

**UNIVERSIDADE FEDERAL DE JUIZ DE FORA**  
**FACULDADE DE ENGENHARIA**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA**

**Robson Rosserrani de Lima**

**Uma abordagem estatística de processamento de sinais para detecção de  
ilhamento**

Juiz de Fora  
2026

**Robson Rosserrani de Lima**

**Uma abordagem estatística de processamento de sinais para detecção de  
ilhamento**

Tese apresentada ao Programa de  
Pós-Graduação em Engenharia  
Elétrica da Universidade Federal de Juiz  
de Fora como requisito parcial à obtenção  
do título de Doutor em Engenharia  
Elétrica. Área: Sistemas Eletrônicos.

Orientador: Prof. Dr. Augusto Santiago Cerqueira

Coorientador: Prof. Dr. Paulo Fernando Ribeiro

Juiz de Fora

2026

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

Lima, Robson Rosserrani de.

Uma abordagem estatística de processamento de sinais para detecção de ilhamento / Robson Rosserrani de Lima. -- 2026.  
164 p. : il.

Orientador: Augusto Santiago Cerqueira

Coorientador: Paulo Fernando Ribeiro

Tese (doutorado) - Universidade Federal de Juiz de Fora, Faculdade de Engenharia. Programa de Pós-Graduação em Engenharia Elétrica, 2026.

1. Detecção de Ilhamento. 2. Geração Distribuída. 3. Estatísticas de Ordem Superior. 4. Processamento Estatístico de Sinais. I. Cerqueira, Augusto Santiago, orient. II. Ribeiro, Paulo Fernando, coorient. III. Título.

**Robson Rosserrani de Lima**

**Uma abordagem estatística de processamento de sinais para detecção de ilhamento**

Tese apresentada ao Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal de Juiz de Fora como requisito parcial à obtenção do título de Doutor em Engenharia Elétrica. Área de concentração: Sistemas Eletrônicos

Aprovada em 14 de janeiro de 2026.

BANCA EXAMINADORA

**Prof. Dr. Augusto Santiago Cerqueira** - Orientador  
Universidade Federal de Juiz de Fora

**Prof. Dr. Paulo Fernando Ribeiro** - Coorientador  
Universidade Federal de Itajubá

**Prof. Dr. Carlos Augusto Duque**  
Universidade Federal de Juiz de Fora

**Prof. Dr. Leandro Rodrigues Manso Silva**  
Universidade Federal de Juiz de Fora

**Prof. Dr. Flavio Bezerra Costa**  
Michigan Technological University

**Prof. Dr. Denis Vinicius Coury**  
Universidade de São Paulo - Campus de São Carlos

Juiz de Fora, 12/12/2025.



Documento assinado eletronicamente por **Augusto Santiago Cerqueira, Professor(a)**, em 14/01/2026, às 13:15, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Carlos Augusto Duque, Professor(a)**, em 14/01/2026, às 13:16, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Leandro Rodrigues Manso Silva, Professor(a)**, em 14/01/2026, às 16:13, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **PAULO FERNANDO RIBEIRO, Usuário Externo**, em 14/01/2026, às 19:43, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).

Documento assinado eletronicamente por **DENIS VINICIUS COURY, Usuário Externo**, em 06/02/2026, às 15:23, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



[de 2020.](#)



Documento assinado eletronicamente por **Flavio Bezerra Costa, Usuário Externo**, em 23/04/2026, às 09:32, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no Portal do SEI-Ufjf ([www2.ufjf.br/SEI](http://www2.ufjf.br/SEI)) através do ícone Conferência de Documentos, informando o código verificador **2791342** e o código CRC **A4F468F5**.

*À minha mãe.  
Com todo o meu amor, sempre.*

*“Não sou nada.  
Nunca serei nada.  
Não posso querer ser nada.  
À parte isso, tenho em mim todos os sonhos do mundo.”*

— Álvaro de Campos (heterônimo de Fernando Pessoa), *Tabacaria*

## RESUMO

A detecção de ilhamento não intencional em redes de distribuição com geração distribuída permanece desafiadora, sobretudo em condições críticas associadas à Zona de Não-Detecção (ZND), nas quais o desbalanço entre geração e carga pode ser mínimo. Esta tese propõe e valida um método passivo de detecção de ilhamento baseado em processamento estatístico de sinais e reconhecimento de padrões. O fluxo é estruturado em duas etapas: (i) calibração offline, na qual se estima uma projeção linear fixa por Análise de Componentes Principais (PCA) orientada ao evento, comprimindo as tensões trifásicas do ponto de acoplamento comum em baixa dimensão; em seguida, extraem-se cumulantes (estatísticas de ordem superior) em atrasos não nulos, selecionando-se de forma parcimoniosa os atributos via Razão Discriminante de Fisher e treinando-se os decisores; e (ii) execução online em janelas deslizantes, na qual apenas os atrasos previamente selecionados são computados e aplicados aos decisores treinados. O método foi avaliado em uma microrrede fotovoltaica simulada, contemplando três classes de eventos (operação nominal, ilhamento e faltas) e cenários na ZND, incluindo validação independente com perfis não vistos. No regime de janelamento deslizante, o classificador MLP selecionado atingiu acurácia global de 98,52% em teste, com alta seletividade para ilhamento (precisão de 98,49% e sensibilidade de 97,15%) e ausência de falsos alarmes relevantes na classe nominal. Como alternativa linear de menor complexidade, ajustou-se também um decisor por mínimos quadrados (Least Squares – LS) no mesmo espaço de atributos; embora apresente elevada precisão, seu desempenho global e, principalmente, a sensibilidade ao ilhamento foram inferiores (acurácia de 79,59% e sensibilidade de 66,38% em teste). Na validação na ZND, a decisão na subjanela mais tardia (pior caso, aproximadamente quatro ciclos) manteve desempenho praticamente ideal para ilhamento (precisão e sensibilidade de 100%) e permaneceu compatível com limites normativos típicos para ambos os decisores adotados. Adicionalmente, a varredura de ruído gaussiano branco aditivo entre 70 e 30 dB evidenciou baixa sensibilidade do vetor de características, preservando altas taxas de detecção, especialmente nas subjanelas mais informativas. Conclui-se que a combinação de projeção fixa orientada ao ilhamento e cumulantes em atrasos selecionados resulta em um detector robusto, de baixa dimensionalidade e com tempo de decisão compatível com requisitos de interconexão, com potencial para implementação prática em funções de proteção de redes de distribuição.

**Palavras-chave:** detecção de ilhamento; geração distribuída; estatísticas de ordem superior; cumulantes; pca; proteção de sistemas de distribuição.

## ABSTRACT

Unintentional islanding detection in distribution networks with distributed generation remains challenging, particularly under critical operating conditions associated with the Non-Detection Zone (NDZ), in which the generation–load mismatch may be minimal. This thesis proposes and validates a passive islanding detection method grounded on statistical signal processing and pattern recognition. The workflow is structured in two stages: (i) an offline calibration stage, in which a fixed linear projection based on event-oriented Principal Component Analysis (PCA) is estimated to compress three-phase point-of-common-coupling voltages into a low-dimensional representation; next, cumulants (higher-order statistics) are extracted at nonzero lags, and features are parsimoniously selected using the Fisher Discriminant Ratio, followed by the training of decision-makers; and (ii) an online execution stage using sliding windows, in which only the previously selected lags are computed and fed into the trained decision-makers. The method was assessed on a simulated photovoltaic microgrid, considering three event classes (normal operation, islanding, and faults) and NDZ scenarios, including independent validation with unseen profiles. Under the sliding-window regime, the selected MLP classifier achieved an overall test accuracy of 98,52%, with high islanding selectivity (precision of 98,49% and sensitivity of 97,15%) and no relevant false alarms in the normal class. As a lower-complexity linear reference, a Least Squares (LS) decision-maker was also fitted in the same feature space; although it presented high precision, its overall performance and, most notably, islanding sensitivity were lower under the sliding-window regime (accuracy of 79,59% and islanding sensitivity of 66,38% in test). In the NDZ validation, the decision at the latest subwindow (worst-case delay, approximately four cycles) maintained a near-ideal performance for islanding (precision and sensitivity of 100%) and remained consistent with typical interconnection normative limits for both adopted decision-makers. Additionally, the sweep of additive white Gaussian noise from 70 to 30 dB indicated low sensitivity of the feature vector, preserving high detection rates, especially in the most informative subwindows. It is concluded that combining a fixed, islanding-oriented projection with cumulants at selected lags yields a robust, low-dimensional detector with a decision time compatible with interconnection requirements, with strong potential for practical implementation in protection functions for distribution networks.

**Keywords:** islanding detection; distributed generation; high-order statistics; cumulants; pca; pattern recognition; distribution protection.

## LISTA DE ILUSTRAÇÕES

Figura 1	– Número anual de resultados recuperados no Portal CAPES para <i>islanding AND detection</i> (2010–2025). . . . .	24
Figura 2	– Conceito Simples de Ilhamento. . . . .	25
Figura 3	– Arranjo de teste no PAC . . . . .	26
Figura 4	– Classificação geral dos métodos de detecção de ilhamento . . . . .	29
Figura 5	– Diagrama da matriz PV de 250 kW conectada à rede. . . . .	45
Figura 6	– Curvas I-V e P-V do módulo fotovoltaico . . . . .	46
Figura 7	– Fluxo de geração dos cenários de simulação. . . . .	50
Figura 8	– Espaço P-Q e casos simulados na ZND . . . . .	52
Figura 9	– Particionamento dos dados simulados. . . . .	54
Figura 10	– Pré-processamento da janela fixa de dois ciclos. . . . .	59
Figura 11	– Pré-processamento da janela deslizante (4 ciclos). . . . .	60
Figura 12	– Fluxo de calibração <i>offline</i> e execução <i>online</i> . . . . .	83
Figura 13	– Sinal bruto e janela fixa do janelamento . . . . .	85
Figura 14	– Janela deslizante e sub-janelas de dois ciclos . . . . .	86
Figura 15	– Distribuição dos modos de falta . . . . .	90
Figura 16	– Densidade de $\eta_1$ por classe no conjunto de otimização. . . . .	92
Figura 17	– Mapa 1-D de Razão Discriminante de Fisher (do inglês, <i>Fisher's Discriminant Ratio</i> ) (FDR) de 2ª ordem em $c_1$ . . . . .	95
Figura 18	– Mapa 1-D de FDR de 3ª ordem em $c_1$ . . . . .	96
Figura 19	– Mapa 1-D de FDR de 4ª ordem em $c_1$ . . . . .	97
Figura 20	– Seis pares com maior separabilidade segundo $J_{2D}$ . . . . .	99
Figura 21	– Projeção LDA com 15 atributos padronizados. . . . .	100
Figura 22	– Matrizes de confusão agregadas ( $K^* = 5$ ). . . . .	105
Figura 23	– Matriz de dispersão dos seis atributos cumulantes selecionados (janela fixa). . . . .	113
Figura 24	– Projeção no plano $r_2[70] \times r_4[69]$ (janela fixa). . . . .	114
Figura 25	– Exemplo de janelamento deslizante em uma realização trifásica. . . . .	116
Figura 26	– Matriz de confusão da MLP no teste . . . . .	120
Figura 27	– Matriz de confusão do classificador Mínimos Quadrados (do inglês, <i>Least Squares</i> ) (LS) no teste (janela deslizante). . . . .	121
Figura 28	– Matrizes de confusão da MLP na ZND (3 sub-janelas; 70 dB). . . . .	125
Figura 29	– Matrizes de confusão do classificador LS na ZND. . . . .	126

Figura 30	– Acurácia global de Perceptron Multicamadas (do inglês, <i>Multilayer Perceptron</i> ) (MLP) e LS em função da Relação Sinal-Ruído (do inglês, <i>Signal-to-Noise Ratio</i> ) (SNR). . . . .	129
Figura 31	– Recall e F1-score da classe de ilhamento em função da SNR. . . . .	130
Figura 32	– $P_{miss}^I$ e $P_{FA}^I$ em função da SNR. . . . .	131
Figura 33	– Recall e F1-score da classe Falta em função da SNR.	132

