um alvo. A entrada é uma janela de sensores; o alvo é a vida remanescente associada ao último instante dessa janela.

Considere uma matriz com 30 linhas e 14 colunas. As linhas representam 30 ciclos consecutivos e as colunas representam sensores. O modelo precisa transformar essa matriz em um único número: o RUL estimado no ciclo mais recente. Para isso, deve reconhecer padrões locais em cada sinal, relações entre sensores e evolução temporal ao longo da janela.

O processo de aprendizagem ocorre pela minimização de uma função de perda. A rede produz uma previsão, compara-a com o alvo, calcula o erro e ajusta seus pesos por retropropagação (Goodfellow, Bengio e Courville, 2016). Esse ciclo é repetido em vários lotes de amostras. A qualidade do modelo não deve ser avaliada nos mesmos exemplos usados para ajustar os pesos, pois isso mede memorização, não generalização. Em séries de falha, o cuidado deve ser ainda maior: pontos próximos de uma mesma trajetória são correlacionados e não podem ser distribuídos livremente entre treino e teste.

2.3.1 Da janela de sensores ao aprendizado do modelo

Para compreender os modelos apresentados a seguir, é útil acompanhar o caminho completo percorrido por uma amostra. A janela temporal não é apenas um recorte gráfico: ela é a unidade entregue à rede. Se a janela possui 30 instantes e 14 sensores, cada exemplo contém uma matriz com 420 valores organizados no tempo e por variável. O alvo associado a essa matriz é o RUL do último instante. O modelo, portanto, não tenta prever uma falha a partir de uma medição isolada; ele procura reconhecer, no histórico recente, uma configuração compatível com determinada vida remanescente.

A primeira passagem da janela pelo modelo é chamada de propagação direta. Cada camada transforma a representação recebida e produz outra. Uma convolução pode destacar variações locais; uma camada recorrente pode preservar informação de instantes anteriores; um mecanismo de atenção pode atribuir pesos diferentes a sensores ou períodos; e uma camada de regressão converte a representação final em um único valor de RUL. No início do treinamento, os parâmetros são inicializados sem conhecimento específico da tarefa, e as previsões tendem a apresentar erros elevados.

A função de perda transforma a diferença entre previsão e valor real em um número que orienta o ajuste. A retropropagação calcula quanto cada parâmetro contribuiu para esse erro, e o otimizador usa essa informação para atualizar os pesos (Goodfellow, Bengio e Courville, 2016). A taxa de aprendizado controla o tamanho dessas atualizações. Valores excessivos podem fazer o treinamento oscilar ou divergir; valores muito pequenos podem tornar a convergência lenta ou impedir que o modelo saia de uma solução inadequada.

As amostras são normalmente processadas em lotes. Um lote, ou *batch*, reúne um subconjunto de janelas antes de uma atualização dos parâmetros (Goodfellow, Bengio e Courville, 2016). Uma época corresponde a uma passagem completa pelas amostras de treinamento. Repetir épocas permite refinar o modelo, mas continuar ajustando após a perda de validação deixar de melhorar pode produzir sobreajuste: o modelo passa a representar particularidades do conjunto de treino que não se repetem em novas trajetórias.

Por esse motivo, treinamento, validação e teste cumprem funções diferentes (Goodfellow, Bengio e Courville, 2016). O conjunto de treinamento ajusta os pesos; a validação orienta escolhas como ponto de parada e hiperparâmetros; e o teste estima o desempenho fora do ajuste. Em séries temporais de falha, essa separação deve ocorrer também entre trajetórias. Misturar

janelas vizinhas da mesma falha entre treino e teste facilitaria a memorização do formato daquela trajetória e produziria uma avaliação artificialmente favorável.

Regularização é o conjunto de medidas usadas para reduzir essa memorização. O *dropout* remove aleatoriamente parte das ativações durante o treinamento; a parada antecipada interrompe o ajuste quando a validação deixa de melhorar; a limitação da norma do gradiente evita atualizações extremas; e a redução programada da taxa de aprendizado permite ajustes progressivamente menores. Nenhuma dessas técnicas garante generalização isoladamente. Elas apenas reduzem riscos que devem ser controlados por um protocolo de avaliação coerente (Goodfellow, Bengio e Courville, 2016).

Um exemplo numérico simples ajuda a interpretar o processo. Suponha que uma janela possua RUL real de 20 dias e que o modelo inicialmente estime 34 dias. O erro assinado é positivo, indicando superestimação da vida disponível. Após comparar muitas janelas e ajustar os parâmetros, a rede pode reduzir a previsão para 22 dias. O aprendizado não significa que ela memorizou a data da falha; significa que encontrou combinações de tendências e relações sensoriais que, nos dados de treinamento, estiveram associadas a horizontes próximos de 20 dias. A validade dessa associação só pode ser avaliada em uma trajetória que não participou do ajuste.

2.3.2 Redes neurais artificiais e aprendizado profundo

A formulação contemporânea das redes neurais deriva de uma trajetória histórica. McCulloch e Pitts propuseram, em 1943, um modelo lógico de neurônio artificial; Rosenblatt apresentou, em 1958, o perceptron como um sistema capaz de ajustar associações a partir de exemplos (McCulloch e Pitts, 1943; Rosenblatt, 1958). As camadas densas atuais generalizam essa ideia para vetores, múltiplas unidades e funções de ativação diferenciáveis. Portanto, a expressão a seguir é uma síntese moderna, e não uma transcrição literal de uma única publicação.

Uma rede neural artificial é formada por unidades que combinam entradas, pesos e uma função de ativação. Em uma camada densa, cada saída depende de todas as entradas da camada anterior. De forma compacta, uma camada pode ser representada pela Equação 5 (adaptado de Goodfellow, Bengio e Courville (2016)):

$$\mathbf{h} = \phi(\mathbf{W}\mathbf{x} + \mathbf{b}), \quad (5)$$

Na Equação 5, $\mathbf{x}$ é o vetor de entrada; $\mathbf{W}$ é a matriz de pesos; $\mathbf{b}$ é o vetor de vieses; $\phi(\cdot)$ é a função de ativação; e $\mathbf{h}$ é a representação produzida. Os pesos determinam quanto cada entrada influencia cada unidade. O viés permite deslocar o ponto de ativação, e a função não linear torna possível aprender relações mais complexas do que uma regressão linear.

O termo aprendizado profundo refere-se ao uso de várias camadas de representação (Goodfellow, Bengio e Courville, 2016). As primeiras camadas podem detectar padrões simples; camadas posteriores combinam esses padrões em estruturas mais abstratas. Em uma série de bomba, uma camada inicial pode responder a pequenas oscilações, outra pode relacionar essas

oscilações com temperatura e pressão, e uma camada posterior pode associar o padrão conjunto à proximidade da falha.

Profundidade, entretanto, não garante bom desempenho. Redes profundas podem sobreajustar bases pequenas, apresentar instabilidade de gradiente e aprender relações espúrias. Por isso, são necessários conjuntos de validação, regularização, parada antecipada e separação rigorosa entre trajetórias de treino e teste (Bengio, Simard e Frasconi, 1994; He *et al.*, 2016).

2.3.3 Redes convolucionais

Redes neurais convolucionais, ou CNNs (do inglês *Convolutional Neural Networks*), aplicam filtros compartilhados ao longo da entrada (LeCun *et al.*, 1998; Krizhevsky, Sutskever e Hinton, 2012). Em séries temporais unidimensionais, o filtro percorre o eixo do tempo e procura um mesmo padrão em posições diferentes. Isso permite detectar, por exemplo, um pico, uma oscilação curta ou uma mudança de inclinação independentemente do ponto exato em que ocorre na janela.

Para um sensor x e um filtro de tamanho K , uma convolução temporal simplificada pode ser escrita conforme a Equação 6 (adaptado de LeCun *et al.* (1998)) :

$$y_t = \phi \left(b + \sum_{k=0}^{K-1} w_k x_{t-k} \right) \quad (6)$$

Na Equação 6, y_t é a saída no instante t ; w_k são os pesos do filtro; x_{t-k} são as observações cobertas pelo filtro; b é o viés; K é o tamanho do *kernel*; e ϕ é a ativação. O mesmo conjunto de pesos w_k é reutilizado ao longo de toda a sequência.

Na convolução *depthwise*, cada canal é filtrado separadamente antes da combinação entre canais. A operação foi consolidada em arquiteturas modernas de convolução separável, como a Xception (Chollet, 2017), e foi empregada no AGATT por Zeng, Wu e Cheng (Zeng, Wu e Cheng, 2023). No AGATT, essa operação permite extrair padrões locais de cada sensor sem misturar imediatamente variáveis de naturezas distintas. A combinação entre sensores ocorre depois, nos mecanismos de atenção espacial e de grafo.

2.3.4 Redes recorrentes, LSTM e GRU

Redes neurais recorrentes, ou RNNs, processam a sequência mantendo um estado que resume informações anteriores (Elman, 1990; Werbos, 1990). A ideia é semelhante a uma memória atualizada a cada passo: o estado no instante t depende da observação atual e do estado anterior. RNNs simples, contudo, podem ter dificuldade para preservar relações distantes devido ao desaparecimento ou à explosão dos gradientes (Bengio, Simard e Frasconi, 1994).

A LSTM (do inglês *Long Short-Term Memory*) foi proposta por Hochreiter e Schmidhuber para reduzir a perda de informação em dependências longas por meio de uma célula de memória e mecanismos de controle de entrada e saída (Hochreiter e Schmidhuber, 1997). A porta de

esquecimento, hoje presente na forma convencional da LSTM, foi introduzida posteriormente por Gers, Schmidhuber e Cummins (Gers, Schmidhuber e Cummins, 2000). A Equação 7 (adaptado de Hochreiter e Schmidhuber (1997) e Gers, Schmidhuber e Cummins (2000)) reúne esses elementos em notação vetorial moderna; portanto, ela deve ser entendida como síntese da arquitetura histórica, e não como reprodução literal de uma única equação:

$$\begin{aligned}
 i_t &= \sigma(W_i[x_t, h_{t-1}] + b_i), \\
 f_t &= \sigma(W_f[x_t, h_{t-1}] + b_f), \\
 \tilde{c}_t &= \tanh(W_c[x_t, h_{t-1}] + b_c), \\
 c_t &= f_t \odot c_{t-1} + i_t \odot \tilde{c}_t, \\
 o_t &= \sigma(W_o[x_t, h_{t-1}] + b_o), \\
 h_t &= o_t \odot \tanh(c_t).
 \end{aligned} \tag{7}$$

Na Equação 7, x_t é a entrada atual; h_{t-1} é o estado oculto anterior; c_{t-1} é a memória anterior; i_t , f_t e o_t são as portas de entrada, esquecimento e saída; $\tilde{c}_t$ é o conteúdo candidato; W e b são parâmetros treináveis; σ é a função sigmoide; $\tanh$ é a tangente hiperbólica; e $\odot$ é a multiplicação elemento a elemento.

A interpretação das portas é mais importante do que memorizar a equação. A porta de esquecimento decide quanto da memória anterior deve ser mantido; a porta de entrada decide quanto da nova informação deve ser incorporado; e a porta de saída controla o conteúdo disponibilizado para a próxima etapa. Em uma trajetória de degradação, a rede pode aprender a preservar uma tendência lenta e reduzir a influência de oscilações pontuais.

A GRU (do inglês *Gated Recurrent Unit*) simplifica essa estrutura, utilizando duas portas e um único estado. Sua formulação é apresentada na Equação 8 (adaptado de Cho *et al.* (2014)):

$$\begin{aligned}
 r_t &= \sigma(W_r x_t + U_r h_{t-1} + b_r), \\
 z_t &= \sigma(W_z x_t + U_z h_{t-1} + b_z), \\
 \tilde{h}_t &= \tanh(W_h x_t + U_h (r_t \odot h_{t-1}) + b_h), \\
 h_t &= z_t \odot h_{t-1} + (1 - z_t) \odot \tilde{h}_t.
 \end{aligned} \tag{8}$$

Na Equação 8, r_t é a porta de reinicialização; z_t é a porta de atualização; $\tilde{h}_t$ é o estado candidato; e h_t é o estado atualizado. A GRU possui menos estados e, em geral, menos parâmetros que uma LSTM comparável. Não existe superioridade universal entre as duas; o desempenho depende do problema, do volume de dados e da configuração experimental.

2.3.5 Autocodificadores

Autocodificadores aprendem a reconstruir a própria entrada por meio de uma representação intermediária, denominada espaço latente (Bourlard e Kamp, 1988; Hinton e Salakhutdinov,

2006). O codificador comprime a informação; o decodificador tenta reconstruí-la. Se a representação latente for menor ou regularizada, a rede precisa preservar os aspectos mais relevantes da entrada.

Em manutenção preditiva, autocodificadores podem ser usados para redução de dimensionalidade, remoção de ruído ou detecção de anomalias (Vincent *et al.*, 2008; Sakurada e Yairi, 2014). Um modelo treinado principalmente com comportamento normal tende a reconstruir bem padrões conhecidos e apresentar erro maior diante de uma condição diferente. Nesta dissertação, os autocodificadores aparecem na revisão e na PBS, mas não compõem a arquitetura experimental principal.

2.3.6 Mecanismos de atenção e Transformers

Mecanismos de atenção permitem que o modelo atribua pesos diferentes às partes da entrada. Em vez de condensar toda a sequência em um único estado recorrente, a rede compara consultas, chaves e valores e seleciona dinamicamente quais elementos são mais relevantes para cada representação (Vaswani *et al.*, 2017).

Na atenção por produto interno escalado, a operação é definida pela Equação 9 (Vaswani *et al.*, 2017):

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^\top}{\sqrt{d_k}}\right) V. \quad (9)$$

Na Equação 9, Q contém as consultas; K , as chaves; V , os valores; $K^\top$ é a transposta de K ; e d_k é a dimensão das chaves. O produto $QK^\top$ mede compatibilidades; a divisão por $\sqrt{d_k}$ controla a magnitude; a função *softmax* transforma os escores em pesos; e a multiplicação por V produz a combinação ponderada.

Em uma janela de sensores, a atenção pode dar maior peso aos instantes em que surge uma mudança persistente e menor peso aos trechos estáveis (Vaswani *et al.*, 2017). A atenção multi-cabeças repete o processo em subespaços diferentes, permitindo que uma cabeça se concentre em tendências lentas, outra em transições locais e outra em relações entre variáveis.

O Transformer organiza atenção, conexões residuais, normalização e redes densas em blocos empilhados (Vaswani *et al.*, 2017). Como não possui recorrência, precisa de codificação posicional para representar a ordem dos elementos. Essa arquitetura facilita o processamento paralelo e a modelagem de dependências distantes, mas também exige volume de dados, regularização e controle de sobreajuste.

2.3.7 Redes neurais em grafos e atenção em grafos

Redes neurais em grafos, ou GNNs, representam o sistema por nós e arestas (Kipf e Welling, 2017; Gilmer *et al.*, 2017). Em uma aplicação multissensor, cada sensor pode ser tratado como um nó e as relações entre sensores como arestas. Em uma aplicação temporal, os instantes

podem formar os nós. A vantagem é modelar explicitamente que a informação de uma variável depende das demais e não apenas de sua evolução isolada.

A Rede de Atenção em Grafos, ou GAT (do inglês *Graph Attention Network*), aprende a importância relativa de cada vizinho. Primeiro, calcula-se um escore não normalizado entre os nós i e j , conforme a Equação 10 (adaptado de Veličković *et al.* (2018)):

$$e_{ij} = \text{LeakyReLU}\left(\mathbf{a}^\top [W\mathbf{h}_i \parallel W\mathbf{h}_j]\right). \quad (10)$$

Na Equação 10, $\mathbf{h}_i$ e $\mathbf{h}_j$ são as características dos nós; W é uma transformação linear; $\mathbf{a}$ é o vetor de atenção; $\parallel$ indica concatenação; e e_{ij} é o escore de compatibilidade.

Os escores são normalizados entre os vizinhos por meio da Equação 11 (adaptado de Veličković *et al.* (2018)):

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k \in \mathcal{N}_i} \exp(e_{ik})}. \quad (11)$$

Na Equação 11, α_{ij} é o peso normalizado atribuído ao vizinho j ; $\mathcal{N}_i$ é o conjunto de vizinhos de i ; e $\exp(\cdot)$ é a função exponencial. A soma dos pesos dos vizinhos de um nó é igual a 1.

A nova representação do nó é então obtida pela agregação ponderada, conforme a Equação 12 (adaptado de Veličković *et al.* (2018)):

$$\mathbf{h}'_i = \sigma \left(\sum_{j \in \mathcal{N}_i} \alpha_{ij} W\mathbf{h}_j \right). \quad (12)$$

Na Equação 12, $\mathbf{h}'_i$ é a característica atualizada; σ é a ativação; e os demais termos mantêm os significados anteriores. Em linguagem operacional, o nó combina a informação de seus vizinhos, dando maior peso às relações consideradas mais úteis para a tarefa.

2.3.8 Modelo híbrido AGATT

O AGATT, proposto por Zeng, Wu e Cheng (2023), combina convolução *depthwise*, atenção espacial e temporal, GATs e blocos Transformer. Cada componente responde a um aspecto diferente do problema. A convolução identifica padrões locais em cada sensor. O canal espacial procura relações entre sensores. O canal temporal compara instantes da janela. O Transformer integra as representações e modela dependências de maior alcance.

Essa combinação é pertinente para prognóstico porque a degradação raramente se manifesta por uma única variável isolada (Zeng, Wu e Cheng, 2023). Um problema de vedação, por exemplo, pode alterar pressões, temperatura, esforço elétrico e vibração de maneira conjunta. O modelo precisa observar tanto a trajetória de cada sinal quanto a mudança de relação entre eles.

A revisão apresenta apenas a lógica dos blocos. As dimensões, a sequência de operações, os hiperparâmetros e a reprodução do AGATT são detalhados no Capítulo 4. Essa separação evita repetir a arquitetura completa em dois capítulos: aqui se explica por que os componentes existem; no capítulo experimental, explica-se como foram implementados.

O AGATT não deve ser interpretado como um detector universal de falhas. O domínio FD001 contém degradação progressiva de motores simulados. Assim, o conhecimento aprendido tende a ser mais transferível quando o domínio industrial também apresenta evolução gradual observável nos sensores. Eventos abruptos, mecanismos não instrumentados ou falhas de natureza distinta podem exigir outros sensores, rótulos e modelos.

A Tabela 1 sintetiza o papel das principais famílias apresentadas. A comparação deve ser lida como orientação conceitual, e não como uma classificação absoluta: arquiteturas híbridas combinam essas capacidades, e o desempenho depende dos dados, do protocolo e da tarefa.

Tabela 1 – Síntese didática das famílias de redes discutidas na revisão.

Família	O que procura capturar	Vantagem principal	Limitação relevante neste estudo
CNN / convolução	Padrões locais que podem surgir em diferentes posições da janela.	Compartilhamento de pesos e boa detecção de variações curtas.	Sozinha, não representa explicitamente dependências longas nem relações globais entre sensores.
LSTM / GRU	Evolução sequencial e memória de observações anteriores.	Controle do que deve ser mantido ou descartado ao longo do tempo.	O processamento recorrente pode ser sensível a sequências longas e ao pequeno volume de trajetórias.
Transformer / atenção	Relações entre posições distantes e importância relativa dos instantes.	Integra contexto amplo e permite processamento paralelo.	Exige regularização e quantidade de dados suficiente para evitar sobreajuste.
GNN / GAT	Relações estruturadas entre sensores ou instantes representados como nós.	Aprende quais vizinhos contribuem mais para cada representação.	A qualidade depende de relações ou candidatos de vizinhança coerentes com o problema.
Autocodificador	Representação latente capaz de reconstruir a entrada.	Compressão, remoção de ruído e detecção de padrões diferentes do comportamento aprendido.	Não produz diretamente RUL sem uma etapa supervisionada adicional.
AGATT	Padrões locais, relações entre sensores, relações temporais e dependências de longo alcance.	Combina capacidades complementares em uma única arquitetura.	A transferência depende da compatibilidade entre degradação, sensores e domínio de origem.

Fonte: Elaborada pelo autor (2026), com base em LeCun *et al.* (1998), Greff *et al.* (2017), Vaswani *et al.* (2017), Veličković *et al.* (2018) e Zeng, Wu e Cheng (2023).

A tabela também esclarece por que a dissertação não compara as arquiteturas apenas por

complexidade. Cada família introduz um tipo de viés indutivo, isto é, uma forma de organizar o problema antes do aprendizado. No AGATT, a convolução favorece localidade, a GAT favorece relações estruturadas e o Transformer favorece dependências amplas. A hipótese experimental é que essa combinação, pré-treinada em trajetórias progressivas, produza uma representação mais útil do que uma rede treinada somente com o pequeno histórico industrial.

2.3.9 Transferência de aprendizado entre domínios

Transferência de aprendizado reutiliza conhecimento obtido em um domínio de origem para apoiar uma tarefa em um domínio de destino (Pan e Yang, 2010). A motivação é especialmente forte quando o destino possui poucos exemplos rotulados, como ocorre em ativos industriais nos quais falhas são raras e a manutenção interrompe trajetórias antes da falha funcional.

Há duas estratégias principais. Na extração de características, as camadas pré-treinadas permanecem congeladas e apenas uma nova saída é ajustada. No ajuste fino, parte dos pesos do modelo de origem também é atualizada. Camadas iniciais tendem a aprender padrões mais gerais, enquanto camadas finais são mais específicas da tarefa (Yosinski *et al.*, 2014).

Nesta dissertação, o FD001 constitui o domínio de origem e as bombas B-0302 constituem o domínio de destino. O AGATT pré-treinado é usado como extrator congelado, e uma nova cabeça de regressão aprende a converter as características em RUL industrial. O congelamento reduz a quantidade de parâmetros ajustados com o pequeno histórico de bombas.

A transferência só é defensável quando existe compatibilidade mínima entre domínios; caso contrário, o conhecimento de origem pode não produzir ganho e pode ocorrer transferência negativa (Pan e Yang, 2010; Yosinski *et al.*, 2014). Essa compatibilidade não exige que turbofan e bomba sejam fisicamente iguais, mas exige que as representações aprendidas possam responder a padrões temporais e multivariados relevantes no destino. A compatibilidade deve ser avaliada pelo mecanismo de degradação, pela estrutura dos sinais e pela tarefa, e não apenas pela semelhança nominal entre equipamentos.

Por isso, o trabalho não conecta diretamente os sensores da bomba ao modelo. É necessária uma etapa de harmonização físico-estatística que projeta os canais industriais para as 14 dimensões esperadas pelo AGATT. A projeção e sua validação sem uso da série externa são apresentadas no Capítulo 5.

2.4 BASE C-MAPSS COMO REFERÊNCIA EM PROGNÓSTICO

A C-MAPSS é uma base sintética desenvolvida pela NASA para simular a operação e a degradação de motores turbofan (Saxena *et al.*, 2008b; Saxena *et al.*, 2008a). Ela se tornou referência em RUL porque fornece trajetórias multivariadas até a falha, permitindo comparar modelos sob um protocolo comum.

Cada trajetória representa uma unidade que inicia em condição saudável e evolui até o fim de vida (Saxena *et al.*, 2008b; Saxena *et al.*, 2008a). Em cada ciclo, são registradas variáveis operacionais e sensores. O conjunto de treinamento contém a trajetória completa; no conjunto de teste, a sequência é interrompida antes da falha, e o objetivo é estimar o RUL no último ciclo observado.

A base possui quatro subconjuntos (Saxena *et al.*, 2008b). O FD001 apresenta uma condição operacional e um modo de falha; o FD002, seis condições e um modo; o FD003, uma condição e dois modos; e o FD004, seis condições e dois modos. O FD001 é um ponto de partida controlado: a menor variabilidade operacional facilita verificar se a arquitetura aprende a progressão de degradação antes de enfrentar cenários mais heterogêneos.

A simplicidade do FD001 também impõe limites. O modelo aprende representações relacionadas a um processo progressivo em um ambiente simulado. As bombas B-0302 operam em rodízio, possuem sensores de naturezas diferentes e apresentam modos de falha heterogêneos. Portanto, o FD001 é usado como domínio de origem e *benchmark*, não como réplica física das bombas.

O Capítulo 4 detalha a estrutura dos dados, a construção do alvo, o janelamento, a normalização e a reprodução do AGATT. Nesta revisão, o papel da C-MAPSS é estabelecer a ponte conceitual entre a literatura de prognóstico e a proposta de transferência para o ativo industrial.

2.5 LEVANTAMENTO EXPLORATÓRIO

Antes da execução da Pesquisa Bibliográfica Sistemática, foi realizado um levantamento exploratório da literatura sobre previsão de RUL em bases de referência, especialmente a C-MAPSS. Esse levantamento teve caráter inicial e serviu para identificar arquiteturas recorrentes, métricas utilizadas e resultados reportados em estudos recentes. Além disso, a etapa exploratória ajudou a delimitar quais famílias de modelos deveriam receber maior atenção na revisão: redes convolucionais, redes recorrentes, autocodificadores, mecanismos de atenção, *Transformers* e arquiteturas híbridas.

O levantamento indicou que a C-MAPSS é o *benchmark* mais consolidado em previsão de RUL para equipamentos de grande porte, especialmente por permitir comparação direta entre modelos em sub-bases padronizadas, como a FD001 (Saxena *et al.*, 2008b). Embora essa base modele motores *turbofan*, e não bombas industriais, os padrões de degradação em séries temporais multivariadas tornam possível investigar estratégias de transferência de aprendizado para ativos reais, desde que sejam reconhecidas as diferenças físicas, operacionais e instrumentais entre os domínios (Lei *et al.*, 2022).

Entre os trabalhos inicialmente identificados, Ragab *et al.* (2021) propôs o modelo *ATS2S* (*Attention-based Sequence-to-Sequence model*, modelo sequência para sequência com atenção), combinando perda de reconstrução e perda de RUL. O modelo superou diferentes abordagens

em bases reais e reforçou a relevância de mecanismos de atenção em cenários com múltiplas condições operacionais. No mesmo contexto, Soni, Kansara e Joshi (2020) comparou redes convolucionais e modelos LSTM para previsão de RUL, observando desempenho competitivo das CNNs em determinadas configurações experimentais.

Outros estudos reforçaram a diversidade de estratégias em aprendizado profundo para prognóstico. Wang *et al.* (2021) empregaram pares adversários com o algoritmo *RMSprop* para estabilizar o treinamento de redes profundas, enquanto Namuduri *et al.* (2020) revisaram algoritmos de *deep learning* aplicados à manutenção preditiva. Wang, Zhao e Addepalli (2020) destacaram o uso de autocodificadores, DBN, CNN e RNN como alternativas a modelos puramente físicos na previsão do RUL, e Pang *et al.* (2021) analisaram técnicas de aprendizado profundo para detecção de anomalias em dados complexos.

Trabalhos posteriores passaram a combinar diferentes blocos arquiteturais. Al-Khazraji *et al.* (2022) integraram *autoencoder* e DBN para extrair características abstratas em motores aeronáuticos. Li, Wang e Li (2022) combinaram CNN, LSTM e o módulo CBAM (*Convolutional Block Attention Module*), reduzindo o RMSE e melhorando o *score* em relação a referências anteriores. Hu, Zhao e Ren (2023) combinaram LSTM, CNN e *Transformer*, reportando RMSE de 3,52 ciclos na FD001. Por estar muito abaixo da faixa observada em vários estudos semelhantes, esse valor passou a demandar análise estatística mais cuidadosa, motivando a formulação das perguntas da PBS apresentadas na seção seguinte.

Também foi identificado o modelo AGATT, proposto por Zeng, Wu e Cheng (2023), com RMSE reportado de 11,45 ciclos na FD001 da C-MAPSS. O AGATT mostrou-se particularmente relevante para esta dissertação por combinar mecanismos de atenção, redes de atenção em grafos e *Transformers*, permitindo capturar dependências temporais e espaciais em dados multissensores.

2.6 PESQUISA BIBLIOGRÁFICA SISTEMÁTICA (PBS)

Ao ser realizada uma busca exploratória inicial na área, apresentada na seção anterior, permaneceram em aberto algumas questões centrais para a delimitação do estado da arte em previsão de RUL na base C-MAPSS, especialmente na sub-base FD001. Assim, esta Pesquisa Bibliográfica Sistemática (PBS) foi conduzida com o objetivo de responder às seguintes perguntas primárias de pesquisa:

- O valor de 3,52 ciclos de RMSE reportado na sub-base FD001 da base C-MAPSS, conforme reportado por Hu, Zhao e Ren (2023), pode ser considerado representativo da fronteira recente de desempenho encontrada na literatura?;
- Esse valor constitui um *outlier* estatístico em relação aos demais resultados reportados?;
- e
- A utilização de arquiteturas baseadas em *Transformers* reflete uma das principais frentes do estado da arte recente para previsão de RUL na FD001?

2.6.1 Palavras-chave, fraseamento e campo a ser pesquisado

As palavras-chave básicas utilizadas nesta pesquisa foram “remaining useful life” (vida útil remanescente) e “cmapss” ou “c-mapss”. As expressões C-MAPSS e CMAPSS foram consideradas equivalentes para fins de busca, pois ambas aparecem na literatura para se referir ao mesmo conjunto de dados.

A frase completa utilizada foi:

[“remaining useful life”] AND [“c-mapss” OR “cmapss”]

Essas palavras e expressões foram definidas a partir da busca exploratória inicial, uma vez que permitem identificar publicações voltadas à previsão de RUL utilizando a base C-MAPSS. A busca foi realizada no campo “Abstract” das publicações.

2.6.2 Bases de busca de referências, justificativa e período de busca

A busca foi realizada em cinco sistemas de referência amplamente utilizados em ciência da computação e engenharia: Association for Computing Machinery (ACM), SCOPUS, Google Scholar¹, Portal de Periódicos CAPES e IEEE Xplore.

A escolha dessas bases se justifica por sua ampla utilização em ciência da computação, engenharia e áreas correlatas, aumentando a probabilidade de recuperação de publicações relevantes e metodologicamente qualificadas.

A busca abrangeu publicações realizadas a partir de janeiro de 2017, ano em que o artigo “Attention Is All You Need”, de Vaswani *et al.* (2017), apresentou a arquitetura Transformer, até 2025. A busca foi realizada em cinco de fevereiro de 2026.

2.6.3 Critérios de seleção, inclusão, exclusão e forma de combinação

Os critérios objetivos para seleção dos artigos foram: possuir a frase de busca em seu *abstract*; ser artigo de periódico ou congresso; e ser possível acessar o material de forma gratuita pelo autor, considerando também os acessos institucionais disponibilizados pela universidade. Foram admitidos apenas artigos em inglês, português ou francês, idiomas dominados pelo autor.

O critério subjetivo de inclusão foi admitir trabalhos considerados promissores do ponto de vista de desempenho em RMSE, relevância arquitetural ou explicabilidade do modelo.

Os critérios de exclusão foram os seguintes:

- Caso o número de artigos encontrados fosse superior a 500, restringiu-se o período de publicação aos anos mais recentes até que o volume fosse reduzido para aproximadamente 500 artigos;

¹ Foi utilizado um script que busca automaticamente a frase escolhida na base do Google Scholar e retorna as 100 publicações com maior número de citações. A partir dessa planilha, foram filtradas as 10 publicações mais citadas em termos absolutos, as 10 mais citadas em termos relativos e as 10 mais recentes. Como esses três conjuntos apresentaram intersecções, o número final de referências foi menor que 30. O script utilizado está disponível em <<https://github.com/WittmannF/sort-google-scholar>>.

- A justificativa foi que artigos mais recentes tendem a refletir melhor o estado da arte vigente. Além disso, considerando a capacidade de análise do autor e o tempo disponível, o número aproximado de 500 artigos foi considerado adequado para a viabilidade dos passos seguintes.
- Artigos com acesso restrito, que exigiam custos financeiros adicionais para consulta, não foram considerados;
- Após a redução para aproximadamente 500 artigos, realizou-se uma análise para identificar quais trabalhos apresentavam resultados de RMSE na sub-base FD001 do conjunto de dados C-MAPSS;
 - Artigos que não apresentavam resultados de RMSE especificamente para a sub-base FD001 foram excluídos.
- Foram selecionados apenas artigos que utilizaram *Transformers* ou arquiteturas híbridas nas quais mecanismos de atenção do tipo Transformer desempenhassem papel central na previsão de RUL;
 - Essa filtragem foi realizada para permitir a análise estatística da representatividade dessa família de modelos em relação às demais e comparar seu desempenho com outras abordagens de aprendizado de máquina.
- Por fim, foram selecionados apenas os artigos com RMSE menor ou igual ao valor máximo do primeiro quartil (Q_1) e maior ou igual ao menor valor acima da fronteira inferior de *outliers*², definida por $Q_1 - 1,5 \cdot IQR$ ³, em que $IQR = Q_3 - Q_1$.

Primeiramente, a pesquisa foi executada nas bases indicadas. Em seguida, adicionaram-se os artigos identificados como relevantes pelo critério subjetivo de inclusão. Depois disso, aplicaram-se os critérios de exclusão descritos.

2.6.4 Execução da PBS

A Figura 2 apresenta o resultado da busca nas bases definidas. Ao todo, foram encontradas 827 publicações com a frase de busca em seu *abstract*, desde 2017 até 2025. O diagrama de Venn apresenta de quais bases de pesquisa vieram as publicações analisadas. Também são apresentadas, abaixo do diagrama, a frase de pesquisa e a quantidade de referências encontradas. A figura ilustra o cenário inicial da PBS, antes da aplicação dos critérios de inclusão e exclusão.

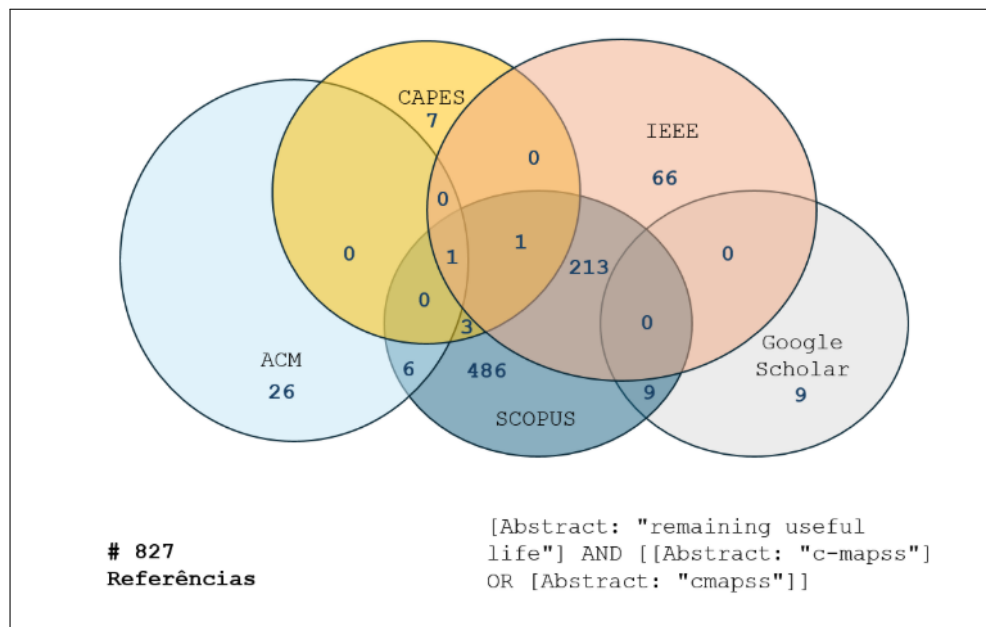
² *Outliers* são observações que apresentam valores significativamente distantes da maior parte dos dados, podendo resultar de variabilidade natural, erros de medição ou eventos raros. Sua identificação é importante, pois esses valores podem influenciar análises estatísticas e o desempenho de modelos de aprendizado de máquina.

³ O IQR (*Interquartile Range*, ou Amplitude Interquartil) é uma medida de dispersão definida pela diferença entre o terceiro quartil (Q_3) e o primeiro quartil (Q_1), isto é, $IQR = Q_3 - Q_1$. É amplamente utilizado para identificar *outliers*, considerando-se, em geral, como valores atípicos aqueles inferiores a $Q_1 - 1,5 \times IQR$ ou superiores a $Q_3 + 1,5 \times IQR$.

O artigo de (Zeng, Wu e Cheng, 2023) foi posteriormente incluído por critério subjetivo, pois apresenta o AGATT, modelo que combina mecanismos de atenção, redes de atenção em grafos e Transformer, oferecendo maior potencial de explicação sobre a influência de sensores e ciclos na previsão do RUL.

Conforme definido no primeiro critério de exclusão, como foram encontradas mais de 500 referências na primeira seleção, considerou-se apenas o período necessário para reduzir o conjunto a aproximadamente 500 artigos. Assim, a análise aprofundada concentrou-se nas publicações de 2023 a 2025.

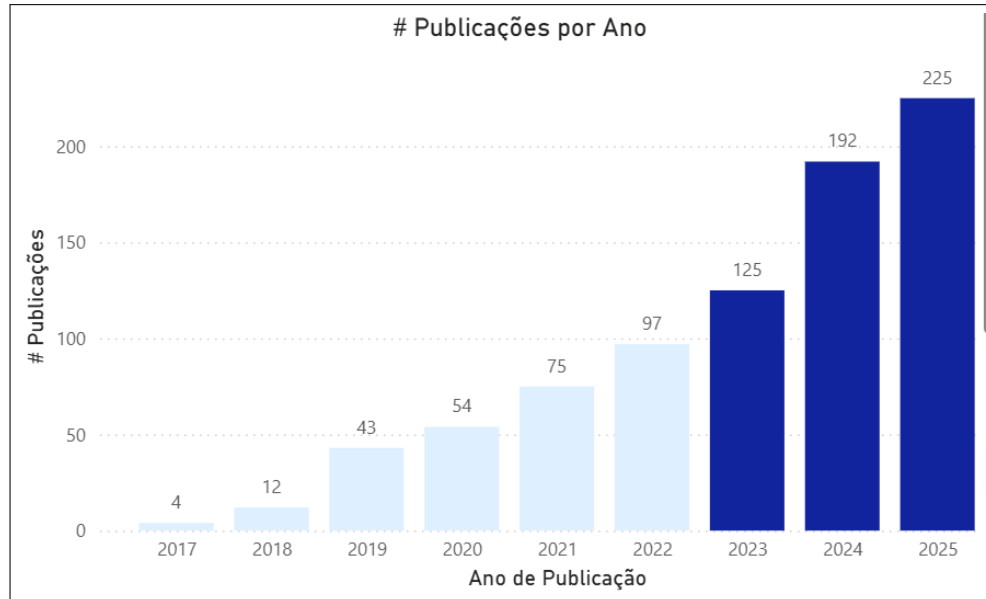
Figura 2 – O diagrama de Venn apresenta de quais bases de pesquisa vieram as publicações analisadas. São apresentadas, abaixo do diagrama, a frase de pesquisa e a quantidade de referências encontradas.