## LISTA DE TABELAS

Tabela 1 – Padrões para DI . . . . .	24
Tabela 2 – Comparação geral (tendências) dos Métodos de Detecção de Ilhamentos (MDIs). . . . .	28
Tabela 3 – Resumo comparativo de desempenho. . . . .	37
Tabela 4 – Visão geral dos cenários de simulação. . . . .	49
Tabela 5 – Perfis de potência usados na análise determinística (grade $3 \times 3$ ). . . . .	51
Tabela 6 – Dimensões dos arquivos *.mat gerados. . . . .	54
Tabela 7 – Visão geral dos conjuntos de dados. . . . .	55
Tabela 8 – Parâmetros efetivamente usados na execução. . . . .	87
Tabela 9 – Resumo de contagens por janelamento e classe. . . . .	88
Tabela 10 – Estratificação e balanceamento por conjunto. . . . .	89
Tabela 11 – Cenários de SNR por bloco de análise. . . . .	89
Tabela 12 – Energia explicada pelos componentes principais ( $\eta_j$ ) em porcentagem. . . . .	91
Tabela 13 – Acurácia média (%) $\pm$ desvio-padrão em validação cruzada 10-fold por componente principal (ensaio preliminar). . . . .	94
Tabela 14 – Poder discriminativo dos picos por ordem (FDR bruta e normalizada). . . . .	97
Tabela 15 – Importância normalizada dos atributos na Análise Discriminante Linear (do inglês, <i>Linear Discriminant Analysis</i> ) (LDA). . . . .	101
Tabela 16 – Atrasos selecionados por ordem (picos de FDR) em $c_1$ . . . . .	102
Tabela 17 – Acurácia (%) na validação cruzada em função de $K$ ( $B = 10$ atributos). . . . .	103
Tabela 18 – Métricas por classe em $K = 5$ (valores em %). . . . .	105
Tabela 19 – Mapeamento das posições do conjunto de atributos $r[\tau]$ . . . . .	106
Tabela 20 – Melhor subconjunto $\mathcal{S}_{B^*}^*$ por cardinalidade e desempenho em validação cruzada ( $K^* = 5$ ). . . . .	107
Tabela 21 – Arquiteturas MLP selecionadas por cardinalidade $B^*$ . . . . .	109
Tabela 22 – Desempenho em validação cruzada das arquiteturas MLP( $B^*, n_h^*$ ) reavaliadas na Rodada 4. . . . .	111
Tabela 23 – Contagem de janelas por classe nos conjuntos de treino e teste (janela deslizante). . . . .	118
Tabela 24 – Acurácia global da MLP em treino e teste . . . . .	119
Tabela 25 – Métricas por classe da MLP(6,2,3) no teste. . . . .	119
Tabela 26 – Acurácia global do classificador LS em treino e teste (janela deslizante). . . . .	121

Tabela 27 – Métricas por classe do LS no teste . . . . .	121
Tabela 28 – Desempenho da MLP na validação independente na ZND.	125
Tabela 29 – Desempenho do classificador por mínimos quadrados (LS) na validação independente na Zona de Não-Detecção (ZND) (SNR=70 dB). . . . .	127
Tabela 30 – Desempenho dos classificadores MLP e LS em três níveis de SNR (sub-janela 3). . . . .	131
Tabela 31 – Número aproximado de operações matemáticas na fase operacional do MDI por sub-janela de dois ciclos, em função do comprimento $L$ da janela. . . . .	139

## LISTA DE ABREVIATURAS E SIGLAS

<b>ANFIS</b>	Sistemas Adaptativos de Inferência Neuro-Fuzzy (do inglês, <i>Adaptative Neuro-Fuzzy Inference Systems</i> )
<b>AWGN</b>	Ruído Branco Gaussiano Aditivo (do inglês, <i>Additive White Gaussian Noise</i> )
<b>CNN</b>	Redes Neurais Convolucionais (do inglês, <i>Convolutional Neural Networks</i> )
<b>DI</b>	Detecção de Ilhamento
<b>DL</b>	Aprendizado Profundo, (do inglês, <i>Deep Learning</i> )
<b>DNN</b>	Redes Neurais Profundas (do inglês, <i>Deep Neural Networks</i> )
<b>DSP</b>	Processamento Digital de Sinais (do inglês, <i>Digital Signal Processing</i> )
<b>EDP</b>	Espaço de Desbalanços de Potências
<b>ELM</b>	Máquina de Aprendizado Extremo (do inglês, <i>Extreme Learning Machine</i> )
<b>EMD</b>	Decomposição de Modo Empírica (do inglês, <i>Empirical Mode Decomposition</i> )
<b>FDR</b>	Razão Discriminante de Fisher (do inglês, <i>Fisher's Discriminant Ratio</i> )
<b>FIR</b>	Resposta ao Impulso Finita (do inglês, <i>Finite Impulse Response</i> )
<b>GD</b>	Geração Distribuída
<b>HIL</b>	<i>Hardware-in-the-Loop</i>
<b>HHT</b>	Transformada de Hilbert-Huang (do inglês, <i>Hilbert-Huang Transform</i> )
<b>HOS</b>	Estatísticas de Ordem Superior (do inglês, <i>High-Order Statistics</i> )
<b>HSD</b>	Diferença Significativa Honesta (do inglês, <i>Honestly Significant Difference</i> )
<b>IC</b>	Inteligência Computacional
<b>IEEE</b>	Instituto de Engenheiros Eletricistas e Eletrônicos (do inglês, <i>Institute of Electrical and Electronics Engineers</i> )
<b>IGBT</b>	Transistor Bipolar de Porta Isolada (do inglês, <i>Insulated-gate bipolar transistor</i> )
<b>ITD</b>	Decomposição Intrínseca do Tempo (do inglês, <i>Intrinsic Time Decomposition</i> )
<b>LDA</b>	Análise Discriminante Linear (do inglês, <i>Linear Discriminant Analysis</i> )
<b>LS</b>	Mínimos Quadrados (do inglês, <i>Least Squares</i> )
<b>LSTM</b>	Rede de Memória de Curto Longo Prazo (do inglês, <i>Long Short-Term Memory Network</i> )

<b>MC</b>	Matriz de Confusão
<b>MDI</b>	Métodos de Detecção de Ilhamento
<b>MLP</b>	Perceptron Multicamadas (do inglês, <i>Multilayer Perceptron</i> )
<b>MPPT</b>	Rastreador de Ponto de Máxima Potência (do inglês, <i>Maximum Power Point Tracking</i> )
<b>MSE</b>	Erro Médio Quadrático (do inglês, <i>Mean Squared Error</i> )
<b>NPC</b>	<i>Neutral Point Clamped</i>
<b>PAC</b>	Ponto de Acoplamento Comum
<b>PCA</b>	Análise de Componentes Principais (do inglês, <i>Principal Component Analysis</i> )
<b>PLL</b>	Circuito de Sincronismo (do inglês, <i>Phase-Locked Loop</i> )
<b>PV</b>	Fotovoltaico (do inglês, <i>Photovoltaic</i> )
<b>PWM</b>	Modulação por Largura de Pulso (do inglês, <i>Pulse Width Modulation</i> )
<b>QEE</b>	Qualidade da Energia Elétrica
<b>ROCOF</b>	Taxa de Variação de Frequência (do inglês, <i>Rate of Change of Frequency</i> )
<b>SCR</b>	Razão de Curto-Circuito (do inglês, <i>Short-Circuit Ratio</i> )
<b>SEP</b>	Sistema Elétrico de Potência
<b>SNR</b>	Relação Sinal-Ruído (do inglês, <i>Signal-to-Noise Ratio</i> )
<b>SRF-PLL</b>	Circuito PLL Baseado no Sistema de Coordenadas Síncrono (do inglês, <i>Synchronous Reference Frame Phase-Locked Loop</i> )
<b>SVD</b>	Decomposição em Valores Singulares (do inglês, <i>Singular Value Decomposition</i> )
<b>SVM</b>	Máquinas de Vetores de Suporte (do inglês, <i>Support Vector Machines</i> )
<b>VMD</b>	Decomposição Variacional de Modos (do inglês, <i>Variational Mode Decomposition</i> )
<b>ZND</b>	Zona de Não-Detecção

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>19</b>
1.1	CONTEXTUALIZAÇÃO ENERGÉTICA E CLIMÁTICA	19
1.2	BENEFÍCIOS E DESAFIOS DA GERAÇÃO DISTRIBUÍDA	19
1.3	O PROBLEMA DO ILHAMENTO NÃO INTENCIONAL	20
1.4	MÉTODOS DE DETECÇÃO DE ILHAMENTO E SUAS LIMITAÇÕES	20
1.5	PROPOSTA, OBJETIVOS E CONTRIBUIÇÕES DO TRABALHO	21
1.6	ORGANIZAÇÃO DA TESE	22
<b>2</b>	<b>REVISÃO BIBLIOGRÁFICA</b>	<b>23</b>
2.1	PANORAMA BIBLIOMÉTRICO E EVIDÊNCIA DE INTERESSE RECENTE	23
2.2	PADRÕES, SISTEMAS E CONDIÇÕES DE TESTE PARA ILHAMENTO	23
2.2.1	<b>Padrões e requisitos normativos (anti-ilhamento)</b>	27
2.2.2	<b>Critérios de avaliação e casos críticos</b>	27
2.2.2.1	<i>Zona de não detecção (ZND)</i>	27
2.2.2.2	<i>Porcentagem de acionamento por ruído e condições de medição</i>	27
2.2.2.3	<i>Aplicabilidade em sistemas com múltiplas GD e efeitos na microrrede</i>	28
2.3	CLASSIFICAÇÃO DOS MÉTODOS DE DETECÇÃO DE ILHAMENTO	28
2.4	REVISÃO DOS MÉTODOS MODERNOS	29
2.4.1	<b>Abordagens baseadas em processamento de sinais</b>	29
2.4.1.1	<i>Transformadas temporais e espectrais</i>	30
2.4.1.2	<i>Técnicas de decomposição adaptativa</i>	30
2.4.1.3	<i>Análise estatística, filtragem avançada e descritores não lineares</i>	30
2.4.1.4	<i>Análise multivariada e medidas espalhadas geograficamente</i>	31
2.4.2	<b>Abordagens passivas baseadas em inteligência computacional</b>	31
2.4.2.1	<i>Aprendizado supervisionado</i>	31
2.4.2.2	<i>Árvores de decisão, Random Forest e gradiente boosting</i>	32
2.4.2.3	<i>Aprendizado não-supervisionado</i>	32
2.4.2.4	<i>Redes neurais e Deep Learning</i>	32
2.4.2.5	<i>Seleção de atributos e parcimônia</i>	32
2.5	VANTAGENS, DESVANTAGENS E LIMITAÇÕES OBSERVADAS	33
2.5.1	<b>Implementabilidade e custo do bloco de representação/extração</b>	33
2.5.1.1	<i>Transformadas tempo–frequência e wavelets: poder descritivo vs. custo</i>	33

2.5.1.2	<i>Decomposições adaptativas/variacionais e índices: compactação vs. parametrização iterativa</i> . . . . .	33
2.5.1.3	<i>Operadores no tempo e morfologia: simplicidade operacional vs. tuning</i> . . . . .	33
2.5.1.4	<i>Fechamento</i> . . . . .	34
2.5.2	<b>Estratégia do decisor (limiares, regras e modelos aprendidos)</b>	34
2.5.2.1	<i>Índices determinísticos e critérios por resíduo/assinaturas (fases, frequência e harmônicos)</i> . . . . .	34
2.5.2.2	<i>Supervisionados clássicos e decisão modular</i> . . . . .	34
2.5.2.3	<i>Ensembles/boosting e seleção explícita de atributos</i> . . . . .	34
2.5.2.4	<i>Aprendizado profundo (representações intermediárias e sequências)</i> . . . . .	35
2.5.2.5	<i>Não supervisionado (agrupamento) como caso especial</i> . . . . .	35
2.5.3	<b>Qualidade de evidência e limitações recorrentes (ZND, ruído, confundidores, generalização)</b> . . . . .	35
2.6	COMPARAÇÃO DE DESEMPENHO ENTRE ABORDAGENS . . . . .	36
2.7	LACUNAS RECORRENTES E REQUISITOS DE PROJETO . . . . .	41
2.7.1	<b>Lacunas recorrentes e requisitos derivados</b> . . . . .	41
2.7.2	<b>Síntese operacional dos requisitos</b> . . . . .	42
2.8	CONCLUSÕES PARCIAIS DO CAPÍTULO . . . . .	43
<b>3</b>	<b>AMBIENTE EXPERIMENTAL E CONJUNTOS DE DADOS</b> . . . . .	<b>44</b>
3.1	MODELAGEM DO SISTEMA DE POTÊNCIA . . . . .	44
3.1.1	<b>Arquitetura geral do sistema</b> . . . . .	45
3.1.2	<b>Geração fotovoltaica</b> . . . . .	46
3.1.3	<b>Conversor CC-CA e Filtro de Saída</b> . . . . .	47
3.1.4	<b>Transformadores e Sistema de Distribuição</b> . . . . .	47
3.1.5	<b>Cargas</b> . . . . .	47
3.1.6	<b>Estrutura de Controle Digital</b> . . . . .	48
3.1.7	<b>Cenários de Falta e Ilhamento</b> . . . . .	48
3.2	CENÁRIOS DE SIMULAÇÃO . . . . .	49
3.2.1	<b>Estratégia geral</b> . . . . .	49
3.2.2	<b>Variabilidade de potência ativa e reativa</b> . . . . .	49
3.2.2.1	<i>Formalização do domínio de operação na ZND</i> . . . . .	51
3.2.3	<b>Modos de operação</b> . . . . .	53
3.2.4	<b>Tamanho e armazenamento do conjunto de dados</b> . . . . .	53
3.2.4.1	<i>Reprodutibilidade</i> . . . . .	54
3.3	DIVISÃO DOS DADOS . . . . .	54
3.3.1	<b>Conjuntos para otimização</b> . . . . .	55
3.3.2	<b>Conjuntos para treinamento e teste</b> . . . . .	55

3.3.3	<b>Validação independente</b> . . . . .	56
<b>4</b>	<b>METODOLOGIA</b> . . . . .	<b>57</b>
4.1	ABORDAGENS DE PRÉ-PROCESSAMENTO E ANÁLISE DE SI- NAIS . . . . .	57
4.1.1	<b>Fluxo de processamento para janela fixa</b> . . . . .	58
4.1.2	<b>Fluxo de processamento para janela deslizante</b> . . . . .	60
4.2	TRANSFORMAÇÃO PCA . . . . .	61
4.2.1	<b>Fundamentação</b> . . . . .	62
4.2.2	<b>Estratégia centrada no ilhamento</b> . . . . .	62
4.2.3	<b>Critério para escolha da dimensionalidade</b> . . . . .	63
4.3	EXTRAÇÃO DE ATRIBUTOS VIA ESTATÍSTICAS DE ORDEM SUPERIOR . . . . .	64
4.3.1	<b>Sinal de entrada e convenções</b> . . . . .	64
4.3.2	<b>Definições e estimadores adotados</b> . . . . .	65
4.3.3	<b>Arranjo de saída e notação</b> . . . . .	65
4.3.4	<b>Propriedades úteis ao problema</b> . . . . .	66
4.3.5	<b>Complexidade computacional</b> . . . . .	67
4.3.6	<b>Resumo operacional</b> . . . . .	67
4.4	RANQUEAMENTO E SELEÇÃO DE CARACTERÍSTICAS . . . . .	67
4.4.1	<b>Rearranjo dos cumulantes</b> . . . . .	67
4.4.2	<b>FDR multiclasse</b> . . . . .	68
4.4.3	<b>Exploração de simetria em <math>\kappa_2</math></b> . . . . .	68
4.4.4	<b>Seleção por picos locais e concatenação final</b> . . . . .	68
4.4.5	<b>Viabilidade em Tempo Real e Complexidade Online</b> . . . . .	70
4.5	MAPA DE ATRIBUTOS NO INTERIOR DA ZND . . . . .	70
4.6	CLASSIFICADORES E OTIMIZAÇÃO . . . . .	71
4.6.1	<b>Preparação dos dados e alvos</b> . . . . .	71
4.6.2	<b>Arquiteturas consideradas</b> . . . . .	72
4.6.2.1	<i>LS multiclasse via pseudoinversa.</i> . . . . .	73
4.6.2.2	<i>MLP com uma camada escondida.</i> . . . . .	73
4.6.2.2.1	Arquitetura e ativações. . . . .	73
4.6.2.2.2	Treinamento por Levenberg–Marquardt. . . . .	74
4.6.2.2.3	Predição e decisão. . . . .	74
4.6.3	<b>Modelo de decisão no espaço de atributos</b> . . . . .	75
4.6.4	<b>Protocolo de Otimização</b> . . . . .	75
4.6.4.1	<i>Validação Cruzada</i> . . . . .	76
4.6.4.2	<i>Atributos</i> . . . . .	76
4.6.4.3	<i>Camada escondida</i> . . . . .	78
4.6.4.4	<i>Arquiteturas e vetor de características finais</i> . . . . .	78

4.6.5	<b>Protocolo de Validação e Controle de Viés</b> . . . . .	79
4.6.5.1	<i>Treinamento Final e Avaliação de Desempenho</i> . . . . .	80
4.7	<b>FLUXO EM EXECUÇÃO</b> . . . . .	81
<b>5</b>	<b>RESULTADOS</b> . . . . .	<b>84</b>
5.1	<b>CONFIGURAÇÃO EXPERIMENTAL E REPRODUTIBILIDADE</b> . . . . .	84
5.1.1	<b>Dados e janelamento</b> . . . . .	84
5.1.2	<b>Reamostragem</b> . . . . .	84
5.1.3	<b>Ruído e normalização</b> . . . . .	86
5.1.4	<b>Projeção PCA</b> . . . . .	87
5.1.5	<b>Atributos</b> . . . . .	87
5.1.6	<b>Partição e decisão</b> . . . . .	87
5.1.7	<b>Implementação e organização dos scripts</b> . . . . .	87
5.2	<b>CARACTERIZAÇÃO DO CONJUNTO</b> . . . . .	88
5.2.1	<b>Composição e Estratificação</b> . . . . .	88
5.2.2	<b>Níveis de ruído (SNR) avaliados</b> . . . . .	89
5.2.3	<b>Modos de falta e checagens de aleatoriedade</b> . . . . .	89
5.3	<b>VALIDAÇÃO DA PROJEÇÃO FIXA (RBAR)</b> . . . . .	89
5.3.1	<b>Erro de ortogonalidade de RBAR</b> . . . . .	90
5.3.2	<b>Energia explicada por componente</b> . . . . .	91
5.3.3	<b>Distribuição de <math>\eta_1</math> por classe</b> . . . . .	92
5.3.4	<b>Separabilidade por componente</b> . . . . .	93
5.4	<b>SELEÇÃO DE ATRIBUTOS POR FDR</b> . . . . .	94
5.4.1	<b>Mapas de FDR e leitura inicial</b> . . . . .	94
5.4.1.1	<i>Poder discriminativo geral</i> . . . . .	96
5.4.1.2	<i>Pares 2D mais discriminativos</i> . . . . .	98
5.4.1.3	<i>Projeção linear e contribuições individuais</i> . . . . .	98
5.4.1.4	<i>Atrasos em amostras, tempo e ângulo</i> . . . . .	101
5.5	<b>OTIMIZAÇÃO EM JANELA FIXA</b> . . . . .	102
5.5.1	<b>Rodada 1 — escolha do tamanho do particionamento</b> . . . . .	103
5.5.1.1	<i>Matrizes de confusão e métricas por classe</i> . . . . .	104
5.5.2	<b>Rodada 2 - melhor subconjunto para cada <math>B^*</math></b> . . . . .	105
5.5.2.1	<i>Resultados por cardinalidade</i> . . . . .	106
5.5.3	<b>Rodada 3 — ajuste de arquitetura da MLP</b> . . . . .	108
5.5.4	<b>Rodada 4 — consolidação da arquitetura</b> . . . . .	110
5.5.4.1	<i>Visualização do espaço de atributos da arquitetura final</i> . . . . .	112
5.6	<b>TREINAMENTO EM JANELA DESLIZANTE</b> . . . . .	114
5.6.1	<b>Formação do conjunto em janela deslizante</b> . . . . .	115
5.6.2	<b>Projeção PCA global e cálculo de cumulantes por janela</b> . . . . .	116
5.6.3	<b>Vetor de atributos e treinamento <i>hold-out</i> 70/30</b> . . . . .	117

5.6.4	<b>Desempenho da MLP (6,2,3) em janela deslizante</b> . . . . .	118
5.6.4.1	<i>Matrizes de confusão e métricas por classe</i> . . . . .	119
5.6.5	<b>Comparação com o classificador LS</b> . . . . .	120
5.6.6	<b>Síntese e preparação para a validação</b> . . . . .	122
5.7	VALIDAÇÃO INDEPENDENTE . . . . .	122
5.7.1	<b>Conjuntos de validação e protocolo experimental</b> . . . . .	123
5.7.2	<b>Resultados para SNR de 70 dB na ZND</b> . . . . .	124
5.7.3	<b>Influência do nível de ruído (30–70 dB)</b> . . . . .	127
5.7.3.1	<i>Interpretação geométrica da validação fora da grade</i> . . . . .	133
5.7.3.2	<i>Extensão qualitativa além da ZND e limitações</i> . . . . .	133
5.7.4	<b>Análise comparativa entre MLP e LS</b> . . . . .	134
5.7.5	<b>Síntese da validação independente</b> . . . . .	136
5.8	COMPLEXIDADE COMPUTACIONAL E CONSIDERAÇÕES DE TEMPO DE EXECUÇÃO . . . . .	137
5.9	ANÁLISE DE ERROS, INCERTEZA ESTATÍSTICA E CASOS LIMITE . . . . .	141
5.10	CONCLUSÕES PARCIAIS DO CAPÍTULO . . . . .	144
<b>6</b>	<b>CONCLUSÃO</b> . . . . .	<b>149</b>
6.1	SÍNTESE DO TRABALHO E RESPOSTA AOS OBJETIVOS . . . . .	149
6.2	PRINCIPAIS CONTRIBUIÇÕES CIENTÍFICAS E TÉCNICAS . . . . .	149
6.2.1	<b>Contribuições metodológicas</b> . . . . .	150
6.2.2	<b>Contribuições experimentais</b> . . . . .	150
6.2.3	<b>Contribuições em decisão, comparação de modelos e modos de operação</b> . . . . .	150
6.2.4	<b>Contribuições em implementabilidade e transferibilidade</b> . . . . .	151
6.3	RESULTADOS CONSOLIDADOS E IMPLICAÇÕES PARA PROTEÇÃO . . . . .	151
6.3.1	<b>Desempenho em classificação e seletividade</b> . . . . .	151
6.3.2	<b>Desempenho na ZND e tempo de detecção</b> . . . . .	152
6.3.3	<b>Robustez a ruído (SNR) e estabilidade operacional</b> . . . . .	152
6.3.4	<b>Síntese: por que o método funciona bem</b> . . . . .	152
6.4	LIMITAÇÕES . . . . .	152
6.5	TRABALHOS FUTUROS . . . . .	153
6.6	CONSIDERAÇÕES FINAIS . . . . .	154
	<b>REFERÊNCIAS</b> . . . . .	<b>155</b>

## 1 INTRODUÇÃO

### 1.1 CONTEXTUALIZAÇÃO ENERGÉTICA E CLIMÁTICA

A mitigação das mudanças climáticas e a necessidade de reduzir emissões de gases de efeito estufa têm acelerado uma transição energética global baseada na descarbonização e na eletrificação de usos finais [1]. No setor elétrico, esse movimento se materializa na crescente participação de fontes renováveis e na modernização da infraestrutura por meio de dois vetores complementares: a descentralização da geração (com a inserção de recursos energéticos distribuídos) e a digitalização das redes, com maior instrumentação, automação e inteligência embarcada.

Entre as tecnologias que mais catalisaram essa transformação, a geração solar fotovoltaica destaca-se por sua queda de custos, modularidade e facilidade de implantação, impulsionando a disseminação da Geração Distribuída (GD) em diferentes escalas. No Brasil, o crescimento do segmento é evidenciado por sua dimensão atual: dados setoriais recentes indicam dezenas de gigawatts de potência instalada e milhões de unidades consumidoras com sistemas de GD em operação, consolidando a fotovoltaica como um dos pilares da expansão renovável na última década [2].

### 1.2 BENEFÍCIOS E DESAFIOS DA GERAÇÃO DISTRIBUÍDA

No contexto brasileiro, a GD ganhou tração a partir de marcos regulatórios que viabilizaram o acesso de micro e minigeradores aos sistemas de distribuição, evoluindo posteriormente para um arcabouço legal específico. Em particular, a REN ANEEL nº 482/2012 instituiu as condições para micro e minigeração distribuída e o Sistema de Compensação de Energia Elétrica (SCEE), com aprimoramentos relevantes introduzidos pela REN nº 687/2015; mais recentemente, a Lei nº 14.300/2022 consolidou o marco legal da micro e minigeração distribuída e do SCEE, motivando adequações regulatórias subsequentes [3, 4, 5, 6].

Além de favorecer a diversificação da matriz e a redução de perdas associadas à transmissão, a GD contribui para a democratização do acesso a tecnologias renováveis e para o estímulo à eficiência no consumo, com efeitos positivos também sobre emprego e investimentos na cadeia produtiva. Segundo a ABSOLAR, a fonte solar já alcançou 62,3 GW de potência instalada no Brasil, e a geração distribuída contabilizava 3.838.765 sistemas em operação (dados atualizados em 11 nov. 2025) [2].

Apesar desses benefícios, a inserção massiva de geração baseada em inversores em redes originalmente projetadas para fluxo unidirecional de potência impõe desafios técnicos relevantes. Em particular, a GD pode alterar perfis de tensão, promover fluxos reversos e impactar a coordenação de proteções. Em cenários com elevada

penetração, surgem ainda novas exigências de seletividade e confiabilidade: a proteção deve distinguir, de forma rápida e robusta, eventos críticos (como o ilhamento) de distúrbios operacionais comuns (como faltas e variações transitórias), evitando tanto a não atuação quanto desligamentos indevidos.

### 1.3 O PROBLEMA DO ILHAMENTO NÃO INTENCIONAL

Denomina-se ilhamento não intencional a condição em que uma porção da rede de distribuição se isola do sistema elétrico principal, mas permanece energizada localmente por uma ou mais unidades de GD. Esse fenômeno é indesejável por razões de segurança e de integridade operacional: pode manter trechos supostamente desenergizados sob tensão, comprometer o religamento automático e degradar os níveis de qualidade de energia na ilha.

Em função desses riscos, normas e práticas de interconexão exigem que a GD cesse a energização do trecho ilhado em um intervalo curto (tipicamente em até alguns segundos). No contexto normativo, guias de aplicação do IEEE Std 1547 registram a diretriz de que ilhas não intencionais sejam detectadas e que o recurso distribuído cesse a energização do trecho em até dois segundos [7].