Fonte: Elaborada pelo autor (2026).

A Figura 3 apresenta a evolução da quantidade de publicações na área ao longo dos anos. Foram selecionadas as publicações mais recentes, a partir de 2023, por se entender que essas tenderiam a refletir melhor o estado da arte vigente. Das 827 publicações iniciais, 542 foram selecionadas para a etapa seguinte.

Figura 3 – Distribuição das publicações ao longo dos anos. Ao se selecionar o período de 2023 em diante, obteve-se um total de 542 publicações.



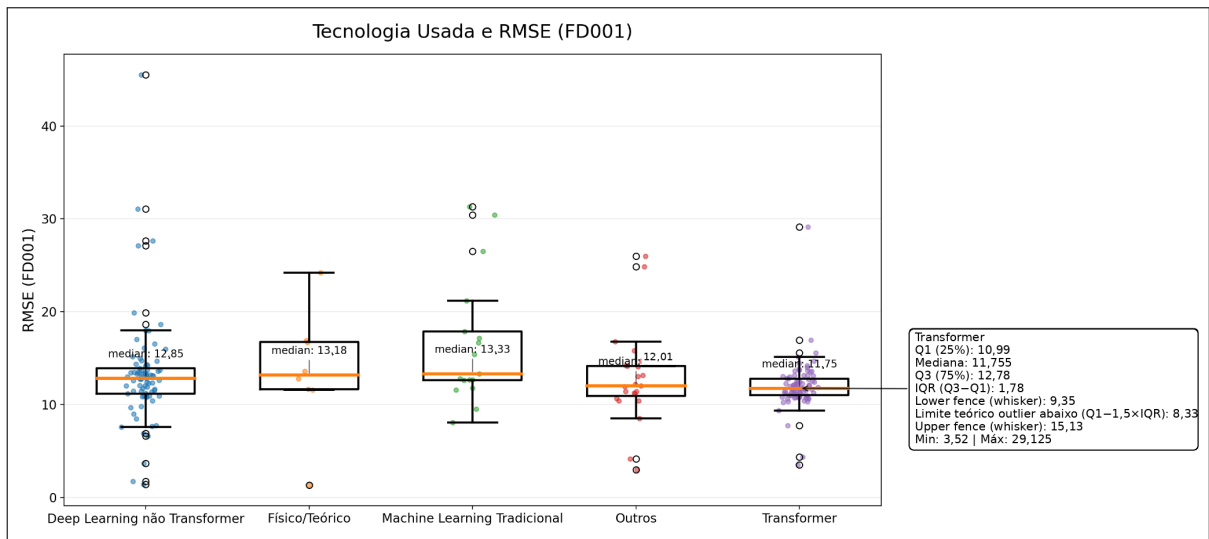
Fonte: Elaborada pelo autor (2026).

Das 542 publicações filtradas, após a leitura e tabulação dos resultados obtidos na sub-base FD001 da base C-MAPSS, 310 estavam disponíveis sem custo, e dessas, 227 apresentaram resultados de RMSE.

A Figura 4 apresenta a distribuição dos valores de RMSE na FD001, em que valores menores indicam melhor desempenho, estratificada por categoria de *Tecnologia Usada*. Observa-se que o grupo *Transformer* concentra os menores erros e a menor dispersão: a mediana é aproximadamente 11,76, com intervalo interquartil estreito ($Q1 \approx 10,99$ e $Q3 \approx 12,78$), indicando maior consistência entre os trabalhos. Ainda assim, o *boxplot* evidencia a presença de *outliers*, com valores extremos acima do limite superior e também abaixo do limite inferior, como o valor de 3,52 ciclos reportado em (Hu, Zhao e Ren, 2023). Isso sugere que parte dos estudos reporta resultados significativamente diferentes do padrão do grupo.

Em contraste, as demais categorias, como *Deep Learning não Transformer*, *Machine Learning Tradicional*, *Físico/Teórico* e *Outros*, apresentam medianas tipicamente mais altas e maior variabilidade, com diversos pontos de *outlier*. Esses padrões reforçam que, no conjunto analisado, abordagens baseadas em *Transformers* tendem a alcançar melhor desempenho mediano e maior robustez relativa na estimativa de RUL na FD001, embora existam casos isolados de desempenho atípico em todas as categorias.

Figura 4 – Observa-se que o uso de *Transformers* proporciona resultados superiores na previsão de falhas na base FD001. A mediana do RMSE obtido em modelos baseados em *Transformers* é de 11,76 ciclos. Q1 = 10,99 ciclos, Q3 = 12,78 ciclos e o menor valor não *outlier* é 9,35 ciclos.



Fonte: Elaborada pelo autor (2026).

A Figura 5 apresenta a distribuição de publicações por *Tecnologia Usada* no período de 2023 a 2025. Observa-se que os trabalhos foram organizados em cinco grupos: *Transformer*, que inclui arquiteturas baseadas em Transformers e variantes com mecanismos de atenção; *Deep Learning não Transformer*, que reúne modelos de redes neurais profundas sem Transformer, como CNN, RNN, GRU, LSTM, BiLSTM e *Autoencoders*; *Outros*, com abordagens diversas que não se enquadram nos grupos anteriores; *Machine Learning Tradicional*, com métodos clássicos, como regressões, SVM e ensembles; e *Físico/Teórico*, com modelagens baseadas em princípios físicos ou formulações teóricas.

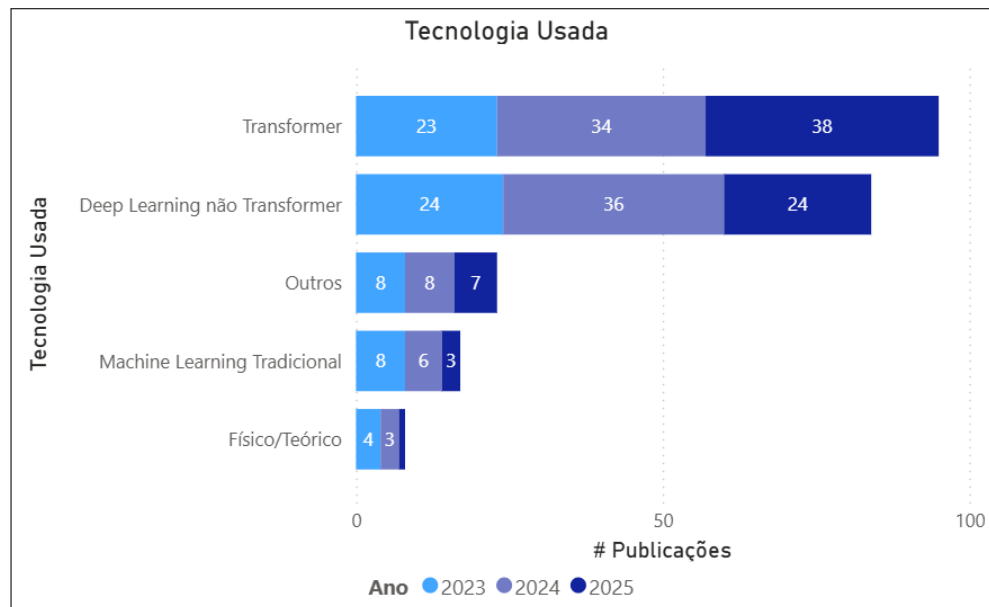
Em termos de volume, o grupo *Transformer* totaliza 95 artigos (23 em 2023, 34 em 2024 e 38 em 2025), enquanto o grupo *Deep Learning não Transformer* soma 84 artigos (24 em 2023, 36 em 2024 e 24 em 2025). As categorias *Outros*, *Machine Learning Tradicional* e *Físico/Teórico* contabilizam, respectivamente, 23 artigos, 17 artigos e oito artigos.

Quanto à evolução temporal, nota-se crescimento consistente do grupo *Transformer* ao longo dos anos (23 → 34 → 38), indicando aumento contínuo de adoção dessa família de modelos. Já *Deep Learning não Transformer* apresenta pico em 2024 (24 → 36), seguido de redução em 2025 (36 → 24), retornando ao nível de 2023. O grupo *Outros* permanece relativamente estável. Por fim, *Machine Learning Tradicional* e *Físico/Teórico* exibem tendência de diminuição ao longo do período, sugerindo perda de participação relativa frente às abordagens baseadas em *deep learning*, sobretudo Transformers.

Dos 95 artigos classificados como *Transformers*, 20 foram selecionados por apresentarem RMSE entre 9,35 ciclos, menor valor não *outlier*, e 10,99 ciclos, correspondente ao primeiro quartil (Q1), conforme mostrado na Figura 6. Adicionalmente, o AGATT (Zeng, Wu e Cheng, 2023) foi incluído por sua relevância arquitetural e interpretabilidade, totalizando 21 modelos

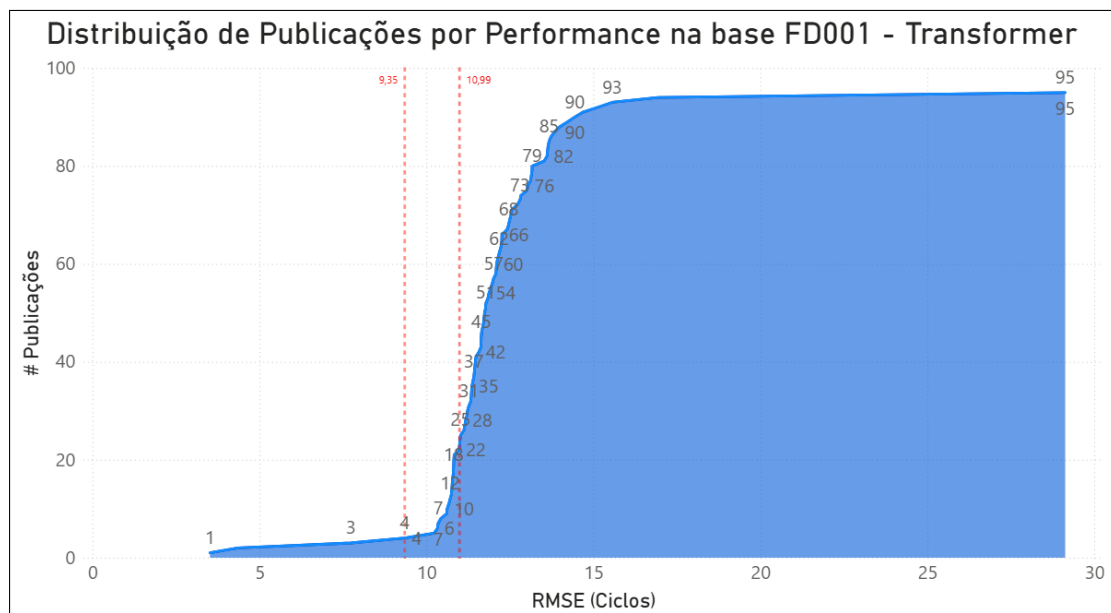
analisados na síntese final.

Figura 5 – Distribuição das publicações por tecnologia e ano. *Transformers* representam 95 publicações no total, com número crescente de publicações ao longo dos anos.



Fonte: Elaborada pelo autor (2026).

Figura 6 – Ao selecionar apenas as publicações com resultados de RMSE entre 9,35 e 10,99 ciclos, obtiveram-se 20 publicações, ampliadas para 21 modelos considerando o critério subjetivo de inclusão do AGATT.

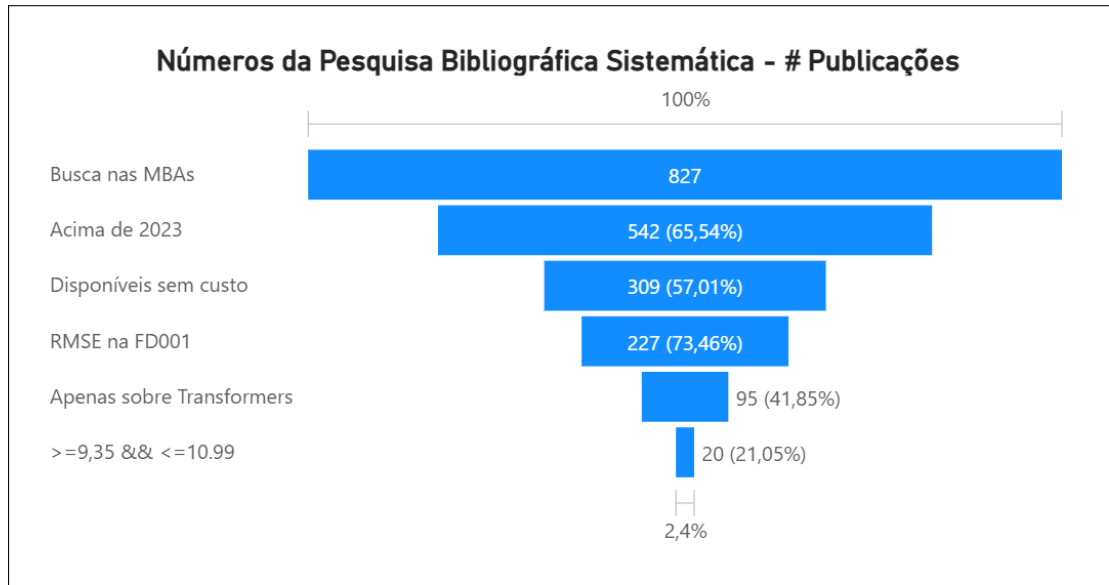


Fonte: Elaborada pelo autor (2026).

A Figura 7 apresenta um resumo de como ocorreu toda a PBS. Ao final, de 827 publicações iniciais, foram selecionadas 20 publicações recentes, baseadas em *Transformers*, com resultados de RMSE situados na faixa não *outlier* entre 9,35 e 10,99 ciclos. Adicionalmente,

o AGATT foi incluído por critério de relevância arquitetural e explicabilidade, totalizando 21 modelos analisados na síntese final.

Figura 7 – Resumo da pesquisa bibliográfica sistemática, com o número de publicações remanescentes em cada etapa. A pirâmide mostra o total de publicações e a porcentagem correspondente em relação à etapa anterior.



Fonte: Elaborada pelo autor (2026).

2.6.5 Síntese Analítica dos Modelos Selecionados na PBS

Ao concluir a fase de seleção e filtragem da PBS, foram identificados 21 modelos que satisfazem critérios quantitativos de desempenho e representam a fronteira recente da aplicação de redes neurais profundas, especialmente arquiteturas baseadas em mecanismos de atenção, para prognóstico de falhas em motores aeronáuticos na base FD001.

A análise consolidada evidencia que os modelos baseados em mecanismos de atenção, especialmente arquiteturas fundamentadas em *Transformers*, apresentam, no recorte analisado, os menores valores medianos de RMSE e menor dispersão relativa. Tal constatação reforça a consolidação do paradigma atencional como uma das principais frentes do estado da arte recente para modelagem de dependências temporais de longo alcance em séries multivariadas degradantes.

A Tabela 2 apresenta uma comparação estruturada entre os 21 modelos selecionados, incluindo métricas de desempenho, características arquiteturais e principais decisões experimentais. LR⁴ significa taxa de aprendizado, do inglês *Learning Rate* e cap_RUL⁵ pode ser entendido

⁴ A *Learning Rate* (LR), ou taxa de aprendizado, é um hiperparâmetro que controla a magnitude das atualizações dos pesos do modelo durante o treinamento. Valores muito elevados podem impedir a convergência ou tornar o treinamento instável, enquanto valores muito baixos podem resultar em convergência lenta ou em mínimos locais indesejados.

⁵ O *cap_RUL* é uma técnica de pré-processamento que estabelece um limite superior para os valores da Vida Útil Remanescente (RUL). Assim, todas as observações com RUL superior ao valor definido são truncadas para esse limite, reduzindo a influência de valores muito elevados e concentrando o aprendizado do modelo na região

como limite superior.

Tabela 2 – Comparação consolidada dos 21 modelos avaliados na base FD001 do C-MAPSS.

Modelo	RMSE	Arquitetura	Heads	Camadas	Janela	LR	cap_RUL	Referência
AGATT	11,45	GAT + Transformer	4	2	30	1e-4	125	(Zeng, Wu e Cheng, 2023)
UQ-DeepRUL	10,99	CNN + Transformer	4	2	30	1e-4	125	(Yang <i>et al.</i> , 2025)
BiLSTM-DAE-TF	10,98	DAE + BiLSTM + Transformer	4	2	30	1e-4	125	(Fan, Li e Chang, 2023)
DDP-NN	10,83	Multi-scale + Frequency features	-	2	30	1e-4	125	(Zhang <i>et al.</i> , 2025b)
HMDAM	10,82	CNN + Dual Attention	4	2	30	1e-4	125	(He <i>et al.</i> , 2025)
DATCN-PSOSEN	10,82	TCN + PSO	-	-	30	1e-4	125	(Fan <i>et al.</i> , 2024)
Global-Attn DL	10,81	Global Attention	-	-	30	1e-4	125	(Xu <i>et al.</i> , 2023)
Bayesian-CPT	10,80	Probabilistic Transformer	4	2	30	1e-4	125	(Xie <i>et al.</i> , 2025)
CBAM	10,78	CNN + CBAM	-	-	30	1e-4	125	(Li, Cao e Ma, 2023)
Dual-Scale Transformer	10,77	Dual-scale Transformer	4	2	30	1e-4	125	(Li <i>et al.</i> , 2024a)
Light-TF-DSC	10,76	Depthwise-separable Conv + Transformer	4	2	30	1e-4	125	(Li <i>et al.</i> , 2024b)
Res-TCEANet	10,75	Semi-supervised TCN + Attention	4	2	30	1e-4	125	(Wang e Huang, 2025)
Dual-Path TSD	10,71	Time-series Decomposition + Transformer	4	2	30	1e-4	125	(Wu <i>et al.</i> , 2025)
DL-Stoch-Ens	10,67	Stochastic Ensemble	-	-	30	1e-4	125	(Faizanbasha e Rizwan, 2025)
AM-TCN	10,62	Attention Multi-scale TCN	-	-	30	2e-4	125	(Xu, Zhang e Miao, 2024)
TS-H-TF	10,61	Hierarchical Transformer	8	2	30	1e-4	125	(Fan, Li e Chang, 2024)
TCRSCANet	10,43	TCN + Recurrent Skip component	8	-	30	1e-4	125	(Wahid, Breslin e Intizar, 2024)
TEMHSA	10,35	Time-enhanced Multi-head Self-attention	8	2-4	40	tuning	125	(Wang <i>et al.</i> , 2023)
STF2	10,33	Spatio-temporal Feature Fusion	4	2	30	1e-4	125	(Du <i>et al.</i> , 2025)
DS-GNN	10,23	Dynamic Spatio-temporal Graph	4	3	50	1e-2	125	(Zhang <i>et al.</i> , 2025a)
MCA-DTCN	9,35	Dual-task TCN + Multi-channel Attention	-	-	30	1e-3	125	(Fu <i>et al.</i> , 2024)

Fonte: Elaborada pelo autor (2026), com base nos estudos selecionados na PBS.

A fronteira tecnológica delineada pelos 21 modelos selecionados na PBS evidencia que a previsão de RUL, particularmente na base FD001 do C-MAPSS, deixou de ser tratada apenas como um problema de regressão temporal direta para tornar-se, na prática, um problema de modelagem de sistemas complexos com dependências espaço-temporais, variação operacional, ruído e não estacionariedade. A evolução observada na literatura não se resume à substituição incremental de arquiteturas; ela reflete uma mudança de pressupostos: enquanto abordagens clássicas e parte das redes recorrentes priorizavam a sequência histórica como eixo principal de explicação, os modelos contemporâneos assumem que a degradação emerge de interações entre variáveis e subsistemas, frequentemente mediadas por regimes operacionais e acoplamentos físicos não lineares.

Nesse contexto, o AGATT destaca-se não por apresentar o menor RMSE do conjunto, mas por materializar uma postura metodológica alinhada à tendência dominante observada na literatura: a modelagem explícita do espaço, representada pela correlação entre sensores e subsistemas, combinada com a modelagem do tempo, representada pela evolução das assinaturas de degradação. A atratividade do AGATT, do ponto de vista científico e aplicado, decorre de dois fatores complementares: primeiro, o modelo reconhece que, em séries multivariadas de sensores, a informação de degradação raramente está isolada em um único canal; segundo, oferece um caminho natural para interpretabilidade operacional, pois a atenção em grafos permite investigar quais conexões entre sensores ganham ou perdem importância ao longo da trajetória de degradação.

Ao observar os demais modelos da Tabela 2, torna-se evidente que a hibridização é a norma para atingir RMSEs competitivos na FD001. A razão central para isso é que Transformers, embora potentes, são arquiteturas com baixa indutividade estrutural quando aplicadas diretamente a séries temporais industriais. Assim, a literatura recorre a módulos com vieses indutivos

próxima à falha, que normalmente apresenta maior relevância prática.

claros para estabilizar e orientar o aprendizado. Camadas convolucionais (CNN) e convoluções temporais (TCN) aparecem como filtros de indução local: capturam padrões locais, suavizam ruído e impõem uma noção de localidade temporal, ajudando a rede a distinguir variações de alta frequência de tendências persistentes de degradação.

Modelos como o MCA-DTCN, ao combinar TCN com mecanismos de atenção em múltiplos canais e estruturar o problema como tarefa dupla, isto é, detecção do primeiro instante preditivo de degradação e regressão do RUL, introduzem uma forma explícita de regularização semântica. O fato de o MCA-DTCN alcançar o menor RMSE não *outlier* observado, 9,35 ciclos, sugere que ganhos relevantes podem ocorrer quando o modelo é forçado a aprender uma decomposição mais estruturada do problema, em vez de tentar resolver toda a tarefa por regressão direta.

Outra vertente recorrente é a decomposição e a modelagem multiescala. Em aplicações reais, engenheiros observam o comportamento dos equipamentos em múltiplas escalas: picos curtos podem sinalizar anomalias transitórias, enquanto tendências longas podem refletir degradação contínua. Modelos hierárquicos e multiescala tentam internalizar essa leitura, avaliando simultaneamente sinais de curto prazo e tendências acumuladas. Essa classe de modelos é especialmente relevante quando se considera a generalização fora do domínio sintético, pois a habilidade de separar tendências robustas de ruído e variações de regime é um pré-requisito para o uso de *transfer learning* em ativos industriais reais.

A exploração do domínio da frequência também aparece como direção promissora para detecção precoce e robustez. Muitos fenômenos de desgaste em sistemas rotativos têm assinaturas espectrais antes de refletirem alterações no valor médio ou em tendências simples no tempo. Quando modelos incorporam características no domínio da frequência, ampliam a base de evidência utilizada para inferência de degradação. Entretanto, essa escolha também traz desafios, pois componentes espectrais podem ser sensíveis a janelamento, resolução e variações de regime operacional.

A presença de modelos probabilísticos e de quantificação de incerteza representa outro amadurecimento do estado da arte. Quando o objetivo deixa de ser apenas minimizar RMSE e passa a incluir confiabilidade da decisão, a incerteza torna-se parte relevante do produto científico. Em ambientes críticos, duas previsões com RMSE semelhante podem ter valores operacionais distintos se uma delas for calibrada e reconhecer situações de maior incerteza. Esse ponto reforça que métricas pontuais de erro, isoladamente, não capturam prontidão industrial.

Os hiperparâmetros da Tabela 2 permitem identificar recorrências no corpus, mas não estabelecer valores universalmente superiores. Entre os estudos que informaram essas escolhas, são frequentes configurações com poucas camadas, múltiplas cabeças de atenção, janelas próximas de 30 ciclos e truncamento do RUL em 125 ciclos. Essas recorrências podem refletir tanto propriedades do FD001 quanto práticas herdadas de trabalhos anteriores. Assim, não se conclui que quatro a oito cabeças, taxa de aprendizado próxima de 10^{-4} ou janela de 30 ciclos sejam ótimos em geral. A leitura defensável é que esses valores formam uma região experimental

recorrente, dentro da qual cada artigo realizou escolhas adicionais de arquitetura, regularização e validação. Nesta dissertação, os hiperparâmetros são justificados pelo artigo-base, pela reprodução realizada e pelas análises de sensibilidade, e não apenas por frequência de ocorrência na literatura.

Sob uma lente crítica, o conjunto de modelos selecionados mostra que a fronteira tecnológica em FD001 está menos associada a uma arquitetura vencedora única e mais a um conjunto de princípios: impor vieses indutivos adequados, usar atenção para integrar dependências de longo alcance, estabilizar o aprendizado por protocolos padronizados e, crescentemente, incorporar incerteza, explicabilidade e robustez. Nesse panorama, a escolha do AGATT como modelo de referência torna-se tecnicamente justificada não por maximizar desempenho em RMSE, mas por oferecer equilíbrio entre desempenho competitivo e interpretabilidade orientada ao domínio, característica relevante para a transferência posterior ao contexto das bombas B-0302.

2.6.6 Discussão e conclusão sobre a PBS

- **O valor de 3,52 ciclos de RMSE reportado na sub-base FD001 da base C-MAPSS, conforme reportado por Hu, Zhao e Ren (2023), pode ser considerado representativo da fronteira recente de desempenho da literatura?**
 - **Resposta:** O valor de 3,52 ciclos pode ser interpretado como um dos menores valores pontuais reportados no conjunto analisado, porém não deve ser tratado como resultado representativo da fronteira recente de desempenho. Pela análise interquartil, esse valor fica abaixo da fronteira inferior de *outliers*, sendo estatisticamente discrepante em relação à distribuição dos modelos baseados em *Transformers*. Assim, a comparação principal desta PBS considera a faixa não *outlier*, cujo menor valor observado foi 9,35 ciclos.
- **Esse valor constitui um *outlier* estatístico em relação aos demais resultados reportados?**
 - **Resposta:** Sim. A mediana de RMSE obtida pelos 95 artigos selecionados nos últimos três anos que utilizam a tecnologia *Transformer* foi de aproximadamente 11,76 ciclos. O primeiro quartil foi $Q1 = 10,99$ ciclos e o terceiro quartil foi $Q3 = 12,78$ ciclos, resultando em $IQR = 1,79$. A fronteira inferior de *outliers*, calculada por $Q1 - 1,5 \cdot IQR$, foi aproximadamente 8,31 ciclos. Como o valor de 3,52 ciclos está abaixo dessa fronteira, ele foi tratado como *outlier*.
- **A utilização de arquiteturas baseadas em *Transformers* reflete uma das principais frentes do estado da arte recente para previsão de RUL na FD001?**
 - **Resposta:** Sim. Conforme os resultados apurados e apresentados na Figura 4, os modelos baseados em *Transformers* apresentaram, no recorte analisado, os

menores valores medianos de RMSE e menor dispersão relativa. Isso indica que essa família de arquiteturas representa uma das principais frentes do estado da arte recente para previsão de RUL na FD001 da C-MAPSS.

2.6.7 Riscos, limitações e formas de validação da PBS

Algumas publicações excluídas por exigirem acesso pago podem ter obtido resultados significativos de RMSE na sub-base FD001. Também é possível que publicações em idiomas não cobertos, como chinês, ou estudos que não apresentaram resultados específicos para a FD001, mas que poderiam ser aplicados a ela, tenham alcançado desempenho relevante.

Outra limitação é que a PBS se baseia em valores de RMSE reportados pelos próprios autores. Muitos estudos apresentam apenas resultados pontuais, sem repetir execuções, divulgar sementes aleatórias, intervalos de confiança ou estatísticas de variabilidade. Assim, diferenças pequenas de RMSE entre modelos devem ser interpretadas com cautela, pois podem refletir sensibilidade de implementação, escolha de hiperparâmetros, seleção da melhor execução ou diferenças sutis de protocolo experimental.

Como forma de mitigação, esta PBS adotou análise interquartil para evitar que valores extremos fossem tratados automaticamente como representativos do estado da arte. Além disso, a seleção final considerou não apenas o menor RMSE, mas também a coerência estatística do resultado, a relevância arquitetural do modelo e sua adequação ao objetivo aplicado desta dissertação. Essa estratégia justifica a inclusão do AGATT mesmo com RMSE acima do primeiro quartil, pois sua arquitetura combina mecanismos temporais, espaciais e de atenção em grafos, alinhando-se à proposta de transferência de aprendizado para bombas industriais.

3 ENGENHARIA DE CONFIABILIDADE NA B-0302

Este capítulo caracteriza, de forma exploratória, o histórico das Bombas B-0302A, B-0302B, B-0302C e B-0302D a partir das ordens de manutenção registradas no sistema de gestão. O objetivo é comparar intervalos entre eventos, modos de ocorrência e ajustes de confiabilidade, fornecendo a base física para interpretar a transferência de aprendizado do Capítulo 5.

3.1 FONTE DOS DADOS E PROCESSAMENTO DOS REGISTROS

Os dados utilizados neste capítulo foram extraídos de ordens de manutenção do sistema corporativo de gestão. Essas ordens contêm informações como equipamento, data, descrição da ocorrência, tipo de intervenção, comentários técnicos, início e fim da manutenção e registros associados à execução do serviço.

Como registros de manutenção industrial são frequentemente escritos em linguagem natural, com variações de termos, abreviações e níveis diferentes de detalhamento, foi desenvolvido um procedimento automatizado em Python para varrer a base de ordens e identificar registros relacionados ao conjunto B-0302. O processamento incluiu filtros por equipamento, seleção de ordens pertinentes e classificação textual das ocorrências.

A classificação dos registros foi realizada com apoio de expressões regulares (*regex*). Essa técnica permitiu identificar padrões recorrentes nos campos de descrição e motivo da ordem, associando cada ocorrência a categorias técnicas previamente mapeadas, como falhas elétricas, falhas de vedação, falhas de vibração, falhas de controle, falhas de instrumentação e intervenções preventivas. O uso de regras automatizadas reduziu a necessidade de classificação manual e aumentou a consistência do tratamento dos dados.

Foram consideradas tanto ordens corretivas quanto preventivas, desde que associadas a eventos relevantes para o comportamento de confiabilidade do conjunto. Essa decisão permite uma visão mais ampla do histórico operacional, mas também exige cautela na interpretação: nem toda ordem preventiva representa uma falha funcional, e nem toda intervenção corretiva possui o mesmo grau de severidade. Por isso, as análises apresentadas neste capítulo devem ser entendidas como uma caracterização exploratória e aplicada da confiabilidade do ativo, e não como um ensaio estatístico controlado.

Após a classificação, foram consolidados por bomba os modos de ocorrência, os tempos entre eventos tratados como TTF, sua média operacional, os tempos de reparo disponíveis, as confiabilidades estimadas para 90 e 180 dias e os parâmetros dos modelos Exponencial e Weibull. Esses indicadores não devem ser confundidos com os modos de falha: TTF, MTBF e MTTR são medidas temporais, enquanto o modo de falha descreve como o equipamento perde ou reduz sua função.

3.2 INDICADORES DE CONFIABILIDADE

A análise de confiabilidade deste capítulo utiliza como base os tempos até a falha observados para cada bomba. Para cada unidade do conjunto B-0302, as datas de ocorrência foram ordenadas cronologicamente e os intervalos entre eventos consecutivos foram calculados em dias. Esses intervalos foram tratados como valores de TTF.

Neste estudo, o TTF representa o intervalo de calendário entre dois eventos consecutivos classificados como relevantes para o comportamento funcional da bomba. A média desses intervalos foi usada como aproximação exploratória do MTBF operacional. As bombas possuem horímetros, mas não foi possível obter, para todo o período, um histórico contínuo e sincronizado de horas por unidade que pudesse ser correlacionado de forma segura às séries de sensores e às datas das ocorrências. Por isso, a escala comum foi o dia de calendário. Em razão do rodízio, essa escala não equivale a horas efetivas de operação e constitui uma limitação da análise.

Sob a hipótese de taxa de falha constante, a confiabilidade no horizonte t é calculada pela Equação 13 (NIST/SEMATECH, 2012a):

$$R(t) = e^{-\lambda t}. \quad (13)$$

Na Equação 13, $R(t)$ é a probabilidade de sobrevivência além do horizonte t , e é a base do logaritmo natural e λ é a taxa de falha constante, em eventos por dia. Neste estudo, λ foi estimada pela Equação 14 (NIST/SEMATECH, 2012a):

$$\lambda = \frac{1}{\overline{TTF}}. \quad (14)$$

Na Equação 14, $\overline{TTF}$ é a média dos intervalos de calendário observados, em dias. As previsões de 90 e 180 dias foram obtidas substituindo t por esses horizontes na Equação 13.

O modelo Exponencial é útil como referência inicial porque assume taxa de falha constante ao longo do tempo. Essa hipótese é simples e frequentemente utilizada em análises preliminares de confiabilidade. Entretanto, equipamentos sujeitos a desgaste, envelhecimento ou degradação acumulada podem apresentar taxa de falha crescente, o que exige modelos mais flexíveis.

Para permitir taxa de falha variável, também foi ajustado o modelo de Weibull, cuja confiabilidade é dada pela Equação 15 (Weibull, 1951; NIST/SEMATECH, 2012d):

$$R(t) = e^{-\left(\frac{t}{\eta}\right)^\beta}. \quad (15)$$

Na Equação 15, $R(t)$ e t mantêm os significados anteriores; $\beta > 0$ é o parâmetro adimensional de forma; e $\eta > 0$, em dias, é o parâmetro de escala. O parâmetro β permite interpretar o regime de falha:

- $\beta < 1$: taxa de falha decrescente, associada a falhas precoces ou mortalidade infantil¹;
- $\beta = 1$: taxa de falha constante, caso equivalente ao modelo Exponencial;
- $\beta > 1$: taxa de falha crescente, associada a desgaste, envelhecimento ou degradação progressiva.

O parâmetro η , medido em dias, representa a escala temporal da distribuição. Em termos práticos, quando $t = \eta$, aproximadamente 63,2% dos eventos esperados já ocorreram, pois $R(\eta) = e^{-1}$. Assim, valores maiores de η indicam maior tempo característico até a falha.

Os parâmetros dos modelos foram estimados por máxima verossimilhança. Como o número de eventos por bomba é reduzido, especialmente nas bombas A e D, as curvas ajustadas devem ser interpretadas como apoio exploratório à decisão e não como estimativas definitivas da distribuição de vida dos equipamentos. Ainda assim, os modelos são úteis para comparar tendências relativas entre as bombas e identificar comportamentos compatíveis com falhas aleatórias ou desgaste progressivo.

3.3 RESULTADOS POR BOMBA

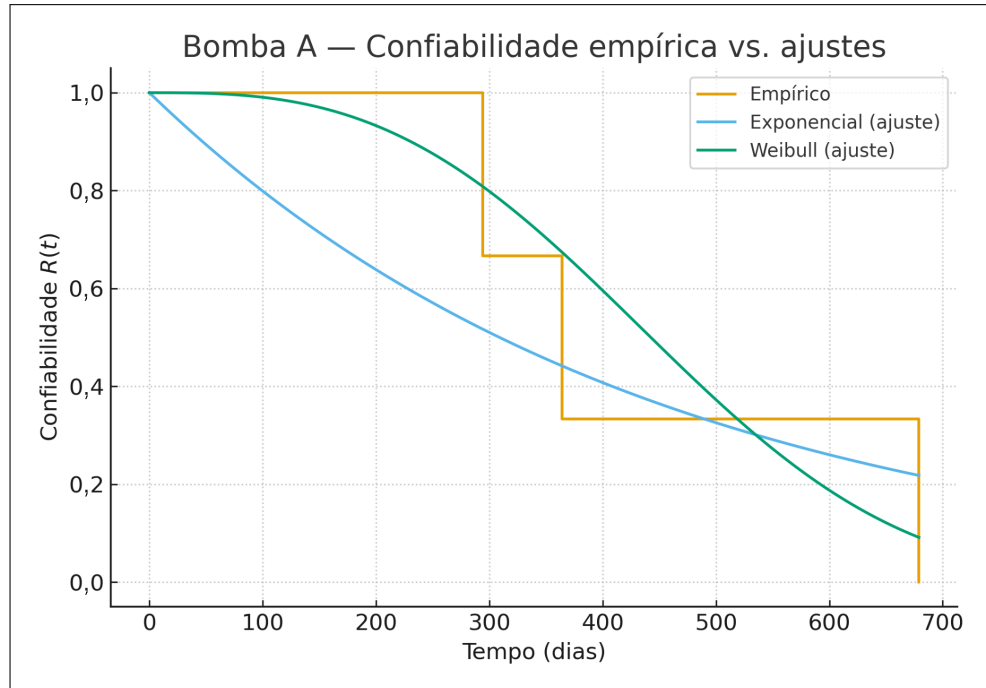
Esta seção apresenta os resultados individuais das bombas A, B, C e D. Para cada bomba, são mostradas as curvas de confiabilidade empírica, os ajustes Exponencial e Weibull, a tabela de datas de falha e TTFs e uma interpretação técnica dos registros SAP associados.

3.3.1 Bomba A

A curva de confiabilidade e os intervalos entre eventos da Bomba B-0302A são apresentados, respectivamente, na Figura 8 e na Tabela 3.

¹ O regime $\beta < 1$ caracteriza a chamada “mortalidade infantil”, associada a uma taxa de falha decrescente e à ocorrência de falhas prematuras, geralmente relacionadas a defeitos de fabricação, instalação ou condições iniciais de operação.

Figura 8 – Curvas de aderência da Bomba A: comparação entre confiabilidade empírica e ajustes Exponencial e Weibull.



Fonte: Elaborada pelo autor (2026).

Tabela 3 – Datas de falha e respectivos tempos até a falha (TTF), em dias, para a Bomba A.

Data da falha	TTF (dias)
25/10/2021	294
24/10/2022	364
02/09/2024	679

Fonte: Elaborada pelo autor (2026), com base nas ordens de manutenção do SAP.

A Figura 8 mostra que a curva empírica da Bomba A apresenta declínio suave, refletindo intervalos relativamente longos entre os eventos registrados. Os TTFs de 294, 364 e 679 dias indicam baixa frequência de falhas no período analisado. O modelo Exponencial acompanha a tendência média dos dados, enquanto o ajuste Weibull, com $\beta > 1$, sugere comportamento compatível com aumento gradual da taxa de falhas ao longo do tempo.

A interpretação da Bomba A deve considerar o número reduzido de eventos ($n = 3$). Por esse motivo, não se pode afirmar de forma conclusiva que a bomba opera em regime de desgaste. O que os dados indicam é uma trajetória compatível com boa estabilidade operacional, com eventos espaçados e ausência de concentração de falhas em intervalos curtos.