O desafio se intensifica quando o desequilíbrio entre a potência gerada e a demanda local é pequeno; nessa situação, grandezas elétricas usuais (tensão e frequência) podem permanecer próximas às faixas nominais após a separação, caracterizando a Zona de Não-Detecção (ZND) de um método de detecção. Reduzir a ZND e, simultaneamente, manter seletividade frente a distúrbios não relacionados ao ilhamento são requisitos centrais para Métodos de Detecção de Ilhamentos (MDIs) aplicáveis em redes modernas.

### 1.4 MÉTODOS DE DETECÇÃO DE ILHAMENTO E SUAS LIMITAÇÕES

Os métodos de detecção de ilhamento são tradicionalmente classificados em passivos, ativos, híbridos e remotos. Abordagens passivas monitoram variáveis elétricas no Ponto de Acoplamento Comum (PAC) (por exemplo, tensão, frequência e Taxa de Variação de Frequência (do inglês, *Rate of Change of Frequency*) (ROCOF)) e as comparam a limiares. São atrativas por sua simplicidade e baixo custo; porém podem apresentar ZND relevante e suscetibilidade a falso disparo, sobretudo em regimes de baixo desequilíbrio de potência e na presença de faltas e transitórios.

Métodos ativos, por sua vez, injetam pequenas perturbações e exploram a resposta do sistema para revelar a desconexão da rede. Embora reduzam a ZND, podem introduzir distúrbios na Qualidade da Energia Elétrica (QEE) e demandar cuidado na coordenação com múltiplas unidades de GD. Técnicas híbridas buscam combinar as

vantagens de abordagens passivas e ativas, enquanto métodos remotos (baseados em comunicação) podem atingir ZND nula, mas com custo e infraestrutura que frequentemente limitam sua aplicação em larga escala.

Nesse cenário, métodos modernos baseados em processamento de sinais e inteligência computacional têm recebido atenção por permitirem extrair assinaturas mais informativas de formas de onda medidas no PAC, potencialmente reduzindo a ZND sem perturbar a rede. Contudo, para serem transferíveis e implementáveis, tais métodos precisam equilibrar desempenho com parcimônia: (i) generalizar para perfis de carga não vistos no projeto; (ii) manter robustez a ruído e incertezas de medição; (iii) distinguir ilhamento de eventos como faltas e (iv) operar com baixa complexidade computacional em janelas curtas.

## 1.5 PROPOSTA, OBJETIVOS E CONTRIBUIÇÕES DO TRABALHO

Diante desse panorama, esta tese propõe um MDI estritamente passivo, orientado por processamento estatístico de sinais e reconhecimento de padrões, com ênfase em seletividade e implementabilidade. O método parte de medições trifásicas no PAC e organiza o processo em duas fases complementares:

- **Calibração e consolidação offline:** estima-se uma projeção linear fixa por Análise de Componentes Principais (do inglês, *Principal Component Analysis*) (PCA) (orientada a realçar a estrutura do transitório associada ao ilhamento), ranqueiam-se e selecionam-se atributos de forma parcimoniosa via Razão Discriminante de Fisher (do inglês, *Fisher's Discriminant Ratio*) (FDR) multiclasse, e treina-se o decisor final (incluindo comparação entre um decisor linear por mínimos quadrados e um decisor não linear do tipo Perceptron Multicamadas (do inglês, *Multilayer Perceptron*) (MLP)).
- **Execução online em janela deslizante:** aplica-se a normalização, projeta-se o sinal na base fixa e computam-se os cumulantes em janelas curtas, permitindo decisão rápida e compatível com os requisitos normativos.

Como diferencial metodológico, a tese combina: (i) projeção linear fixa para compactar a dinâmica trifásica e aumentar separabilidade, (ii) cumulantes como descritores do transitório e (iii) seleção parcimoniosa de atributos para reduzir risco de sobreajuste e facilitar implementação. Do ponto de vista experimental, constrói-se e organiza-se um banco de dados balanceado em três classes (nominal, ilhamento e falta), incluindo: (a) exploração sistemática de perfis de carga em uma grade regular, (b) validação independente em perfis fora da grade, (c) avaliação explícita sob condições associadas à ZND (baixo desequilíbrio de potência) e (d) varredura de Relação Sinal-Ruído (do inglês, *Signal-to-Noise Ratio*) (SNR), analisando robustez a ruído.

Na configuração proposta, a etapa de seleção visa produzir um vetor compacto de atributos (cumulantes de 2ª e 4ª ordens em atrasos não nulos), e o estudo inclui uma comparação sistemática entre decisores lineares e não lineares para detecção binária e discriminação multiclasse entre condições nominal, ilhamento e falta. Nessa análise de robustez, o ruído é varrido em uma faixa ampla de SNR (30–70 dB), de modo a caracterizar a degradação de desempenho em cenários realistas de medição.

Os resultados apresentados ao longo da tese corroboram a viabilidade da abordagem, bem como sua adequação aos requisitos de seletividade e de implementabilidade, com atuação rápida em janelamento curto e comportamento consistente em cenários críticos de ZND.

Além disso, a solução final preserva características desejáveis para a aplicação prática: número reduzido de atributos, projeção fixa e processamento estatístico simples, os quais são compatíveis com execução em tempo real e com a baixa demanda computacional.

## 1.6 ORGANIZAÇÃO DA TESE

O restante do texto está organizado da seguinte forma: o Capítulo 2 apresenta a fundamentação teórica e a revisão dos métodos de detecção de ilhamento; o Capítulo 3 descreve o sistema elétrico modelado, os cenários de simulação e a construção dos conjuntos de dados; o Capítulo 4 detalha a metodologia proposta (pré-processamento, projeção PCA, cumulantes, ranqueamento/seleção e classificadores); o Capítulo 5 discute os resultados experimentais, incluindo validação independente, robustez a ruído e análise de tempo de detecção; por fim, a Conclusão sintetiza as contribuições e aponta trabalhos futuros.

## 2 REVISÃO BIBLIOGRÁFICA

Este capítulo apresenta uma revisão bibliográfica sobre Detecção de Ilhamento (DI), com ênfase em métodos passivos modernos apoiados por Processamento Digital de Sinais (do inglês, *Digital Signal Processing*) (DSP) e Inteligência Computacional (IC). O objetivo é sintetizar o estado da arte, destacando os critérios de avaliação que mais condicionam o desempenho reportado e as lacunas que motivam o método desenvolvido nos Capítulos 3, 4 e 5.

O ilhamento ocorre quando uma porção do sistema permanece energizada pela GD após a desconexão do Sistema Elétrico de Potência (SEP), com impactos sobre segurança, coordenação de proteção e qualidade de energia no PAC. Por isso, normas de interconexão estabelecem limites de atuação e condições mínimas de ensaio para funções anti-ilhamento [7]. No recorte metodológico, a literatura distingue métodos passivos, ativos, híbridos e remotos; nesta tese, o foco recai sobre métodos passivos recentes (2018–2025), sobretudo quando buscam reduzir a sensibilidade a ruído, melhorar a seletividade frente a confundidores e ampliar a robustez perto da ZND sem introduzir perturbações deliberadas no sistema [8].

### 2.1 PANORAMA BIBLIOMÉTRICO E EVIDÊNCIA DE INTERESSE RECENTE

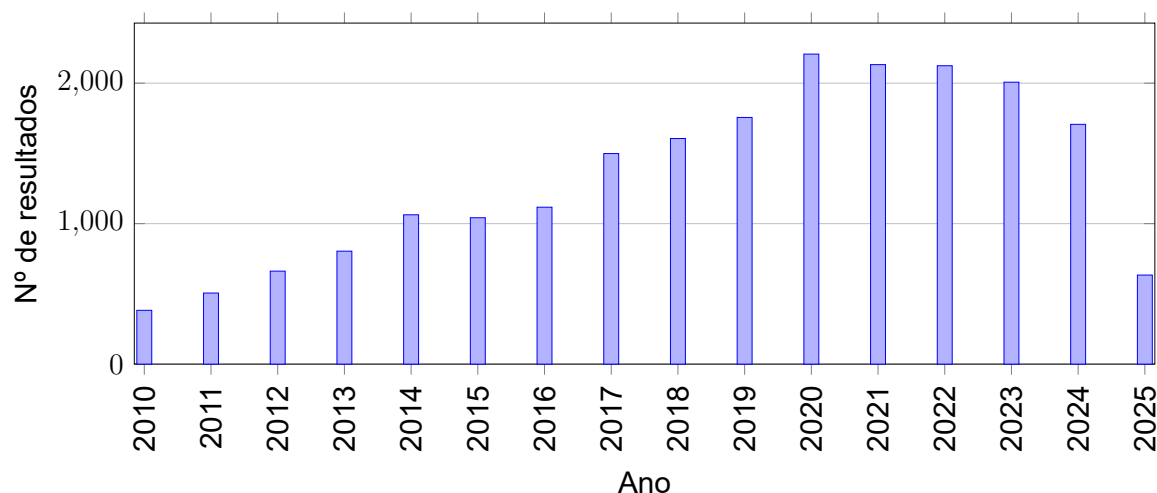
Uma sondagem bibliométrica não sistemática no Portal de Periódicos CAPES, com a consulta reproduzível `islanding AND detection`, reforça que a pesquisa em detecção de ilhamento permaneceu ativa ao longo da última década. As contagens anuais mostram crescimento acentuado até 2020 e posterior estabilização em patamar elevado até 2024; o valor de 2025 deve ser interpretado com cautela por provável parcialidade de indexação. A Figura 1 a seguir serve como evidência complementar de atualidade e é coerente com revisões recentes que apontam a persistência de desafios em seletividade, ruído e validação de MDIs passivos [9, 10].

### 2.2 PADRÕES, SISTEMAS E CONDIÇÕES DE TESTE PARA ILHAMENTO

A comparação entre MDIs depende fortemente do protocolo de ensaio. Topologia, interface da GD, desbalanço entre geração e carga e qualidade da medição alteram diretamente a assinatura elétrica observável; por isso, esta seção resume o arcabouço normativo, explicita um sistema genérico de testes e destaca os casos críticos que mais influenciam a interpretação dos resultados.

No contexto desta tese, o interesse recai sobre o ilhamento não intencional em microrredes, isto é, a permanência de uma porção do sistema energizada após a desconexão do SEP. Normas como IEEE 1547 e ABNT NBR IEC 62116 orientam

Figura 1 – Número anual de resultados recuperados no Portal CAPES para islanding AND detection (2010–2025).



Legenda: Contagens anuais por faixa  $[Y, Y+1)$ , exibida pelo Portal como  $Y-Y+1$ .

Fonte: Elaborado pelo autor a partir de consulta no Portal de Periódicos CAPES (CAFe), 2026.

limites de tensão e frequência, tempos máximos de desconexão e condições mínimas de ensaio, ao passo que cenários com múltiplas GDs exigem cautela adicional na extrapolação de resultados obtidos em configurações unitárias [7, 11, 12, 13]. Em termos conceituais, a microrrede pode operar conectada ou ilhada, mas o interesse aqui está na formação não planejada da ilha, conforme a representação simplificada da Figura 2 [14].

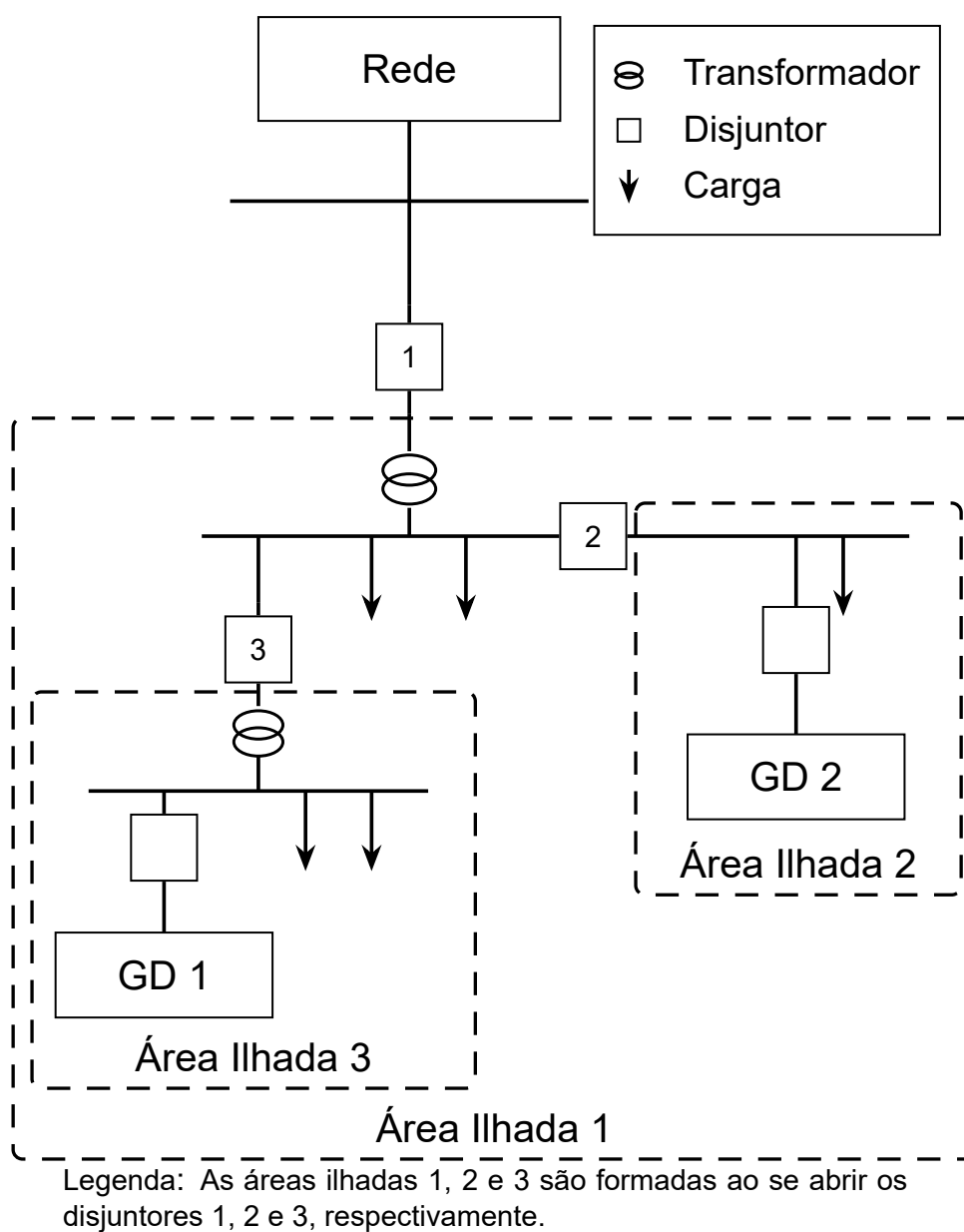
Ilhamentos não intencionais costumam decorrer de aberturas de dispositivos de proteção, contingências ou manobras, e por isso demandam atuação anti-ilhamento. A Tabela 1 resume os limites de operação e o tempo máximo de detecção adotados como referência pelas normas IEEE 1547 e ABNT NBR IEC 62116.

Tabela 1 – Padrões para DI

Parâmetro	IEEE 1547[12]	NBR-IEC 62116[11]
Tempo para Detecção	$t < 2s$	$t < 2s$
Variação de Frequência Permitida	$59,3 \leq f < 60,5$	$(f_0 - 1,5Hz) \leq f < (f_0 + 1,5Hz)$
Variação de Tensão Permitida	$88\% \leq V \leq 110\%$	$85\% \leq V \leq 110\%$
Fator de Qualidade	$Q_f = 1$	$Q_f = 1$

Embora esses documentos estabeleçam requisitos de desempenho, eles não impõem uma única topologia canônica para estudos comparativos. Por isso, a literatura frequentemente adota sistemas genéricos com equivalente simplificado da rede

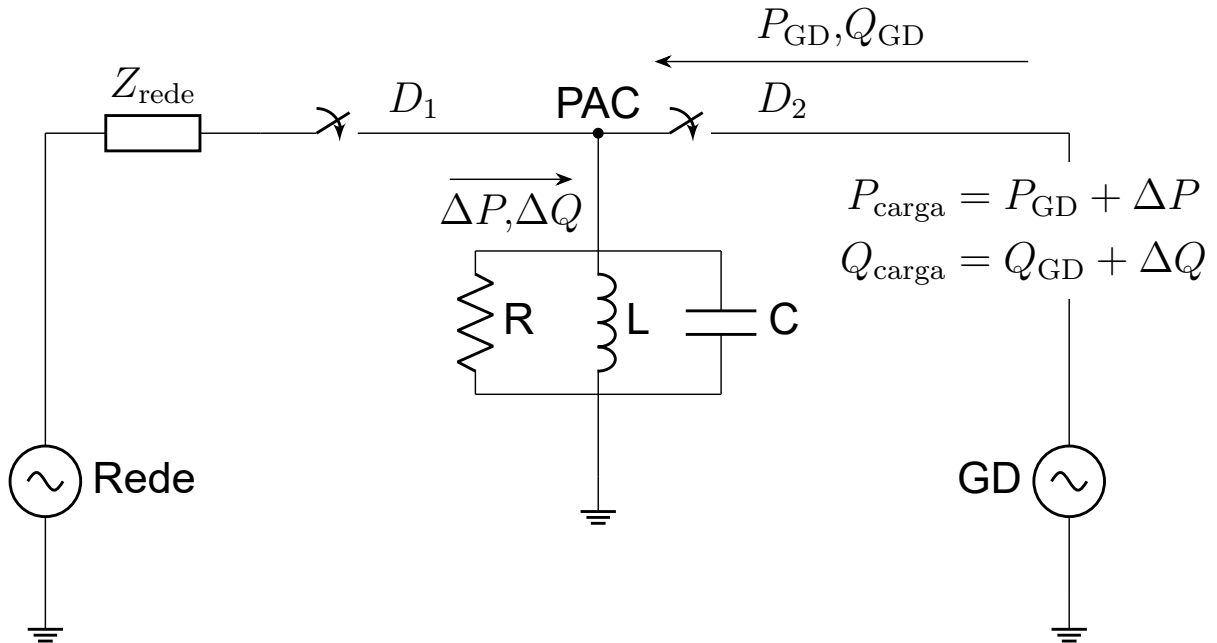
Figura 2 – Conceito Simples de Ilhamento.



Fonte: Adaptado de [8]

e carga RLC em paralelo ajustada próxima à ressonância, pois esse arranjo reproduz condições críticas associadas à ZND e permanece alinhado aos procedimentos de ensaio usuais [11, 15]. A Figura 3 resume esse arranjo de referência.

Figura 3 – Arranjo de teste no PAC com carga RLC e disjuntores D1/D2.



Legenda: Neste sistema a rede e a GD são modeladas como fontes senoidais e a carga ilhada como uma carga RLC em paralelo.

Como caso crítico, a carga local é usualmente representada por uma RLC paralela sintonizada à frequência da rede. Nessas condições, os parâmetros  $R$ ,  $L$  e  $C$  podem ser escritos pelas Equações (2.1)–(2.5) [8, 16].

$$R = \frac{V^2}{P_{carga}} \quad (2.1)$$

$$L = \frac{V^2}{2\pi f Q_f P_{carga}} \quad (2.2)$$

$$C = \frac{Q_f P_{carga}}{2\pi f V^2} \quad (2.3)$$

$$f = \frac{1}{2\pi\sqrt{LC}} \quad (2.4)$$

$$Q_f = R\sqrt{\frac{C}{L}} \quad (2.5)$$

### 2.2.1 Padrões e requisitos normativos (anti-ilhamento)

Normas de interconexão e procedimentos de ensaio convertem o problema de DI em requisitos verificáveis de segurança e desempenho. Em termos práticos, isso significa avaliar o atendimento a limites admissíveis de tensão e frequência, ao tempo máximo de desconexão e a condições críticas de balanço entre geração e carga local, normalmente com cargas ressonantes representativas [12, 7, 11, 8]. Para métodos passivos, esse enquadramento é particularmente relevante porque a busca por maior sensibilidade não pode ser dissociada de seletividade frente a distúrbios e de robustez sob incertezas de medição [17].

### 2.2.2 Critérios de avaliação e casos críticos

Além da conformidade normativa, a avaliação de um MDI deve refletir pior caso e operação realista. Neste capítulo, três eixos organizam a leitura comparativa: ZND, robustez a ruído/condições de medição e comportamento em sistemas com múltiplas GDs. Esses critérios condicionam a observabilidade do ilhamento, a seletividade frente a confundidores e a força de evidência dos resultados reportados [12, 7, 11].

#### 2.2.2.1 Zona de não detecção (ZND)

A ZND permanece como a referência central de pior caso para MDIs passivos. Ela corresponde ao conjunto de condições em que, após a desconexão do SEP, tensão e frequência no PAC permanecem por tempo suficiente dentro de limites aceitáveis, tipicamente quando o desbalanço entre geração e carga é pequeno e a carga local está próxima da ressonância [12, 7, 11, 15, 18]. Em trabalhos recentes, a redução aparente da ZND depende fortemente do protocolo de validação e do tipo de assinatura explorada, o que reforça a necessidade de relatar claramente criticidade, eventos confundidores e condições de ensaio [8, 19, 20].

#### 2.2.2.2 Porcentagem de acionamento por ruído e condições de medição

Em aplicações reais, medições no PAC sofrem influência de ruído, quantização, distorções harmônicas e limitações da cadeia de instrumentação. Por isso, além de tempo de detecção e ZND, torna-se relevante quantificar a propensão do método a disparar sob condições de não-ilhamento apenas por degradação da medição. Estudos recentes tratam explicitamente esse problema e mostram que a robustez a ruído deve ser avaliada de modo controlado, idealmente com varreduras de SNR e cenários que representem a variabilidade do sistema [21, 22].

### 2.2.2.3 Aplicabilidade em sistemas com múltiplas GD e efeitos na microrrede

Sistemas com múltiplas GDs conectadas ao mesmo PAC ou distribuídas ao longo da microrrede constituem outro caso crítico recorrente. Nesses cenários, a divisão de potência entre fontes e a ação combinada de controladores podem amortecer as assinaturas elétricas observáveis e alterar substancialmente o desempenho de detectores ajustados em configuração unitária [17, 13]. A literatura recente já traz propostas voltadas a esse contexto, tanto em recortes mais gerais quanto em aplicações a GD fotovoltaica trifásica e microrredes híbridas, o que reforça a necessidade de validação explícita sob condições multi-GD [23, 24, 25].

## 2.3 CLASSIFICAÇÃO DOS MÉTODOS DE DETECÇÃO DE ILHAMENTO

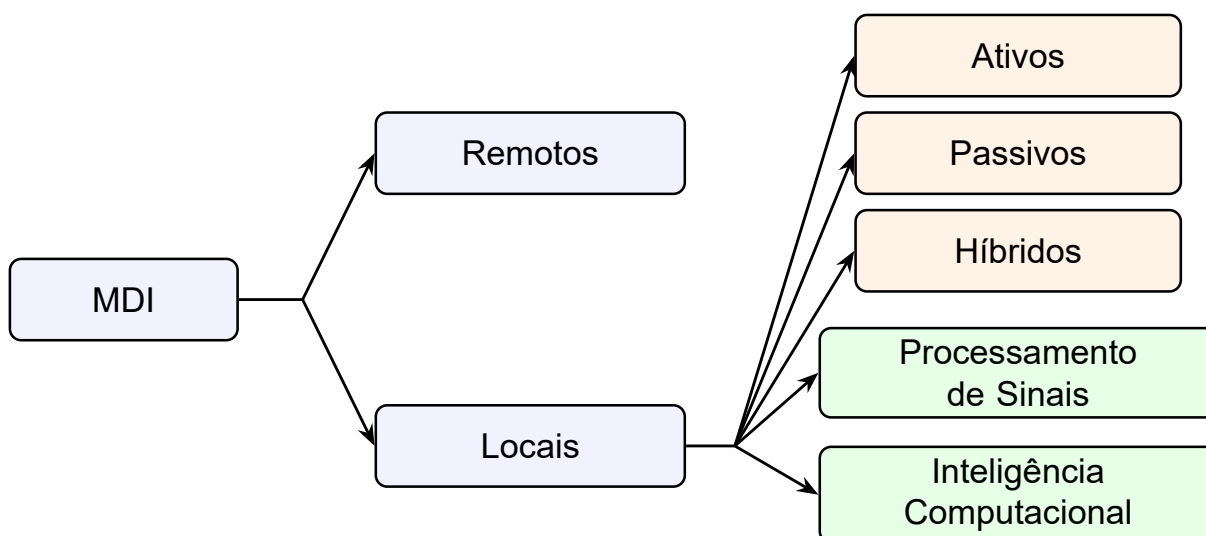
A DI em sistemas com GD é usualmente organizada em técnicas locais e remotas. Nas técnicas locais, a decisão decorre de medições no PAC ou na própria unidade geradora; nas remotas, depende de supervisão e comunicação externas [26, 8, 9]. Dentro do escopo local, a taxonomia clássica distingue métodos passivos, ativos e híbridos.

Nesta tese, a separação entre abordagens clássicas e modernas é tratada menos como corte temporal e mais como critério de projeto. Métodos clássicos operam predominantemente com limiares diretos de grandezas elétricas; métodos modernos procuram ampliar a evidência disponível com DSP e, em seguida, aprimorar a decisão com IC, sobretudo para lidar com ruído, confundidores e criticidade próxima à ZND [8, 9, 27]. Métodos ativos e híbridos podem reduzir a ZND, mas à custa de maior intrusividade e complexidade; métodos remotos tendem a oferecer alta confiabilidade, porém exigem infraestrutura dedicada [26, 17]. Por isso, o recorte principal deste capítulo recai sobre métodos passivos modernos. A Figura 4 e a Tabela 2 sintetizam esse enquadramento.

Tabela 2 – Comparação geral (tendências) dos MDIs.