Os registros SAP associados às datas da Tabela 3 mostram ocorrências em subsistemas distintos. Em **25/10/2021**, o registro refere-se à manutenção preventiva elétrica na bomba B-0302A, com necessidade de correção de status operacional e inspeção elétrica de rotina, sem identificação de falha mecânica significativa. Em **24/10/2022**, os registros posteriores indicam vazamento de produto e necessidade de ajuste de gaxeta da válvula FCV-03201A, caracterizando

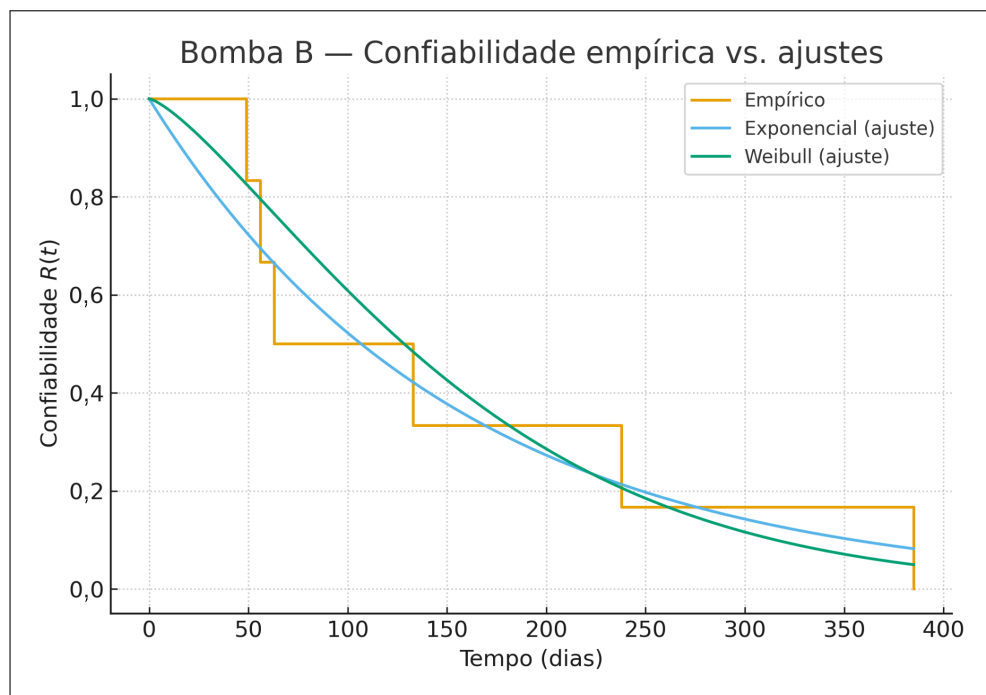
falha de estanqueidade recorrente. Em **02/09/2024**, a ocorrência está relacionada a falha elétrica no motor da B-0302A, com erro de leitura no sensor de temperatura do mancal, desequilíbrio térmico, desligamento automático por proteção e posterior substituição do motor.

De forma geral, os registros indicam que as falhas da Bomba A não se concentram em um único mecanismo dominante. Há eventos elétricos, de vedação e de motor, o que sugere comportamento heterogêneo. Essa heterogeneidade é relevante para os capítulos seguintes, pois falhas de naturezas distintas podem apresentar diferentes graus de previsibilidade por sensores operacionais.

3.3.2 Bomba B

A curva de confiabilidade e os intervalos entre eventos da Bomba B-0302B são apresentados, respectivamente, na Figura 9 e na Tabela 4.

Figura 9 – Curvas de aderência da Bomba B: comparação entre confiabilidade empírica e ajustes Exponencial e Weibull.



Fonte: Elaborada pelo autor (2026).

Tabela 4 – Datas de falha e respectivos tempos até a falha (TTF), em dias, para a Bomba B.

Data da falha	TTF (dias)
17/05/2021	133
12/07/2021	56
01/08/2022	385
27/03/2023	238
15/05/2023	49
17/07/2023	63

Fonte: Elaborada pelo autor (2026), com base nas ordens de manutenção do SAP.

A Figura 9 revela uma curva de confiabilidade com quedas acentuadas e espaçamento irregular entre os eventos de falha. Os TTFs alternam períodos curtos, como 56, 49 e 63 dias, com intervalos mais longos, como 385 e 238 dias. Esse comportamento indica maior instabilidade em comparação às demais bombas do conjunto.

O ajuste Exponencial fornece uma referência média para a taxa de falhas, enquanto a Weibull, com $\beta \approx 1,3$, sugere comportamento próximo de taxa constante, com leve tendência de aumento do risco ao longo do tempo. Como a dispersão dos TTFs é elevada, a interpretação mais prudente é que a Bomba B apresenta comportamento irregular, possivelmente influenciado por falhas de natureza elétrica e de controle, que nem sempre seguem uma degradação lenta e contínua.

Os registros SAP associados à Bomba B mostram predominância de ocorrências em subsistemas de acionamento, comando e controle. Em **17/05/2021**, o apontamento refere-se à necessidade de instalação de novos atuadores elétricos nas linhas de sucção e descarga. Em **12/07/2021**, houve nova avaliação elétrica, com indícios de falha em placa controladora de comando. Em **01/08/2022**, os registros indicam falha na partida do motor principal, com necessidade de revisão de relés e circuitos de comando.

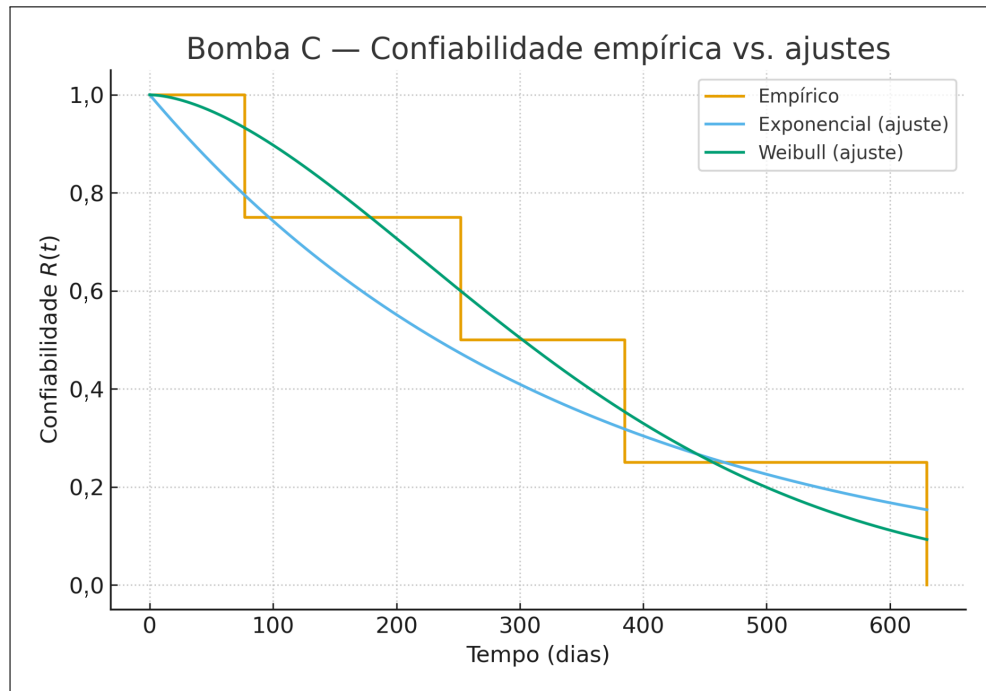
Em **27/03/2023**, a bomba não entrou em operação por falha no disjuntor principal, com persistência do defeito em inspeções posteriores. Em **15/05/2023**, houve limpeza de filtro de sucção, sem parada total, mas com indicação de anomalia operacional leve. Por fim, em **17/07/2023**, a ocorrência está associada a falha elétrica no motor da bomba principal, falha de partida e provável defeito no disjuntor, exigindo substituição e reprogramação do circuito de comando.

A Bomba B apresenta histórico dominado por eventos elétricos e de controle. Isso não significa que falhas elétricas sejam, em geral, imprevisíveis: técnicas específicas podem usar grandezas elétricas e de proteção adequadas. No recorte desta pesquisa, porém, placas, disjuntores e circuitos de comando não apresentaram uma trajetória de degradação compatível com o FD001 nem eram descritos por sensoriamento específico suficiente, o que limitou a transferência do AGATT.

3.3.3 Bomba C

A curva de confiabilidade e os intervalos entre eventos da Bomba B-0302C são apresentados, respectivamente, na Figura 10 e na Tabela 5.

Figura 10 – Curvas de aderência da Bomba C: comparação entre confiabilidade empírica e ajustes Exponencial e Weibull.



Fonte: Elaborada pelo autor (2026).

Tabela 5 – Datas de falha e respectivos tempos até a falha (TTF), em dias, para a Bomba C.

Data da falha	TTF (dias)
26/09/2022	630
12/12/2022	77
01/01/2024	385
09/09/2024	252

Fonte: Elaborada pelo autor (2026), com base nas ordens de manutenção do SAP.

A Figura 10 mostra que a curva empírica da Bomba C apresenta declínio gradual, com boa proximidade visual em relação ao ajuste Weibull. O primeiro intervalo observado, de 630 dias, indica longo período sem falhas, seguido por intervalos menores. Esse padrão é compatível com a hipótese de aumento gradual do risco após período prolongado de operação.

O ajuste Weibull apresentou $\beta = 1,68$, indicando taxa de falhas crescente. Ainda assim, como a bomba possui apenas quatro eventos registrados, essa interpretação deve ser tratada como evidência exploratória de desgaste, e não como conclusão definitiva. Os dados sugerem que a Bomba C combina eventos mecânicos e elétricos, com possível influência de degradação acumulada em motor, alimentação elétrica e instrumentação.

Os registros da Bomba C indicam eventos predominantemente elétricos e de vibração. Em **26/09/2022**, a falha foi registrada após detecção de vibração elevada pelo sistema de monitoramento, com identificação de danos ao motor e necessidade de envio para reparo externo. Em **12/12/2022**, os registros indicam desequilíbrio de tensão e necessidade de ajustes no TAP do transformador, com assinatura de corrente irregular no barramento.

Em **01/01/2024**, houve registro de queda de tensão elevada no desumidificador e valores anormais nos parâmetros IP e IA², com impacto indireto sobre a operação e necessidade de novos testes de partida após retorno do motor. Em **09/09/2024**, foram relatados níveis baixos de IP e IA e falha de comunicação MODBUS, associados a erro no sinal do sensor de corrente e necessidade de substituição de cabeamento.

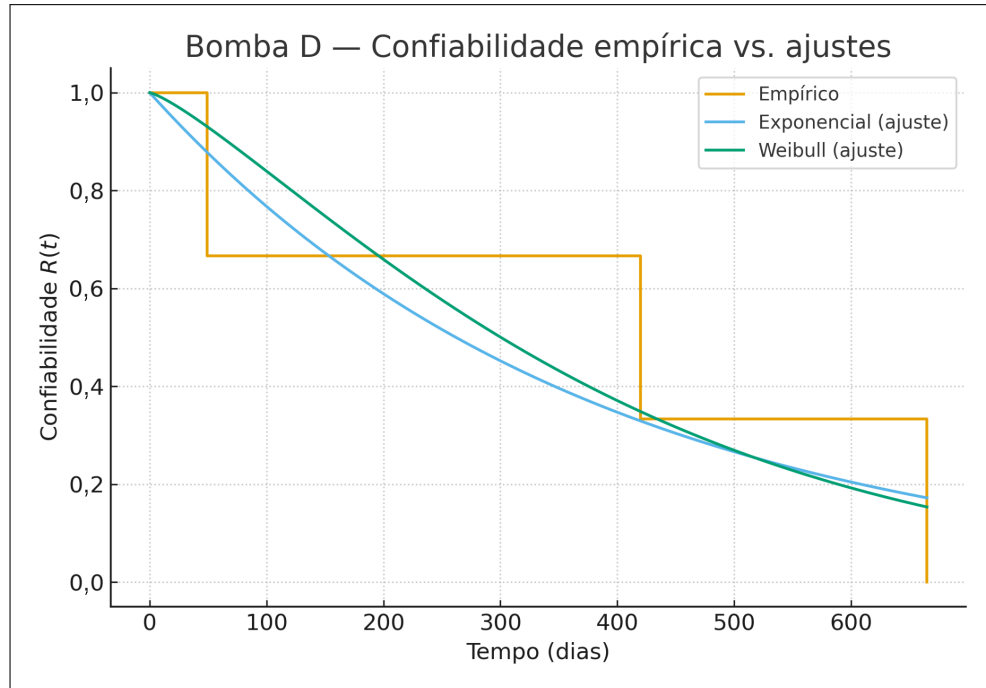
Esses registros mostram que a Bomba C apresenta interação entre falhas de vibração, motor, alimentação elétrica e instrumentação. Essa combinação pode gerar padrões parciais de degradação observáveis por sensores, mas também inclui eventos de instrumentação e comunicação, que podem reduzir a previsibilidade do RUL quando analisados apenas por sinais físicos de operação.

3.3.4 Bomba D

A curva de confiabilidade e os intervalos entre eventos da Bomba B-0302D são apresentados, respectivamente, na Figura 11 e na Tabela 6.

² IP e IA correspondem, respectivamente, aos parâmetros de corrente associados ao circuito de potência e ao circuito de alimentação/acionamento do equipamento, conforme a nomenclatura adotada nos registros operacionais.

Figura 11 – Curvas de aderência da Bomba D: comparação entre confiabilidade empírica e ajustes Exponencial e Weibull.



Fonte: Elaborada pelo autor (2026).

Tabela 6 – Datas de falha e respectivos tempos até a falha (TTF), em dias, para a Bomba D.

Data da falha	TTF (dias)
31/10/2022	665
25/12/2023	420
12/02/2024	49

Fonte: Elaborada pelo autor (2026), com base nas ordens de manutenção do SAP.

A Figura 11 evidencia queda de confiabilidade após o segundo evento de falha, refletindo redução acentuada do intervalo entre falhas. Os TTFs de 665, 420 e 49 dias formam uma sequência decrescente, compatível com um processo de degradação progressiva. Essa característica torna a Bomba D particularmente relevante para esta dissertação, pois falhas com trajetória temporal mais clara tendem a ser mais adequadas à previsão por modelos de RUL baseados em sensores.

O ajuste Exponencial aproxima a tendência média dos dados, enquanto a Weibull, com $\beta \approx 1,25$, sugere leve aumento da taxa de falhas. Novamente, o número reduzido de eventos exige cautela. Entretanto, diferentemente de bombas com falhas predominantemente elétricas ou abruptas, a Bomba D apresenta uma sequência tecnicamente mais coerente com desgaste de componente específico.

Os registros SAP associados à Bomba D indicam predominância de problemas de vedação e calibração operacional, com recorrência de falhas no selo mecânico. Em **31/10/2022**, o comentário técnico descreve necessidade de aplicação de selante na válvula de sucção da

bomba e calibração da válvula FCV-201D, que não acompanhava corretamente o controle PID em modo automático. Em **25/12/2023**, houve registro de monitoramento do selo mecânico, com observação de gotejamento e abertura de ordem de manutenção para substituição do componente. Em **12/02/2024**, há novo relato de selo gotejando, reforçando a recorrência do problema.

A recorrência de problemas de vedação e a redução dos TTFs tornam a B-0302D o caso mais compatível com o padrão progressivo representado no FD001. Essa compatibilidade, combinada à maior instrumentação, fundamenta o aprofundamento do Capítulo 5; não constitui evidência de que mecanismos elétricos sejam intrinsecamente imprevisíveis.

3.4 PREVISÕES DE CONFIABILIDADE EM 90 E 180 DIAS

A Tabela 7 apresenta os parâmetros ajustados de cada bomba e as probabilidades estimadas de operação sem falha em 90 e 180 dias. Os parâmetros foram estimados por Máxima Verossimilhança para os modelos Exponencial e Weibull.

Cada coluna da tabela possui o seguinte significado:

- **Bomba:** identificação da unidade do conjunto B-0302;
- **n:** quantidade de falhas observadas, isto é, amostras de TTF;
- **TTF médio (d):** média aritmética dos tempos para falha observados, em dias;
- **λ :** taxa média de falhas, em 1/dia, utilizada no modelo Exponencial;
- **$R(90)$ e $R(180)$ – Exponencial:** probabilidades de operação sem falha em 90 e 180 dias, calculadas por $R(t) = e^{-\lambda t}$;
- **β :** parâmetro de forma da Weibull, utilizado para interpretar tendência de falha;
- **η :** parâmetro de escala da Weibull, em dias;
- **$R(90)$ e $R(180)$ – Weibull:** probabilidades estimadas pela distribuição Weibull, calculadas por $R(t) = e^{-(t/\eta)^\beta}$.

Tabela 7 – Parâmetros ajustados e confiabilidades previstas para 90 e 180 dias.

Bomba	n	TTF médio (d)	λ (1/d)	Exponencial		β	η (d)	Weibull	
				$R(90)$	$R(180)$			$R(90)$	$R(180)$
A	3	445,7	0,00224	81,7%	66,8%	2,893	502,2	99,3%	95,0%
B	6	154,0	0,00649	55,7%	31,1%	1,336	169,0	65,0%	33,7%
C	4	336,0	0,00298	76,5%	58,5%	1,678	376,0	91,3%	74,8%
D	3	378,0	0,00265	78,8%	62,1%	1,252	402,6	85,8%	69,4%

Fonte: Elaborada pelo autor (2026).

Os valores da Tabela 7 são coerentes com as curvas das Figuras 8–11. A B-0302B apresenta o menor TTF médio e as menores confiabilidades previstas. Todos os ajustes Weibull resultaram em $\beta > 1$, sugerindo taxa crescente no pequeno conjunto observado. Na B-0302A, $\beta = 2,893$ indica crescimento mais acentuado da taxa no ajuste, mas essa interpretação é especialmente incerta porque há somente três intervalos. A B-0302D combina β moderadamente superior a 1 com uma sequência decrescente de TTFs e recorrência do mesmo mecanismo de vedação.

A comparação entre Exponencial e Weibull também reforça diferenças importantes. O modelo Exponencial resume o comportamento por uma taxa média constante. A Weibull, por sua vez, permite capturar tendências de falha crescentes ou decrescentes. Em bombas com poucos eventos, a Weibull pode se ajustar visualmente melhor à curva empírica, mas também pode ser mais sensível à pequena quantidade de amostras. Assim, os dois modelos devem ser usados de forma complementar: o Exponencial como referência simples e robusta, e a Weibull como ferramenta interpretativa para investigar indícios de desgaste.

3.5 SÍNTESE DO CAPÍTULO

A análise mostrou que as quatro bombas não devem ser tratadas como casos equivalentes. Os ajustes são exploratórios, pois se baseiam em poucos intervalos de calendário e não em horas efetivas de exposição. A B-0302D apresentou a combinação mais coerente de recorrência do mecanismo de vedação, TTFs decrescentes e disponibilidade sensorial; por isso, foi selecionada para a transferência de aprendizado. Eventos elétricos e de controle das demais bombas exigiriam sinais e modelos mais diretamente relacionados aos respectivos mecanismos.

4 REPRODUÇÃO DE BENCHMARKING NA C-MAPSS

Este capítulo apresenta a reprodução do modelo AGATT na base C-MAPSS, com foco no subdataset FD001. O objetivo não é apenas relatar a execução computacional, mas explicar de forma encadeada por que a C-MAPSS é utilizada como base de referência em previsão de RUL, como seus dados são estruturados, quais decisões de pré-processamento são necessárias, como o AGATT é composto internamente e quais resultados foram obtidos na reprodução (Zeng, Wu e Cheng, 2023; Saxena *et al.*, 2008b).

A reprodução do AGATT cumpre duas funções nesta dissertação. A primeira é de benchmarking: verificar, em uma base pública e amplamente utilizada, se a arquitetura escolhida é capaz de aprender padrões de degradação em séries temporais multivariadas. A segunda é metodológica: construir uma base técnica para posterior adaptação por *transfer learning* ao domínio das bombas industriais B-0302. Assim, a etapa na C-MAPSS funciona como uma ponte entre a literatura de prognóstico baseada em motores aeronáuticos e o problema aplicado de previsão de falhas em bombas de transporte dutoviário.

Ressalta-se que os resultados apresentados neste capítulo correspondem ao experimento principal de reprodução e *benchmarking* do AGATT, realizado com janelas de 30 ciclos. Essa execução foi utilizada para avaliar a arquitetura na base FD001 e analisar sua reprodutibilidade em múltiplos treinamentos. Posteriormente, para a etapa de transferência de aprendizado, foi treinada uma nova instância da mesma família arquitetural, com janela efetiva de 35 pontos e configuração específica para uso como *backbone*. O treinamento e a utilização desse *backbone* são descritos no Capítulo 5. Assim, os resultados deste capítulo não devem ser interpretados como as métricas do arquivo de pesos utilizado diretamente na transferência, mas como a reprodução experimental que fundamentou a escolha e a compreensão da arquitetura AGATT.

4.1 BASE C-MAPSS

A C-MAPSS é uma base sintética desenvolvida pela NASA para simular degradação progressiva de motores turbofan (Saxena *et al.*, 2008b). Cada *ciclo* corresponde a um passo discreto da simulação no qual são registrados o regime operacional e as leituras dos sensores; não representa, necessariamente, uma unidade física fixa como hora ou dia. No treinamento, as trajetórias seguem até a falha; no teste, são truncadas antes dela.

Cada trajetória da C-MAPSS representa um motor. No conjunto de treinamento, cada motor é acompanhado desde um estado inicial considerado saudável até a falha. Portanto, o RUL verdadeiro pode ser calculado em cada ciclo como a distância até o ciclo final da trajetória. No conjunto de teste, por outro lado, as trajetórias são interrompidas antes da falha. Para cada motor de teste, a base fornece separadamente o RUL verdadeiro no ponto em que a sequência foi truncada. Dessa forma, o modelo deve observar apenas a parte disponível da série e estimar o número de ciclos restantes.

A base é composta por quatro *subdatasets*: FD001, FD002, FD003 e FD004. Eles diferem

quanto ao número de condições operacionais e ao número de modos de falha. A Tabela 8 resume essas diferenças.

Tabela 8 – Resumo dos subdatasets da C-MAPSS.

Subdataset	Motores treino	Motores teste	Condições operacionais	Modos de falha
FD001	100	100	1	1
FD002	260	259	6	1
FD003	100	100	1	2
FD004	248	249	6	2

Fonte: Elaborada pelo autor (2026), com base em Saxena et al. (2008).

O FD001 é o subconjunto mais simples e mais utilizado como referência inicial, pois possui apenas uma condição operacional e um modo de falha. Essa característica reduz a variabilidade externa e torna o problema mais controlado. Por esse motivo, o FD001 é adequado para avaliar se uma arquitetura consegue aprender a progressão de degradação sem a interferência de múltiplos regimes de operação ou múltiplos mecanismos simultâneos de falha.

Essa simplicidade, entretanto, não torna o problema trivial. Mesmo em FD001, a degradação não aparece necessariamente de forma explícita em um único sensor. Ela emerge de padrões sutis ao longo do tempo e de relações entre variáveis. Por isso, modelos capazes de explorar simultaneamente dependências temporais e relações entre sensores tendem a ser particularmente adequados ao problema.

4.2 POR QUE UTILIZAR A C-MAPSS NESTA DISSERTAÇÃO

A escolha da C-MAPSS nesta dissertação se justifica por quatro razões principais.

Primeiro, a C-MAPSS é uma base pública, padronizada e amplamente reconhecida na literatura de RUL. Isso permite comparar resultados com trabalhos anteriores e verificar se a implementação segue uma referência experimental conhecida. Em problemas industriais reais, como o das bombas B-0302, os dados costumam ser restritos, ruidosos, incompletos e específicos de uma instalação. A C-MAPSS, ao contrário, oferece um ambiente controlado para desenvolvimento e validação inicial.

Segundo, a base possui trajetórias *run-to-failure*, isto é, séries completas até a falha. Esse tipo de dado é raro em ambientes industriais reais, porque falhas críticas são evitadas por manutenção preventiva, substituições, intervenções corretivas intermediárias ou mudanças operacionais. Para treinar modelos de RUL, entretanto, conhecer a distância até a falha é essencial. A C-MAPSS fornece esse conhecimento de forma estruturada.

Terceiro, o FD001 permite isolar decisões de modelagem. Como há uma condição operacional e um modo de falha, diferenças de desempenho tendem a estar mais associadas ao pré-processamento, à arquitetura e ao treinamento do modelo do que à heterogeneidade do domínio. Isso torna a base adequada para reproduzir o AGATT e compreender o comportamento da arquitetura.

Quarto, a C-MAPSS pode servir como domínio de origem em uma estratégia de *transfer learning*. A tarefa final desta dissertação envolve bombas industriais, não turbofans aeronáuticos. Mesmo assim, ambos os contextos possuem séries temporais multivariadas, sensores físicos, degradação progressiva e interesse em prever o tempo restante até um evento de falha. Assim, a C-MAPSS não é utilizada porque turbofans e bombas sejam fisicamente idênticos, mas porque a base permite aprender representações gerais de degradação em sistemas monitorados por sensores.

4.3 ESTRUTURA DOS DADOS

Cada linha da C-MAPSS representa uma leitura de um motor em um determinado ciclo. A estrutura geral contém o identificador do motor, o ciclo de operação, variáveis operacionais e leituras de sensores. No FD001, cada motor possui uma sequência temporal própria, e o número de ciclos até a falha varia de motor para motor.

A tarefa de RUL consiste em transformar essas séries em pares de entrada e saída. A entrada é uma janela de leituras recentes dos sensores, e a saída é o RUL associado ao último ciclo dessa janela. Assim, em vez de entregar ao modelo um único ciclo isolado, utiliza-se uma sequência curta de ciclos anteriores para permitir que o modelo observe tendências, variações e padrões temporais.

A estrutura original da C-MAPSS contém, além do identificador do motor e do ciclo de operação, três variáveis que descrevem as condições operacionais e 21 variáveis correspondentes a sensores simulados do motor. As condições operacionais não devem ser confundidas com os sensores s_1 , s_2 e s_3 . A Tabela 9 apresenta separadamente esses dois grupos de variáveis.

Embora a base disponibilize 21 sensores, nem todos apresentam variabilidade relevante no FD001. Seguindo o protocolo adotado na reprodução do AGATT, foram removidos os sensores s_1 , s_5 , s_6 , s_{10} , s_{16} , s_{18} e s_{19} , por apresentarem comportamento constante ou praticamente constante nesse subdataset. Variáveis com variação nula ou muito reduzida carregam pouca informação para distinguir diferentes estágios de degradação.

Consequentemente, seguindo a seleção da reprodução do AGATT (Zeng, Wu e Cheng, 2023), foram mantidos os sensores s_2 , s_3 , s_4 , s_7 , s_8 , s_9 , s_{11} , s_{12} , s_{13} , s_{14} , s_{15} , s_{17} , s_{20} e s_{21} . Essa nomenclatura é usada nas análises, na arquitetura e na projeção do domínio industrial.

Tabela 9 – Condições operacionais e sensores disponíveis na base C-MAPSS.

ID	Variável	Descrição
Condições operacionais		
op1	Altitude	Condição operacional associada à altitude simulada.
op2	Mach	Condição operacional associada ao número de Mach.
op3	TRA	Ângulo do acionador de potência, do inglês <i>Throttle Resolver Angle</i> .
Sensores		
s1	T2	Temperatura total na entrada do ventilador.
s2	T24	Temperatura total na saída do compressor de baixa pressão.
s3	T30	Temperatura total na saída do compressor de alta pressão.
s4	T50	Temperatura total na saída da turbina de baixa pressão.
s5	P2	Pressão total na entrada do ventilador.
s6	P15	Pressão total na saída do compressor de baixa pressão.
s7	P30	Pressão total na saída do compressor de alta pressão.
s8	Nf	Velocidade física do ventilador.
s9	Nc	Velocidade física do núcleo do motor.
s10	EPR	Razão de pressão do motor.
s11	Ps30	Pressão estática na saída do compressor de alta pressão.
s12	ϕ	Razão entre o fluxo de combustível e a pressão <i>Ps30</i> .
s13	NRf	Velocidade corrigida do ventilador.
s14	NRC	Velocidade corrigida do núcleo do motor.
s15	BPR	Razão de derivação, do inglês <i>Bypass Ratio</i> .
s16	farB	Razão combustível–ar na câmara de combustão.
s17	htBleed	Entalpia do ar sangrado do compressor.
s18	Nf_dmd	Velocidade demandada do ventilador.
s19	PCNfR_dmd	Velocidade corrigida demandada do ventilador.
s20	W31	Vazão de ar de refrigeração da turbina de alta pressão.
s21	W32	Vazão de ar de refrigeração da turbina de baixa pressão.

Fonte: Elaborada pelo autor (2026), com base em Saxena et al. (2008).

4.4 CONSTRUÇÃO DO ALVO RUL

No treinamento, o RUL é a distância entre o ciclo¹ atual e o ciclo final da trajetória, conforme a Equação 16 (Saxena *et al.*, 2008b; Saxena *et al.*, 2008a):

$$RUL(t) = T - t. \quad (16)$$

Na Equação 16, T é o último ciclo da trajetória, t é o ciclo observado e $RUL(t)$ é a quantidade de ciclos remanescentes. Seguindo o protocolo reproduzido por Zeng, Wu e Cheng (Zeng, Wu e Cheng, 2023), o alvo foi limitado a 125 ciclos pela Equação 17 (adaptado de Heimes (2008), Zheng *et al.* (2017) e Zeng, Wu e Cheng (2023)):

$$RUL_{cap}(t) = \min(RUL(t), 125), \quad (17)$$

Na Equação 17, $\min(\cdot)$ seleciona o menor valor e 125 é o limite superior adotado no artigo-base. O truncamento representa uma fase inicial aproximadamente saudável, na qual diferenças

¹ Na base C-MAPSS, um ciclo representa uma unidade discreta de tempo de operação do motor, na qual são registrados os valores dos sensores e das condições operacionais. A sequência de ciclos representa a evolução temporal do motor ao longo de sua trajetória até a falha.

muito elevadas de RUL não possuem assinatura sensorial proporcionalmente distinguível.

4.5 JANELAMENTO TEMPORAL

Cada trajetória foi convertida em janelas deslizantes de $w = 30$ ciclos e 14 sensores, isto é, matrizes $X \in \mathbb{R}^{30 \times 14}$. A janela avança um ciclo por vez; portanto, janelas consecutivas compartilham 29 ciclos. O alvo de cada amostra é o RUL do último ciclo da janela. A sobreposição aumenta o número de exemplos, mas também produz dependência entre amostras vizinhas, razão pela qual a avaliação respeitou a separação por motores.

4.6 NORMALIZAÇÃO DOS SENSORES

Como os sensores possuem escalas distintas, os dados foram normalizados no intervalo $[0, 1]$ pela Equação 18 (Pedregosa *et al.*, 2011):

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}}. \quad (18)$$

Na Equação 18, x é o valor original, $x_{\min}$ e $x_{\max}$ são o mínimo e o máximo estimados no treinamento, e x' é o valor normalizado. Os parâmetros do conjunto de teste não foram usados para ajustar a transformação.

4.7 ANÁLISE EXPLORATÓRIA DOS SENSORES

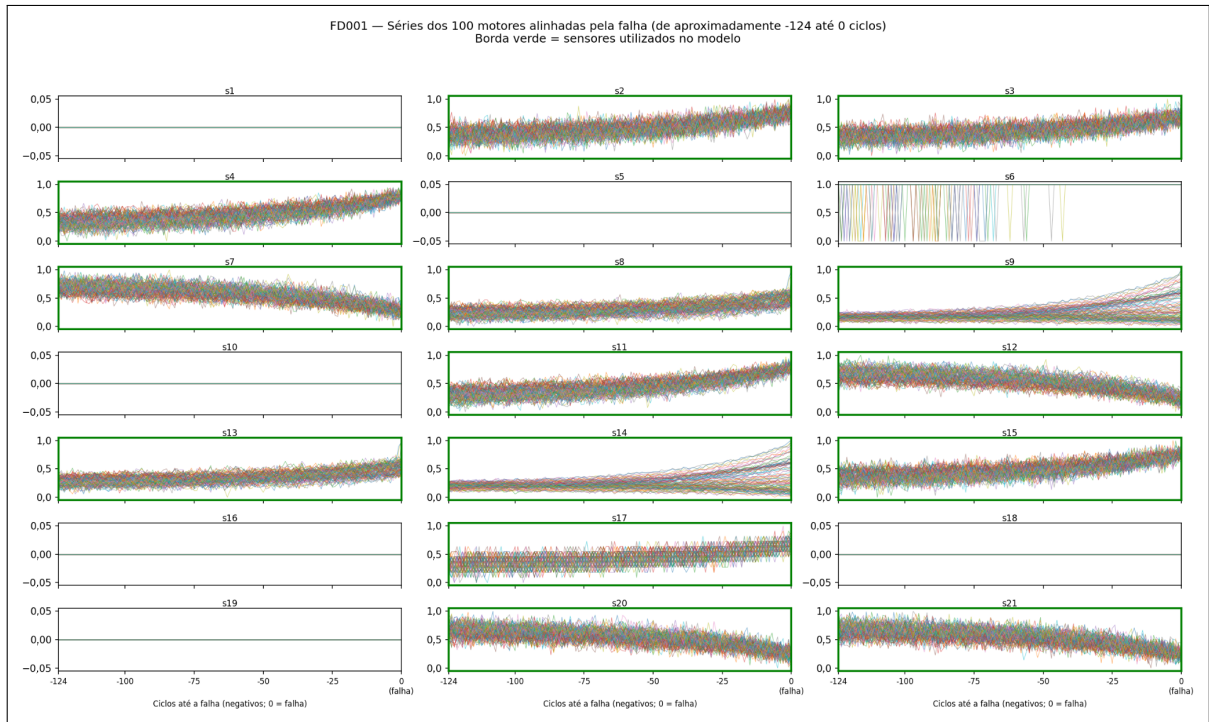
Antes do treinamento, foi realizada uma análise exploratória para verificar a contribuição dos sensores e a coerência da seleção dos 14 canais ativos. A Figura 12 apresenta as séries dos 21 sensores alinhadas à falha, considerando os últimos 125 ciclos dos motores do FD001. Os sensores mantidos no modelo são destacados, enquanto os sensores removidos apresentam pouca ou nenhuma variação.

Para quantificar essa observação visual, a Figura 13 apresenta a variância dos sensores após normalização. Sensores com variância próxima de zero não contribuem para distinguir estados de degradação. A baixa variabilidade dos sensores removidos reforça a decisão de excluí-los do treinamento.

A Figura 14 apresenta a matriz de correlação entre os 14 sensores utilizados. Essa análise é relevante porque o AGATT não trata os sensores como variáveis completamente independentes. Ao contrário, uma de suas premissas é que a degradação pode ser expressa por relações entre sensores. Assim, observar blocos de correlação ajuda a justificar o uso de mecanismos capazes de modelar dependências espaciais.

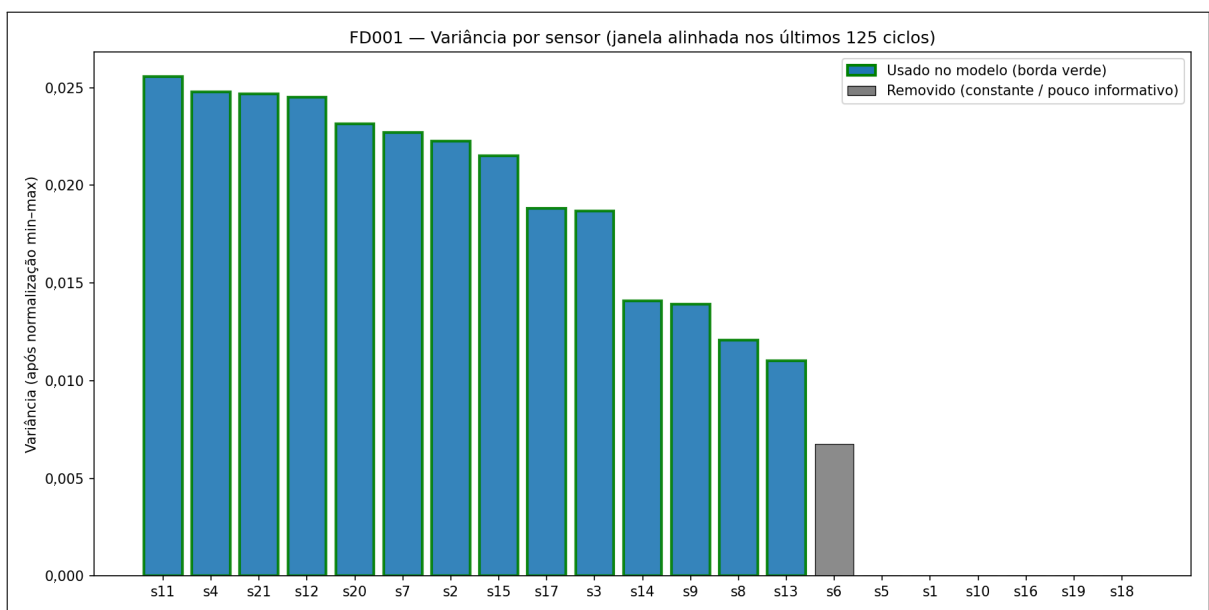
A Figura 15 apresenta a correlação média absoluta por sensor. Sensores com maior correlação média tendem a participar mais intensamente de relações multivariadas, enquanto sensores com menor correlação média podem carregar informação mais específica. Ambos os

Figura 12 – FD001: séries dos 21 sensores alinhadas à falha nos últimos 125 ciclos. Os sensores mantidos no modelo apresentam variações mais compatíveis com o processo de degradação, enquanto os sensores removidos exibem baixa variabilidade.