<b>MDI</b>	<b>Princípio</b>	<b>ZND</b>	<b>Efeito na QEE</b>
Remoto	Comunicação/supervisão entre SEP e GD	Muito pequena	Nenhum
Ativo	Injeção de distúrbios controlados	Pequena	Degradação
Híbrido	Passivo + ativo	Muito pequena	Degradação
Passivo (limiares)	Monitoramento por limiares	Grande	Nenhum
Passivo (DSP)	Extração de características	Reduzida	Nenhum
Passivo (IC)	Classificação baseada em características	Reduzida	Nenhum

Figura 4 – Classificação geral dos MDI.



Legenda: Classificação dos MDIs em abordagens remotas e locais; e, no escopo local, em métodos passivos, ativos e híbridos. Também se destaca o recorte moderno de métodos passivos baseados em DSP e IC.

## 2.4 REVISÃO DOS MÉTODOS MODERNOS

A revisão a seguir concentra-se em métodos passivos modernos porque o desempenho de detectores convencionais por limiar continua limitado pelo compromisso entre sensibilidade, seletividade e robustez próximo à ZND. Em resposta a esse problema, a literatura recente converge para duas frentes complementares: extração de representações mais informativas por DSP e uso de modelos de decisão mais flexíveis por IC [8, 9, 17]. A exposição abaixo organiza os trabalhos por famílias de técnicas enfatizando princípios, contribuições recorrentes e limitações comuns.

### 2.4.1 Abordagens baseadas em processamento de sinais

Nas abordagens baseadas em processamento de sinais, o objetivo central é tornar mais separáveis os padrões de ilhamento e não-ilhamento por meio de representações no tempo, na frequência ou no domínio tempo–frequência. Em linhas gerais, a cadeia empregada na literatura envolve aquisição no PAC, pré-processamento, transformação ou decomposição do sinal e extração de atributos que alimentam um decisor por limiar ou por aprendizado [28, 29, 30]. Essa estratégia amplia a informação disponível em relação a métricas clássicas de  $V$  e  $f$ , mas introduz custos computacionais, dependência de hiperparâmetros e maior sensibilidade à qualidade da instrumentação [31, 22, 32, 33].

#### 2.4.1.1 *Transformadas temporais e espectrais*

Transformadas temporais, espectrais e tempo–frequência formam uma das famílias mais recorrentes de MDIs modernos. Em comum, essas abordagens buscam evidenciar transientes, conteúdo harmônico, componentes simétricas ou distribuições de energia por banda que tendem a ser pouco visíveis em proteções convencionais baseadas apenas em  $V$  e  $f$ . Nessa família aparecem soluções apoiadas em Fourier e janelas curtas, wavelets e medidas entrópicas, *S-transform* e variantes rápidas/esparsas, além de formulações paramétricas ou fasoriais subciclo [34, 35, 36, 27, 37, 38, 29, 28, 39, 40, 41, 10].

Em termos comparativos, o ganho dessa família está no maior poder descritivo e na possibilidade de alimentar classificadores com atributos mais informativos. O custo é a dependência de janelamento, escolhas de transformada e parâmetros internos, além de latência e esforço computacional nem sempre explicitados com clareza. Por isso, o ponto decisivo deixa de ser apenas a boa acurácia e passa a ser a relação entre separabilidade, robustez e implementabilidade.

#### 2.4.1.2 *Técnicas de decomposição adaptativa*

Técnicas de decomposição adaptativa procuram isolar modos oscilatórios relevantes sem assumir estacionariedade estrita do sinal. Nesse grupo situam-se abordagens baseadas em Transformada de Hilbert-Huang (do inglês, *Hilbert-Huang Transform*) (HHT), Decomposição de Modo Empírica (do inglês, *Empirical Mode Decomposition*) (EMD), Decomposição Variacional de Modos (do inglês, *Variational Mode Decomposition*) (VMD), Decomposição Intrínseca do Tempo (do inglês, *Intrinsic Time Decomposition*) (ITD) e combinações com índices energéticos, entropias, árvores por regras, agrupamento ou classificadores supervisionados [30, 31, 42, 43, 44, 45, 46, 8]. Em geral, essas propostas oferecem boa capacidade de adaptação a sinais não estacionários e costumam relatar ganhos em cenários críticos.

O principal limite dessa família está na parametrização e no custo do bloco de extração. Número de modos, critérios de parada, sensibilidade a ruído e latência computacional passam a influenciar fortemente a viabilidade prática, o que faz com que resultados promissores dependam de reporte claro de janelas, parâmetros e condições de teste.

#### 2.4.1.3 *Análise estatística, filtragem avançada e descritores não lineares*

Outra linha relevante reúne métodos que procuram robustez por meio de estatísticas compactas, filtragem avançada e descritores não lineares. Nessa categoria aparecem soluções baseadas em resíduos de reconstrução, análise morfológica, estimação fasorial/frequencial, filtros robustos e indicadores derivados de séries temporais, mui-

tas vezes orientados por simplicidade operacional e menor custo de implementação [47, 48, 49, 50, 20, 21, 51]. O mérito dessa família é oferecer cadeias mais leves e mais previsíveis computacionalmente; a limitação recorrente é que a separação entre ilhamento e certos confundidores pode depender fortemente do ajuste fino do índice e do protocolo de validação.

#### 2.4.1.4 *Análise multivariada e medidas espalhadas geograficamente*

Quando há disponibilidade de medições mais ricas ou distribuídas espacialmente, cresce a exploração de análise multivariada, sincrofasores e medidas espalhadas geograficamente. Essas abordagens buscam elevar a observabilidade do fenômeno por correlação entre múltiplos canais e por descritores espaciais ou morfológicos, como ilustram trabalhos com processamento multivariado, PMUs e arranjos distribuídos [32, 52, 53, 54, 55]. O ganho potencial é maior seletividade; o custo é a dependência de instrumentação, sincronismo e, por vezes, infraestrutura adicional.

#### 2.4.2 **Abordagens passivas baseadas em inteligência computacional**

Em métodos passivos baseados em IC, a fronteira de decisão deixa de depender exclusivamente de limiares fixos e passa a ser aprendida a partir de atributos físicos ou de representações derivadas do sinal. Na prática, a literatura combina extração por DSP com classificadores supervisionados, ensembles, aprendizado não supervisionado e arquiteturas profundas, buscando maior seletividade sob ruído e confundidores [26, 34, 27, 31]. O benefício é aumentar a flexibilidade da decisão; o custo é a dependência de dados, particionamento, balanceamento e documentação reproduzível do treinamento.

##### 2.4.2.1 *Aprendizado supervisionado*

O aprendizado supervisionado clássico domina grande parte da literatura recente. Nessa família aparecem soluções baseadas em Sistemas Adaptativos de Inferência Neuro-Fuzzy (do inglês, *Adaptative Neuro-Fuzzy Inference Systems*) (ANFIS), Máquinas de Vetores de Suporte (do inglês, *Support Vector Machines*) (SVM), Máquina de Aprendizado Extremo (do inglês, *Extreme Learning Machine*) (ELM), boosting e comitês, geralmente operando sobre atributos já selecionados por alguma etapa prévia de DSP [29, 56, 57, 58, 59, 19, 60]. O padrão recorrente é claro: quando a representação do sinal é informativa e o protocolo experimental é favorável, esses modelos podem alcançar alta acurácia e baixa latência; porém a evidência de generalização depende fortemente da cobertura de cenários, da transparência do particionamento e da presença de confundidores realmente desafiadores.

#### 2.4.2.2 *Árvores de decisão, Random Forest e gradiente boosting*

Árvores de decisão, *Random Forest* e gradiente boosting ocupam um espaço intermediário entre interpretabilidade e poder discriminativo. Na tradição de classificação por árvores, já consolidada em problemas de aprendizado de máquina [61], trabalhos recentes aplicam variantes modernas como CatBoost a conjuntos compactos de atributos elétricos, com resultados fortes quando o protocolo é bem controlado [62]. Ainda assim, o desempenho dessa família continua sensível à seleção de atributos, à composição dos cenários e à forma como a validação é estruturada.

#### 2.4.2.3 *Aprendizado não-supervisionado*

O aprendizado não supervisionado aparece com menor frequência, mas é metodologicamente interessante por dispensar rótulos completos. O exemplo mais representativo desta revisão usa agrupamento *K-means* sobre atributos derivados de EMD+HHT para separar ilhamento, faltas e chaveamentos [46]. A principal limitação é que a utilidade prática dessa estratégia ainda depende de evidência mais ampla sob ruído, mudanças de topologia e variações de controle.

#### 2.4.2.4 *Redes neurais e Deep Learning*

As redes neurais e o Aprendizado Profundo, (do inglês, *Deep Learning*) (DL) ampliam ainda mais a expressividade da decisão. Essa família inclui redes rasas sobre atributos construídos manualmente, Redes Neurais Convolucionais (do inglês, *Convolutional Neural Networks*)s (CNNs) sobre imagens ou sequências, Rede de Memória de Curto Longo Prazo (do inglês, *Long Short-Term Memory Network*)s (LSTMs) para dependências temporais e arquiteturas híbridas que combinam DSP com modelos profundos [63, 13, 64, 65, 66, 67, 27, 68]. Em geral, essas propostas mostram bom poder discriminativo e flexibilidade para cenários complexos, mas cobram um preço em necessidade de dados, custo de treinamento, risco de sobreajuste e menor transparência quanto ao papel do bloco de representação [33]. Assim, o ganho aparente do DL só é tecnicamente convincente quando acompanhado de protocolos robustos de validação e de reporte claro de latência e custo.

#### 2.4.2.5 *Seleção de atributos e parcimônia*

A seleção de atributos aparece como contrapeso importante à complexidade crescente das cadeias modernas. Em vez de maximizar o número de descritores, essa linha busca manter apenas o subconjunto mais informativo para preservar desempenho com menor custo e maior interpretabilidade, como exemplifica a combinação de seleção multiobjetivo com K-ELM em [69]. Em termos de projeto, a mensagem cen-

tral é simples: parcimônia bem justificada costuma ser mais valiosa do que riqueza descritiva pouco transparente.

## 2.5 VANTAGENS, DESVANTAGENS E LIMITAÇÕES OBSERVADAS

Esta seção consolida a leitura crítica da revisão anterior. O foco passa a ser a comparação entre famílias de métodos em torno de três perguntas: quão custoso é o bloco de representação/extração, como a decisão é construída e quão forte é a evidência experimental que sustenta o desempenho reportado.

### 2.5.1 Implementabilidade e custo do bloco de representação/extração

O custo do bloco de representação/extração é um dos fatores que mais separa métodos academicamente promissores de métodos efetivamente implementáveis. A distinção principal está entre cadeias compactas e previsíveis, baseadas em atributos simples, e cadeias mais ricas, porém mais pesadas, baseadas em transformadas, decomposições iterativas ou geração de representações intermediárias de alta dimensionalidade.

#### 2.5.1.1 *Transformadas tempo–frequência e wavelets: poder descritivo vs. custo*

Transformadas tempo–frequência e wavelets oferecem representações ricas e, em geral, elevam a separabilidade entre ilhamento e não-ilhamento. Esse ganho aparece em diferentes variantes de ST, SST, DWT/CWT e combinações com atributos derivados [28, 39, 30, 36, 35, 29, 63]. Em contrapartida, a latência mínima passa a depender mais fortemente de janelamento, e o custo do bloco de extração nem sempre é reportado de forma decomposta, o que dificulta avaliar sua viabilidade em IEDs e relés digitais.

#### 2.5.1.2 *Decomposições adaptativas/variacionais e índices: compactação vs. parametrização iterativa*

Decomposições adaptativas e variacionais conseguem capturar componentes relevantes de sinais não estacionários e costumam produzir bons índices discriminativos, como ilustram trabalhos com VMD, EMD, ITD, ANN e classificadores multiclasse [42, 43, 44, 45, 46, 13, 19, 31]. O preço dessa flexibilidade é a dependência de parametrização iterativa, critérios de parada e escolhas de modos, o que torna o custo fim-a-fim menos previsível e mais difícil de reproduzir entre estudos.

#### 2.5.1.3 *Operadores no tempo e morfologia: simplicidade operacional vs. tuning*

Operadores no tempo e técnicas morfológicas ocupam o extremo mais leve do espectro de implementação. Eles favorecem cadeias curtas e custo previsível, mas frequen-

temente dependem mais do ajuste fino de limiares e podem perder robustez quando eventos confundidores compartilham assinaturas semelhantes [50].

#### 2.5.1.4 *Fechamento*

Em síntese, o bloco de representação/extração acrescenta poder descritivo, mas também pode se tornar o principal custo do método. Essa tensão aparece tanto em cadeias baseadas em transformadas quanto em decomposições adaptativas, e precisa ser tratada explicitamente quando se discute implementabilidade [28, 39, 30, 43, 13, 19].

### 2.5.2 **Estratégia do decisor (limiares, regras e modelos aprendidos)**

Além do bloco de extração, a forma de decisão altera diretamente o compromisso entre interpretabilidade, seletividade e generalização. A literatura recente distribui-se entre critérios determinísticos por índice ou resíduo, modelos supervisionados clássicos, ensembles e DL, cada qual com vantagens distintas em custo, transparência e dependência de dados.

#### 2.5.2.1 *Índices determinísticos e critérios por resíduo/assinaturas (fasores, frequência e harmônicos)*

Decisores determinísticos continuam relevantes porque são interpretáveis, compatíveis com implementação enxuta e frequentemente operam com poucos sinais e poucas etapas. Nessa classe entram critérios baseados em resíduos, frequência, sequência negativa, conteúdo harmônico, filtros e assinaturas específicas [39, 48, 22, 40, 41, 34, 49, 36, 57, 58]. O desafio é que a simplicidade da decisão costuma deslocar a dificuldade para a calibração do índice e para a seletividade frente a confundidores com assinaturas próximas.

#### 2.5.2.2 *Supervisionados clássicos e decisão modular (SVM/ANFIS/ELM/ANN rasa)*

Modelos supervisionados clássicos e arquiteturas modulares ocupam um ponto intermediário entre interpretabilidade e flexibilidade. Eles aparecem em formulações com SVM, ANFIS, ELM e ANN rasas, normalmente sobre atributos físicos ou estatísticos já condensados [29, 56, 57, 58]. Em geral, são atraentes quando se deseja boa capacidade de separação com custo moderado, mas a robustez fora do cenário de treino depende mais da qualidade do conjunto de dados do que da sofisticação do classificador em si.

#### 2.5.2.3 *Ensembles/boosting e seleção explícita de atributos*

Ensembles e técnicas de boosting reforçam a ideia de decisão por comitê e costumam aparecer associados a seleção explícita de atributos. Trabalhos com AdaBoost, Cat-

Boost, RF multiclasse e seleção multiobjetivo ilustram essa linha, na qual boa parte do ganho decorre da combinação entre atributos compactos e decisores robustos a variabilidade local [59, 62, 45, 19, 69]. Ainda assim, o desempenho dessa família também depende de protocolos transparentes e de cobertura adequada de cenários críticos.

#### 2.5.2.4 *Aprendizado profundo (representações intermediárias e sequências)*

O DL desloca parte do esforço de projeto da escolha manual de limiares e atributos para o desenho da arquitetura e para a curadoria dos dados. Isso aparece em combinações de wavelets, imagens, sequências temporais e modelos como CNN, LSTM e Redes Neurais Profundas (do inglês, *Deep Neural Networks*) (DNN) [27, 64, 65, 67, 66, 70, 38]. O ganho pode ser expressivo em cenários complexos, mas a transparência da decisão, o custo do bloco de entrada e a evidência de generalização tornam-se ainda mais críticos.

#### 2.5.2.5 *Não supervisionado (agrupamento) como caso especial*

Como caso especial, métodos não supervisionados por agrupamento mostram que é possível detectar ilhamento sem depender integralmente de rótulos, mas a evidência ainda é mais limitada do que nos cenários supervisionados e profundos [46].

### 2.5.3 **Qualidade de evidência e limitações recorrentes (ZND, ruído, confundidores, generalização)**

A leitura transversal dos trabalhos mostra um problema metodológico recorrente: o ganho reportado depende fortemente do protocolo de avaliação. A caracterização de criticidade e ZND varia de varreduras mais alinhadas a ensaios normativos a grades discretas muito limitadas; a robustez a ruído, quando presente, costuma ser mostrada em poucos níveis de SNR; e a seletividade frente a confundidores nem sempre vem acompanhada de particionamento reproduzível entre treino, validação e teste [28, 42, 43, 41, 48, 59, 66, 39, 30, 69, 64, 67, 57, 58, 56, 62, 45, 34]. Somam-se a isso a pouca transparência do custo fim-a-fim quando o bloco de extração domina o processamento e a dificuldade de extrapolar resultados para implantação em IEDs ou relés [36, 31, 27, 64, 70, 46].

Em consequência, acurácia isolada raramente basta para comparar métodos. O que diferencia evidência forte de evidência fraca é a combinação entre criticidade explicitada, robustez a ruído, cobertura de confundidores, documentação de hiperparâmetros e reporte operacional de latência e custo.

## 2.6 COMPARAÇÃO DE DESEMPENHO ENTRE ABORDAGENS

A Tabela 3 reúne as referências discutidas no capítulo e serve como régua comparativa mínima. O objetivo não é ranquear métodos, mas registrar, de forma compacta, o bloco principal de representação/extração, o tipo de decisor, a quantidade de parâmetros e os indicadores de desempenho quando explicitados.

A leitura comparativa confirma um padrão recorrente: representações mais ricas tendem a oferecer maior poder discriminativo, mas frequentemente à custa de maior latência, maior sensibilidade a hiperparâmetros e menor transparência do custo de implementação. Em contrapartida, métodos compactos podem ser mais previsíveis computacionalmente, embora também possam ser mais sensíveis a confundidores e a ajustes finos de limiar.

Tabela 3 – Resumo comparativo de desempenho.

Citação	Método principal	Parâmetros (lista ou nº)	Acurácia ou ZND	Tempo de detecção (janela + proc.)
[28]	SST + limiar	$V_{abc} \rightarrow$ modal; SST 75–450 Hz; 3 feat ( $E_g, S_d, E_t$ ); $\tau$ por simulação; SNR:30(1)	ZND: parc. (poucos $\Delta P, \Delta Q$ ); acurácia: n.r.	Jan.: n.r.; Proc.: $\sim$ ms (rep.); Tot.: dezenas de ms (rep.)
[39]	DFST + ELM	NSV/NSC (terminal GD); DFST; 4 feat; ELM (R/B); SNR:30/20	Acc/F1/MC: rep. (n.r. aqui); ZND: n.r.	Jan.: 1 ciclo; Proc.: n.r.; Tot.: $\approx$ 78,5 ms (rep.)
[30]	HHT + índice/limiar	HHT em $V, I$ ; IMF1; 1 feat (razão); $\tau = 1, 1$ ; SNR:40(1)	Acur./ZND: n.r. (evid. limitada)	Jan.: n.r.; Proc.: n.r.; Tot.: poucos ms (rep.)
[36]	RTCWT + regras (+ injeção)	RTCWT (Morlet) p/ gatilho; injeção inter-harm.; 6 feat + regras (normas+duração); SNR: n.r.	Acur./ZND: n.r.	Jan.: n.r.; Proc.: n.r.; Tot.: n.r.
[35]	WT + entropia + SVM/KNN	WT 6 níveis; entropia Rényi; SVM/KNN; (QEE)	Acurácia (QEE): rep.; DI/ZND: n.a.	Tempo: n.r.
[29]	ST + ANFIS	ST em seq. negativa; 2 feat ( $Z_1, Z_2$ ); ANFIS; $N=20$ ; SNR/ZND: n.r.	Acur./ZND: n.r. (amostra pequena)	Tempo: n.r.
[63]	ST + mRMR + ANN	ST em $V_{abc}$ ; energias harm. 1–32; mRMR $\rightarrow$ 7 feat; gatilho MA/ROCMA; 64 samp/ciclo; SNR: n.r.	Acur./ZND: n.r. (ZND indireta)	Jan.: n.r.; Proc.: n.r.; Tot.: $\approx$ 26 ms (trip) e $<$ 2 ciclos (rep.)
[42]	VMD + SVD + entropia ( $\tau$ )	$V_{abc} \rightarrow V_m$ ; VMD $K = 4$ ; SVD; entropia Shannon; $\tau = 0,015$ ; UL-1741; AWGN (sem curva SNR)	ZND/criticidade: UL-1741 (boa); acurácia: n.r.	Tempo: n.r.
[43]	VMD + IsIIdx ( $\tau$ )	VMD $K = 4$ ; $E_k + \Delta$ (jan. 24 amostras); $\tau$ empírico; SNR:20	Acurácia: 98,7% (SNR=20); ZND: varredura $\Delta P$ (rep.)	Jan.: 24 amostras (ms n.r.); Proc.: n.r.; Tot.: $\sim$ 10 ms (rep.)
[44]	ITD + índice/limiar	ITD (base+resíduo); 1 feat (energia); demais (jan./ $\tau$ /SNR/métricas): n.r.	Acur./ZND: n.r.	Tempo: n.r.
[45]	EMD + RF (multiclasse)	EMD em $V_a$ ; 4 IMFs (usa IMF2); 4 feat; RF multi (ilh. vs 2 chaveamentos)	Acurácia: 97,59%; sens.: 98,88%; ZND: n.r.	Tempo: n.r.
[46]	EMD+HHT + K-means	$Q$ (dq); 2 feat ( $\delta$ IMF <sub>1</sub> , $E_{\max}$ ); K-means $k = 3$ ; $\Delta P, \Delta Q = \pm 10\%$ (discreto); $Q_f \in [0,5,2,5]$ ; SNR: n.r.	ZND: “quase nula” (sim.; grade discreta); acurácia: n.r.	Jan.: 1–2 ciclos (rep.); Proc.: n.r.; Tot.: 20–40 ms (@50 Hz) / 20 ms (jan. desliz.)

**Notas:** n.r.=não reportado; parc.=evidência parcial (poucos SNR, poucos pontos  $\Delta P, \Delta Q$ , ou protocolo incompleto); n.a.=não aplicável; *feat*=nº de atributos/entradas;  $\tau$ =limiar; SNR em dB; *Jan.* = atraso mínimo por janela (quando existe); *Proc.* = tempo de processamento quando explicitado; quando só há tempo agregado, indica-se *Tot.* e marca-se janela/proc. como n.r. (ou embut.).

Tabela 3 (continuação).

Citação	Método principal	Parâmetros	Acurácia ou ZND	Tempo de detecção (janela + proc.)
[13]	VMD+HT + ANN (GWO)	Modal; VMD+HT; 6 feat; ANN (GWO); SNR:35(1)	Acurácia: n.r.; ZND: varredura $\Delta P, \Delta Q$ (rep.)	Jan.: 5 ciclos; Proc.: n.r.; Tot.: n.r.
[19]	VMD+HSA + AMKELM	$I_{abc}$ ; VMD; HSA (energia) + $\tau$ ; AMKELM (DF-BIO); multi(17); $f_s = 100$ kHz; SNR:20/30/50	Acur./ZND: n.r. (no resumo); SNR: sim (3 níveis)	Jan.: n.r.; Proc.: n.r.; Tot.: 10–20 ciclos (qualit.; rep.)
[31]	VMD(2x) + árvore por regras	VMD (2 etapas); feat: MCF, RMER, ZC, envol-tória IA; regras normativas; (QEE); SNR:45–30	Acurácia (QEE): rep.; DI/ZND: n.a.	Jan.: 200 ms; Proc.: $\sim 190$ ms (MATLAB); Tot.: $\sim 390$ ms (impl.)
[50]	Morfologia (DEDF) + $\tau$	RMS( $V_{abc}$ ); MM (dil./eros.); $\tau$ por tentativa-e-erro; SNR: n.r.	Acur./ZND: n.r.	Jan.: n.r.; Proc.: n.r.; Tot.: < 10 ms (mín. 6,09; máx. 17 ms)
[40]	PLL + MP+SVD + regra	$f(t)$ via PLL; buffer 1000@10 kHz; MP+SVD; $\tau$ (Th1/Th2)	Acur./ZND: n.r.	Jan.: 0,1 s; Proc.: n.r.; Tot.: $\sim 0,2$ s (sim) / $\sim 0,1$ s (RT)
[41]	Phaselets + $ Z_{-} $ ( $\tau$ )	Phaselets (jan. 1/20 a 1 ciclo); seq. negativa; $ Z_{-}  =  V_{-} / I_{-} $ ; $\tau \approx 0,7$ ; UL-1741/Instituto de Engenheiros Eletricistas e Eletrônicos (do inglês, <i>Institute of Electrical and Electronics Engineers</i> ) (IEEE)-1547	ZND: cobre $\Delta P = \Delta Q = 0$ (boa); acurácia: n.r.	Jan.: 1/20 ciclo (mín.)–1 ciclo; Proc.: n.r.; Tot.: tip. < 2 ciclos (rep.)
[48]	LS+FT + resíduo ( $\tau$ )	LS (fund.); reconstr. FT; TMF/TMFI; $\tau = 0,5$ ; 64 samp/ciclo (crit.); SNR:20/40	ZND: < $\pm 1\%$ ( $\Delta P, \Delta Q$ ) (rep.); acurácia: n.r.	Jan.: n.r. (64 samp/ciclo); Proc.: n.r.; Tot.: $\sim 55$ ms (RTDS)
[22]	KF(harm.) + SHD ( $\tau$ )	Modal $V$ ; KF p/ harm.; SHD ( $3^{\circ}/5^{\circ}/7^{\circ}$ ); $\tau$ via PCHIP (=3)	Acurácia: 91,6%; ZND: qualit. (casos críticos)	Jan.: n.r.; Proc.: n.r.; Tot.: 1–2 ciclos (ex. 50–70 ms)
[49]	LS(TIV) + direcional (+ comm)	$\zeta$ (LS); ângulo (seq. positiva); $\zeta_{th} = 0,09$ ; 2 pontos + comunicação	Acur./ZND: n.r.	Jan.: n.r.; Proc.: $\sim 2$ ms (impl.); Tot.: 7 ms (incl. 5 ms comm)
[34]	DFT(2ª harm.) + KNN	DFT; 2ª harm. ( $V, I$ ); comp. simétricas + $V \times I$ ; 9 feat; KNN (K=2); jan. 10 amostras; 128 samp/ciclo (50 Hz); SNR:20/40	Acur./ZND: n.r. (nota: 2ª harm. pode confundir)	Jan.: 10/128 ciclo $\approx 1,56$ ms (@50 Hz); Proc.: n.r.; Tot.: 4–9 ms (sem ruído) / 8–16 ms (SNR=20)
[57]	SVM modular + PCA + lógica	RMS/THD/ $f/P, Q$ ; amostra 5 ms; FIFO 8 (40 ms); PCA 95% (por descritor)	Acurácia: até 100% (rep.; protocolo parc.); ZND: n.r.	Jan.: 40 ms; Proc.: n.r. (embut.); Tot.: $\sim 50$ –55 ms (UL-1741)

**Notas:** n.r.=não reportado; parc.=evidência parcial (poucos SNR, poucos pontos  $\Delta P, \Delta Q$ , ou protocolo incompleto); n.a.=não aplicável; *feat*=nº de atributos/entradas;  $\tau$ =limiar; SNR em dB; *Jan.* = atraso mínimo por janela (quando existe); *Proc.* = tempo de processamento quando explicitado; quando só há tempo agregado, indica-se *Tot.* e marca-se janela/proc. como n.r. (ou embut.).