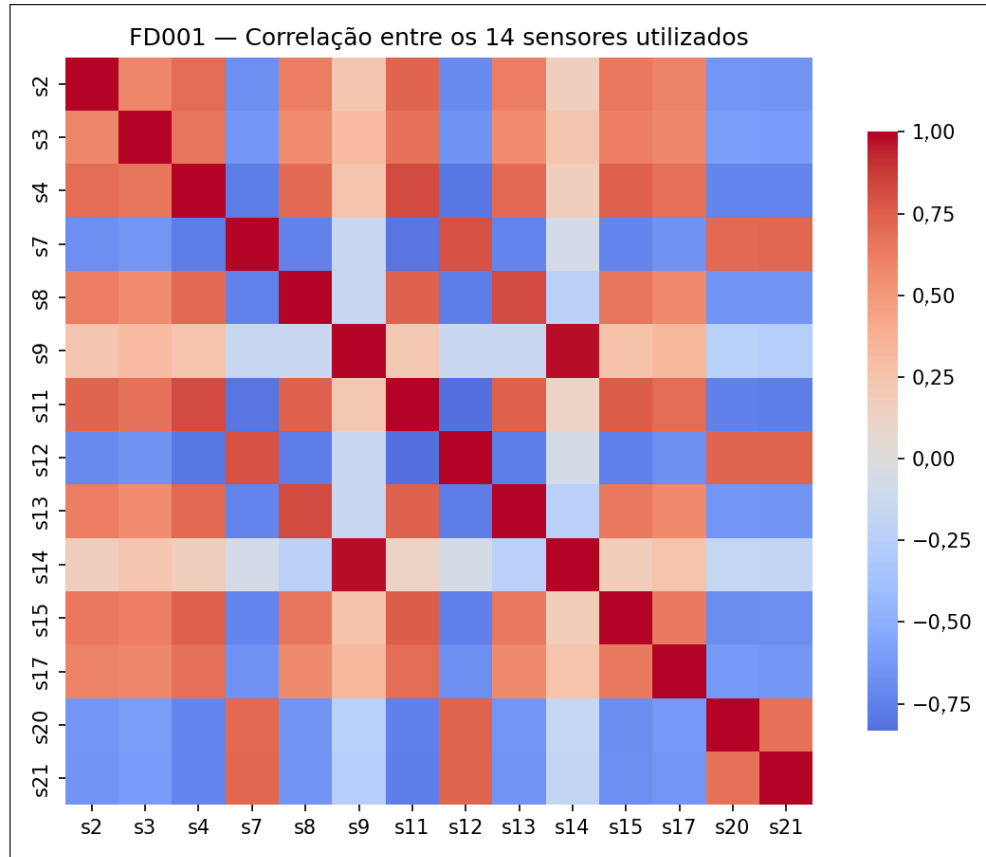
Fonte: Elaborada pelo autor (2026).

Figura 13 – Variância por sensor no FD001 após normalização. Sensores com baixa variância tendem a carregar pouca informação sobre a degradação.



Fonte: Elaborada pelo autor (2026).

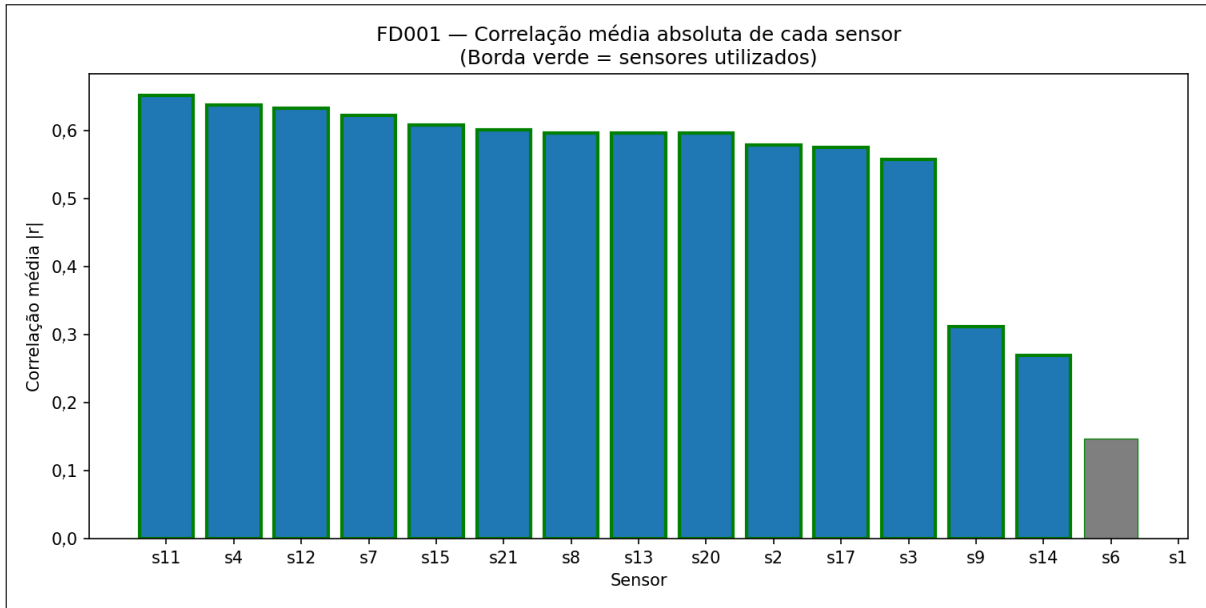
Figura 14 – Matriz de correlação dos 14 sensores utilizados no treinamento. A existência de relações entre sensores justifica o uso de componentes capazes de capturar dependências espaciais.



Fonte: Elaborada pelo autor (2026).

casos são úteis: sensores centrais ajudam a modelar padrões globais de degradação, e sensores menos redundantes podem ajudar a distinguir regimes específicos.

Figura 15 – Correlação média absoluta por sensor. Sensores com maior valor tendem a se mover de forma mais associada aos demais canais, enquanto sensores menos correlacionados podem carregar informação complementar.



Fonte: Elaborada pelo autor (2026).

4.8 ARQUITETURA AGATT

O AGATT é um modelo híbrido projetado para previsão de RUL em séries temporais multivariadas. Sua principal característica é combinar três ideias: convolução unidimensional, atenção em grafos e Transformer. A convolução extrai características locais de cada sensor. As atenções espacial e temporal identificam, respectivamente, sensores e instantes mais relevantes. As GATs modelam dependências entre sensores e entre passos temporais. Por fim, o Transformer integra as representações extraídas e produz a previsão final de RUL.

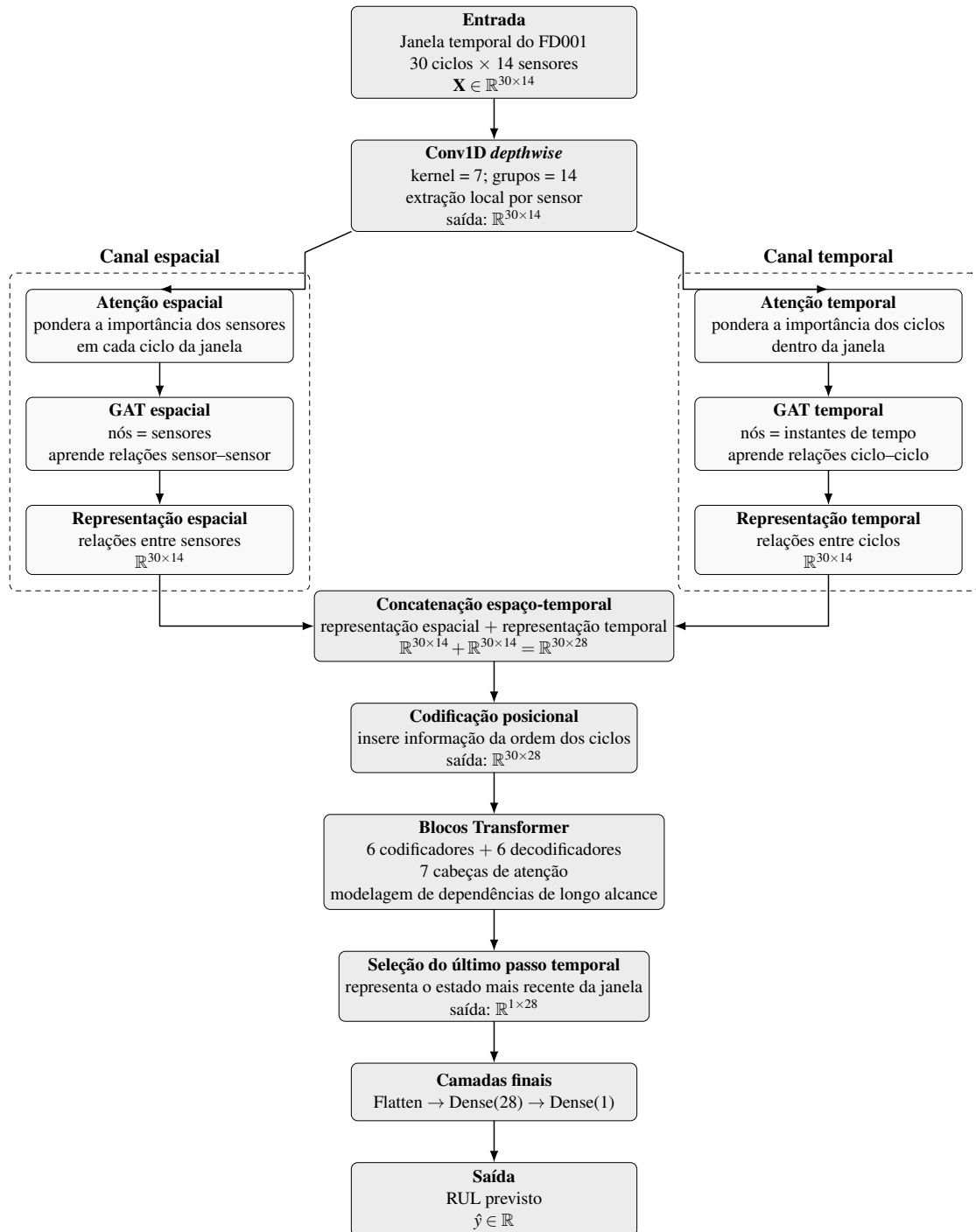
A motivação do AGATT é que a degradação não está apenas na trajetória temporal de cada sensor isolado. Ela também pode estar na relação entre sensores. Por exemplo, uma elevação de temperatura pode ter significado diferente dependendo do comportamento simultâneo de pressão, rotação ou vazão. Do mesmo modo, nem todos os ciclos de uma janela têm a mesma importância: alguns representam operação estável, outros podem indicar transição ou aceleração da degradação.

A Figura 16 apresenta o fluxo reproduzido. A convolução *depthwise* extrai padrões locais por sensor; os ramos espacial e temporal modelam, respectivamente, relações entre sensores e entre ciclos; as representações são concatenadas e processadas por blocos Transformer; por fim, o último passo temporal alimenta as camadas de regressão.

4.8.1 Entrada do modelo

A entrada é uma matriz $X_t \in \mathbb{R}^{w \times N}$, em que t identifica o ciclo final da amostra, $w = 30$ é a quantidade de ciclos e $N = 14$ é a quantidade de sensores. Cada linha reúne os sensores de um

Figura 16 – Fluxo da arquitetura AGATT reproduzida para uma entrada de 30 ciclos e 14 sensores.



Fonte: Elaborada pelo autor (2026).

ciclo e cada coluna contém a evolução de um sensor.

4.8.2 Camada convolucional *depthwise*

A primeira camada aplica uma convolução unidimensional separadamente em cada sensor, com 14 grupos. Antes de combinar sensores, ela procura padrões locais em cada canal, como picos, oscilações e alterações de inclinação. Para um canal c , a operação pode ser representada pela Equação 19 (Sifre, 2014), (Chollet, 2017) e (Zeng, Wu e Cheng, 2023):

$$H_{t,c} = \phi \left(b_c + \sum_{k=0}^{K-1} w_{c,k} X_{t-k,c} \right). \quad (19)$$

Na Equação 19, $X_{t-k,c}$ são as observações do sensor c cobertas pelo filtro no entorno do instante t ; $w_{c,k}$ são os pesos exclusivos desse canal; K é o tamanho do *kernel*; b_c é o viés; ϕ é a ativação; e $H_{t,c}$ é a característica local produzida. O uso de *padding* preserva o comprimento temporal da sequência. A diferença para uma convolução convencional é que, nesta etapa, o filtro de um sensor não é aplicado aos demais canais.

A separação é útil porque temperatura, pressão, rotação e fluxo possuem significados e escalas diferentes. O modelo primeiro identifica como cada sinal se comporta localmente e somente depois aprende relações entre sensores. A quantidade de parâmetros da camada é calculada pela Equação 20 (Zeng, Wu e Cheng, 2023):

$$N_\theta = F \left(K \frac{C_{\text{in}}}{G} + 1 \right). \quad (20)$$

Na Equação 20, N_θ é a quantidade de parâmetros; F é o número de filtros; K é o tamanho do *kernel*; C_{in} é a quantidade de canais de entrada; G é a quantidade de grupos; e o termo 1 representa um viés por filtro. Com $F = 14$, $K = 7$ e $C_{\text{in}} = G = 14$, cada filtro observa um único canal e a camada possui 112 parâmetros.

4.8.3 Canal espacial

O canal espacial busca responder à seguinte pergunta: quais sensores são mais importantes e como eles se relacionam entre si? Para isso, a matriz de características convolucionais passa primeiro por um mecanismo de atenção espacial. Essa atenção atribui pesos aos sensores em cada instante de tempo, permitindo que a rede enfatize canais mais informativos para a previsão do RUL.

Em seguida, aplica-se uma GAT espacial. Nesse caminho, os sensores são tratados como nós de um grafo. As arestas representam possíveis relações entre sensores. Como a estrutura física exata de dependência entre os sensores não é imposta manualmente, a GAT aprende pesos de atenção que indicam quais relações sensor-sensor são mais relevantes. O resultado é uma matriz de características espaciais de dimensão (30, 14).

Essa etapa é importante porque a degradação de um sistema mecânico geralmente não aparece isoladamente em uma variável. Ela pode surgir como mudança de relação entre temperatura, pressão, rotação e vazão. O canal espacial permite capturar esse tipo de dependência.

4.8.4 Canal temporal

O canal temporal busca responder a outra pergunta: quais ciclos dentro da janela são mais relevantes para estimar o RUL? Nem todos os 30 ciclos carregam a mesma informação. Alguns podem representar comportamento estável, outros podem conter indícios de mudança de regime, e outros podem estar mais próximos de uma fase de degradação acelerada.

Primeiro, o mecanismo de atenção temporal atribui pesos aos passos temporais. Depois, a GAT temporal trata os ciclos como nós de um grafo, aprendendo relações entre instantes da janela. A saída também possui dimensão (30, 14), mas agora representa características temporais extraídas da sequência.

Esse caminho complementa o canal espacial. Enquanto o canal espacial observa relações entre sensores, o canal temporal observa relações entre ciclos. A separação em dois canais evita misturar prematuramente essas duas fontes de informação.

4.8.5 Concatenação espaço-temporal

As saídas espacial e temporal, ambas com dimensão (30, 14), são reunidas pela concatenação da Equação 21 (Zeng, Wu e Cheng, 2023):

$$Z_t = Z_t^{\text{esp}} \parallel Z_t^{\text{temp}} \in \mathbb{R}^{30 \times 28}. \quad (21)$$

Na Equação 21, Z_t^{esp} contém as 14 características produzidas pelo canal espacial; Z_t^{temp} contém as 14 características do canal temporal; e $\parallel$ representa concatenação na dimensão de atributos. Cada um dos 30 ciclos passa, assim, a reunir 28 características: metade derivada das relações entre sensores e metade das relações entre instantes.

4.8.6 Codificação posicional

Como o Transformer não possui recorrência, ele não reconhece naturalmente a ordem dos elementos da sequência. Por isso, aplica-se uma codificação posicional à matriz concatenada. Essa codificação informa ao modelo a posição de cada ciclo dentro da janela.

Sem codificação posicional, o Transformer poderia saber quais padrões existem na janela, mas não em que ordem eles aparecem. Em RUL, a ordem é fundamental: uma elevação gradual de temperatura, por exemplo, tem significado diferente de uma sequência em que a temperatura sobe e depois retorna ao padrão normal.

4.8.7 Blocos Transformer

Após a codificação posicional, a sequência (30, 28) é processada por seis blocos codificadores e seis decodificadores, com sete cabeças de atenção. Como $D_{\text{model}} = 28$, cada cabeça opera em um subespaço de dimensão $D_{\text{head}} = 28/7 = 4$.

A atenção multi-cabeças permite que o modelo observe diferentes tipos de relação simultaneamente. Uma cabeça pode se concentrar em tendências lentas, outra em transições abruptas, outra em relações específicas entre ciclos e outra em padrões globais da janela. A presença de múltiplas cabeças aumenta a capacidade de representar relações complementares.

Após os blocos Transformer, seleciona-se o último passo temporal da sequência. Essa decisão é coerente com a formulação da tarefa: o RUL previsto corresponde ao estado no ciclo final da janela, depois de o modelo ter processado todo o contexto anterior. A transformação final é sintetizada pela Equação 22 (Zeng, Wu e Cheng, 2023):

$$\hat{y}_t = g\left(Z_{t,\text{ultimo}}^{(L)}; \theta_g\right). \quad (22)$$

Na Equação 22, $Z_{t,\text{ultimo}}^{(L)} \in \mathbb{R}^{28}$ é a representação do último passo após o último bloco Transformer; $g(\cdot)$ representa as camadas finais de regressão; θ_g são seus parâmetros; e $\hat{y}_t$ é o RUL previsto. A operação produz um único valor porque o problema é de regressão escalar, e não de classificação.

4.9 RESUMO DA ARQUITETURA REPRODUZIDA

A Tabela 10 resume o fluxo da arquitetura reproduzida.

Tabela 10 – Fluxo da arquitetura AGATT reproduzida.

Etapa	Dimensão principal	Função
Entrada	(30, 14)	Janela com 30 ciclos e 14 sensores ativos
Conv1D <i>depthwise</i>	(30, 14)	Extrair padrões locais em cada sensor separadamente
Atenção espacial	(30, 14)	Ponderar a importância dos sensores
GAT espacial	(30, 14)	Capturar relações sensor-sensor
Atenção temporal	(30, 14)	Ponderar a importância dos ciclos
GAT temporal	(30, 14)	Capturar relações tempo-tempo
Concatenação	(30, 28)	Combinar informação espacial e temporal
Codificação posicional	(30, 28)	Inserir informação de ordem temporal
Transformer	(30, 28)	Modelar dependências de longo alcance
Seleção do último passo	(1, 28)	Representar o estado atual da janela
Camadas densas	$28 \rightarrow 1$	Produzir a estimativa final de RUL

Fonte: Elaborada pelo autor (2026), com base em Zeng, Wu e Cheng (2023).

O total de parâmetros treináveis da implementação foi de aproximadamente 144.111, coerente com o porte esperado para uma arquitetura híbrida com convolução, atenção, GAT e Transformer.

4.10 HIPERPARÂMETROS E DECISÕES DE TREINAMENTO

A Tabela 11 apresenta os principais hiperparâmetros utilizados na reprodução.

Tabela 11 – Hiperparâmetros e configuração de treinamento do AGATT na C-MAPSS FD001.

Parâmetro	Valor
Subdataset	FD001
Sensores utilizados	14
Sensores removidos	s1, s5, s6, s10, s16, s18 e s19
Tamanho da janela	30 ciclos
Formato da entrada	(30, 14)
Normalização	Min-max em [0, 1]
<i>cap_RUL</i>	125 ciclos
<i>Batch size</i>	256
Taxa de aprendizado inicial	1×10^{-3}
Otimizador	Adam
<i>clipnorm</i>	1,0
<i>Dropout</i>	0,2
Cabeças de atenção	7
Blocos Transformer	6 codificadores + 6 decodificadores
Ambiente	Google Colab Pro, GPU NVIDIA A100 40 GB

Fonte: Elaborada pelo autor (2026).

O tamanho de janela igual a 30 ciclos foi mantido por ser compatível com o artigo-base e por representar um equilíbrio entre contexto temporal e custo computacional. Janelas muito curtas podem não conter sinais suficientes de degradação; janelas muito longas aumentam custo, redundância e risco de incluir trechos pouco informativos.

O *batch size* = 256 permitiu boa utilização da GPU A100 sem saturação de memória. Lotes maiores reduzem a variância do gradiente, mas podem dificultar a generalização; lotes menores aumentam ruído de treinamento. O valor 256 apresentou bom equilíbrio entre estabilidade e desempenho.

A taxa de aprendizado inicial 1×10^{-3} foi adotada com o otimizador Adam. Como arquiteturas com atenção podem ser sensíveis a atualizações muito grandes, foi utilizado *clipnorm*=1,0 para limitar a norma dos gradientes e reduzir instabilidades. O *dropout*=0,2 desativa aleatoriamente 20% das ativações durante o treinamento, reduzindo coadaptação e sobreajuste; na inferência, todas as unidades permanecem ativas com o escalonamento aprendido.

Também foram utilizados mecanismos de controle de treinamento. O mecanismo *Reduce LR on Plateau* reduz a taxa de aprendizado quando a perda de validação estaciona, permitindo refinamento progressivo. O *EarlyStopping* interrompe o treinamento quando não há ganho relevante na validação, restaurando os melhores pesos encontrados. Esses mecanismos evitam treinamento excessivo e reduzem o risco de sobreajuste.

4.11 MÉTRICAS DE AVALIAÇÃO

A métrica principal foi o RMSE, definido pela Equação 23 (Saxena *et al.*, 2008a; Zeng, Wu e Cheng, 2023):

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}. \quad (23)$$

Na Equação 23, n é a quantidade de amostras; i é o índice; y_i é o RUL real; e $\hat{y}_i$ é o RUL previsto. O RMSE eleva os erros ao quadrado antes de calcular a média; por isso, erros grandes exercem influência maior que erros pequenos.

Também foram calculados MAE, MAPE, MSE, R^2 e Score NASA. O Score NASA é assimétrico e utiliza o erro assinado $d_i = \hat{y}_i - y_i$, conforme a Equação 24 (Saxena *et al.*, 2008):

$$s_i = \begin{cases} \exp(-d_i/13) - 1, & d_i < 0, \\ \exp(d_i/10) - 1, & d_i \geq 0. \end{cases} \quad S = \sum_{i=1}^n s_i. \quad (24)$$

Na Equação 24, s_i é a penalização da amostra i , S é o Score total e d_i é positivo quando o modelo superestima o RUL. O denominador 10 faz a penalização crescer mais rapidamente na superestimação do que na subestimação, cujo denominador é 13. Em termos operacionais, prever vida maior que a real pode postergar uma intervenção; prever vida menor tende a antecipá-la. Quanto menor o Score, melhor o desempenho.

4.12 RESULTADOS DA REPRODUÇÃO

A Tabela 12 apresenta os resultados da melhor execução do AGATT reproduzido na FD001.

Tabela 12 – Métricas da melhor execução do AGATT na C-MAPSS FD001.

Métrica	Valor
RMSE	13,88
MAE	10,19
MAPE	18,45%
MSE	192,56
R^2	0,8885
Score	275,56

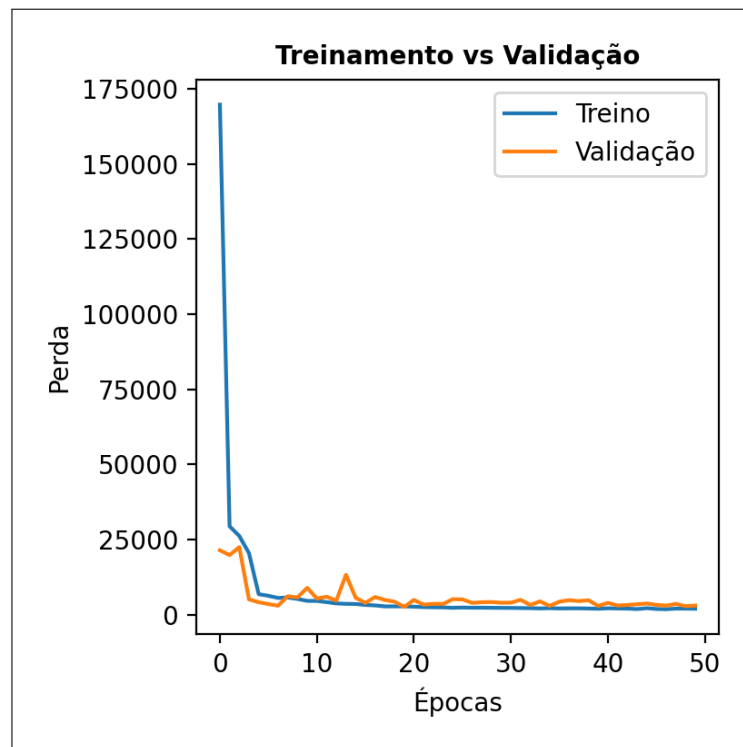
Fonte: Elaborada pelo autor (2026).

O RMSE de 13,88 ciclos indica que o erro médio quadrático da previsão ficou em torno de 14 ciclos. O MAE de 10,19 ciclos, menor que o RMSE, sugere que a maior parte dos erros é moderada, mas que alguns erros maiores elevam a média quadrática. O $R^2 = 0,8885$ indica que

o modelo explicou aproximadamente 88,85% da variabilidade do RUL no conjunto de teste, o que representa bom ajuste global.

A Figura 17 mostra as curvas de perda durante treinamento e validação. A convergência estável indica que o modelo conseguiu reduzir o erro sem apresentar sobreajuste evidente nas curvas observadas.

Figura 17 – Curvas de perda no treinamento e validação do AGATT reproduzido.

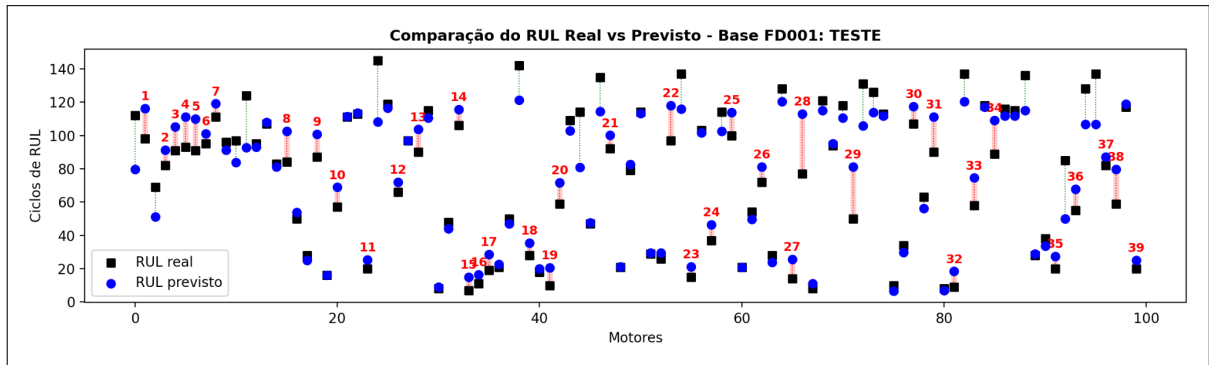


Fonte: Elaborada pelo autor (2026).

A Figura 18 compara o RUL real e o RUL previsto no conjunto de teste. Pontos próximos à diagonal indicam boa aderência entre previsão e valor real. Erros positivos, em que o modelo prevê RUL maior do que o real, são particularmente importantes porque representam uma situação operacionalmente mais arriscada: o modelo indicaria que ainda há mais vida útil disponível quando, na realidade, a falha está mais próxima.

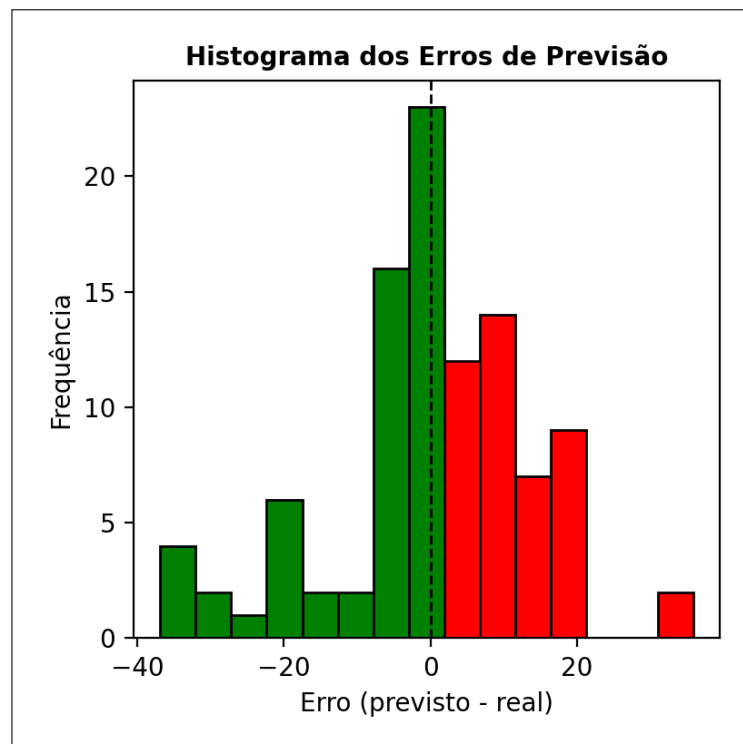
A Figura 19 apresenta a distribuição dos erros. A concentração próxima de zero indica que o modelo produziu previsões coerentes para a maior parte dos motores. A assimetria da análise é relevante porque, em manutenção preditiva, errar para menos e errar para mais não têm o mesmo impacto: subestimar o RUL pode antecipar uma intervenção, enquanto superestimar o RUL pode atrasar uma manutenção necessária.

Figura 18 – Comparação entre RUL real e RUL previsto no conjunto de teste. Superestimações do RUL são destacadas por sua relevância operacional.



Fonte: Elaborada pelo autor (2026).

Figura 19 – Distribuição dos erros de previsão do AGATT reproduzido. A concentração próxima de zero sugere boa aderência geral.



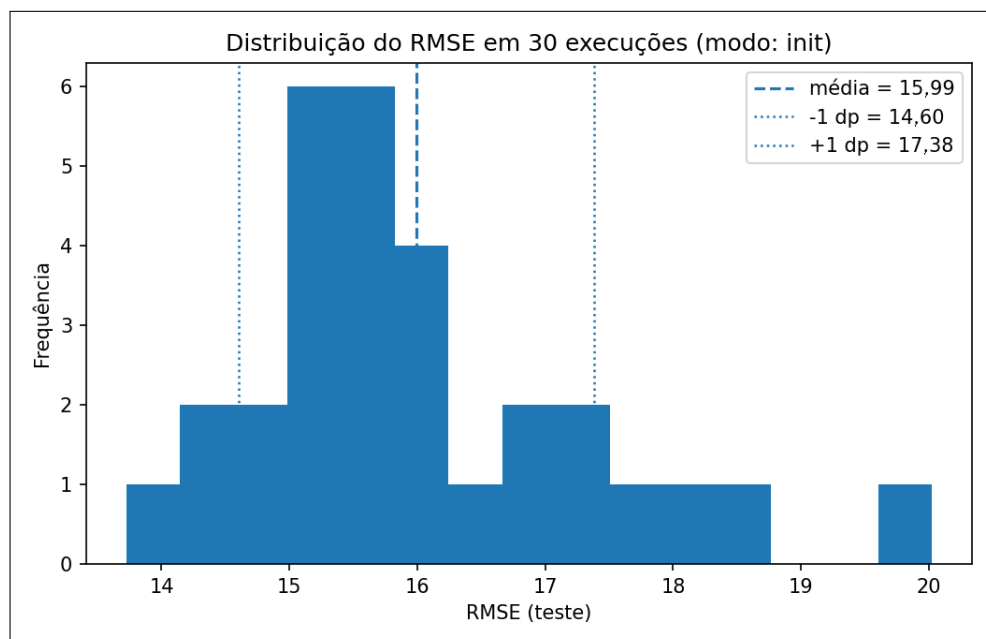
Fonte: Elaborada pelo autor (2026).

4.13 ANÁLISE DE REPRODUTIBILIDADE

Além da execução principal, foi realizada uma análise com 30 execuções independentes para avaliar a estabilidade do treinamento. Esse procedimento é importante porque redes profundas possuem componentes estocásticos, como inicialização dos pesos, embaralhamento dos lotes, divisão de validação e dinâmica de otimização. Assim, uma única execução pode não representar adequadamente o comportamento típico da arquitetura.

A Figura 20 apresenta a distribuição do RMSE nas 30 execuções independentes.

Figura 20 – Distribuição do RMSE em 30 execuções independentes no modo de inicialização avaliado. As linhas verticais indicam a média de 15,99 ciclos e os limites de um desvio-padrão: 14,60 e 17,38 ciclos.



Fonte: Elaborada pelo autor (2026).

A análise das 30 execuções resultou em RMSE médio de $15,99 \pm 1,39$ ciclos. O melhor resultado observado foi $\text{RMSE} = 13,88$, utilizado como referência de desempenho da melhor execução. A diferença entre a melhor execução e a média das execuções evidencia a sensibilidade do treinamento a fatores estocásticos, como inicialização dos pesos, ordenação dos lotes, dinâmica de otimização, seleção do melhor ponto de parada e possíveis pequenas variações numéricas do ambiente computacional.

Essa variação é esperada em modelos profundos, especialmente em arquiteturas com mecanismos de atenção, grafos e função de perda assimétrica. Em termos científicos, ela reforça que a comparação entre modelos de RUL deve considerar não apenas o menor RMSE obtido em uma execução, mas também a estabilidade do método ao longo de múltiplas sementes.

Essa constatação é particularmente relevante para a literatura da FD001, na qual diferenças pequenas de RMSE são frequentemente interpretadas como avanço de estado da arte. Quando vários modelos ocupam uma faixa estreita de desempenho, a variabilidade entre execuções

pode ser da mesma ordem de grandeza das diferenças reportadas entre arquiteturas. Portanto, a avaliação por múltiplas execuções é mais informativa do que a apresentação de um único valor pontual.

4.14 COMPARAÇÃO COM O ARTIGO ORIGINAL

Zeng, Wu e Cheng (2023) reportam RMSE de 11,45 ciclos na FD001. A melhor execução alcançou RMSE de 13,88 ciclos, enquanto a média das 30 execuções independentes foi $15,99 \pm 1,39$ ciclos. Embora a reprodução não tenha recuperado numericamente o valor reportado no artigo-base, os resultados demonstram que a arquitetura foi capaz de aprender padrões relevantes de degradação e alcançar desempenho competitivo na FD001.

A diferença numérica pode decorrer de múltiplos fatores: inicialização aleatória, versões de bibliotecas, detalhes de implementação das camadas de atenção e GAT, critério de parada, composição da validação, função de perda, seleção de *checkpoint* e pequenas diferenças no pré-processamento. Em modelos profundos, especialmente arquiteturas com atenção e grafos, esses detalhes podem alterar o resultado final mesmo quando a estrutura geral da rede é preservada.

Portanto, a conclusão principal não é que o AGATT original esteja incorreto, mas que sua reprodução exata exige controle rigoroso de detalhes que nem sempre são completamente especificados em artigos. Essa observação reforça uma contribuição metodológica da dissertação: em problemas de RUL, principalmente quando os resultados são usados como base para transferência de aprendizado, a reprodutibilidade deve ser tratada como parte central da avaliação.

4.15 INTERPRETAÇÃO DOS RESULTADOS

Os resultados obtidos indicam que o AGATT reproduzido conseguiu aprender uma relação consistente entre padrões multivariados de sensores e RUL no FD001. O bom valor de R^2 sugere que a rede capturou parte substancial da variabilidade do alvo. O RMSE e o MAE indicam erro médio compatível com a literatura, ainda que superior ao valor pontual reportado no artigo original.

Do ponto de vista arquitetural, o resultado confirma a utilidade de combinar três níveis de representação. A convolução *depthwise* extrai padrões locais de cada sensor. Os canais espacial e temporal separam dois tipos de dependência: relações entre sensores e relações entre ciclos. O Transformer, por fim, integra essas informações e modela dependências de longo alcance dentro da janela.

Essa estrutura é adequada para posterior adaptação ao contexto das bombas B-0302. Em bombas industriais, assim como em *turbofans*, a degradação pode se manifestar por padrões temporais e por relações entre sensores. A diferença é que os sensores, modos de falha e regimes operacionais não são os mesmos. Por isso, o AGATT não é transferido como solução pronta, mas como uma arquitetura de referência capaz de representar degradação em séries multivariadas.

4.16 SÍNTESE DO CAPÍTULO

Este capítulo apresentou a base C-MAPSS, justificou a escolha do FD001, detalhou o pré-processamento dos dados, explicou a arquitetura AGATT reproduzida e discutiu os resultados obtidos. A base C-MAPSS foi utilizada por ser pública, padronizada, amplamente aceita na literatura e adequada para treinamento inicial de modelos de RUL. O FD001 foi escolhido por oferecer um cenário controlado, com uma condição operacional e um modo de falha, facilitando a reprodução e interpretação dos resultados.

A reprodução do AGATT utilizou 14 sensores, janela temporal de 30 ciclos, normalização min-max e $cap_RUL = 125$. A arquitetura reproduzida combinou convolução *depthwise*, atenção espacial, GAT espacial, atenção temporal, GAT temporal, concatenação espaço-temporal, codificação posicional e blocos Transformer. O melhor resultado obtido foi RMSE de 13,88 ciclos e $R^2 = 0,8885$, enquanto a análise de 30 execuções independentes resultou em $RMSE = 15,99 \pm 1,39$.

Mais importante do que o valor isolado de RMSE, a reprodução mostrou que a arquitetura AGATT é capaz de representar padrões de degradação em séries temporais multivariadas no FD001. Esse experimento, conduzido com janela de 30 ciclos, serviu como etapa de *benchmarking*, análise de reprodutibilidade e compreensão dos componentes da arquitetura. Para a transferência de aprendizado ao domínio industrial, foi posteriormente treinada uma nova instância do AGATT com janela efetiva de 35 pontos e configuração própria para utilização como *backbone*. Essa versão, seus parâmetros de treinamento e sua forma de utilização com os sensores das bombas B-0302 são detalhados no Capítulo 5.

5 TRANSFER LEARNING PARA B-0302

Este capítulo descreve a adaptação do AGATT pré-treinado no subconjunto FD001 da C-MAPSS ao domínio industrial do conjunto B-0302. Como os domínios diferem quanto à natureza dos equipamentos, aos sensores, à quantidade de trajetórias e aos mecanismos de falha, a transferência exigiu organizar as séries industriais, harmonizar os canais de entrada e treinar uma nova cabeça de regressão sobre o *backbone* congelado.

As Bombas B-0302A, B-0302B, B-0302C e B-0302D foram inicialmente avaliadas sob o mesmo protocolo. O aprofundamento concentrou-se na B-0302D porque ela reuniu maior riqueza sensorial, eventos mais compatíveis com degradação progressiva e melhor desempenho exploratório. Após essa delimitação, o capítulo apresenta o protocolo de validação, o *backbone*, a projeção físico-estatística, as análises de sensibilidade, o resultado principal, as linhas de base e a calibração operacional.

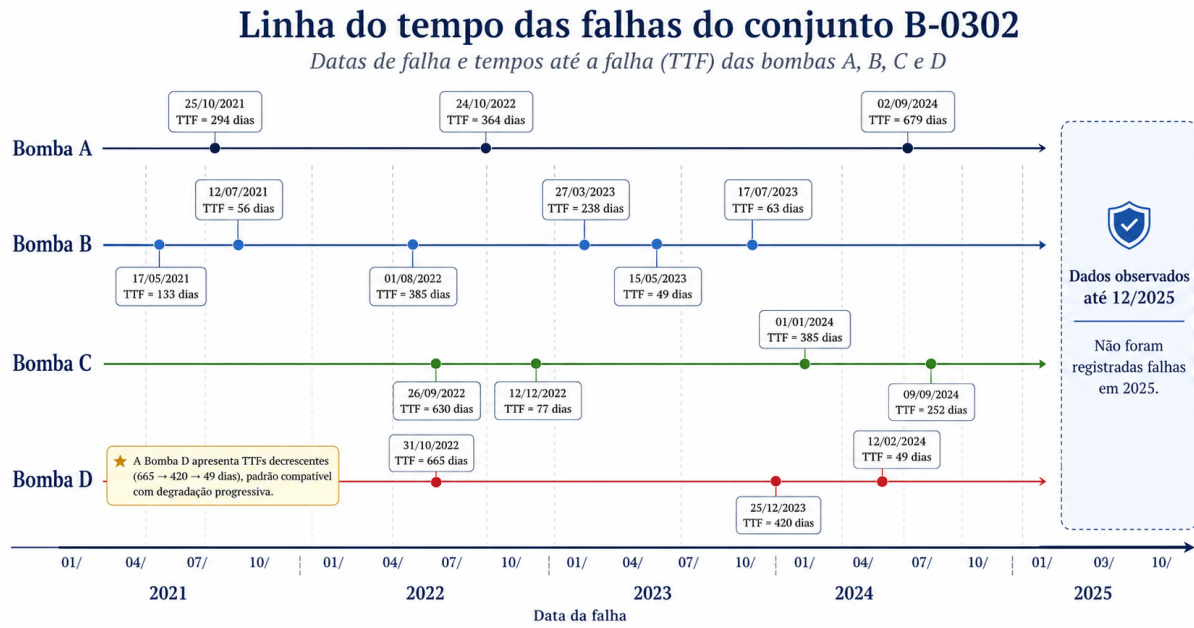
5.1 CARACTERIZAÇÃO DO DOMÍNIO INDUSTRIAL DE DESTINO

A etapa de transferência de aprendizado foi estruturada inicialmente para as quatro bombas do conjunto B-0302: A, B, C e D. Essa escolha preserva a coerência com a análise de Engenharia de Confiabilidade apresentada no Capítulo 3, na qual se observou que as unidades possuem históricos de falha, modos de ocorrência e graus de previsibilidade distintos.

As bombas operam em paralelo e em regime de rodízio. Embora possuam horímetros, não foi possível obter, para todo o período estudado, uma série histórica contínua e temporalmente sincronizada de horas de funcionamento por bomba que pudesse ser correlacionada de modo seguro às leituras dos sensores e às datas das falhas. Por isso, adotaram-se dias de calendário como escala comum. Essa escolha permite alinhar observações e eventos, mas não mede diretamente a exposição operacional: duas bombas com o mesmo intervalo de calendário podem ter acumulado horas de funcionamento diferentes.

A compatibilidade entre o mecanismo de falha e o domínio de origem também condiciona a transferência. Existem técnicas específicas para diagnóstico e prognóstico de falhas elétricas quando são monitoradas grandezas adequadas, como corrente por fase, tensão, isolamento, harmônicos, partidas e atuações de proteção (Nandi, Toliyat e Li, 2005). A limitação observada nesta dissertação não implica que falhas elétricas sejam imprevisíveis. O FD001 representa trajetórias de degradação progressiva, e o AGATT pré-treinado tende, portanto, a transferir melhor quando a falha industrial também produz uma evolução gradual observável nos sensores disponíveis. Eventos de comando, proteção, comunicação ou instrumentação sem trajetória compatível exigiriam sensores, rótulos ou modelos especificamente direcionados a esses mecanismos.

Figura 21 – Linha do tempo das falhas registradas para as bombas A, B, C e D do conjunto B-0302. Os marcadores indicam as datas de falha e os respectivos tempos até a falha, em dias. A base foi observada até dezembro de 2025, sem registro de novas falhas em 2025.



Fonte: Elaborada pelo autor (2026), com apoio de inteligência artificial generativa e prompt do autor.

5.1.1 Histórico de falhas e construção das séries

A primeira etapa consistiu em organizar o histórico de falhas das bombas A, B, C e D. Para cada evento de falha conhecido, foi construída uma série temporal retrospectiva, composta pelos dados de sensores disponíveis antes da data do evento. O RUL foi calculado como a distância, em dias, entre cada observação da série e a respectiva data de falha. Dessa forma, quanto mais próximo o ponto está da data de falha, menor é seu RUL.

A Figura 21 apresenta a linha do tempo das falhas consideradas para cada bomba. Os marcadores indicam as datas de falha e os respectivos TTFs. Observa-se que as bombas possuem quantidades diferentes de eventos e intervalos de TTF bastante heterogêneos. A base foi observada até dezembro de 2025 e não houve registro de novas falhas em 2025.

A leitura da Figura 21 é relevante por três razões. Primeiro, ela mostra que o conjunto industrial é pequeno do ponto de vista de aprendizado supervisionado, pois cada bomba possui poucas trajetórias completas até a falha. Segundo, ela evidencia que os TTFs não seguem um padrão homogêneo entre as bombas. Terceiro, ela permite observar se há indícios de encurtamento progressivo dos intervalos entre falhas, o que pode sugerir um processo de degradação acumulada.

Na Bomba A, os TTFs observados foram 294, 364 e 679 dias. Esses valores não indicam encurtamento progressivo do tempo entre falhas. Além disso, os registros associados aos eventos são heterogêneos, envolvendo elétrica, vedação e motor. Na Bomba B, os TTFs foram 133, 56, 385, 238, 49 e 63 dias. Embora haja maior quantidade de eventos, a natureza predominante das ocorrências está associada a falhas elétricas, comando, controle e disjuntor, o que reduz a

expectativa de que os sensores de processo capturem uma trajetória gradual de degradação. Na Bomba C, os TTFs foram 630, 77, 385 e 252 dias, com eventos mistos envolvendo vibração, motor, alimentação elétrica, instrumentação e comunicação. Já a Bomba D apresentou TTFs de 665, 420 e 49 dias, com problemas recorrentes de vedação, selo mecânico e gotejamento, formando a sequência mais compatível com degradação progressiva entre as quatro bombas.

Essa heterogeneidade motivou a adoção de uma validação por série de falha, em vez de uma divisão aleatória simples dos pontos. Em problemas de RUL, pontos vizinhos de uma mesma trajetória temporal são fortemente correlacionados. Uma divisão aleatória poderia colocar pontos muito próximos da mesma falha simultaneamente no treino e no teste, gerando uma avaliação artificialmente otimista. Por isso, o protocolo foi estruturado para avaliar a capacidade do modelo em generalizar para pontos ocultados dentro das séries de falha, preservando a interpretação temporal do problema.

5.1.2 Protocolo de validação por série de falha e pontos ocultados

O protocolo de avaliação foi desenhado para simular uma situação operacional em que o modelo recebe medições parciais de uma bomba e precisa estimar quanto tempo falta até a falha, sem acesso à informação futura. Para isso, cada falha conhecida foi tratada como uma série independente. A partir de uma data de falha, construiu-se a trajetória retrospectiva dos sensores e calculou-se o RUL de cada ponto da série.

Em seguida, foram selecionados pontos ocultados no final de cada série de falha, dentro de um intervalo de RUL de interesse. No protocolo principal, esse intervalo foi de 5 a 60 dias. Isso significa que o teste não foi feito em qualquer ponto da trajetória, mas em pontos próximos o suficiente da falha para terem relevância operacional e, ao mesmo tempo, distantes o suficiente para representar uma previsão antecipada. O limite superior de 60 dias define o horizonte máximo de interesse nesta aplicação, enquanto o limite inferior de 5 dias evita concentrar a avaliação em pontos de falha praticamente iminente.

O termo “ocultar pontos” significa que esses pontos são retirados do processo de treinamento e reservados para avaliação. O modelo é treinado sem acesso a esses exemplos e, posteriormente, deve prever o RUL correspondente a eles. Essa estratégia é mais adequada do que avaliar o modelo em todos os pontos da série, pois reproduz o problema real de prognóstico: no momento da previsão, o futuro ainda não é conhecido.

Além da ocultação dos pontos, foi adotado um rodízio por série de falha. Em cada rodada de validação, uma série de falha é separada para teste e não participa do treinamento. As demais séries são usadas para treinar a nova cabeça de regressão sobre as características produzidas pelo *backbone* congelado. Na rodada seguinte, outra série é separada para teste, e o processo se repete. Esse procedimento reduz o risco de vazamento temporal entre treino e teste, pois evita que a própria trajetória avaliada contribua para o ajuste das camadas treináveis, para os limites de normalização ou para as distribuições usadas na projeção sensorial.

Esse desenho experimental é particularmente importante porque a quantidade de falhas

industriais é pequena. Se pontos de uma mesma trajetória fossem misturados livremente entre treino e teste, o modelo poderia aprender padrões específicos daquela falha e produzir uma estimativa excessivamente otimista. Com o rodízio por série, a avaliação torna-se mais exigente: o modelo precisa aplicar à série separada o que aprendeu nas demais trajetórias.

Também foi testada uma estratégia adicional de ponderação temporal entre séries, na qual séries de falha temporalmente mais próximas entre si recebiam maior peso no treinamento. A hipótese era que falhas mais próximas no tempo poderiam refletir condições operacionais, ambientais ou de manutenção mais semelhantes, favorecendo a transferência interna entre séries. Entretanto, essa penalização/ponderação temporal não produziu melhoria consistente nos resultados. Por isso, ela não foi mantida na configuração principal. Esse resultado é relevante porque indica que a proximidade cronológica entre eventos, isoladamente, não foi suficiente para melhorar a capacidade preditiva; a natureza física da falha e a riqueza dos sensores mostraram-se mais determinantes.

Ressalta-se que os pontos de teste utilizados neste protocolo não correspondem a falhas independentes adicionais. Eles são obtidos a partir das séries temporais associadas às falhas reais disponíveis no histórico de manutenção. Para cada série de falha, são ocultados diferentes pontos temporais antes da data do evento, simulando situações retrospectivas nas quais o modelo deve estimar a Vida Útil Remanescente sem observar diretamente a falha futura. Assim, quando se reporta um conjunto de 30 pontos de avaliação, trata-se de 30 pontos de ocultamento temporal derivados das trajetórias de falha existentes, e não de 30 eventos de falha estatisticamente independentes. Essa distinção é importante para a interpretação das métricas, que devem ser entendidas como evidência retrospectiva de aderência do modelo às trajetórias observadas, e não como validação populacional ampla do comportamento de todas as bombas.

5.1.3 Síntese comparativa das bombas A, B, C e D

A síntese técnico-operacional a seguir relaciona quatro dimensões que influenciam diretamente a viabilidade do *transfer learning*: quantidade de eventos, padrão dos TTFs, natureza predominante das falhas e expectativa de previsibilidade por sensores. Embora todas as bombas tenham sido consideradas na análise inicial, a Bomba D apresentou a combinação mais favorável de histórico de falhas, coerência física e observabilidade da degradação.

A Bomba A apresenta poucos eventos, TTFs relativamente longos e registros heterogêneos. Como os eventos envolvem naturezas diferentes, incluindo elétrica, vedação e motor, o modelo não encontra uma assinatura sensorial única e recorrente que possa ser associada ao encurtamento do RUL. Isso não significa que a Bomba A não possua risco operacional, mas sim que seu histórico disponível não favorece a modelagem supervisionada de RUL pelo protocolo proposto.

A Bomba B apresenta mais eventos, o que inicialmente poderia sugerir maior potencial para aprendizado. Entretanto, quantidade de falhas não é suficiente para caracterizar previsibilidade. A predominância de falhas elétricas, de comando, controle e disjuntor indica com-

portamento mais abrupto e irregular. Nesses casos, o equipamento pode operar com sinais aparentemente normais até a ocorrência de um evento de proteção, comando ou alimentação. Assim, mesmo com mais exemplos, a relação entre trajetória sensorial e RUL tende a ser menos estável.

A Bomba C representa um caso intermediário. A triagem por horários indicou algum sinal de previsibilidade, especialmente quando se observa a redução de RMSE e Score NASA em relação às Bombas A e B. Entretanto, os eventos permanecem mistos, envolvendo vibração, motor, alimentação elétrica, instrumentação e comunicação. Essa mistura dificulta a defesa de um mecanismo dominante de degradação progressiva. Por esse motivo, a Bomba C foi considerada promissora em relação às Bombas A e B, mas ainda insuficiente para o aprofundamento principal deste capítulo.

A Bomba D, por sua vez, reuniu a melhor combinação de condições. Seus TTFs decrescentes, de 665 para 420 e depois 49 dias, sugerem encurtamento progressivo dos intervalos entre falhas. Os registros recorrentes de vedação, selo mecânico e gotejamento são fisicamente mais compatíveis com degradação acumulada do que falhas abruptas de comando ou proteção elétrica. Além disso, a Bomba D possui maior riqueza sensorial, com sensores adicionais de pressão e temperatura. Assim, ela apresentou o cenário mais favorável para testar se o AGATT pré-treinado no FD001 poderia ser adaptado com sucesso ao domínio industrial.

Portanto, embora as quatro bombas tenham sido submetidas a tentativas de transferência de aprendizado, apenas a Bomba D apresentou desempenho e coerência física suficientes para justificar o detalhamento completo das escolhas metodológicas. Essa delimitação é importante para a interpretação da dissertação: o objetivo não é afirmar que o modelo funciona igualmente para qualquer bomba, mas demonstrar que a transferência é viável quando existe compatibilidade mínima entre mecanismo de falha, sensoriamento disponível e estrutura temporal dos dados.

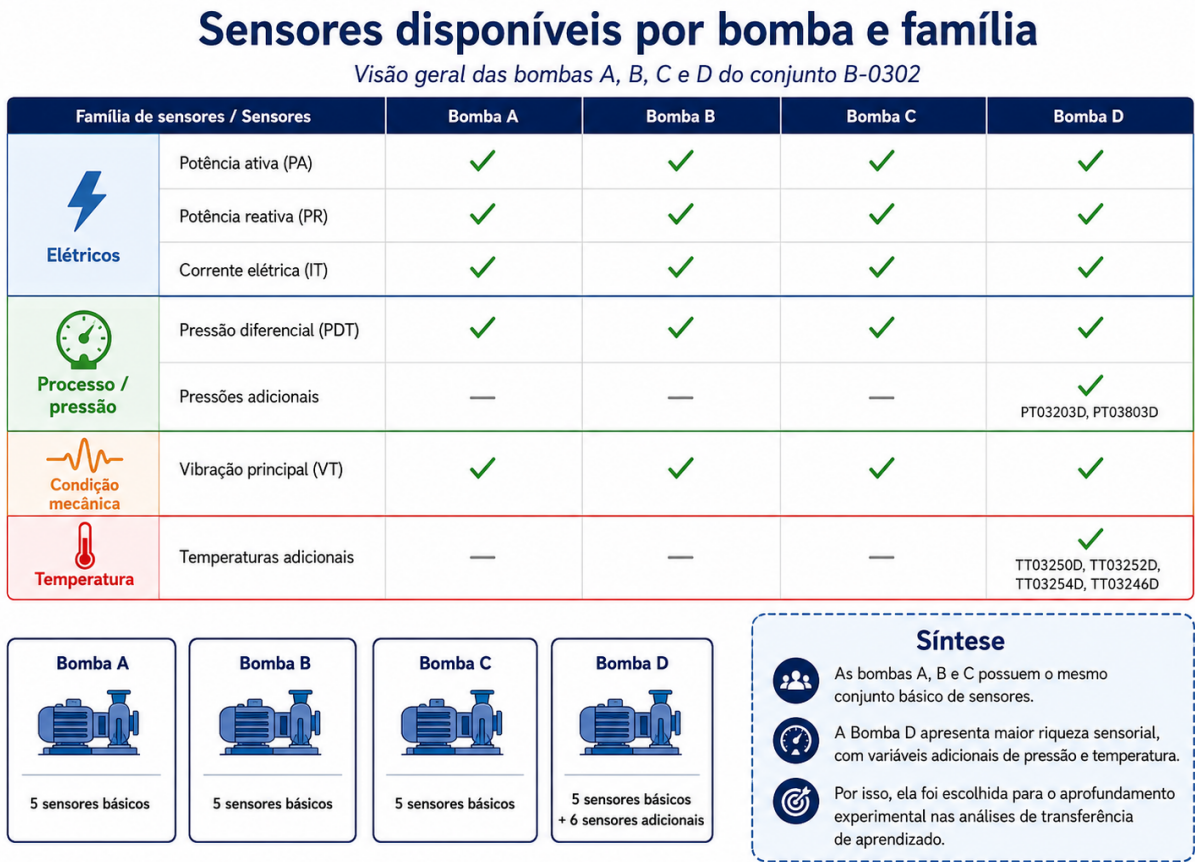
5.1.4 Sensores disponíveis por bomba

Além do histórico de falhas, foi necessário levantar os sensores disponíveis em cada bomba. Essa etapa é essencial porque a transferência de aprendizado depende da qualidade, da diversidade e da continuidade das variáveis monitoradas. No conjunto analisado, as bombas A, B e C possuem essencialmente o mesmo conjunto básico de sensores, composto por variáveis elétricas, de processo e de condição mecânica. A Bomba D, entretanto, apresenta maior riqueza sensorial, incluindo sensores adicionais de pressão e temperatura.

A Figura 22 resume a distribuição dos sensores por bomba e por família. Todas as bombas possuem medições de potência ativa, potência reativa, corrente elétrica, pressão diferencial e vibração principal. A Bomba D dispõe ainda de variáveis adicionais de pressão e temperatura, o que amplia a capacidade de observar diferentes aspectos do processo físico associado à degradação.

A Figura 22 complementa a análise de falhas ao mostrar a capacidade de observação disponível para cada equipamento. Em modelos de RUL, sensores não são apenas variáveis

Figura 22 – Sensores disponíveis por bomba e por família no conjunto B-0302. As bombas A, B e C compartilham o mesmo conjunto básico de sensores, enquanto a Bomba D apresenta sensores adicionais de pressão e temperatura.



Fonte: Elaborada pelo autor (2026), com apoio de inteligência artificial generativa e prompt do autor.

explicativas genéricas; eles são a forma pela qual o processo físico de degradação se torna observável. Uma falha mecânica progressiva pode ser parcialmente percebida por mudanças em vibração, temperatura, pressão ou corrente. Porém, se o conjunto de sensores for reduzido ou se a falha ocorrer em um subsistema pouco monitorado, a trajetória de degradação pode não estar suficientemente representada nos dados.

Nesse sentido, a Bomba D apresenta vantagem metodológica. A presença de sensores adicionais de pressão e temperatura permite representar melhor variações hidráulicas e térmicas associadas a problemas de vedação, selo mecânico e gotejamento. Esses fenômenos podem alterar condições de sucção, descarga, diferencial de pressão, esforço do conjunto e aquecimento localizado. Assim, a riqueza sensorial da Bomba D aumenta a probabilidade de que o modelo encontre padrões temporais associados à aproximação da falha.

5.1.5 Análise exploratória dos sensores industriais da Bomba D

Antes da aplicação do pipeline de transferência de aprendizado, foi realizada uma análise exploratória dos sensores industriais da Bomba D. O objetivo dessa etapa não foi produzir um modelo preditivo, mas verificar se os sinais disponíveis apresentavam relações físicas e

estatísticas minimamente coerentes com o problema de RUL. Essa análise é relevante porque, em dados industriais reais, a simples existência de sensores não garante que eles contenham informação útil sobre a aproximação da falha.

Foram avaliadas duas perspectivas complementares. Na análise de correlação, foram mantidos 16 sensores com variabilidade e disponibilidade suficientes no conjunto preprocessado, buscando identificar relações lineares entre variáveis de mesma família física ou de famílias operacionalmente relacionadas, como pressão, corrente, vibração e temperatura. A segunda consistiu no alinhamento das séries temporais pela data de falha, permitindo observar o comportamento dos sinais em função dos dias restantes até o evento.

A matriz de correlação auxilia a identificar redundâncias e complementaridades entre sensores. Sensores altamente correlacionados podem carregar informação semelhante, enquanto sensores pouco correlacionados podem representar aspectos distintos do estado operacional do equipamento. Essa análise também é importante para justificar a etapa posterior de projeção físico-estatística com diversidade condicional, pois a reutilização excessiva de sensores semelhantes poderia reduzir a diversidade da entrada projetada para o espaço FD001.

A Figura 23 apresenta as correlações entre os sensores disponíveis na B-0302D.

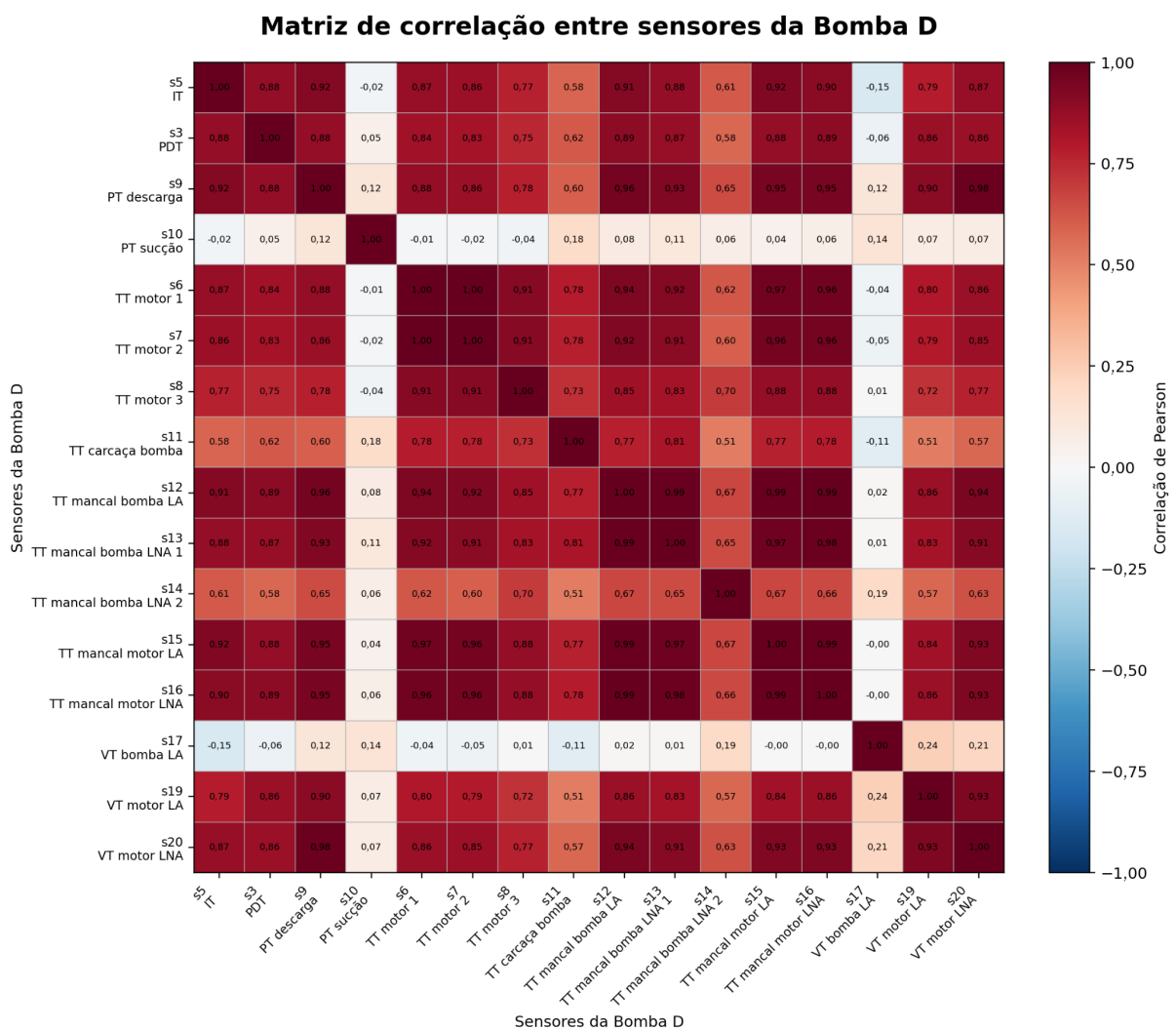
A Figura 23 indica a presença de blocos de associação positiva entre sensores hidráulicos, elétricos, térmicos e de vibração. Observa-se que variáveis como pressão de descarga, corrente elétrica, temperaturas de carcaça e temperaturas de mancais apresentam correlações elevadas entre si. Esse comportamento é fisicamente coerente, pois aumentos de carga hidráulica tendem a elevar o esforço do conjunto motobomba, refletindo-se em maior demanda elétrica, alterações de pressão e aumento de temperatura em componentes estruturais e mancais.

Também se observa um bloco térmico relativamente coeso, formado por temperaturas de carcaça do motor, carcaça da bomba e mancais. Esse agrupamento sugere que parte dos sensores térmicos compartilha um componente comum associado ao regime operacional e à dissipação de calor do equipamento. Tal achado reforça a pertinência de considerar famílias físicas na etapa de projeção sensorial, pois sensores de mesma natureza podem carregar informação redundante, enquanto sensores de famílias diferentes podem contribuir com aspectos complementares do estado da bomba.

A pressão de sucção apresentou comportamento menos associado a parte dos demais sensores quando comparada à pressão de descarga, o que é compatível com sua maior dependência das condições hidráulicas a montante. Já sensores de vibração e temperaturas de mancais apresentaram correlações positivas com variáveis térmicas e hidráulicas, sugerindo acoplamento entre regime de operação e condição mecânica. Essa relação é relevante no contexto da Bomba D, cujas falhas estão associadas a vedação, selo mecânico e gotejamento, fenômenos que podem alterar simultaneamente pressão, esforço mecânico, aquecimento e vibração.

Assim, a matriz de correlação não foi utilizada como critério único de seleção de sensores, mas como evidência exploratória de que os sinais da Bomba D possuem estrutura física interpretável. A presença de blocos correlacionados também justifica o uso da penalização de

Figura 23 – Matriz de correlação de Pearson entre os 16 sensores da Bomba D que apresentaram variabilidade e disponibilidade suficientes para a análise exploratória. Observam-se blocos de associação entre pressões, corrente elétrica, temperaturas de carcaça, temperaturas de mancais e vibrações, sugerindo acoplamento entre regime hidráulico, esforço eletromecânico e condição térmica e mecânica do equipamento.



Fonte: Elaborada pelo autor (2026).

diversidade na projeção físico-estatística: sem esse controle, sensores muito correlacionados poderiam ser reutilizados excessivamente, reduzindo a diversidade informacional da entrada projetada para o espaço FD001.

Além da correlação, as séries da Bomba D foram alinhadas em relação à data de falha. Nesse tipo de visualização, o eixo horizontal representa o tempo restante até o evento, de modo que $RUL = 0$ corresponde à data da falha. Essa representação permite comparar trajetórias associadas a diferentes eventos em uma mesma escala temporal, evidenciando se há alterações progressivas nos sensores à medida que a falha se aproxima.

Para evitar uma visualização excessivamente densa, a Figura 24 apresenta sensores selecionados da Bomba D. A seleção priorizou sensores com variação observável na janela analisada e com relevância física para o problema, incluindo pressão diferencial/filtro, temperatura de carcaça do motor, pressão de descarga, pressão de sucção, temperatura da carcaça da bomba e temperatura do mancal da bomba. Sensores constantes ou sem variabilidade útil no período, como alguns canais básicos do conjunto, foram mantidos no pipeline quando aplicável, mas não foram destacados nessa visualização exploratória.

A Figura 24 complementa a análise de falhas e sensores ao mostrar a forma das trajetórias antes do evento. Mesmo quando a degradação não aparece como uma tendência perfeitamente monotônica em todos os sensores, a visualização permite avaliar a estabilidade dos sinais, a presença de regiões de maior oscilação e a diferença entre séries de falha. Essa análise reforça que a Bomba D possui riqueza sensorial e histórico físico mais compatíveis com a tentativa de prognóstico do que as demais bombas.

Os dados industriais empregados na configuração principal foram organizados com uma observação por sensor e por dia, no horário selecionado pelo protocolo experimental. Para a B-0302D, adotou-se a extração às 00:00. Os 20 canais usados nas três trajetórias apresentaram 100% de observações válidas, sem valores sentinela e sem lacunas temporais; portanto, não foi aplicada imputação, interpolação ou preenchimento estatístico.

5.1.6 Normalização e suavização temporal sem uso de observações futuras

Como os canais industriais possuem unidades e amplitudes distintas, eles foram colocados em uma escala comum. Essa etapa é necessária porque corrente, pressão, vibração e temperatura possuem unidades e amplitudes muito diferentes. Sem normalização, uma variável numericamente maior poderia dominar o treinamento apenas por sua escala, e não por conter mais informação sobre a degradação.

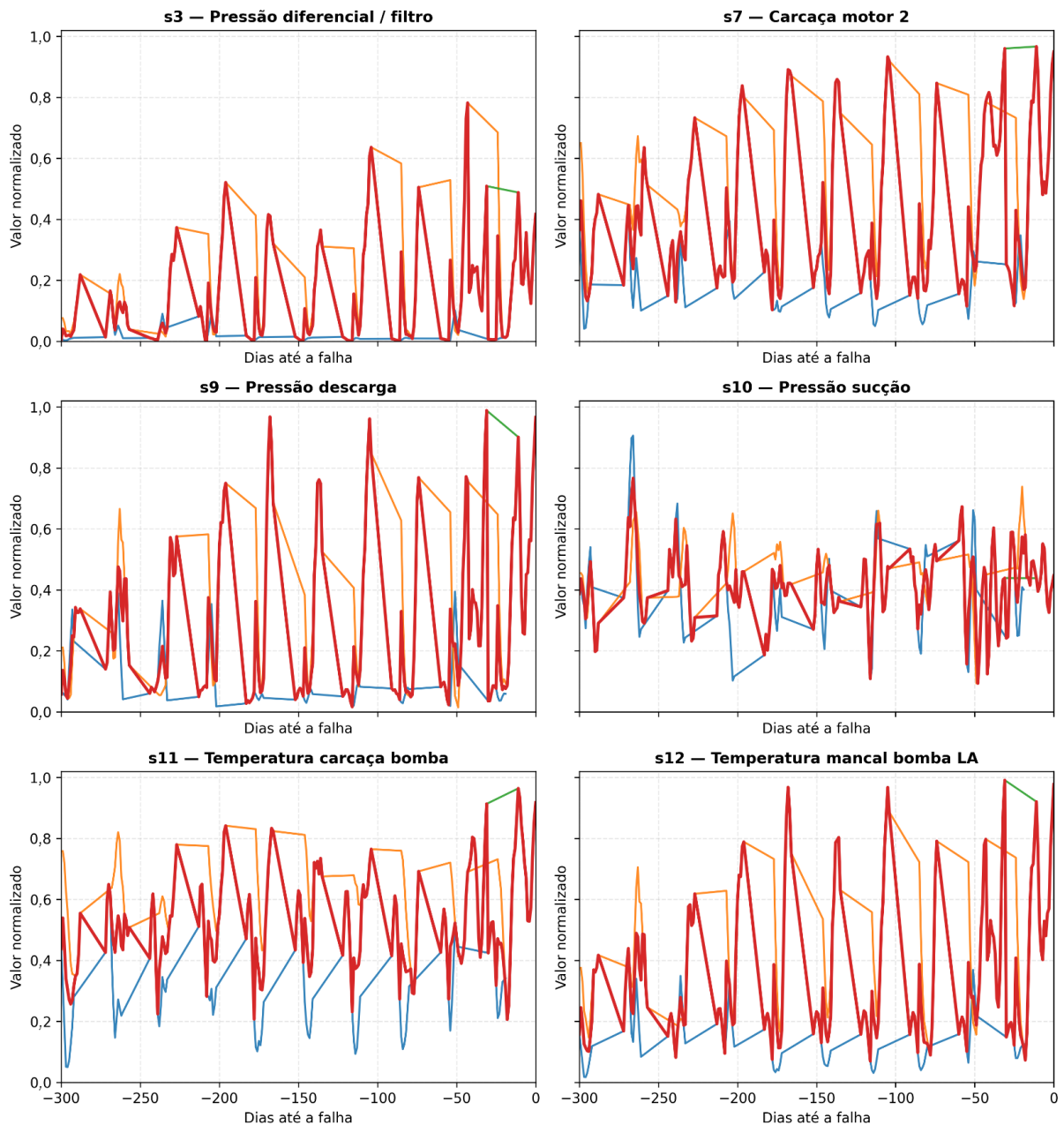
Em cada rodada externa de validação, os limites de normalização foram estimados nas séries reservadas para treinamento. Para reduzir a influência de valores extremos, cada canal foi inicialmente limitado entre os percentis 1 e 99 calculados no conjunto de treinamento. Em seguida, aplicou-se a transformação definida na Equação 25 para limitar e normalizar cada canal

Figura 24 – Séries temporais selecionadas da Bomba D alinhadas pela data de falha. O eixo horizontal representa os dias até a falha, com $RUL = 0$ na data do evento, e foi limitado a até 300 dias antes da falha para padronização visual. As séries possuem comprimentos diferentes; por exemplo, D-3 possui TTF de 49 dias e aparece apenas no trecho final da janela. Foram destacados sensores com variação observável e relevância física para a análise exploratória: pressão diferencial/filtro, carcaça motor 2, pressão de descarga, pressão de sucção, temperatura da carcaça da bomba e temperatura do mancal da bomba LA.

Bomba D — séries selecionadas alinhadas à falha

Pré-processamento: hybrid_time_median_bomba, normalização min-max e suavização MM=3

— D-1 — D-2 — D-3 — média



Nota: o eixo foi limitado a até 300 dias antes da falha. Séries com TTF menor, como D-3, aparecem apenas no trecho disponível.

Fonte: Elaborada pelo autor (2026).

no intervalo $[0, 1]$:

$$x'_t = \frac{\text{clip}(x_t, p_1, p_{99}) - x_{\min}}{x_{\max} - x_{\min} + \varepsilon}. \quad (25)$$

Na Equação 25, x_t é a observação original no instante t ; $\text{clip}(\cdot)$ limita o valor ao intervalo definido pelos percentis p_1 e p_{99} , calculados no treinamento; $x_{\min}$ e $x_{\max}$ são os extremos após essa limitação; x'_t é o valor normalizado; e ε é uma constante positiva pequena que evita divisão por zero. Os mesmos parâmetros foram então aplicados à série externa de teste. Dessa forma, a escala da série avaliada não foi utilizada para ajustar a transformação.

Após a normalização, foi aplicada uma média móvel causal de três observações, indicada por $\text{smooth_w} = 3$. A suavização causal é definida pela Equação 26, que usa a observação atual e até duas anteriores:

$$\tilde{x}_t = \frac{1}{k_t} \sum_{j=0}^{k_t-1} x'_{t-j}. \quad k_t = \min(3, t+1). \quad (26)$$

Na Equação 26, $\tilde{x}_t$ é o valor suavizado no instante t ; x'_{t-j} são as observações normalizadas atual e anteriores; j é o índice retrospectivo do somatório; e $k_t = \min(3, t+1)$ é a quantidade de observações disponíveis, limitada a três. A janela é retrospectiva e não centralizada; portanto, nenhuma observação posterior ao instante previsto participa da suavização. Entre os níveis efetivamente avaliados durante o desenvolvimento, $\text{smooth_w} = 3$ apresentou o melhor comportamento na configuração reproduzida da Bomba D. Sua interpretação é a de uma suavização moderada: evita tanto a ausência de filtragem quanto o uso de médias móveis longas, que poderiam eliminar variações locais potencialmente informativas. Assim, o valor 3 foi mantido como o melhor compromisso empírico observado entre estabilidade e preservação temporal.

5.1.7 Influência do horário de coleta dos sensores

Como as bombas do conjunto B-0302 operam em paralelo e com alternância de carga, o instante diário de extração dos sensores pode capturar regimes operacionais distintos. Uma mesma bomba pode estar em operação, em reserva ou em transição de regime conforme o horário observado. Por esse motivo, o horário de coleta foi tratado como uma variável metodológica, e não como uma escolha meramente operacional. Foram comparadas extrações diárias às 00:00, 06:00, 10:00, 12:00 e 15:00, todas fora do período de ponta considerado pela operação.

Essa análise foi conduzida com um protocolo homogêneo de triagem, utilizando *backbone* com janela efetiva de 35 pontos, janela industrial de 35 pontos, *loss* NASA, projeção físico-estatística com diversidade condicional, taxa de aprendizado de 2×10^{-5} , $\lambda_D = 0,25$ e seleção de 10 pontos ocultados por série no intervalo de 5 a 60 dias. O objetivo foi verificar se o horário de coleta alterava de forma relevante a qualidade preditiva do *transfer learning* nas quatro bombas do conjunto B-0302.

A comparação quantitativa completa é apresentada na Tabela 13, que inclui R^2 , MAE, MedAE, viés, RMSE e Score NASA. A Mediana do Erro Absoluto (MedAE) foi incluída por ser menos sensível a valores extremos do que o MAE.

Tabela 13 – Resultados de transferência de aprendizado para as bombas A, B, C e D em diferentes horários de coleta, utilizando *backbone* com janela efetiva de 35 pontos, janela industrial de 35 pontos e *loss* NASA.

Horário	Bomba	n	R^2	MAE	MedAE	Viés	RMSE	Score NASA
00:00	A	30	-1,76	21,92	19,77	+7,35	25,57	556,55
00:00	B	60	-2,15	22,24	20,88	-4,49	26,04	699,54
00:00	C	40	-0,50	17,25	17,63	+1,11	19,82	239,16
00:00	D	30	0,44	9,02	9,34	+0,43	10,67	48,30
06:00	A	30	-1,55	20,94	18,61	+8,47	24,58	533,46
06:00	B	60	-1,79	21,10	19,80	-1,41	24,70	673,27
06:00	C	40	-0,33	14,84	14,17	-2,34	18,68	196,04
06:00	D	30	-0,18	13,34	9,84	+4,68	15,56	126,53
10:00	A	30	-1,61	22,09	21,46	+4,57	24,83	353,44
10:00	B	60	-1,64	20,26	18,90	-3,37	24,01	571,63
10:00	C	40	-0,38	15,60	15,89	-1,71	19,00	195,25
10:00	D	30	-0,15	12,09	9,57	-7,65	15,39	78,04
12:00	A	30	-1,41	21,20	20,92	+5,05	23,87	330,56
12:00	B	60	-1,32	18,68	16,39	-7,80	22,54	438,78
12:00	C	40	-0,21	14,73	12,78	-0,08	17,82	180,70
12:00	D	30	-0,52	14,24	10,81	-5,85	17,69	114,45
15:00	A	30	-1,71	21,69	20,73	+5,04	25,34	431,16
15:00	B	60	-1,23	18,38	18,30	-8,30	22,09	397,27
15:00	C	40	-0,41	15,83	15,11	+3,84	19,25	242,93
15:00	D	30	0,10	10,27	7,14	+1,95	13,63	99,05

Fonte: Elaborada pelo autor (2026).

Os resultados mostraram que o horário de coleta influencia o desempenho preditivo, mas esse efeito não é uniforme entre as bombas. As Bombas A e B apresentaram desempenho insatisfatório em todos os horários avaliados, com R^2 negativos e erros elevados, o que reforça a interpretação de que seus históricos de falha, mais heterogêneos ou dominados por eventos abruptos, não produziram trajetórias sensoriais suficientemente consistentes para sustentar a previsão de RUL pelo protocolo adotado.

A Bomba C apresentou comportamento intermediário, com melhora relativa em vários horários e melhor resultado em 12:00. Nesse horário, a bomba obteve $R^2 = -0,21$, MAE de 14,73 dias, RMSE de 17,82 dias e Score NASA de 180,70. Esse resultado indica que a Bomba C contém algum sinal preditivo útil e é sensível ao horário de coleta. Ainda assim, seus resultados permaneceram inferiores aos observados para a Bomba D.

A Bomba D foi a que apresentou o comportamento mais consistente e superior no conjunto dos horários avaliados. Em todos os cinco horários testados, a Bomba D apresentou o menor RMSE e o menor Score NASA entre as quatro bombas. O melhor resultado absoluto

ocorreu no horário de 00:00, com $R^2 = 0,44$, MAE de 9,02 dias, RMSE de 10,67 dias e Score NASA de 48,30. Esse resultado, somado à maior riqueza sensorial da bomba e à maior coerência física de seus modos de falha, reforçou sua seleção como caso de aprofundamento neste capítulo.

Portanto, a adoção de 00:00 no aprofundamento não decorreu de uma suposição de que a bomba necessariamente estaria operando nesse instante. Ela resultou de uma comparação empírica entre horários operacionalmente plausíveis. No protocolo avaliado, 00:00 produziu a melhor combinação de R^2 , RMSE, MAE e Score NASA para a Bomba D. A Tabela 13 documenta essa escolha e mostra que o horário de amostragem pode alterar substancialmente a condição capturada pelos sensores.

A análise quantitativa do horário de coleta também fortaleceu a justificativa metodológica para o aprofundamento da Bomba D. Embora a Bomba C tenha apresentado desempenho intermediário e algum potencial, apenas a Bomba D reuniu, de forma simultânea, superioridade preditiva, riqueza sensorial e coerência física suficiente para sustentar a continuidade das análises de transferência de aprendizado.

5.1.8 Da reprodução com janela 30 ao *backbone* com janela 35

O Capítulo 4 apresentou o experimento de reprodução e *benchmarking* do AGATT com janela de 30 ciclos. Essa execução teve como finalidade verificar o comportamento da arquitetura na FD001 e estudar sua reprodutibilidade. Ela não deve ser confundida com o arquivo de pesos utilizado na melhor configuração industrial.

As primeiras tentativas de transferência partiram da configuração de janela 30, por ser a reprodução inicialmente disponível e por constituir a referência apresentada no Capítulo 4. Esses testes mostraram, entretanto, que um bom resultado no domínio de origem não garantia automaticamente a melhor adaptação ao domínio industrial. Foram então avaliadas novas combinações de semente, comprimento de janela do modelo de origem, janela de geração industrial e suavização. Dentro desse conjunto de experimentos, a instância do AGATT treinada com janela efetiva de 35 pontos apresentou o comportamento mais favorável no protocolo da Bomba D.

A janela 35 não foi escolhida por uma hipótese de superioridade universal nem para substituir retrospectivamente o experimento de *benchmarking*. Ela identifica o arquivo de pesos que, entre os *backbones* efetivamente testados, forneceu as características mais úteis para a adaptação industrial. Por isso, o Capítulo 4 preserva os resultados de reprodução com janela 30, enquanto este capítulo descreve separadamente a geração e o uso do *backbone* de janela 35.

Por essa razão, foi treinada uma nova instância da mesma família AGATT no FD001, especificamente para atuar como *backbone*. Essa execução utilizou semente 1007, janela 35, passo de deslocamento igual a 1, os 14 sensores ativos do FD001 e truncamento do RUL de treinamento em 125 ciclos. A normalização min-max foi ajustada sobre o conjunto de treinamento. O treinamento utilizou Adam, taxa inicial de aprendizado 10^{-3} , *clipnorm* igual a 1, lote de 256 amostras, até 400 épocas, *EarlyStopping* com paciência 30 e *Reduce LR on Plateau* com paciência 10 e taxa mínima 10^{-5} .

Tabela 14 – Configuração utilizada para gerar o *backbone* AGATT de janela 35 empregado na transferência.

Elemento	Configuração
Base de origem	C-MAPSS FD001
Semente	1007
Janela efetiva	35 ciclos
Sensores	14 sensores ativos do FD001
RUL de treinamento	Truncado em 125 ciclos
Normalização	Min–max ajustada no treinamento
Otimizador	Adam, taxa inicial 10^{-3} , <i>clipnorm</i> =1
Treinamento	Lote 256, até 400 épocas
Controle de treinamento	<i>EarlyStopping</i> =30; <i>Reduce LR on Plateau</i> =10; taxa mínima 10^{-5}
Função de perda	Score NASA assimétrico
Uso posterior	Extrator de características congelado no domínio industrial

Fonte: Elaborada pelo autor (2026).

A arquitetura permaneceu coerente com a reprodução apresentada no capítulo anterior: convolução *depthwise*, atenções espacial e temporal, GAT nos dois eixos, seis blocos codificadores, seis blocos decodificadores, sete cabeças de atenção e *dropout* de 0,2. O melhor *checkpoint* foi selecionado pela perda de validação, salvo em formato *.keras*, recarregado em um processo limpo e novamente avaliado. Essa verificação de recarregamento confirmou que o arquivo versionado reproduzia as métricas obtidas antes do salvamento.

Na avaliação do próprio FD001, esse *checkpoint* de janela 35 apresentou $R^2 = 0,848$, RMSE de 16,23 ciclos, MAE de 12,17 ciclos e Score NASA de 366,35. Esses valores não substituem o *benchmarking* do Capítulo 4; eles apenas identificam o modelo de origem efetivamente salvo para a transferência. A janela 35 foi mantida porque forneceu melhor comportamento no destino industrial entre os *backbones* testados, e não porque tivesse produzido o menor RMSE no FD001.

A Tabela 14 apresenta a configuração do *backbone* efetivamente usado na transferência.

5.2 PIPELINE GERAL DE TRANSFERÊNCIA DE APRENDIZADO

Após a caracterização do domínio industrial e a definição do *backbone*, o pipeline foi organizado em cinco etapas: construção das séries diárias por falha; normalização e suavização causal; projeção dos canais industriais para 14 dimensões; extração de características pelo AGATT congelado; e treinamento e avaliação da nova cabeça de regressão.

O fluxo começa com uma observação diária por sensor, em horário fixo. Como a B-0302D não apresentou lacunas ou valores sentinela nos canais usados, a etapa inicial consistiu apenas em verificar integridade, normalizar e suavizar causalmente os sinais. As observações foram então organizadas em séries retrospectivas, com RUL em dias de calendário, e os pontos ocultos

Tabela 15 – Fluxo entre os canais industriais, o *backbone* congelado e a cabeça de regressão treinada para a Bomba D.

Etapa	Saída principal	Papel no pipeline
Janela industrial	36×20	Reúne os 20 canais processados da Bomba D.
Projeção fixa do <i>fold</i>	36×14	Compatibiliza os canais industriais com os 14 sensores ativos do FD001.
Adaptação temporal	35×14	Preserva os 35 pontos mais recentes exigidos pelo <i>backbone</i> .
AGATT congelado	Vetor com 28 características	Extraí representações espaço-temporais aprendidas no FD001; seus pesos não são atualizados.
Dense(64), ReLU	64 características	Aprende a combinação industrial das características extraídas.
Dropout(0,10)	64 características	Reduz coadaptação durante o treinamento; não possui parâmetros.
Dense(1), linear	1 valor de RUL	Produz a estimativa contínua, em dias.

Fonte: Elaborada pelo autor (2026).

formaram o conjunto externo de avaliação.

A Bomba D disponibilizou 20 canais de entrada no conjunto processado. Na análise exploratória de correlação, 16 sensores apresentaram variabilidade e disponibilidade suficientes para a visualização. Para alimentar o AGATT, entretanto, o pipeline trabalha com a matriz de canais industriais disponível em cada *fold* e a transforma, por meio da projeção físico-estatística, em uma representação de 14 dimensões compatível com os sensores ativos do FD001.

Quando a janela industrial é maior do que a janela do *backbone*, o pipeline preserva os 35 pontos mais recentes. Assim, na configuração principal, cada amostra é inicialmente formada com 36 observações, projetada para 14 dimensões e, antes de entrar no AGATT, recortada para as 35 observações finais. O *backbone* nunca recebe 36 pontos; sua forma de entrada permanece exatamente aquela utilizada no pré-treinamento.

O AGATT é carregado sem a camada escalar final de RUL. A saída da camada imediatamente anterior, com 28 características, é utilizada como representação aprendida da janela. Todos os pesos desse extrator permanecem congelados. Sobre esse vetor são acrescentadas duas camadas densas treináveis: uma camada com 64 unidades e ativação ReLU e uma camada linear com uma unidade para a previsão do RUL. Entre elas, aplica-se *dropout* de 0,10, que não possui parâmetros e atua apenas como regularização durante o treinamento.

A Tabela 15 detalha a cabeça de regressão treinada sobre o *backbone* congelado.

Como o vetor de entrada da cabeça possui 28 características, a camada Dense(64) contém $28 \times 64 + 64 = 1.856$ parâmetros e a camada Dense(1) contém $64 + 1 = 65$. Portanto, apenas 1.921 parâmetros foram ajustados com os dados industriais, enquanto a parte profunda do AGATT permaneceu preservada. Essa redução é particularmente importante porque, em cada *fold* externo da Bomba D, apenas duas trajetórias de falha ficam disponíveis para treinamento.

A taxa de aprendizado de 2×10^{-5} controla somente a atualização dessas novas camadas.

Tabela 16 – Configuração experimental principal utilizada no aprofundamento da Bomba D.

Elemento	Valor adotado	Justificativa
Bomba de aprofundamento	D	Maior riqueza sensorial, melhor resultado na triagem por horário e histórico mais compatível com degradação progressiva.
<i>Backbone</i>	AGATT FD001, semente 1007	Instância treinada especificamente para a transferência e salva após seleção do melhor <i>checkpoint</i> .
Janela do <i>backbone</i>	35 pontos	Melhor comportamento entre as configurações avaliadas após tentativas iniciais com a reprodução de janela 30.
Estado do <i>backbone</i>	Congelado	Preserva as representações aprendidas no FD001 e evita reajustar uma rede profunda com poucas trajetórias industriais.
Cabeça treinável	Dense(64)–Dropout(0,10)–Dense(1)	Converte as características extraídas pelo AGATT em previsão contínua de RUL.
Janela de geração industrial	36 pontos	Melhor resultado entre as janelas avaliadas; o <i>backbone</i> continua recebendo apenas os 35 pontos mais recentes.
<i>smooth_w</i>	3	Média móvel causal de três observações, escolhida como compromisso entre redução de ruído e preservação das variações locais.
Dados faltantes	Não aplicável na configuração principal	Os 20 canais utilizados apresentaram cobertura completa; foi realizada somente verificação de integridade.
Projeção	Físico-estatística com diversidade condicional	Compatibiliza os canais industriais com os 14 sensores ativos do FD001, combinando coerência física e aderência estatística.
λ_D	0,25	Penalização moderada por reutilização, sem impedir que um sensor seja repetido quando sua aderência é claramente superior.
Intervalo de RUL ocultado	5–60 dias	Melhor equilíbrio observado entre antecipação operacional e desempenho global.
Pontos ocultados	10 por série	Distribui a avaliação por diferentes horizontes dentro de cada trajetória de falha.
Taxa de aprendizado da cabeça	2×10^{-5}	Atualização conservadora das novas camadas treináveis.
Função de perda principal	NASA	Melhor resultado global de regressão na configuração final.
Função de perda operacional	<i>horizon_conservative</i>	Direciona o treinamento para maior cautela em horizontes próximos da falha.

Fonte: Elaborada pelo autor (2026).

O valor baixo foi mantido por apresentar treinamento estável nas tentativas realizadas; ele não modifica os pesos internos do AGATT, pois o *backbone* permanece congelado.

A aplicação do pipeline às quatro bombas teve caráter exploratório e comparativo. Nas Bombas A e B, os resultados permaneceram fracos, coerentemente com a heterogeneidade e a natureza abrupta de seus eventos. A Bomba C apresentou comportamento intermediário e sensível ao horário. A Bomba D reuniu melhor desempenho, maior riqueza de sinais e falhas mais compatíveis com degradação progressiva.

5.2.1 Configuração experimental principal

A configuração experimental principal foi consolidada após as análises de bomba, horário, janela, suavização, projeção e horizonte de RUL. A Tabela 16 reúne os valores adotados e, principalmente, a justificativa de cada escolha.

A janela 36 deve ser interpretada como parâmetro de geração das amostras industriais, e não como dimensão efetiva do AGATT. Para uma amostra com dados suficientes, o recorte

final apresentado ao *backbone* contém os mesmos 35 instantes mais recentes associados ao alvo. Alterar window modifica, contudo, quais amostras podem ser geradas ao longo das séries, a quantidade de janelas disponíveis e, em séries curtas, o tratamento de borda necessário para formar uma janela completa. Esse tratamento de borda pertence à construção das amostras e não corresponde à imputação de dados faltantes. Assim, a vantagem observada para 36 foi empírica no pipeline completo e não deve ser atribuída à entrada de um ponto adicional no AGATT.

O intervalo 5–60 dias também faz parte do desenho experimental. O limite de 60 dias delimita o horizonte máximo de interesse, enquanto o limite inferior de 5 dias evita restringir a análise à falha praticamente iminente. Os testes com limites inferiores alternativos são apresentados adiante.

5.3 PROJEÇÃO DOS SENSORES INDUSTRIAIS PARA O ESPAÇO FD001

Uma etapa central da transferência de aprendizado foi a compatibilização entre os sensores industriais das bombas B-0302 e os sensores do C-MAPSS FD001. O modelo AGATT utilizado como *backbone* foi treinado originalmente com os sensores ativos do FD001, que representam variáveis de motores aeronáuticos, como temperaturas, pressões, rotações, fluxos e variáveis de condição. Já o domínio industrial deste trabalho é formado por bombas centrífugas, monitoradas por sensores de potência, corrente, pressão, vibração e temperatura. Portanto, não existe correspondência física direta e perfeita entre os dois conjuntos de sensores.

Por esse motivo, foi criada uma etapa de projeção sensorial. O objetivo dessa projeção é transformar o vetor de sensores industriais da bomba em uma representação compatível com o espaço de entrada esperado pelo AGATT. Em termos simples, a projeção responde à seguinte pergunta: dado um sensor ativo do FD001, qual sensor industrial da bomba pode representá-lo de forma fisicamente plausível e estatisticamente semelhante?

5.3.1 Agrupamento físico dos sensores e alternativas de projeção

A primeira decisão metodológica foi organizar os sensores dos dois domínios em famílias físicas. Essa organização não impõe uma equivalência direta entre motor aeronáutico e bomba centrífuga, mas restringe o conjunto de candidatos a associações minimamente plausíveis do ponto de vista de engenharia. Assim, sensores térmicos do FD001 foram comparados com sensores térmicos da Bomba D; sensores de pressão foram comparados com pressões industriais; variáveis de fluxo e rotação foram aproximadas por variáveis proxy relacionadas ao carregamento hidráulico e elétrico; e variáveis de condição mecânica foram aproximadas por vibração e temperaturas estruturais.

A Tabela 17 apresenta as famílias físicas usadas para restringir os candidatos da projeção.

Inicialmente, foram avaliadas correspondências manuais entre os sensores do FD001 e os sensores da Bomba D, baseadas exclusivamente em plausibilidade física. Essa alternativa, embora interpretável, apresentou duas limitações. A primeira é que nem todos os sensores

Tabela 17 – Agrupamento físico utilizado para orientar a projeção entre sensores do FD001 e sensores industriais da Bomba D.

Família física	Sensores do FD001	Sensores candidatos na Bomba D
Temperatura	Sensores associados a temperaturas do motor aeronáutico.	Temperaturas de carcaça da bomba, carcaça do motor e mancais.
Pressão	Sensores associados a pressões internas do motor.	Pressão de sucção, pressão de descarga e pressão diferencial.
Fluxo / rotação	Sensores associados a rotação, fluxo e regime operacional.	Corrente elétrica, potência ativa, potência reativa e pressões como proxies de carregamento.
Condição mecânica	Sensores associados à condição operacional no FD001.	Vibração, temperaturas estruturais e outras variáveis relacionadas ao estado mecânico.

Fonte: Elaborada pelo autor (2026).

do FD001 possuem equivalente direto no domínio de bombas centrífugas. A segunda é que a escolha manual não considera a aderência estatística efetiva entre as distribuições dos sensores. Assim, duas variáveis fisicamente próximas poderiam apresentar comportamentos estatísticos muito diferentes, enquanto sensores de famílias correlatas poderiam representar melhor o padrão observado no domínio industrial.

Por esse motivo, a abordagem final adotada foi a projeção físico-estatística com diversidade condicional. Essa estratégia manteve a restrição de plausibilidade física por famílias de engenharia, mas passou a selecionar os sensores candidatos com base em uma distância estatística composta. Além disso, incluiu uma penalização por reutilização para evitar que muitos sensores do FD001 fossem representados pelo mesmo sensor industrial da Bomba D.

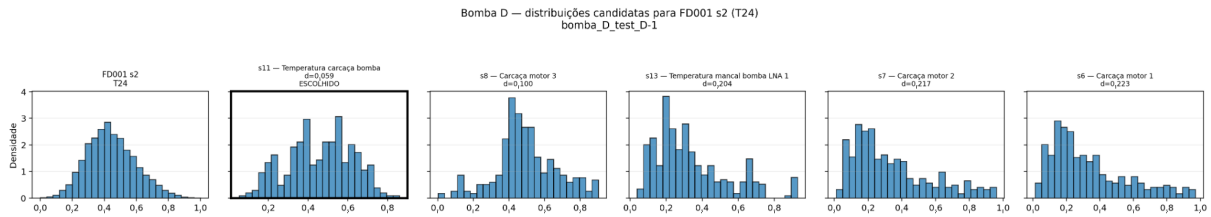
5.3.2 Comparação visual das distribuições na escolha do de-para

Além da análise numérica da distância composta, foi realizada uma inspeção visual das distribuições dos sensores candidatos na Bomba D. O objetivo dessa etapa foi tornar interpretável o processo de seleção do de-para entre os sensores ativos do FD001 e os sensores industriais do domínio de destino.

Essa visualização é importante porque a projeção proposta não deve ser entendida como uma simples associação nominal entre sensores. Em alguns casos, a correspondência física é relativamente direta, como ocorre entre sensores de temperatura do FD001 e temperaturas estruturais da bomba. Em outros casos, a equivalência direta não existe, como nas variáveis de fluxo e rotação do motor aeronáutico, que precisaram ser aproximadas por variáveis industriais capazes de representar regime de carregamento, como corrente, potência ou pressão. Por isso, a inspeção visual das distribuições foi utilizada como evidência complementar à distância composta.

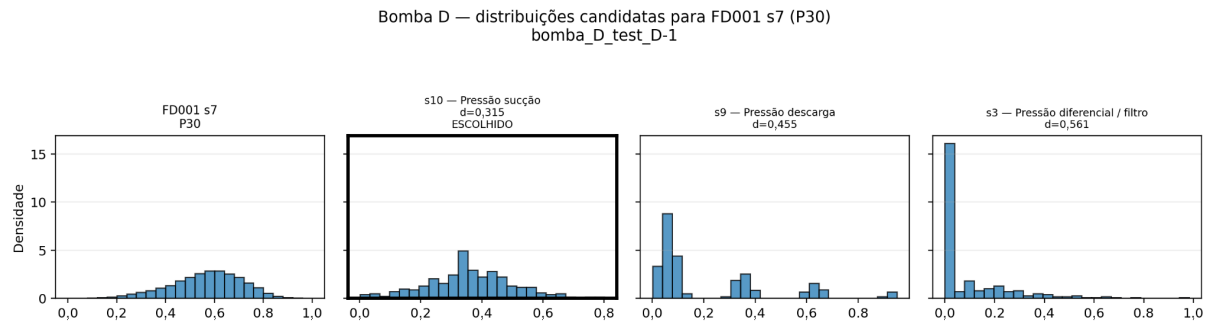
As Figuras 25, 26, 27 e 28 apresentam exemplos representativos dessa comparação. Em cada figura, observa-se um sensor ativo do FD001 e os sensores candidatos da Bomba D pertencentes às famílias físicas compatíveis. A leitura conjunta permite comparar a forma da distribuição, sua dispersão, sua posição central e a aderência relativa entre candidatos. Dessa

Figura 25 – Comparação visual das distribuições para o sensor s2 do FD001 e sensores candidatos da Bomba D. A figura exemplifica o processo de escolha do de-para a partir da combinação entre plausibilidade física e aderência estatística.



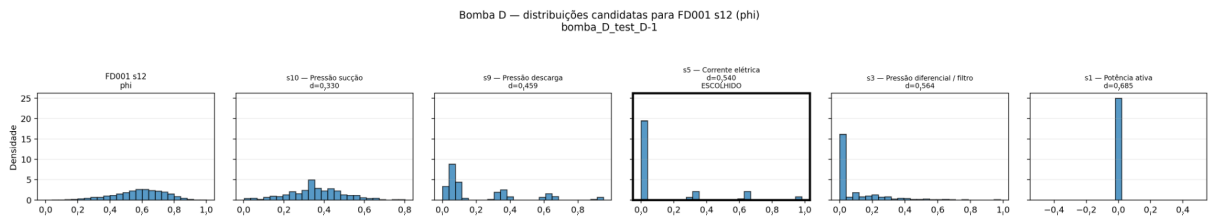
Fonte: Elaborada pelo autor (2026).

Figura 26 – Comparação visual das distribuições para o sensor s7 do FD001 e sensores candidatos da Bomba D. A figura mostra como sensores industriais de famílias compatíveis foram comparados antes da definição do mapeamento final.



Fonte: Elaborada pelo autor (2026).

Figura 27 – Comparação visual das distribuições para o sensor s12 do FD001 e sensores candidatos da Bomba D. O exemplo ilustra a necessidade de utilizar variáveis industriais como aproximações físico-estatísticas quando não há equivalência direta entre os domínios.

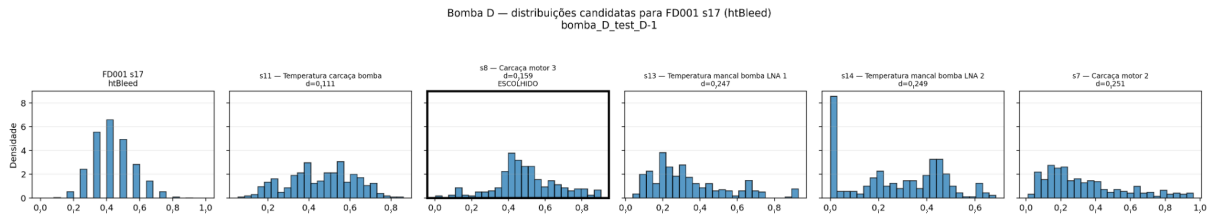


Fonte: Elaborada pelo autor (2026).

forma, os quadrinhos ilustram o princípio da projeção físico-estatística: a escolha do sensor industrial não foi feita apenas por semelhança conceitual, mas pelo melhor compromisso entre coerência física, aderência estatística e controle de diversidade.

A leitura das Figuras 25–28 complementa a análise quantitativa da distância composta. Enquanto as equações sintetizam numericamente a semelhança entre candidatos, os quadrinhos permitem verificar visualmente a coerência da escolha final. Essa visualização também ajuda a justificar por que a solução final baseada em projeção físico-estatística com diversidade condicional foi preferida em relação a uma correspondência puramente manual. A correspondência manual foi útil como referência inicial, mas permaneceu insuficiente como critério final, pois não avaliava a distribuição real dos sinais nem controlava a reutilização excessiva de sensores.

Figura 28 – Comparação visual das distribuições para o sensor s17 do FD001 e sensores candidatos da Bomba D. A figura complementa a análise de aderência ao mostrar a seleção de candidatos associados à condição mecânica e ao estado operacional do equipamento.



Fonte: Elaborada pelo autor (2026).

5.3.3 Distância composta e diversidade condicional

A projeção físico-estatística com diversidade condicional foi implementada em três etapas. Primeiro, para cada sensor ativo do FD001, foram definidos sensores candidatos da bomba com base nas famílias de engenharia descritas na Tabela 17. Essa primeira etapa evita associações estatisticamente convenientes, mas fisicamente incoerentes.

Na segunda etapa, para cada par formado por um sensor FD001 e um sensor industrial candidato, foi calculada uma distância estatística composta proposta neste trabalho. A distância, definida na Equação 27, combina quatro componentes:

$$d_{\text{composta}} = 0,45 \cdot d_{\text{Wasserstein}} + 0,35 \cdot d_{\text{KS}} + 0,10 \cdot |\Delta\mu| + 0,10 \cdot |\Delta\sigma|. \quad (27)$$

Na Equação 27, d_{composta} é a medida usada para ordenar os candidatos; $d_{\text{Wasserstein}}$ mede a distância entre as distribuições dos dois sensores. Essa métrica é útil porque considera o deslocamento global entre distribuições, podendo ser interpretada como o esforço necessário para transformar uma distribuição na outra. A estatística d_{KS} , obtida pelo teste de Kolmogorov–Smirnov, compara as distribuições acumuladas e indica diferenças na forma geral das distribuições. Já $|\Delta\mu|$ mede a diferença absoluta entre as médias, enquanto $|\Delta\sigma|$ mede a diferença absoluta entre os desvios-padrão. Assim, a distância composta considera simultaneamente posição central, dispersão e formato das distribuições.

Os pesos da distância composta foram definidos para dar maior importância à semelhança entre distribuições. A distância de Wasserstein recebeu peso 0,45 por representar a diferença global entre as distribuições. A estatística KS recebeu peso 0,35 por capturar diferenças de forma acumulada. As diferenças de média e desvio-padrão receberam pesos menores, 0,10 cada, pois resumem características importantes, mas parciais, dos sinais. Com isso, evita-se escolher sensores apenas por terem médias próximas, ignorando diferenças relevantes no comportamento estatístico.

Na terceira etapa, foi aplicada uma penalização de diversidade proposta neste trabalho. Essa penalização foi necessária porque, sem restrição adicional, vários sensores do FD001 poderiam ser mapeados para o mesmo sensor industrial. Esse fenômeno reduziria a diversidade

da entrada projetada e faria com que diferentes dimensões do espaço FD001 carregassem praticamente a mesma informação. Para evitar esse colapso, a distância de cada candidato foi ajustada conforme a Equação 28:

$$d_{\text{ajustada}} = d_{\text{composta}} \cdot (1 + \lambda_D \cdot u_j). \quad (28)$$

Na Equação 28, d_{ajustada} é a distância usada na seleção final; d_{composta} é a medida da Equação 27; u_j representa o número de usos prévios do sensor industrial j , e λ_D é o parâmetro de penalização por diversidade. Neste trabalho, adotou-se $\lambda_D = 0,25$. Portanto, se um sensor industrial ainda não foi utilizado, sua distância não sofre penalização. Se já foi utilizado uma vez, sua distância aumenta 25%. Se já foi utilizado duas vezes, aumenta 50%, e assim sucessivamente. Essa regra não proíbe a reutilização de sensores, pois um sensor ainda pode ser escolhido novamente se sua aderência estatística for claramente superior. Entretanto, ela torna a reutilização progressivamente menos favorecida, preservando maior diversidade na representação final.

A seleção foi realizada de forma sequencial, seguindo a ordem dos sensores ativos do FD001. Para cada sensor FD001, escolheu-se o candidato industrial com menor distância ajustada. O resultado final é uma matriz de projeção W , na qual as linhas representam sensores industriais da bomba e as colunas representam sensores do FD001. Quando um sensor industrial é escolhido para representar determinado sensor FD001, a posição correspondente da matriz recebe peso 1. Ao final, as colunas são normalizadas para garantir que cada dimensão projetada possua contribuição válida.

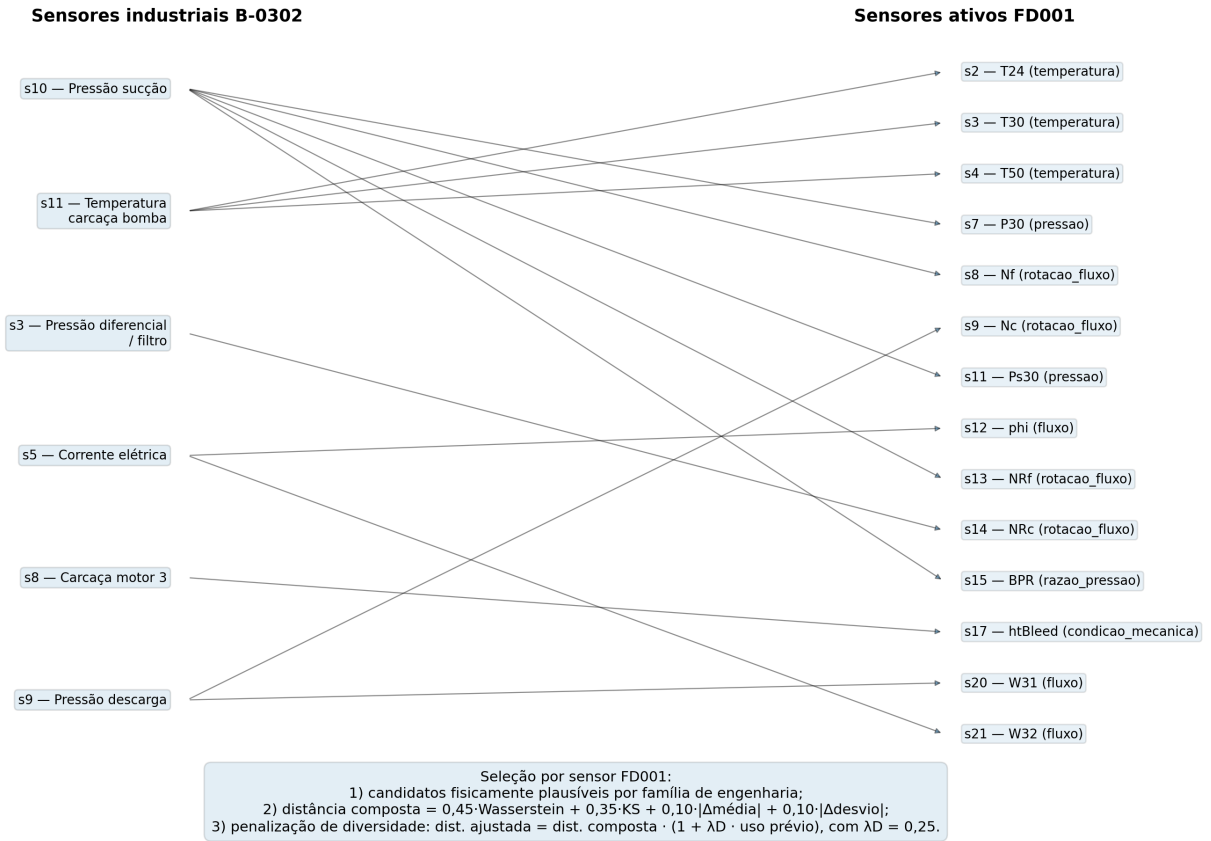
Um aspecto metodológico importante é que a matriz de projeção não foi definida uma única vez para toda a Bomba D utilizando todas as séries disponíveis. Em cada rodada de validação, a projeção foi recalculada exclusivamente a partir das séries utilizadas no treinamento daquele *fold*. A série separada para teste permaneceu fora tanto da adaptação supervisionada do modelo quanto do cálculo das distribuições utilizadas para definir o de-para entre sensores. Essa decisão evita vazamento de informação do conjunto de teste para a etapa de pré-processamento e torna a avaliação mais coerente com uma situação operacional real, na qual a falha futura ainda não é conhecida.

Também foram avaliadas alternativas com uma configuração global de projeção por bomba, isto é, uma matriz única de correspondência sensorial calculada para todo o histórico disponível da bomba. Entretanto, essa estratégia produziu desempenho inferior. A interpretação adotada é que uma projeção fixa para toda a bomba tende a suavizar diferenças entre séries de falha e pode capturar relações estatísticas que não se repetem de forma estável entre os *folds*. Por isso, a configuração final recalcula a projeção em cada rodada de treino, preservando a coerência do protocolo *leave-one-failure-series-out* e permitindo que o mapeamento sensorial reflita apenas a informação disponível no conjunto de treinamento.

A aplicação da matriz de projeção é apresentada na Equação 29:

Figura 29 – Mapeamento da projeção físico-estatística com diversidade condicional para a Bomba D no fold D-1. A matriz de projeção é recalculada em cada rodada de validação utilizando apenas as séries de treinamento, evitando vazamento de informação da série reservada para teste.

Projeção físico-estatística com diversidade condicional — Bomba D (fold D-1)



Fonte: Elaborada pelo autor (2026), com apoio de inteligência artificial generativa e prompt do autor.

$$X_{\text{proj}} = X_{B0302}W. \quad (29)$$

Na Equação 29, X_{B0302} é a matriz de entrada com sensores industriais da bomba no *fold* considerado, W é a matriz de projeção estimada a partir das séries de treinamento daquele *fold* e X_{proj} é a representação projetada no espaço compatível com o FD001. Essa representação projetada é então usada como entrada do *backbone* AGATT.

A Figura 29 apresenta o mapeamento selecionado para a Bomba D em um dos folds de validação.

A Figura 29 permite verificar como a projeção traduziu o espaço sensorial da bomba para o espaço esperado pelo modelo treinado no FD001. Sensores de temperatura do FD001 foram associados principalmente a temperaturas de carcaça da bomba. Sensores relacionados a pressão foram aproximados por pressão de sucção, pressão de descarga e pressão diferencial. Variáveis associadas a fluxo e rotação foram aproximadas por grandezas industriais correlatas, como corrente elétrica e pressões do sistema. Essa combinação preserva a coerência física da

Tabela 18 – Projeção físico-estatística com diversidade condicional selecionada para a Bomba D no fold D-1.

Sensor FD001	Grupo FD001	Sensor B-0302 escolhido	Distância composta	Uso prévio	Distância ajustada
s2 – T24	Temperatura	s11 – Temperatura carcaça bomba	0,059	0	0,059
s3 – T30	Temperatura	s11 – Temperatura carcaça bomba	0,091	1	0,114
s4 – T50	Temperatura	s11 – Temperatura carcaça bomba	0,054	2	0,082
s7 – P30	Pressão	s10 – Pressão sucção	0,315	0	0,315
s8 – Nf	Rotação/fluxo	s10 – Pressão sucção	0,139	1	0,174
s9 – Nc	Rotação/fluxo	s9 – Pressão descarga	0,262	0	0,262
s11 – Ps30	Pressão	s10 – Pressão sucção	0,080	2	0,120
s12 – phi	Fluxo	s5 – Corrente elétrica	0,540	0	0,540
s13 – NRf	Rotação/fluxo	s10 – Pressão sucção	0,106	3	0,186
s14 – NRc	Rotação/fluxo	s3 – Pressão diferencial / filtro	0,320	0	0,320
s15 – BPR	Razão de pressão	s10 – Pressão sucção	0,142	4	0,285
s17 – htBleed	Condição mecânica	s8 – Carcaça motor 3	0,159	0	0,159
s20 – W31	Fluxo	s9 – Pressão descarga	0,417	1	0,521
s21 – W32	Fluxo	s5 – Corrente elétrica	0,518	1	0,647

Fonte: Elaborada pelo autor (2026).

projeção e, ao mesmo tempo, utiliza a semelhança estatística para selecionar os sensores mais aderentes.

A Tabela 18 detalha a projeção selecionada para o fold D-1. A coluna de distância composta mostra a aderência estatística original entre o sensor FD001 e o sensor industrial candidato. A coluna de uso prévio indica quantas vezes aquele sensor industrial já havia sido selecionado anteriormente. A distância ajustada incorpora a penalização de diversidade e foi a métrica efetivamente utilizada para a escolha final.

A Tabela 18 apresenta um exemplo da projeção estimada para o fold D-1. Como a projeção é recalculada em cada rodada, pequenas variações no de-para podem ocorrer entre folds, dependendo das distribuições observadas nas séries de treinamento disponíveis. Essa característica é desejável, pois evita impor uma correspondência global rígida para toda a bomba e preserva a lógica de validação por série de falha.

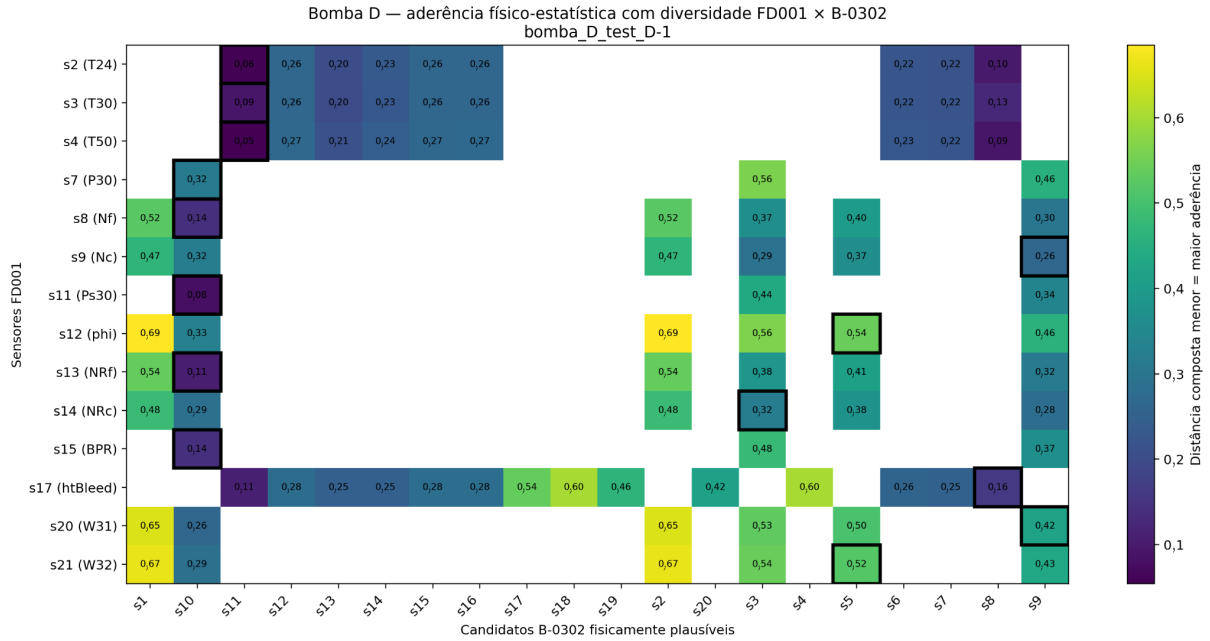
A Tabela 18 também mostra que a diversidade condicional não impede a reutilização de sensores. Por exemplo, um mesmo sensor industrial pode ser escolhido mais de uma vez quando sua aderência estatística e coerência física são superiores às alternativas. No entanto, a cada reutilização, sua distância ajustada aumenta, tornando progressivamente mais provável a escolha de outro sensor adequado. Essa estratégia evita que a projeção colapse em poucas variáveis industriais e preserva uma representação mais rica para o AGATT.

Além do mapeamento final, foram gerados mapas de aderência para verificar visualmente a qualidade das associações. Essas figuras comparam sensores candidatos industriais com sensores FD001, permitindo avaliar se a escolha é plausível não apenas numericamente, mas também visualmente.

A Figura 30 apresenta o mapa de aderência utilizado para inspecionar as alternativas de projeção.

A Figura 30 complementa a Tabela 18. Enquanto a tabela mostra a escolha final de cada sensor, o mapa de aderência permite observar o conjunto de alternativas avaliadas. Essa visualização é importante porque torna transparente a decisão de mapeamento: a escolha não de-

Figura 30 – Exemplo de mapa de aderência entre sensores industriais candidatos e sensores ativos do FD001 para a Bomba D. A figura auxilia a avaliação visual da coerência estatística das associações utilizadas na projeção. Os elementos em branco representam combinações para as quais o mapeamento não foi avaliado, por não apresentarem correspondência física ou funcional plausível entre as variáveis comparadas.



Fonte: Elaborada pelo autor (2026).

pende apenas de uma associação manual, mas de uma comparação entre candidatos compatíveis. Assim, o leitor pode verificar se os sensores selecionados fazem sentido dentro da família física correspondente e se apresentam aderência estatística compatível.

Portanto, a projeção físico-estatística com diversidade condicional foi uma etapa central do estudo. Ela permitiu reaproveitar o AGATT treinado no C-MAPSS sem exigir correspondência física perfeita entre sensores aeronáuticos e sensores industriais. A abordagem final foi escolhida porque as alternativas testadas, mais manuais ou mais puramente estatísticas, não equilibravam tão bem plausibilidade física, aderência estatística e diversidade de informação.

5.4 SENSIBILIDADE À JANELA DE GERAÇÃO INDUSTRIAL NA BOMBA D

Após a definição do pipeline de transferência, foi analisado o efeito do parâmetro *window*, responsável pela geração das amostras industriais. O *backbone* AGATT, entretanto, possui entrada fixa de 35 pontos. Quando $window > 35$, a amostra é primeiro construída com o tamanho solicitado e, em seguida, são preservados apenas os 35 pontos mais recentes antes da passagem pelo AGATT.

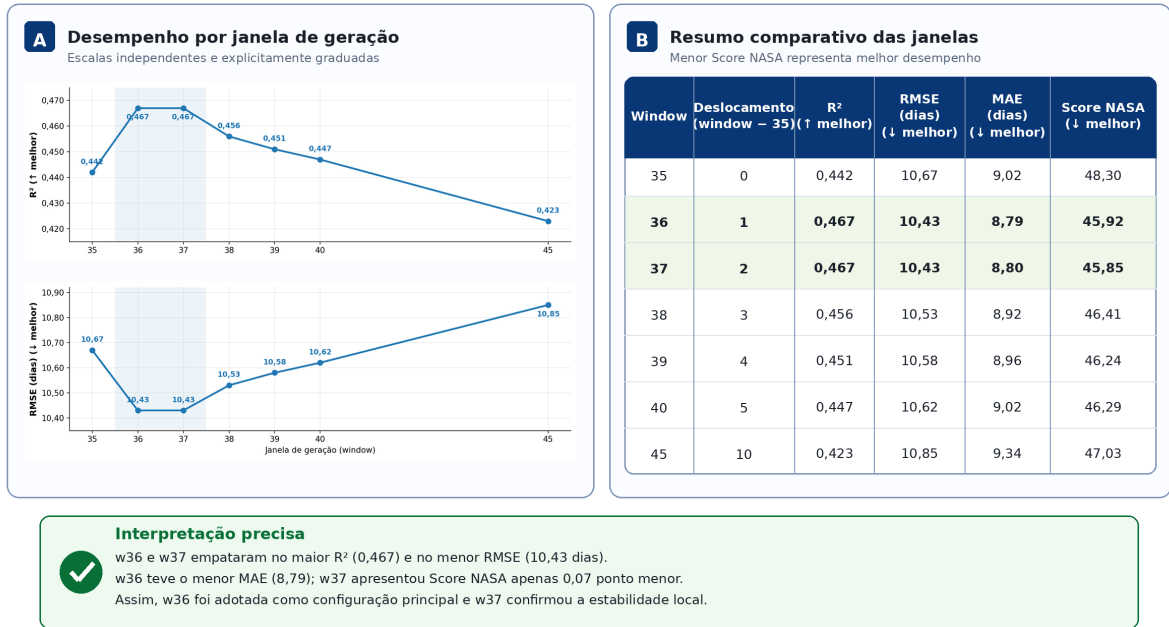
A diferença entre a janela de geração e a entrada efetiva é definida na Equação 30:

$$\Delta w = window - 35. \quad (30)$$

Figura 31 – Sensibilidade à janela de geração industrial para a Bomba D. A entrada efetiva do *backbone* permanece com 35 pontos; o parâmetro de geração foi variado entre 35 e 45. A melhor região experimental ocorreu nas janelas 36 e 37.

Análise de deslocamento da janela (Bomba D, backbone = 35)

Configuração fixa: `smooth_w = 3`, `loss = NASA`, `RUL ocultado = 5-60`,
 projeção = físico-estatística com diversidade condicional, $\lambda_D = 0,25$, $lr = 2e-5$.



Backbone efetivo = 35; deslocamento = window - 35.

Fonte: Elaborada pelo autor (2026), com apoio de inteligência artificial generativa e prompt do autor.

Na Equação 30, Δw é o deslocamento, *window* é o comprimento solicitado na geração industrial e 35 é a quantidade fixa de pontos recebida pelo *backbone*. Esse valor expressa a diferença entre o comprimento usado na geração industrial e a dimensão temporal efetiva do *backbone*. Ele não significa que o AGATT recebe contexto adicional. Para um mesmo instante-alvo e uma série suficientemente longa, a entrada efetiva permanece formada pelos mesmos 35 pontos finais. O parâmetro altera principalmente a quantidade e a posição inicial das janelas que podem ser produzidas ao longo das séries e, em trechos curtos, o tratamento de borda à esquerda usado na construção da amostra. Esse procedimento não representa lacuna na série sensorial.

Foram avaliadas as janelas de geração 35, 36, 37, 38, 39, 40 e 45, mantendo-se fixos os demais parâmetros: `smooth_w = 3`, `loss = NASA`, `RUL ocultado` entre 5 e 60 dias, projeção físico-estatística com diversidade condicional, $\lambda_D = 0,25$ e taxa de aprendizado 2×10^{-5} .

A Figura 31 compara os resultados obtidos para as janelas de geração avaliadas.

A Figura 31 mostra que a janela 35 constitui o caso em que a amostra industrial já possui exatamente o comprimento exigido pelo *backbone*. Esse caso obteve $R^2 = 0,442$ e RMSE de 10,67 dias. As janelas 36 e 37 formaram a melhor região experimental, ambas com $R^2 \approx 0,467$ e RMSE de aproximadamente 10,43 dias. A janela 36 obteve MAE de 8,79 dias, viés médio de +0,04 dia e Score NASA de 45,92. A janela 37 apresentou resultado praticamente equivalente,

com Score NASA próximo de 45,85, o que indica estabilidade local da região selecionada.

A partir da janela 38, verificou-se queda gradual das métricas, e a janela 45 apresentou piora mais acentuada. Como os 35 pontos finais processados pelo AGATT permanecem fixos, esses resultados não permitem concluir que a rede se beneficiou diretamente de um ponto temporal adicional. A explicação mais prudente é que o parâmetro de geração alterou a composição e a quantidade das amostras de treinamento, bem como situações de borda em séries curtas. O efeito observado deve, portanto, ser entendido como propriedade empírica do pipeline completo.

A configuração $\text{window} = 36$ foi adotada por apresentar a melhor combinação observada de R^2 , RMSE, MAE e viés. A janela 37 foi interpretada como evidência de uma região local estável, e não como prova de que 36 seja um comprimento universalmente ótimo.

5.5 SENSIBILIDADE AO INTERVALO DE RUL OCULTADO

Além da janela temporal, avaliou-se a sensibilidade do modelo ao intervalo de RUL utilizado na seleção dos pontos ocultados. Essa análise buscou verificar se a inclusão ou exclusão de pontos muito próximos da falha alterava a capacidade de transferência do modelo.

A configuração estrutural foi mantida fixa na melhor região identificada anteriormente: $\text{window} = 36$, *backbone* efetivo de 35 pontos, $\text{smooth_w} = 3$, $\text{loss} = \text{NASA}$, projeção físico-estatística com diversidade condicional, $\lambda_D = 0,25$ e taxa de aprendizado 2×10^{-5} . Foram avaliados os intervalos de RUL ocultado 3–60, 5–60, 7–60 e 10–60 dias.

A Figura 32 apresenta a sensibilidade ao intervalo de RUL ocultado.

A Figura 32 mostra que o intervalo 5–60 dias apresentou o melhor desempenho global, com $R^2 = 0,467$, RMSE de 10,43 dias, MAE de 8,79 dias, viés médio de apenas +0,04 dia e Score NASA de 45,92. O intervalo 3–60 não superou esse desempenho. A inclusão de pontos demasiadamente próximos da falha iminente tornou a tarefa mais instável e piorou o RMSE e o Score NASA. Já os intervalos 7–60 e 10–60, que removem pontos mais próximos da falha, também reduziram a aderência global do modelo.

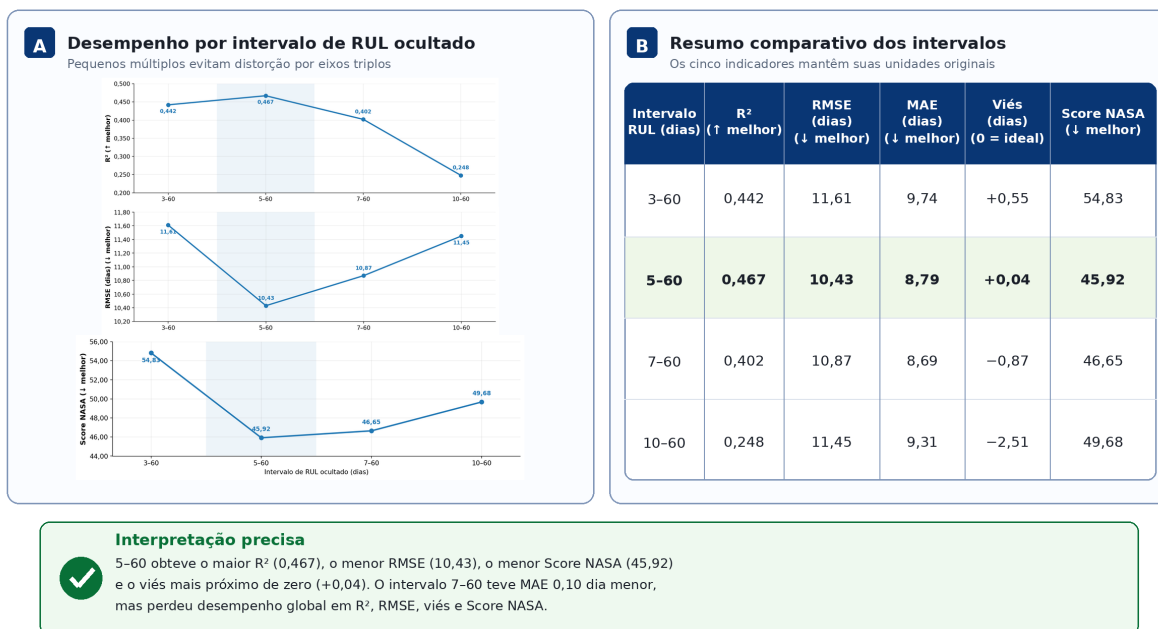
A interpretação desse resultado é importante. Quando o limite inferior é muito baixo, como em 3 dias, o modelo é avaliado em uma região extremamente próxima da falha, na qual pequenas diferenças de data ou ruído sensorial podem produzir grande variação relativa no erro. Por outro lado, quando o limite inferior aumenta para 7 ou 10 dias, os pontos mais próximos da falha são removidos, e o modelo perde parte da informação que caracteriza a transição para condição crítica. O intervalo 5–60 dias representa, portanto, um equilíbrio entre antecipação operacional e estabilidade estatística.

Assim, o intervalo 5–60 foi adotado como protocolo principal. Essa escolha também torna a avaliação mais coerente com o uso prático do modelo: prever falhas com antecedência suficiente para apoiar inspeção ou planejamento, sem restringir a análise apenas à falha iminente.

Figura 32 – Sensibilidade ao intervalo de RUL ocultado na Bomba D. A configuração com RUL entre 5 e 60 dias apresentou o melhor equilíbrio entre antecipação e desempenho preditivo.

Sensibilidade ao intervalo de RUL ocultado (Bomba D)

Configuração fixa: $\text{window} = 36$, $\text{backbone} = 35$, $\text{smooth_w} = 3$, $\text{loss} = \text{NASA}$,
 projeção = físico-estatística com diversidade condicional, $\lambda_D = 0,25$, $\text{lr} = 2e-5$.



Conclusão: 5-60 foi adotado como protocolo principal.

Fonte: Elaborada pelo autor (2026), com apoio de inteligência artificial generativa e prompt do autor.

5.6 RESULTADO PRINCIPAL DA TRANSFERÊNCIA DE APRENDIZADO

A configuração principal da transferência de aprendizado utilizou a Bomba D, *backbone* efetivo de 35 pontos, janela de geração 36, suavização temporal $\text{smooth_w} = 3$, projeção físico-estatística com diversidade condicional, $\lambda_D = 0,25$, RUL ocultado entre 5 e 60 dias, 10 pontos ocultados por série, taxa de aprendizado de 2×10^{-5} e *loss* NASA. Como os 20 canais utilizados apresentaram cobertura completa, não houve etapa de imputação.

Antes de apresentar os resultados, convém retomar o significado das métricas utilizadas na avaliação global de regressão. O coeficiente de determinação, R^2 , compara o erro quadrático do modelo com o erro de uma linha de base que prediz sempre a média empírica dos RULs reais avaliados. Assim, o valor $R^2 = 0,467$ obtido pela configuração principal indica redução aproximada de 46,7% da soma dos erros quadráticos em relação a esse preditor médio. Portanto, além de medir aderência global, o R^2 indica o ganho obtido em relação a uma previsão média trivial.

O RMSE representa a raiz do erro quadrático médio, em dias. Por elevar os erros ao quadrado antes da média, essa métrica penaliza mais fortemente erros grandes. O MAE corresponde ao erro absoluto médio, também em dias, sendo mais diretamente interpretável como o erro médio esperado da previsão. O MedAE é a mediana dos erros absolutos e, por isso,

é menos sensível a valores extremos do que o MAE.

O viés médio indica a tendência sistemática do modelo. Neste trabalho, viés positivo significa que o modelo tende a superestimar o RUL, isto é, prever vida remanescente maior do que a real. Viés negativo significa tendência de subestimar o RUL, ou seja, prever vida remanescente menor do que a real. Do ponto de vista operacional, a superestimação próxima da falha é especialmente crítica, pois pode atrasar uma intervenção de manutenção.

O Score NASA é uma métrica assimétrica tradicionalmente utilizada em problemas de RUL no C-MAPSS (Saxena *et al.*, 2008a; Saxena e Goebel, 2008). Sua principal característica é penalizar de forma diferente atrasos e antecipações na previsão da falha. Assim, além das métricas usuais de erro, o Score NASA permite avaliar a previsão sob uma perspectiva mais próxima da penalização operacional adotada em benchmarks de prognóstico.

Para cada amostra, a penalização do Score NASA é calculada pela Equação 31, conforme o critério do desafio C-MAPSS (Saxena *et al.*, 2008a):

$$s_i = \begin{cases} \exp(-d_i/13) - 1, & d_i < 0, \\ \exp(d_i/10) - 1, & d_i \geq 0. \end{cases} \quad (31)$$

Na Equação 31, y_i é o RUL real da amostra i , $\hat{y}_i$ é o RUL previsto, $d_i = \hat{y}_i - y_i$ é o erro assinado e s_i é sua penalização. A função $\exp(\cdot)$ é a exponencial natural; os denominadores 13 e 10 controlam a assimetria. A perda total corresponde à soma das penalizações no lote. Como o denominador da superestimação é menor, previsões acima do RUL real crescem mais rapidamente em penalidade, refletindo o risco de estimar uma vida remanescente maior do que a efetivamente disponível.

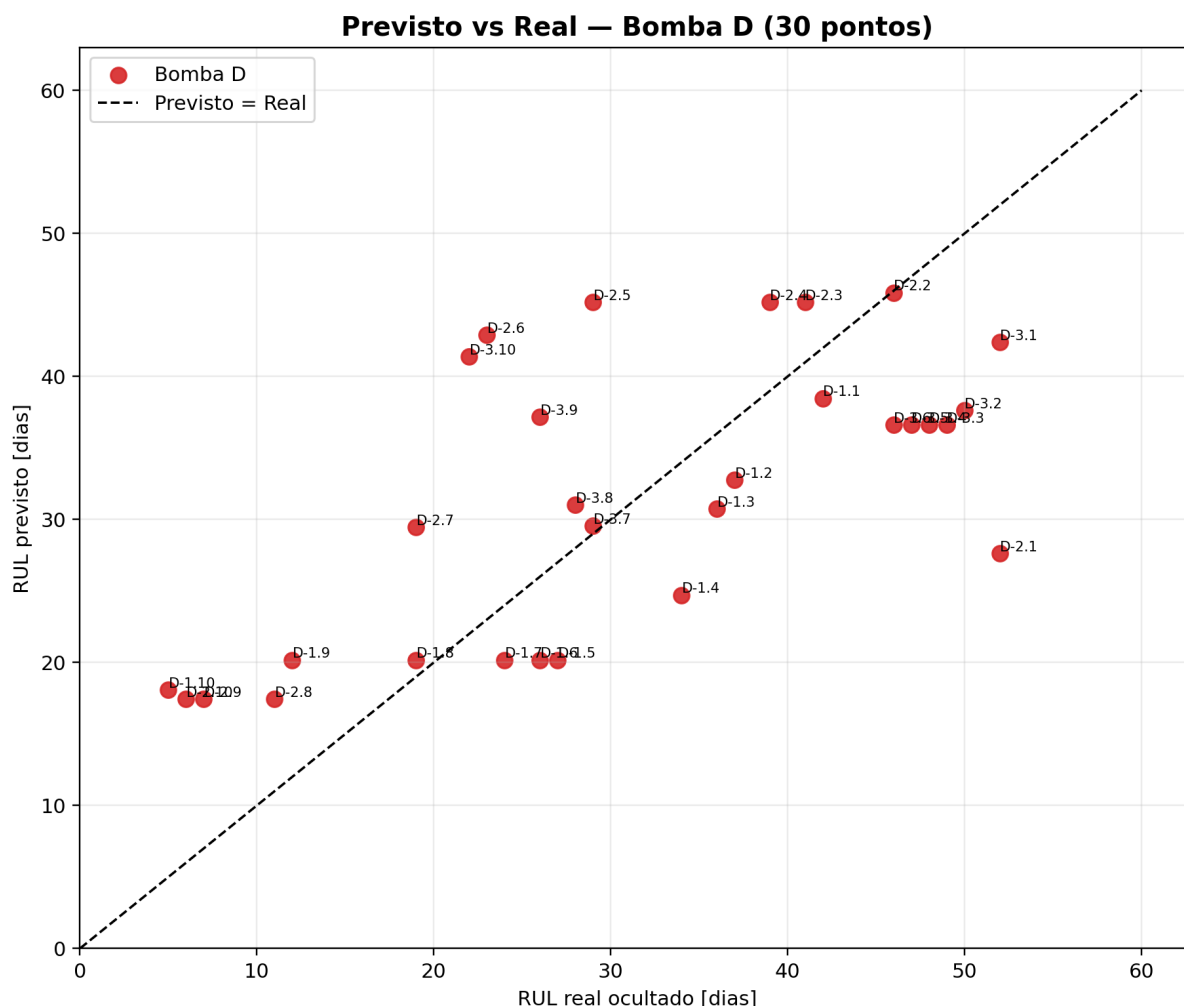
A Figura 33 apresenta o resultado principal. O gráfico de dispersão compara o RUL real com o RUL previsto nos pontos ocultados da Bomba D. A linha diagonal representa a previsão ideal, isto é, o caso em que o valor previsto seria igual ao valor real.

A Figura 33 deve ser interpretada como a evidência principal de viabilidade do *transfer learning* neste capítulo. Cada ponto representa uma previsão feita em um ponto ocultado da série de falha, isto é, em uma observação que não foi usada no treinamento daquela rodada de validação. Quanto mais próximo um ponto está da diagonal, menor é o erro de previsão. Pontos acima da diagonal indicam superestimação do RUL, enquanto pontos abaixo indicam subestimação.

O resultado obtido foi $R^2 = 0,467$, RMSE de 10,43 dias, MAE de 8,79 dias, MedAE de 9,26 dias, viés médio de +0,04 dia e Score NASA de 45,92. O valor de R^2 indica que o modelo produziu uma soma de erros quadráticos aproximadamente 46,7% menor do que a obtida pelo preditor constante igual à média dos RULs reais avaliados. Assim, os resultados sugerem que o modelo transferido foi capaz de capturar parte relevante da variação do RUL no domínio industrial, mesmo tendo sido originalmente treinado em uma base pública de motores aeronáuticos.

O viés médio praticamente nulo é um ponto importante dessa configuração. Ele indica

Figura 33 – Resultado principal de transferência de aprendizado para a Bomba D com *loss* NASA. A configuração utiliza *backbone* efetivo de 35 pontos, janela de geração 36, suavização temporal de ordem 3, RUL ocultado entre 5 e 60 dias e projeção físico-estatística com diversidade condicional.



Fonte: Elaborada pelo autor (2026).

que, no conjunto avaliado, o modelo não apresentou tendência sistemática forte de superestimar ou subestimar a vida remanescente. Assim, a *loss* NASA foi adotada como configuração principal para demonstrar a viabilidade da transferência de aprendizado em termos de regressão global.

Esse resultado deve ser interpretado com cautela metodológica. Não se trata de afirmar que o modelo resolve genericamente o problema de RUL para qualquer bomba industrial. O que se demonstra é que, para a Bomba D, na qual há maior riqueza sensorial, falhas fisicamente mais compatíveis com degradação progressiva e TTFs decrescentes, o conhecimento extraído do FD001 pôde ser adaptado de forma útil ao domínio industrial. Essa é a principal contribuição experimental desta etapa.

As métricas apresentadas nesta seção devem ser interpretadas considerando a estrutura do protocolo experimental. Os pontos de teste resultam do ocultamento retrospectivo de posições temporais ao longo das séries pré-falha disponíveis para a Bomba D. Portanto, embora a avaliação

contenha múltiplos pontos de previsão, esses pontos permanecem vinculados ao número limitado de falhas reais registradas no histórico. Consequentemente, os resultados indicam a capacidade do modelo de reconstruir e antecipar o padrão temporal observado nessas trajetórias, mas não devem ser interpretados como evidência estatística definitiva de generalização para falhas futuras de natureza distinta.

5.7 LINHAS DE BASE E ABLAÇÃO DA TRANSFERÊNCIA

A avaliação da transferência foi complementada por linhas de base com graus crescentes de complexidade. Essa etapa responde a duas perguntas distintas. A primeira é se o modelo transferido supera regras simples que não utilizam sensores. A segunda é se o ganho observado decorre efetivamente do pré-treinamento no FD001, e não apenas da arquitetura AGATT.

O coeficiente de determinação já incorpora uma comparação com um preditor constante. Na formulação utilizada por bibliotecas consolidadas de aprendizado de máquina (Pedregosa *et al.*, 2011), ele é definido pela Equação 32 ((Kvålseth, 1985)):

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2}, \quad (32)$$

Na Equação 32, y_i é o RUL real, $\hat{y}_i$ é o RUL previsto, $\bar{y}$ é a média empírica dos RULs reais avaliados. Assim, $R^2 > 0$ indica redução do erro quadrático em relação a essa previsão média, enquanto $R^2 < 0$ indica desempenho inferior ao preditor constante.

Como referência conceitual adicional, a Equação 33 define uma regra ingênua baseada no ponto médio do horizonte ocultado:

$$\widehat{RUL}_{\text{constante}} = \frac{5 + 60}{2} = 32,5 \text{ dias}. \quad (33)$$

Na Equação 33, $\widehat{RUL}_{\text{constante}}$ é a previsão fixa, e 5 e 60 dias são os limites inferior e superior do intervalo de avaliação. Essa regra ignora os sensores e a trajetória temporal. Ela serve apenas para explicitar a escala de uma previsão constante e não utiliza qualquer informação aprendida previamente.

A análise de confiabilidade apresentada no Capítulo 3 constitui uma terceira referência. Ela é útil para caracterizar horizontes globais de falha, mas não incorpora diretamente o estado recente dos sensores.

5.7.1 LSTM treinada exclusivamente com dados industriais

Como linha de base neural convencional, foi treinada uma LSTM pequena utilizando diretamente os canais industriais da Bomba D. Nenhum peso da C-MAPSS foi carregado, nenhuma distribuição do FD001 foi utilizada e não houve projeção para os 14 sensores do domínio de origem. Essa linha de base responde a uma pergunta distinta da ablação do AGATT: uma rede recorrente simples, treinada apenas com o histórico industrial, consegue aprender a evolução do RUL?

Tabela 19 – Arquitetura da linha de base LSTM treinada somente com os dados industriais da Bomba D.

Etapa	Dimensão de saída	Parâmetros	Função
Entrada	36×20	0	Janela formada pelos canais industriais.
LSTM(32)	32	6.784	Resume a sequência em um vetor temporal; <i>dropout</i> =0,20 e <i>recurrent_dropout</i> =0.
Dense(16), ReLU	16	528	Combina as características temporais extraídas.
Dropout(0,10)	16	0	Regulariza a cabeça de regressão.
Dense(1), linear	1	17	Produce a previsão contínua de RUL.
Total treinável	—	7.329	Todos os parâmetros são ajustados com dados industriais.

Fonte: Elaborada pelo autor (2026).

A entrada da LSTM possui forma 36×20 : 36 observações temporais e 20 canais processados. A rede contém uma única camada recorrente LSTM com 32 unidades. Em uma LSTM, cada unidade possui quatro mecanismos internos — porta de entrada, porta de esquecimento, porta de saída e estado candidato — que controlam a retenção e a atualização da memória. Esses mecanismos fazem parte da mesma camada recorrente e não correspondem a quatro camadas empilhadas.

Após a camada LSTM, o vetor de 32 características passa por uma camada densa de 16 unidades com ativação ReLU, por uma camada de *dropout* de 0,10 e por uma camada densa linear com uma saída. Na própria LSTM foi utilizado *dropout* de 0,20 sobre as entradas; o parâmetro *recurrent_dropout* foi mantido em zero. Portanto, a arquitetura possui uma camada recorrente treinável e duas camadas densas treináveis. As camadas de *dropout* regularizam o treinamento, mas não possuem pesos.

A Tabela 19 apresenta a arquitetura da LSTM industrial usada como linha de base.

O total de parâmetros da camada LSTM é dado por $4u(m + u + 1)$, em que $m = 20$ é o número de canais e $u = 32$ é o número de unidades. Assim, a camada recorrente contém $4 \times 32 \times (20 + 32 + 1) = 6.784$ parâmetros, e a rede completa possui 7.329 parâmetros treináveis. Embora seja muito menor que o AGATT, esse número ainda precisa ser estimado a partir de apenas duas trajetórias de treinamento em cada *fold* externo.

Antes do estudo aprofundado, a mesma LSTM foi aplicada às quatro bombas em uma triagem de configuração fixa, com semente 42, MSE e taxa de aprendizado 10^{-3} . O objetivo dessa rodada não foi otimizar cada bomba, mas verificar se uma arquitetura recorrente simples encontrava sinal preditivo básico no conjunto completo.

A Tabela 20 sintetiza a triagem exploratória da LSTM nas quatro bombas.

Todos os valores de R^2 permaneceram negativos, indicando que a configuração fixa não superou o preditor médio em nenhuma bomba. A Bomba B apresentou o menor RMSE dessa

Tabela 20 – Triagem exploratória da LSTM nas quatro bombas, utilizando uma configuração fixa e sem otimização individual por bomba.

Bomba	n	R^2	RMSE	MAE	MedAE	Viés
A	30	-0,630	19,64	16,56	14,15	+7,26
B	60	-0,142	15,67	12,58	10,81	-5,66
C	40	-0,122	17,16	14,88	13,35	-2,87
D	30	-0,347	16,58	13,66	11,91	+7,43

Fonte: Elaborada pelo autor (2026).

triagem, mas as bombas possuem números diferentes de séries, pontos externos e distribuições de RUL; por isso, os valores não devem ser usados como um ranking direto entre equipamentos. O aprofundamento permaneceu na Bomba D porque a comparação principal precisava reproduzir o domínio já selecionado pelo estudo de transferência, no qual havia maior riqueza de canais e melhor coerência física das falhas.

A LSTM preservou o protocolo externo da Bomba D: janela 36, suavização causal de três observações, RUL entre 5 e 60 dias, 10 pontos ocultados por série e rodízio *leave-one-failure-series-out*. Em cada *fold*, os parâmetros de normalização foram estimados apenas nas séries de treinamento. Como os canais empregados estavam completos, não houve estimação nem aplicação de medianas de preenchimento. A validação interna reservou o trecho temporal final de cada série de treinamento e removeu da parte de ajuste as janelas que compartilhariam observações com a primeira janela de validação. Esse expurgo reduz o otimismo causado pela sobreposição de janelas temporais adjacentes.

Foram avaliadas cinco sementes, três taxas de aprendizado $\{10^{-4}, 5 \times 10^{-4}, 10^{-3}\}$ e duas funções de perda, MSE e NASA. Como cada configuração foi avaliada nos três *folds* externos, o estudo correspondeu a 90 treinamentos. Para cada semente, a combinação de taxa e perda foi selecionada pelo menor RMSE médio da validação interna, sem consultar os 30 pontos externos. Nas cinco sementes, a configuração escolhida foi MSE com taxa 5×10^{-4} .

O resultado médio foi $R^2 = -0,309 \pm 0,010$, RMSE de $16,34 \pm 0,06$ dias, MAE de $13,46 \pm 0,08$ dias, MedAE de $11,79 \pm 0,23$ dias, viés de $+6,97 \pm 0,13$ dias e Score NASA de $154,70 \pm 2,04$. A pequena dispersão entre sementes indica estabilidade numérica entre as execuções, mas não que o modelo tenha aprendido adequadamente a degradação. As previsões se concentraram em patamares quase constantes dentro de cada série, comportamento compatível com um regressor que aproxima valores centrais do alvo e não acompanha sua evolução temporal. Como consequência, nenhuma das cinco sementes emitiu alarmes nos limiares de 15, 20 ou 25 dias.

5.7.2 AGATT treinado do zero no domínio industrial

Para isolar o efeito do pré-treinamento, foi realizada uma ablação com a mesma família AGATT inicializada aleatoriamente e treinada do zero nas séries industriais. A arquitetura, a

Tabela 21 – Comparação entre o AGATT com transferência e as linhas de base neurais treinadas sem pesos da C-MAPSS. Os resultados da LSTM e do AGATT do zero são médias de cinco sementes; o resultado transferido corresponde à configuração principal selecionada.

Modelo	R^2	RMSE	MAE	MedAE	Viés	Score NASA
LSTM industrial	$-0,309 \pm 0,010$	$16,34 \pm 0,06$	$13,46 \pm 0,08$	$11,79 \pm 0,23$	$+6,97 \pm 0,13$	$154,70 \pm 2,04$
AGATT do zero	$-0,636 \pm 0,068$	$18,26 \pm 0,38$	$15,77 \pm 0,30$	$13,85 \pm 0,61$	$+7,40 \pm 0,53$	$206,04 \pm 14,08$
AGATT com transferência	0,467	10,43	8,79	9,26	+0,04	45,92

Fonte: Elaborada pelo autor (2026).

janela industrial 36, a entrada efetiva de 35 pontos, a suavização 3, a função de perda NASA, o intervalo 5–60, os pontos externos e o protocolo de validação foram mantidos. A diferença central é que nenhum peso aprendido no FD001 foi carregado e todo o AGATT permaneceu treinável.

A projeção físico-estatística foi preservada nessa ablação para manter a mesma dimensionalidade de entrada e o mesmo mecanismo de harmonização usado no modelo transferido. Portanto, “sem transferência” significa ausência dos pesos pré-treinados; não significa ausência completa das estatísticas do domínio de origem utilizadas para construir a projeção. Essa escolha torna a comparação mais controlada: o que se retira é justamente o conhecimento codificado nos pesos do AGATT.

Foram testadas cinco sementes e quatro taxas de aprendizado: 2×10^{-5} , 10^{-4} , 5×10^{-4} e 10^{-3} . Cada combinação foi executada nos três *folds* externos, totalizando 60 treinamentos. Para cada semente, a taxa foi selecionada pela menor média da perda de validação interna. Em todas as sementes, a taxa escolhida foi 2×10^{-5} .

O resultado médio foi $R^2 = -0,636 \pm 0,068$, RMSE de $18,26 \pm 0,38$ dias, MAE de $15,77 \pm 0,30$ dias, MedAE de $13,85 \pm 0,61$ dias, viés de $+7,40 \pm 0,53$ dias e Score NASA de $206,04 \pm 14,08$. Assim como na LSTM, as previsões ficaram praticamente constantes dentro de cada série. No limiar de 15 dias, nenhum dos cinco pontos críticos foi identificado em qualquer semente. Esse comportamento é compatível com a dificuldade de estimar uma arquitetura profunda a partir de apenas duas trajetórias de treinamento por *fold*, sem o suporte das representações aprendidas no FD001.

A Tabela 21 compara o modelo transferido às linhas de base sem pesos da C-MAPSS.

Em relação à LSTM, o modelo transferido reduziu o RMSE em aproximadamente 36,2%, o MAE em 34,7% e o Score NASA em 70,3%. Em relação ao AGATT treinado do zero, as reduções foram de aproximadamente 42,9% no RMSE, 44,3% no MAE e 77,7% no Score NASA.

Essa comparação sustenta duas interpretações. Primeiro, uma arquitetura menor, como a LSTM, foi mais adequada do que o AGATT completo quando ambos dependeram exclusivamente do pequeno histórico industrial. Segundo, o melhor resultado foi obtido quando a complexidade do AGATT veio acompanhada das representações aprendidas no FD001. Portanto, o ganho não pode ser atribuído apenas ao número de parâmetros ou à arquitetura; ele está associado ao reaproveitamento do conhecimento de origem e à compatibilização entre os domínios.

A comparação, contudo, não é perfeitamente simétrica. As linhas de base são apresentadas como média e desvio-padrão de cinco sementes, enquanto o AGATT transferido corresponde à configuração principal selecionada durante o desenvolvimento. Assim, a tabela demonstra que as alternativas sem pré-treinamento falharam de maneira estável em reproduzir o desempenho da configuração transferida, mas não substitui uma futura análise multissemente completa do modelo transferido.

5.8 LIMITAÇÕES DA SELEÇÃO EXPERIMENTAL

As análises de horário, janela de geração, suavização, horizonte de RUL, projeção e função de perda foram realizadas retrospectivamente sobre o histórico disponível. A configuração final representa, portanto, a melhor combinação observada entre as alternativas efetivamente testadas; ela não deve ser interpretada como ótimo universal. Como as mesmas três trajetórias da Bomba D sustentaram tanto a exploração de hiperparâmetros quanto a avaliação retrospectiva final, existe possibilidade de viés de seleção experimental e de uma estimativa otimista do desempenho.

O rodízio *leave-one-failure-series-out*, o recálculo da projeção por *fold* e o isolamento dos pontos externos reduzem vazamentos diretos, mas não criam novas trajetórias independentes. A confirmação mais forte do método exige aplicação prospectiva: manter a configuração congelada e avaliá-la em falhas futuras que não tenham participado de nenhuma decisão de modelagem. Também seria desejável repetir o modelo transferido com múltiplas sementes para quantificar sua variabilidade de treinamento de forma simétrica às linhas de base.

Assim, o resultado de RMSE 10,43 dias deve ser entendido como evidência retrospectiva de viabilidade no caso estudado. Ele sustenta a continuidade da investigação e a implantação experimental monitorada, mas não autoriza afirmar desempenho garantido para qualquer bomba, período operacional ou modo de falha.

5.9 CALIBRAÇÃO OPERACIONAL COM FUNÇÃO DE PERDA CONSERVADORA

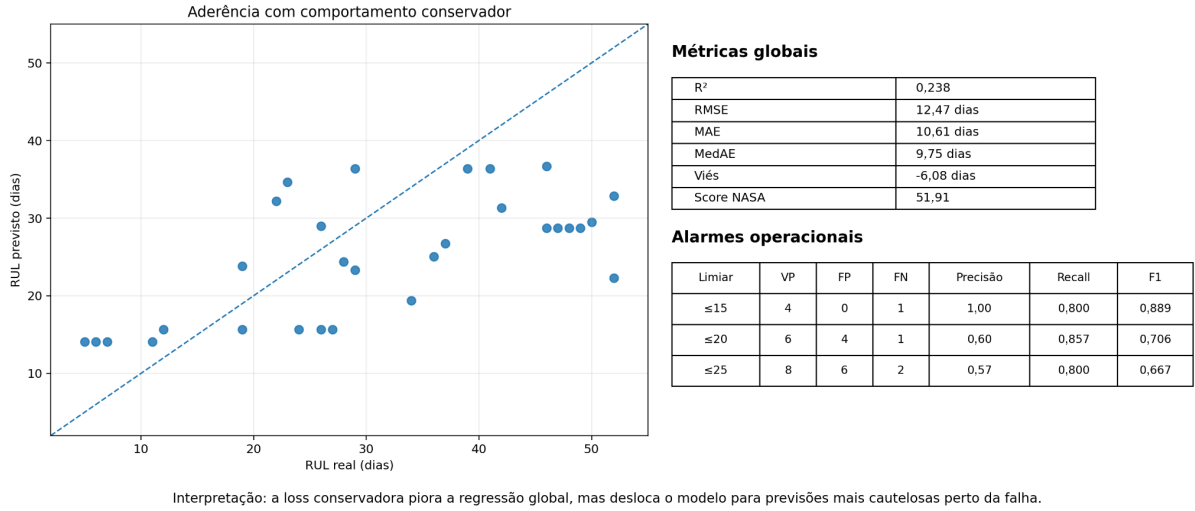
Embora a configuração com *loss* NASA tenha apresentado o melhor desempenho global de regressão, a avaliação de RUL também exige considerar o comportamento operacional do modelo em horizontes críticos. Por isso, foi avaliada a função de perda *horizon_conservative*, voltada a tornar o modelo mais conservador na região próxima à falha.

Essa função altera o objetivo de otimização do modelo. Em vez de buscar apenas o menor erro médio global, ela penaliza mais fortemente erros em horizontes próximos da falha e, principalmente, superestimções do RUL. Uma superestimção ocorre quando o modelo prevê uma vida remanescente maior do que a real. Do ponto de vista operacional, esse é o erro mais perigoso, pois pode atrasar uma intervenção de manutenção.

A configuração estrutural foi mantida igual à configuração principal: *backbone* efetivo de 35 pontos, *window* = 36, *smooth_w* = 3, projeção físico-estatística com diversidade condi-

Figura 34 – Resultado da Bomba D com a função de perda `horizon_conservative`. A configuração reduz o desempenho global de regressão, mas desloca o modelo para um comportamento mais conservador na previsão de casos próximos da falha.

Resultado com nova loss operacional — Bomba D



Fonte: Elaborada pelo autor (2026).

cional, $\lambda_D = 0,25$, RUL ocultado entre 5 e 60 dias e taxa de aprendizado de 2×10^{-5} . A única mudança relevante foi a função de perda.

A função implementada, proposta neste estudo para favorecer previsões conservadoras próximas da falha, é definida pela Equação 34:

$$\mathcal{L}_{\text{cons}} = \frac{1}{n} \sum_{i=1}^n \left[1 + 2 \exp \left(-\frac{y_i}{20} \right) \right] \omega(d_i) d_i^2, \quad (34)$$

Na Equação 34, $\mathcal{L}_{\text{cons}}$ é a perda média do lote; n é a quantidade de amostras; i é o índice da amostra; y_i e $\hat{y}_i$ são, respectivamente, o RUL real e previsto; $d_i = \hat{y}_i - y_i$ é o erro assinado; o termo exponencial aumenta o peso próximo da falha; e $\omega(d_i)$ é o peso direcional definido na Equação 35:

$$\omega(d_i) = \begin{cases} 2,0, & d_i > 0, \\ 0,7, & d_i \leq 0. \end{cases} \quad (35)$$

Na Equação 35, $\omega(d_i)$ vale 2,0 quando o erro é positivo, isto é, quando o modelo superestima o RUL, e 0,7 nos demais casos. O termo exponencial aumenta o peso dos erros quando o RUL real é baixo. O peso direcional penaliza mais a superestimação, associada a atraso operacional, e menos a subestimação, associada a uma postura conservadora.

A Figura 34 apresenta o resultado obtido com a perda conservadora.

A Figura 34 mostra o efeito da mudança de função de perda. Em comparação com a *loss* NASA, os pontos deixam de se distribuir em torno de uma regressão global mais equi-

librada e passam a refletir uma tendência de subestimação do RUL. Essa tendência aparece quantitativamente no viés negativo.

Com a *loss* conservadora, o desempenho global foi inferior ao observado com a *loss* NASA. O modelo obteve $R^2 = 0,238$, RMSE de 12,47 dias, MAE de 10,61 dias, MedAE de 9,75 dias, viés médio de $-6,08$ dias e Score NASA de 51,91. O viés negativo indica tendência de subestimação do RUL, isto é, o modelo passou a prever vidas remanescentes menores que as reais.

Embora essa piora global pareça indesejável em uma avaliação puramente estatística, ela é coerente com o objetivo operacional da função de perda. A subestimação moderada é mais conservadora do que a superestimação quando o equipamento está próximo da falha. Em outras palavras, o modelo passa a errar mais do lado da cautela, antecipando alertas em vez de atrasá-los.

5.10 COMPARAÇÃO ENTRE DESEMPENHO GLOBAL E DESEMPENHO OPERACIONAL

A comparação entre a *loss* NASA e a *loss* horizon_conservative evidencia uma distinção fundamental entre dois objetivos de modelagem: previsão global de RUL e apoio operacional à manutenção. A primeira busca minimizar o erro contínuo das previsões. A segunda busca transformar previsões em decisões úteis, especialmente em regiões próximas da falha.

Para a avaliação operacional, as previsões contínuas de RUL foram convertidas em alarmes a partir de limiares críticos. Por exemplo, no limiar de 15 dias, considera-se que há alarme quando o modelo prevê $RUL \leq 15$ dias. Essa transformação permite avaliar o modelo não apenas como regressor, mas também como sistema de apoio à decisão para manutenção.

Nesse contexto, um verdadeiro positivo (VP) ocorre quando a bomba está realmente em condição crítica e o modelo emite o alarme corretamente. Um falso positivo (FP) ocorre quando o modelo emite alarme, mas a condição real ainda não é crítica. Um falso negativo (FN) ocorre quando a bomba está em condição crítica, mas o modelo não emite alarme. Por fim, um verdadeiro negativo (VN) ocorre quando a bomba não está em condição crítica e o modelo corretamente não emite alarme.

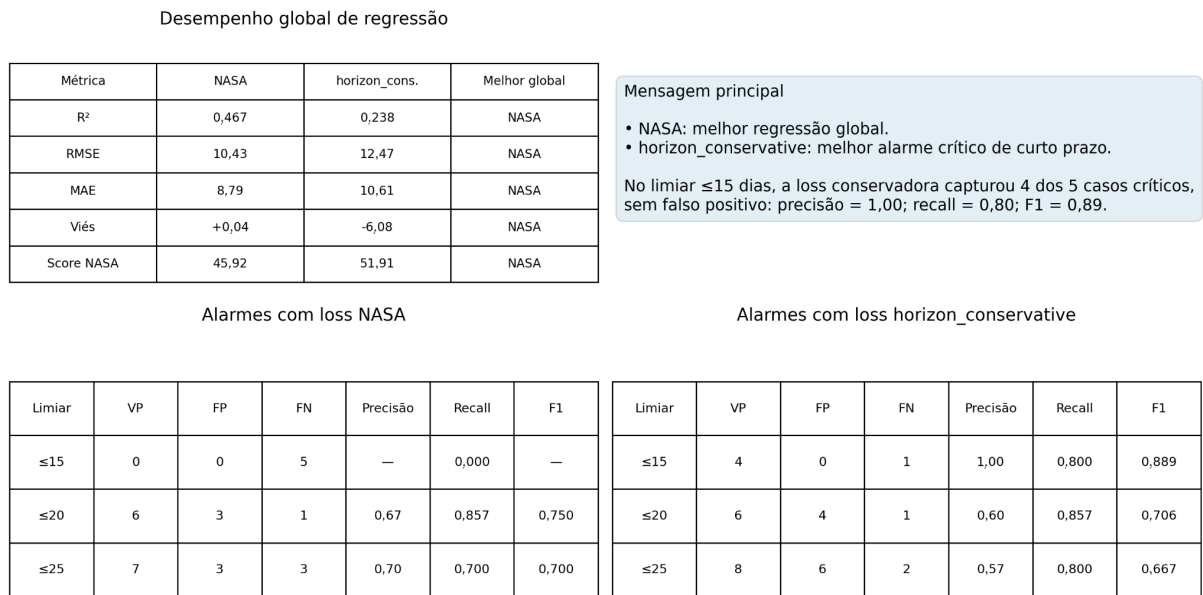
As três métricas operacionais são definidas em conjunto pela Equação 36 ((Rijsbergen, 1979)):

$$\begin{aligned} \text{Prec} &= \frac{VP}{VP + FP}, \\ \text{Rec} &= \frac{VP}{VP + FN}, \\ F_1 &= 2 \frac{\text{Prec} \cdot \text{Rec}}{\text{Prec} + \text{Rec}}. \end{aligned} \tag{36}$$

Na Equação 36, Prec representa a precisão, Rec representa o *recall* e F_1 representa o F1-score. A precisão é a proporção de alarmes emitidos que estavam corretos; portanto, diminui

Figura 35 – Comparativo entre a *loss* NASA e a *loss* horizon_conservative. A *loss* NASA apresentou melhor desempenho global de regressão, enquanto a *loss* conservadora apresentou melhor desempenho operacional no limiar crítico de 15 dias.

Comparativo operacional — NASA vs. horizon_conservative



Fonte: Elaborada pelo autor (2026).

quando há muitos falsos positivos. O recall, também chamado de sensibilidade ou revocação, é a proporção dos casos críticos reais identificados; portanto, diminui quando há falsos negativos. O F1-score é a média harmônica entre precisão e recall e só permanece alto quando ambas são elevadas. Neste estudo, essa distinção é operacionalmente importante: um falso positivo pode antecipar uma inspeção, enquanto um falso negativo deixa de sinalizar uma condição realmente crítica.

A Figura 35 resume a comparação operacional entre as duas funções de perda.

Como mostra a Figura 35, a *loss* NASA apresentou melhor desempenho global, com maior R^2 , menor RMSE, menor MAE, menor Score NASA e viés próximo de zero. Por isso, ela foi adotada como configuração principal para demonstrar a viabilidade do *transfer learning* como problema de regressão.

No limiar de 15 dias, a configuração NASA identificou 3 dos 5 pontos críticos, sem falso positivo: foram 3 verdadeiros positivos, 2 falsos negativos e 25 verdadeiros negativos. Consequentemente, sua precisão foi 1,00, o recall foi 0,60 e o F1-score foi 0,75. Esse resultado mostra que os alarmes emitidos eram confiáveis, mas dois casos críticos permaneceram sem detecção.

Por outro lado, a *loss* horizon_conservative apresentou melhor comportamento no limiar operacional mais crítico. No limiar de 15 dias, a *loss* conservadora obteve 4 verdadeiros positivos, nenhum falso positivo e apenas 1 falso negativo. Em termos práticos, isso significa

que, dos 5 casos realmente críticos, 4 foram corretamente identificados pelo modelo, enquanto nenhum alarme foi emitido indevidamente para casos não críticos. Por isso, a precisão foi igual a 1,00, o recall foi igual a 0,80 e o F1-score foi igual a 0,889.

Nos limiares de 20 e 25 dias, a *loss* conservadora manteve recall elevado, mas passou a gerar mais falsos positivos. No limiar de 20 dias, foram obtidos 6 verdadeiros positivos, 4 falsos positivos e 1 falso negativo, com precisão de 0,60, recall de 0,857 e F1 de 0,706. No limiar de 25 dias, foram obtidos 8 verdadeiros positivos, 6 falsos positivos e 2 falsos negativos, com precisão de 0,57, recall de 0,80 e F1 de 0,667.

Esse comportamento é esperado quando o limiar é ampliado. Ao passar de 15 para 20 ou 25 dias, mais pontos passam a ser classificados como críticos. Isso aumenta a chance de capturar falhas reais, mas também eleva a probabilidade de emitir alarmes em situações que ainda não estavam dentro do horizonte crítico considerado. Portanto, a escolha do limiar deve refletir o equilíbrio desejado entre antecipação e confiabilidade dos alarmes.

Esses resultados sugerem que o limiar de 15 dias é o ponto em que a *loss* conservadora apresentou melhor utilidade prática, pois combinou alta captura de casos críticos com ausência de falsos positivos. Em aplicações industriais, esse tipo de comportamento pode ser desejável quando o custo de perder uma falha iminente é maior do que o custo de antecipar uma inspeção.

Portanto, as duas funções de perda cumprem papéis diferentes. A *loss* NASA é mais adequada para demonstrar a capacidade geral de transferência de conhecimento entre o C-MAPSS e a Bomba D. Já a *loss* horizon_conservative funciona como uma calibração operacional, direcionada à proteção de curto prazo. Essa distinção é relevante para a interpretação final da dissertação: um modelo de RUL pode ser avaliado tanto como regressor quanto como sistema de apoio à decisão, e essas duas perspectivas podem favorecer configurações diferentes.

5.11 SÍNTESE DO CAPÍTULO

A transferência foi tecnicamente viável apenas no caso em que mecanismo de falha, sensoriamento e conhecimento de origem apresentaram compatibilidade suficiente: a B-0302D. Sob validação *leave-one-failure-series-out*, o AGATT com *backbone* congelado superou as linhas de base sem pré-treinamento, mas os 30 pontos de teste derivam de somente três trajetórias e não constituem falhas independentes.

A configuração principal alcançou $R^2 = 0,467$ e RMSE de 10,43 dias. A perda conservadora aumentou a captura de casos críticos no limiar de 15 dias, ao custo de pior desempenho contínuo. Os achados devem, portanto, ser interpretados como evidência retrospectiva de viabilidade para a B-0302D, condicionada à existência de degradação observável e sujeita a confirmação prospectiva com o pipeline congelado.

6 CONCLUSÃO

Esta dissertação avaliou em que condições representações aprendidas no FD001 da C-MAPSS podem apoiar a estimação retrospectiva de RUL nas bombas do conjunto B-0302. A investigação combinou caracterização exploratória de confiabilidade, reprodução do AGATT no domínio de origem e transferência por extração de características para o domínio industrial.

A análise das bombas mostrou que quantidade de eventos, por si só, não determina previsibilidade. A B-0302D reuniu maior riqueza sensorial e recorrência de falhas de vedação compatíveis com degradação progressiva; por isso, foi o caso mais defensável para o aprofundamento. Os eventos elétricos e de controle das demais bombas não devem ser considerados intrinsecamente imprevisíveis: existem técnicas específicas para seu diagnóstico e prognóstico (Nandi, Toliyat e Li, 2005). A limitação observada decorreu da incompatibilidade entre parte desses eventos, os sensores disponíveis e o padrão progressivo representado no FD001.

No domínio de origem, a melhor execução do AGATT alcançou RMSE de 13,88 ciclos e $R^2 = 0,8885$; em 30 execuções, o RMSE médio foi de $15,99 \pm 1,39$ ciclos. Essa dispersão confirma que resultados de redes profundas não devem ser avaliados apenas pela melhor execução.

Na B-0302D, o AGATT congelado, associado à projeção físico-estatística e a uma nova cabeça de regressão, obteve $R^2 = 0,467$, RMSE de 10,43 dias, MAE de 8,79 dias e viés de +0,04 dia. A LSTM industrial e o AGATT treinado sem transferência de aprendizado apresentaram RMSE médios de 16,34 e 18,26 dias. O ganho, portanto, não decorreu somente da arquitetura, mas do reaproveitamento das representações aprendidas no FD001. No limiar de 15 dias, a perda conservadora identificou quatro dos cinco casos críticos sem falsos positivos, mostrando que desempenho contínuo e utilidade operacional podem favorecer configurações diferentes.

Esses resultados constituem evidência de viabilidade, não de generalização irrestrita. Os 30 pontos externos derivam de somente três trajetórias e apresentam dependência intrassérie. A configuração foi selecionada retrospectivamente; as linhas de base foram avaliadas em múltiplas sementes, enquanto o modelo transferido foi apresentado por sua configuração principal; e não houve validação prospectiva. Além disso, o RUL industrial foi expresso em dias de calendário. Embora as bombas possuam horímetros, não foi possível obter uma série histórica contínua e sincronizada de horas de funcionamento correlacionada ao sensoriamento e às falhas. Em um conjunto operado por rodízio, dias de calendário não equivalem a exposição efetiva.

A contribuição principal é um fluxo auditável para testar transferência entre domínios fisicamente distintos: caracterizar os modos de falha, reproduzir o modelo de origem, harmonizar os sensores sem usar a série externa de teste, comparar com linhas de base e interpretar os resultados conforme a decisão de manutenção. As previsões devem apoiar triagem e priorização de inspeções, mantendo o diagnóstico e a decisão final sob responsabilidade da Engenharia de Manutenção.

6.1 TRABALHOS FUTUROS

A prioridade é validar prospectivamente o pipeline congelado, registrando previsões antes de novos eventos e repetindo o modelo transferido em múltiplas sementes sob protocolo simétrico às linhas de base. Também é necessário integrar os horímetros ao histórico sensorial, com identificação por bomba e sincronização temporal, além de incorporar número de partidas, carga, vazão e tempo em reserva.

A ampliação da base deverá distinguir mecanismos de falha e incluir sensoriamento direcionado a eventos elétricos, como grandezas por fase, qualidade da alimentação, isolamento, harmônicos, partidas e atuações de proteção. Com novas trajetórias, poderão ser avaliados modelos específicos por mecanismo, quantificação de incerteza, pré-treinamento auto-supervisionado em dados industriais e estratégias de adaptação menos dependentes de correspondência explícita entre sensores.

Para uso contínuo, serão necessários versionamento de dados e modelos, monitoramento de deriva, auditoria das previsões e registro das intervenções. A implantação deve começar em modo de recomendação, com critérios operacionais definidos e acompanhamento da engenharia.

6.2 CONSIDERAÇÕES FINAIS

A transferência do FD001 para a B-0302D superou as alternativas treinadas apenas com o pequeno histórico industrial disponível, mas seu benefício dependeu de degradação observável, instrumentação relacionada ao mecanismo de falha e harmonização cuidadosa entre os domínios. Assim, a compreensão física do ativo e a seleção adequada do problema mostraram-se tão importantes quanto a arquitetura empregada.

REFERÊNCIAS

- AL-KHAZRAJI, Huthaifa *et al.* Aircraft engines remaining useful life prediction based on a hybrid model of autoencoder and deep belief network. **IEEE Access**, v. 10, p. 82156–82163, 2022.
- BENGIO, Yoshua; SIMARD, Patrice; FRASCONI, Paolo. Learning long-term dependencies with gradient descent is difficult. **IEEE Transactions on Neural Networks**, v. 5, n. 2, p. 157–166, 1994.
- BOURLARD, Hervé; KAMP, Yves. Auto-association by multilayer perceptrons and singular value decomposition. **Biological Cybernetics**, v. 59, p. 291–294, 1988.
- Celesc. **Perguntas frequentes: horário de ponta em Santa Catarina**. 2024. <<https://pee.celesc.com.br/perguntas-frequentes>>. Acesso em: 19 maio 2026.
- CHO, Kyunghyun *et al.* Learning phrase representations using RNN encoder–decoder for statistical machine translation. In: **Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)**. Association for Computational Linguistics, 2014. p. 1724–1734. Disponível em: <<https://aclanthology.org/D14-1179.pdf>>. Acesso em: 4 ago. 2026.
- CHOLLET, François. Xception: Deep learning with depthwise separable convolutions. In: **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**. [s.n.], 2017. p. 1251–1258. Disponível em: <https://openaccess.thecvf.com/content_cvpr_2017/papers/Chollet_Xception_Deep_Learning_CVPR_2017_paper.pdf>. Acesso em: 4 ago. 2026.
- DU, Yulin *et al.* STF2: A novel spatial–temporal feature fusion method for aero-engine remaining useful life prediction. **Expert Systems with Applications**, v. 290, p. 128370, 2025.
- ELMAN, Jeffrey L. Finding structure in time. **Cognitive Science**, v. 14, n. 2, p. 179–211, 1990.
- FAIZANBASHA, A.; RIZWAN, U. Deep learning–stochastic ensemble for RUL prediction and predictive maintenance with dynamic mission abort policies. **Reliability Engineering & System Safety**, v. 259, p. 110919, 2025.
- FAN, Yijing *et al.* A remaining useful life prediction method based on DATCN-PSOSEN for turbofan engines. **Journal of Physics: Conference Series**, v. 2853, n. 1, p. 012068, 2024.
- FAN, Zhengyang; LI, Wanru; CHANG, Kuo-Chu. A bidirectional long short-term memory autoencoder transformer for remaining useful life estimation. **Mathematics**, v. 11, n. 24, p. 4972, 2023. ISSN 2227-7390.
- FAN, Zhengyang; LI, Wanru; CHANG, Kuo-Chu. A two-stage attention-based hierarchical transformer for turbofan engine remaining useful life prediction. **Sensors**, v. 24, n. 3, p. 824, 2024.
- FU, Song *et al.* MCA-DTCN: A novel dual-task temporal convolutional network with multi-channel attention for first prediction time detection and remaining useful life prediction. **Reliability Engineering & System Safety**, v. 241, p. 109696, 2024.
- GERs, Felix A.; SCHMIDHUBER, Jürgen; CUMMINS, Fred. Learning to forget: Continual prediction with LSTM. **Neural Computation**, v. 12, n. 10, p. 2451–2471, 2000. Disponível em: <<https://sferics.idsia.ch/pub/juergen/FgGates-NC.pdf>>. Acesso em: 4 ago. 2026.

GILMER, Justin *et al.* Neural message passing for quantum chemistry. In: **Proceedings of the 34th International Conference on Machine Learning (ICML)**. PMLR, 2017. (Proceedings of Machine Learning Research, v. 70), p. 1263–1272. Disponível em: <<https://proceedings.mlr.press/v70/gilmer17a.html>>. Acesso em: 4 ago. 2026.

GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. **Deep Learning**. Cambridge, MA: MIT Press, 2016. ISBN 9780262035613. Disponível em: <<https://www.deeplearningbook.org/>>. Acesso em: 4 ago. 2026.

GREFF, Klaus *et al.* Lstm: A search space odyssey. **IEEE Transactions on Neural Networks and Learning Systems**, v. 28, n. 10, p. 2222–2232, 2017.

HE, Chenwen *et al.* A hybrid model based on a dual-attention mechanism for the prediction of remaining useful life of aircraft engines. **Sensors**, v. 25, n. 18, p. 5682, 2025.

HE, Kaiming *et al.* Deep residual learning for image recognition. In: **IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**. [S.l.: s.n.], 2016. p. 770–778.

HEIMES, Felix O. Recurrent neural networks for remaining useful life estimation. In: **2008 International Conference on Prognostics and Health Management**. [s.n.], 2008. p. 1–6. Disponível em: <<https://doi.org/10.1109/PHM.2008.4711422>>. Acesso em: 4 ago. 2026.

HENG, A. *et al.* Rotating machinery prognostics: State of the art, challenges and opportunities. **Mechanical Systems and Signal Processing**, v. 23, n. 3, p. 724–739, 2009.

HINTON, Geoffrey E.; SALAKHUTDINOV, Ruslan R. Reducing the dimensionality of data with neural networks. **Science**, v. 313, n. 5786, p. 504–507, 2006.

HOCHREITER, Sepp; SCHMIDHUBER, Jürgen. Long short-term memory. **Neural Computation**, v. 9, n. 8, p. 1735–1780, 1997. Disponível em: <<https://people.idsia.ch/~juergen/lstm1997-2024head.pdf>>. Acesso em: 4 ago. 2026.

HU, Qiankun; ZHAO, Yongping; REN, Lihua. Novel transformer-based fusion models for aero-engine remaining useful life estimation. **IEEE Access**, v. 11, p. 52668–52685, 2023.

IEC. **IEC 60812:2018 — Failure modes and effects analysis (FMEA and FMECA)**. 2018. International Electrotechnical Commission.

ISO. **ISO 14224:2016 — Petroleum, petrochemical and natural gas industries — Collection and exchange of reliability and maintenance data for equipment**. 2016. International Organization for Standardization.

KIPF, Thomas N.; WELING, Max. Semi-supervised classification with graph convolutional networks. In: **5th International Conference on Learning Representations (ICLR)**. OpenReview.net, 2017. Disponível em: <<https://openreview.net/forum?id=SJU4ayYgl>>. Acesso em: 4 ago. 2026.

KRIZHEVSKY, Alex; SUTSKEVER, Ilya; HINTON, Geoffrey E. Imagenet classification with deep convolutional neural networks. In: **Advances in Neural Information Processing Systems (NeurIPS)**. [s.n.], 2012. p. 1097–1105. Disponível em: <<https://proceedings.neurips.cc/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html>>. Acesso em: 4 ago. 2026.

KVÅLSETH, Tarald O. Cautionary note about r^2 . **The American Statistician**, v. 39, n. 4, p. 279–285, 1985. Disponível em: <<https://doi.org/10.1080/00031305.1985.10479448>>. Acesso em: 4 ago. 2026.

LECUN, Yann *et al.* Gradient-based learning applied to document recognition. **Proceedings of the IEEE**, v. 86, n. 11, p. 2278–2324, 1998. Disponível em: <<https://leon.bottou.org/publications/pdf/ieee-1998.pdf>>. Acesso em: 4 ago. 2026.

LEI, Yaguo *et al.* Deep transfer learning for cross-domain machinery remaining useful life prediction using c-mapss knowledge. **Reliability Engineering & System Safety**, v. 218, p. 108114, 2022. Acesso em 05 ago. 2025. Disponível em: <<https://doi.org/10.1016/j.ress.2021.108114>>.

LI, Hao; WANG, Zhuojian; LI, Zhe. An enhanced CNN-LSTM remaining useful life prediction model for aircraft engine with attention mechanism. **PeerJ Computer Science**, v. 8, p. e1084, 2022.

LI, Junhuai *et al.* A dual-scale transformer-based remaining useful life prediction model in industrial internet of things. **IEEE Internet of Things Journal**, v. 11, n. 16, p. 26656–26667, 2024.

LI, Rongzhang *et al.* A lightweight transformer and depthwise separable convolution model for remaining useful life prediction of turbofan engines. **Measurement Science and Technology**, v. 35, n. 2, p. 025020, 2024.

LI, Xiaosheng; CAO, Hui; MA, Zhenhao. Research on remaining useful life prediction of dual-fuel main engine based on CBAM attention mechanism. In: **Proceedings of the 3rd International Conference on Artificial Intelligence, Automation and Algorithms (AI2A 2023)**. New York: ACM, 2023. p. 91–99.

LIU, Yitong; WEN, Jiarui; WANG, Guoqiang. A comprehensive overview of remaining useful life prediction: From traditional literature review to scientometric analysis. **Machine Learning with Applications**, v. 21, p. 100704, 2025.

MCCULLOCH, Warren S.; PITTS, Walter. A logical calculus of the ideas immanent in nervous activity. **The Bulletin of Mathematical Biophysics**, v. 5, p. 115–133, 1943. Disponível em: <<https://link.springer.com/content/pdf/10.1007/BF02478259.pdf>>. Acesso em: 4 ago. 2026.

MITCHELL, Tom M. **Machine Learning**. New York: McGraw-Hill, 1997. ISBN 978-0070428072.

NAMUDURI, Srikanth *et al.* Deep learning methods for sensor based predictive maintenance and future perspectives for electrochemical sensors. **Journal of The Electrochemical Society**, v. 167, n. 3, p. 037552, 2020.

NANDI, Subhasis; TOLIYAT, Hamid A.; LI, Xiaodong. Condition monitoring and fault diagnosis of electrical motors—a review. **IEEE Transactions on Energy Conversion**, v. 20, n. 4, p. 719–729, 2005.

NELSON, Wayne. **Applied Life Data Analysis**. New York: Wiley, 1982. ISBN 9780471094586.

NIST/SEMATECH. **Exponential Distribution**. 2012. Engineering Statistics Handbook, Section 8.1.6.1. Disponível em: <<https://www.itl.nist.gov/div898/handbook/apr/section1/apr161.htm>>. Acesso em: 4 ago. 2026.

NIST/SEMATECH. **Failure (or Hazard) Rate**. 2012. Engineering Statistics Handbook, Section 8.1.2.3. Disponível em: <<https://www.itl.nist.gov/div898/handbook/apr/section1/apr123.htm>>. Acesso em: 4 ago. 2026.

NIST/SEMATECH. **Reliability or Survival Function**. 2012. Engineering Statistics Handbook, Section 8.1.2.2. Disponível em: <<https://www.itl.nist.gov/div898/handbook/apr/section1/apr122.htm>>. Acesso em: 4 ago. 2026.

NIST/SEMATECH. **Weibull Distribution**. 2012. Engineering Statistics Handbook, Section 8.1.6.2. Disponível em: <<https://www.itl.nist.gov/div898/handbook/apr/section1/apr162.htm>>. Acesso em: 4 ago. 2026.

O'CONNOR, Patrick; KLEYNER, Andre. **Practical Reliability Engineering**. 5. ed. Chichester: Wiley, 2012. ISBN 9781119961260.

PAN, Sinno Jialin; YANG, Qiang. A survey on transfer learning. **IEEE Transactions on Knowledge and Data Engineering**, v. 22, n. 10, p. 1345–1359, 2010.

PANG, Guansong *et al.* Deep learning for anomaly detection: A review. **ACM Computing Surveys**, v. 54, n. 2, p. 1–38, 2021.

PEDREGOSA, F. *et al.* Scikit-learn: Machine learning in python. **Journal of Machine Learning Research**, v. 12, p. 2825–2830, 2011. Disponível em: <<https://www.jmlr.org/papers/v12/pedregosa11a.html>>. Acesso em: 4 ago. 2026.

RAGAB, Mohamed *et al.* Attention-based sequence to sequence model for machine remaining useful life prediction. **Neurocomputing**, v. 466, p. 58–68, 2021.

ReliaSoft Corporation. **Introduction to Life Data Analysis**. 2023. <https://help.reliasoft.com/reference/life_data_analysis/lda/introduction_to_life_data_analysis.html>. Acesso em 05 ago 2025.

RIJSBERGEN, Cornelis Joost van. **Information Retrieval**. 2. ed. London: Butterworths, 1979. Disponível em: <<https://www.dcs.gla.ac.uk/Keith/pdf/Chapter7.pdf>>. Acesso em: 4 ago. 2026.

ROSENBLATT, Frank. The perceptron: A probabilistic model for information storage and organization in the brain. **Psychological Review**, v. 65, n. 6, p. 386–408, 1958. Disponível em: <<https://papers.baulab.info/papers/Rosenblatt-1958.pdf>>. Acesso em: 4 ago. 2026.

SAKURADA, Mayu; YAIRI, Takehisa. Anomaly detection using autoencoders with nonlinear dimensionality reduction. In: **Proceedings of the MLSDA 2014 2nd Workshop on Machine Learning for Sensory Data Analysis**. [S.l.: s.n.], 2014. p. 4:1–4:8.

SAXENA, Abhinav *et al.* Metrics for evaluating performance of prognostic techniques. In: **2008 International Conference on Prognostics and Health Management**. [s.n.], 2008. p. 1–17. Disponível em: <<https://data.nasa.gov/dataset/metrics-for-evaluating-performance-of-prognostics-techniques>>. Acesso em: 4 ago. 2026.

SAXENA, Abhinav; GOEBEL, Kai. Phm '08 data challenge competition. In: IEEE. **Proceedings of the International Conference on Prognostics and Health Management**. 2008. Disponível em: <<https://data.nasa.gov/dataset/phm-2008-challenge>>. Acesso em: 4 ago. 2026.

SAXENA, Abhinav *et al.* Damage propagation modeling for aircraft engine run-to-failure simulation. In: **IEEE International Conference on Prognostics and Health Management (PHM)**. [s.n.], 2008. p. 1–9. Disponível em: <<https://data.nasa.gov/dataset/damage-propagation-modeling-for-aircraft-engine-run-to-failure-simulation>>. Acesso em: 4 ago. 2026.

SAXENA, Abhinav *et al.* Turbofan engine degradation simulation data set for prognostics and health management. In: IEEE. **Proceedings of the International Conference on Prognostics and Health Management**. 2008. Disponível em: <<https://ti.arc.nasa.gov/c/6/>>. Acesso em: 5 ago. 2025.

SIFRE, Laurent. **Rigid-Motion Scattering for Image Classification**. Tese (Doutorado) — École Polytechnique, Palaiseau, France, 2014. Disponível em: <https://www.di.ens.fr/data/publications/papers/phd_sifre.pdf>. Acesso em: 4 ago. 2026.

SONI, Hima; KANSARA, Akshay; JOSHI, Tanvi. Predictive maintenance of gas turbine using prognosis approach. **International Research Journal of Engineering and Technology**, v. 7, n. 6, p. 4683–4691, 2020. Disponível em: <<https://www.irjet.net/archives/V7/i6/IRJET-V7I6878.pdf>>. Acesso em: 4 ago. 2026.

VASWANI, Ashish *et al.* Attention is all you need. In: **Advances in Neural Information Processing Systems (NeurIPS)**. [s.n.], 2017. p. 5998–6008. Disponível em: <https://papers.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf>. Acesso em: 4 ago. 2026.

VELIČKOVIĆ, Petar *et al.* Graph attention networks. In: **6th International Conference on Learning Representations (ICLR)**. OpenReview.net, 2018. Disponível em: <<https://arxiv.org/pdf/1710.10903>>. Acesso em: 4 ago. 2026.

VINCENT, Pascal *et al.* Extracting and composing robust features with denoising autoencoders. In: **Proceedings of the 25th International Conference on Machine Learning (ICML)**. [S.l.: s.n.], 2008. p. 1096–1103.

WAHID, Abdul; BRESLIN, John G.; INTIZAR, Muhammad Ali. TCRSCANet: Harnessing temporal convolutions and recurrent skip component for enhanced RUL estimation in mechanical systems. **Human-Centric Intelligent Systems**, v. 4, n. 1, p. 1–24, 2024.

WANG, Feng *et al.* A deep learning based data fusion method for degradation modeling and prognostics. **IEEE Transactions on Reliability**, v. 70, n. 2, p. 775–789, 2021.

WANG, Xin *et al.* Remaining useful life prediction for aero-engines using a time-enhanced multi-head self-attention model. **Aerospace**, v. 10, n. 1, p. 80, 2023.

WANG, Youming; HUANG, Yirun. Res-TCEANet: An expansive attention mechanism with positional correspondence based on semi-supervised temporal convolutional network for RUL estimation. **Measurement**, v. 241, p. 115714, 2025.

WANG, Youdao; ZHAO, Yifan; ADDEPALLI, Sri. Remaining useful life prediction using deep learning approaches: A review. **Procedia Manufacturing**, v. 49, p. 81–88, 2020.

WEIBULL, Waloddi. A statistical distribution function of wide applicability. **Journal of Applied Mechanics**, v. 18, n. 3, p. 293–297, 1951. Disponível em: <https://asmedigitalcollection.asme.org/appliedmechanics/article-pdf/18/3/293/6747088/293_1.pdf>. Acesso em: 4 ago. 2026.

WERBOS, Paul J. Backpropagation through time: What it does and how to do it. **Proceedings of the IEEE**, v. 78, n. 10, p. 1550–1560, 1990.

WU, Chenchen *et al.* A dual-path architecture based on time series decomposition and degradation correction for remaining useful life prediction of aero-engine. **Computers & Industrial Engineering**, v. 203, p. 110964, 2025.

WU, Fan; ZHANG, Tong; LI, Wei. Remaining useful life prediction based on deep learning: A survey. **Sensors**, v. 24, n. 11, p. 3454, 2024.

XIE, Shushuai *et al.* Bayesian cooperative probabilistic transformer for remaining useful life prediction with uncertainty estimation in industrial equipment. **Advanced Engineering Informatics**, v. 67, p. 103515, 2025.

XU, Zhiqiang *et al.* Global attention mechanism based deep learning for remaining useful life prediction of aero-engine. **Measurement**, v. 217, p. 113098, 2023.

XU, Zhiqiang; ZHANG, Yujie; MIAO, Qiang. An attention-based multi-scale temporal convolutional network for remaining useful life prediction. **Reliability Engineering & System Safety**, v. 250, p. 110288, 2024.

YANG, C.-L. *et al.* Uncertainty quantification and remaining useful life prediction using deep learning. **International Journal of Simulation Modelling**, v. 24, n. 4, p. 611–622, 2025.

YOSINSKI, Jason *et al.* How transferable are features in deep neural networks? In: **Advances in Neural Information Processing Systems (NeurIPS)**. [s.n.], 2014. p. 3320–3328. Disponível em: <https://proceedings.neurips.cc/paper_files/paper/2014/hash/532a2f85b6977104bc93f8580abbb330-Abstract.html>. Acesso em: 4 ago. 2026.

ZENG, Ming; WU, Feng; CHENG, Yiwei. Remaining useful life prediction via spatio-temporal channels and transformer. **IEEE Sensors Journal**, v. 23, n. 23, p. 29176–29185, 2023. Disponível em: <<https://doi.org/10.1109/JSEN.2023.3324330>>. Acesso em: 4 ago. 2026.

ZHANG, Qiang *et al.* A dynamic spatio-temporal graph neural network with global and local perception for remaining useful life prediction. **Expert Systems with Applications**, v. 282, p. 127707, 2025.

ZHANG, Xu *et al.* A dual-dimensional parallel neural network integrating multi-scale and frequency-domain features for aircraft engine life prediction. **Arabian Journal for Science and Engineering**, 2025. Publicação online em 2025; paginação final atribuída ao volume 51 em 2026.

ZHENG, Shuai *et al.* Long short-term memory network for remaining useful life estimation. In: **2017 IEEE International Conference on Prognostics and Health Management (ICPHM)**. [s.n.], 2017. p. 88–95. Disponível em: <<https://doi.org/10.1109/ICPHM.2017.7998311>>. Acesso em: 4 ago. 2026.

UNIVERSIDADE DO ESTADO DE SANTA CATARINA – UDESC
BIBLIOTECA UNIVERSITÁRIA
REPOSITÓRIO INSTITUCIONAL

CENTRO DE CIÊNCIAS TECNOLÓGICAS – CCT

ATESTADO DE VERSÃO FINAL

Eu, Yuri Kaszubowski Lopes, professor(a) do curso de Mestrado em Computação Aplicada, declaro que esta é a versão final aprovada pela comissão julgadora da dissertação intitulada: **“PREVISÃO DE FALHAS EM BOMBAS INDUSTRIAIS DE TRANSPORTE DUTOVIÁRIO DE PETRÓLEO E DERIVADOS COM REDES NEURAIS ARTIFICIAIS PROFUNDAS E TRANSFERÊNCIA DE APRENDIZADO ENTRE DOMÍNIOS”** de autoria do(a) acadêmico Gabriel Vinicius Boin Freitas

Joinville, <datado conforme assinatura digital>.

Assinatura digital do(a) orientador(a):

Documento assinado digitalmente
 **YURI KASZUBOWSKI LOPES**
Data: 03/09/2026 18:46:30-0300
Verifique em <https://validar.iti.gov.br>

Yuri Kaszubowski Lopes