Tabela 3 (continuação).

Citação	Método principal	Parâmetros	Acurácia ou ZND	Tempo de detecção (janela + proc.)
[58]	2 blocos SVM + lógica booleana	$2 \times (7 \text{ SVM})$ ; FIFO 8@5 ms (40 ms) + PCA 95%	Acurácia: 100%; ZND: n.mapeada em $\Delta P - \Delta Q$ (nota)	Jan.: 40 ms; Proc.: n.r. (embut.); Tot.: 40–55 ms (rep.)
[56]	ANFIS hierárquico (7+1)	Descritores conv.; FIFO 8@5 ms (40 ms); decisão por faixa; treino offline	Acur./ZND: n.r. (protocolo parc.)	Jan.: 40 ms; Proc.: n.r.; Tot.: 40–45 ms (alguns 15 ms; cond. n.r.)
[59]	AdaBoost (stumps) + pesos	10 feat; suavização por média em 5 ciclos; 30 modelos fracos	Erros: 0 (rep.); ZND: forte (maioria dentro)	Jan.: 5 ciclos; Proc.: n.r.; Tot.: $\leq 0,07$ s (rep.; nota de consistência p/ 50/60 Hz)
[62]	CatBoost + seleção Spearman	10 feat $\rightarrow 5 (f, df/dt, dI/dt, THD_U, THD_I)$ ; média em 3 ciclos pós-evento	Acc/F1/AUC: $\approx 1$ (rep.; protocolo parc.); ZND: n.r.	Jan.: $\geq 3$ ciclos; Proc.: n.r.; Tot.: n.r. (além do atraso da janela)
[69]	Seleção (MMODEA) + K-ELM	45 feat $\rightarrow 3 (V_-/V_+, V_0/V_+, THD_i)$ ; SNR:20/30	Acc/F1: rep. (n.r. aqui); ZND: por variedades (boa)	Jan.: n.r.; Proc.: n.r.; Tot.: $\sim 75$ ms (rep.)
[64]	CWT $\rightarrow$ imagem + CNN	CWT (Morse) em $V_{abc}$ ; janela 6 ciclos; $f_s = 1000$ Hz; SNR:20/30	Acurácia: 98,78%; ZND: parc. (baixo desbalanço)	Jan.: 6 ciclos; Proc.: n.r.; Tot.: 218 ms (médio)
[65]	DFT(2ª harm.) + LSTM	6 feat (seq. pos/neg/zero da 2ª harm. de $V$ e $I$ ); SNR: n.r.	Acurácia: $\approx 99,5-99,6\%$ ; ZND: qualit. (quase-balanço)	Jan.: n.r.; Proc.: n.r.; Tot.: 2–10 ms (sim) / 6 ms (lab)
[66]	1D-CNN/LSTM/CNN-LSTM	3 feat (1 ambíguo); janela 40 ms @1 kHz; quase-balanço $\pm 0,5\%$	Desempenho alto (pior: 2/200 erros); ZND: recorte forte	Jan.: 40 ms; Proc.: $\sim 10$ ms (rep.); Tot.: $\sim 50$ ms
[67]	Estatísticas + DNN	5 feat (RMS, $\sigma$ , curtose, assim., clearance); DNN 2 cam.; SNR:30/40; bin + multi	Acurácia: 99,8% (orig.); degrada em SNR (parc.); ZND: parc.	Jan.: n.r.; Proc.: extração $\approx 25,1$ ms + DNN $\approx 18,3$ ms; Tot.: $\approx 43,4$ ms
[27]	DWT + MR-SSE + SAE-DNN	DWT (Db4, 6 níveis); MR-SSE (SVD/ent.); SAE-DNN; janela 0,2 s @30 kHz; SNR:30(1)	SNR: rep. (1 ponto); acurácia/ZND: n.r.	Jan.: 0,2 s; Proc.: n.r.; Tot.: 0,18 s (rep.; def. potencialmente ambígua)

**Notas:** n.r.=não reportado; parc.=evidência parcial (poucos SNR, poucos pontos  $\Delta P, \Delta Q$ , ou protocolo incompleto); n.a.=não aplicável; *feat*=nº de atributos/entradas;  $\tau$ =limiar; SNR em dB; *Jan.* = atraso mínimo por janela (quando existe); *Proc.* = tempo de processamento quando explicitado; quando só há tempo agregado, indica-se *Tot.* e marca-se janela/proc. como n.r. (ou embut.).

Tabela 3 (continuação).

Citação	Método principal	Parâmetros	Acurácia ou ZND	Tempo de detecção (janela + proc.)
[70]	Denoising + ELM	Seq. negativa; TI-wavelet + morfologia; 4 ou 6 feat; ELM (~100 neur.); janela 1 ciclo; SNR:20 (30)	Acurácia: n.r.; SNR: parc.; ZND: n.r.	Jan.: 1 ciclo; Proc.: n.r.; Tot.: < 2 ciclos (rep.)
[38]	Lifting wavelet + BP	Lifting (HF de $V,I$ ); média $ d $ em 1 período; 6 feat; BP (6–12–1); SNR/ZND: n.r.	Acur./ZND: n.r.	Jan.: n.r.; Proc.: n.r.; Tot.: $\approx 0,04$ s

## 2.7 LACUNAS RECORRENTES E REQUISITOS DE PROJETO

A síntese crítica das seções anteriores mostra que o estado da arte avançou mais rapidamente na diversidade de técnicas do que na padronização da evidência experimental. Nesta seção, as limitações recorrentes são convertidas em requisitos operacionais de projeto para orientar o ambiente experimental, a construção dos cenários e o reporte de desempenho nos capítulos subsequentes.

### 2.7.1 Lacunas recorrentes e requisitos derivados

A seguir, cada lacuna (L) é acompanhada do respectivo requisito (R), formulado em termos operacionais e orientado por critérios de engenharia.

- **L1 — Criticidade/ZND tratada de forma heterogênea na literatura.** Observa-se desde caracterizações próximas a ensaios normativos [41, 48, 42] até avaliações em poucos pontos discretos ou recortes específicos [28, 46, 66], o que reduz comparabilidade e pode ocultar regimes próximos da ZND.  
**R1 — Requisito de criticidade:** adotar uma caracterização operacional explícita do entorno crítico (incluindo quase-balanço), com critérios claros de sucesso/falha e amostragem suficientemente representativa (p.ex., varreduras em  $(\Delta P, \Delta Q)$  ou equivalente operacional).
- **L2 — Robustez a ruído raramente é varrida de modo sistemático.** Quando ruído aparece, costuma ser pontual (1–3 níveis de SNR) [28, 39, 43, 48]. Em outros estudos, ele sequer é considerado [69, 64, 67, 19], dificultando inferências sobre sensibilidade comparativa.  
**R2 — Requisito de robustez:** avaliar sensibilidade sob uma faixa de SNR (não apenas pontos isolados) e explicitar o modelo de ruído/perturbação adotado, reportando degradação de desempenho de forma rastreável (curvas desempenho  $\times$  SNR quando aplicável).
- **L3 — Confundidores e seletividade multievento são tratados de modo irregular.** Muitos trabalhos listam confundidores, mas sem um protocolo reproduzível (composição de classes, separação por cenários etc.) [57, 62, 34]; há exceções com tratamento mais explícito [58, 45, 19], porém ainda com lacunas em métricas de seletividade e definição operacional.  
**R3 — Requisito de seletividade:** definir um conjunto mínimo e reproduzível de eventos confundidores (com ênfase nos mais similares ao ilhamento) e reportar seletividade de forma transparente (por tipo de evento e/ou matriz de confusão quando aplicável), incluindo taxas de falso disparo quando pertinente.

- **L4 — Latência fim-a-fim e custo do bloco de extração nem sempre são transparentes.** Há contraste entre *fluxos* leves com janelas curtas [34, 41, 48] e abordagens com extração dominante (p.ex., CWT→imagem, VMD iterativa) [31, 64, 27, 13], frequentemente com tempo reportado de forma agregada.  
**R4 — Requisito de implementabilidade:** favorecer extração com custo previsível e poucos hiperparâmetros, e reportar latência de forma operacional (idealmente separando *janela mínima* e *processamento*, ou justificando quando inseparável), além de fornecer ao menos uma estimativa objetiva de custo computacional compatível com implantação em IED/relé.
  
- **L5 — Requisitos de medição e dependência de infraestrutura nem sempre são colocados como restrição de projeto.** Algumas propostas assumem comunicação ou módulos adicionais que elevam complexidade [49, 36], o que impacta implantação e comparabilidade.  
**R5 — Requisito de instrumentação:** explicitar o conjunto mínimo de medições e premissas (tensões/correntes, necessidade de sequência, fasores sub-ciclo,  $f_s$ , sincronismo) e, quando houver comunicação, quantificar sensibilidade a atraso/perda e delimitar condições de operação.
  
- **L6 — Generalização fora-do-cenário e reprodutibilidade de validação são frequentemente fracas.** Aparecem riscos de otimismo por bases pequenas [29], particionamento pouco transparente [57, 62] e ausência de separação por cenários/topologias; além disso, alguns modelos são explicitamente dependentes do sistema [59].  
**R6 — Requisito de validação/reprodutibilidade:** documentar protocolo de geração de dados e conjuntos (treino/validação/teste) de forma completa, buscando reduzir vazamento entre cenários, e incluir ao menos um teste fora-do-cenário (p.ex., mudança de topologia/controle/carga ou parâmetros fora da faixa de treino), com checklist mínimo de reporte (janelas,  $f_s$ , hiperparâmetros, critérios de decisão e métricas).

### 2.7.2 Síntese operacional dos requisitos

Em conjunto, R1–R6 formam uma régua mínima para avaliação de MDIs passivos: criticidade/ZND explícita, robustez a ruído com modelo declarado, seletividade frente a confundidores, latência e custo reportados de forma operacional, instrumentação claramente delimitada e validação reprodutível com alguma evidência de generalização. Esses requisitos orientam diretamente os capítulos seguintes.

## 2.8 CONCLUSÕES PARCIAIS DO CAPÍTULO

Este capítulo reuniu uma revisão crítica dos MDIs, com foco em abordagens passivas modernas apoiadas por DSP e IC, situando-as frente a requisitos normativos e condições de ensaio anti-ilhamento [7, 11, 8, 71]. A análise mostrou que o avanço recente é real, mas fortemente condicionado por escolhas de protocolo: criticidade/ZND, ruído, confundidores, janelamento, particionamento dos dados e transparência do custo fim-a-fim influenciam decisivamente a interpretação dos resultados [28, 36, 29, 66, 60, 38].

Em síntese, métodos passivos continuam atraentes por sua baixa intrusividade, mas a literatura ainda sustenta melhor o potencial dessas abordagens do que sua comparabilidade rigorosa em campo. Por isso, as lacunas convertidas em requisitos R1–R6 estabelecem a régua mínima adotada nesta tese para desenho experimental, organização dos dados e reporte de desempenho. Os Capítulos 3 e 4 dão sequência direta a essa agenda ao formalizar o ambiente de teste e o método proposto de forma alinhada a essas exigências.

### 3 AMBIENTE EXPERIMENTAL E CONJUNTOS DE DADOS

Este capítulo descreve o ambiente experimental adotado para validar o MDI proposto. Nele, detalha-se desde a modelagem do sistema de potência, que reproduz as principais características de uma microrrede conectada à rede de distribuição, até a geração dos conjuntos de dados que sustentam as etapas de otimização, treinamento e teste dos classificadores.

O estudo reporta uma avaliação numérica controlada de uma microrrede fotovoltaica de 250 kW modelada no *MATLAB/Simulink*, em que cada ensaio possui duração total de  $t_{\text{sim}} = 0,5 \text{ s}$  e o evento principal é programado em  $t \approx 0,35 \text{ s}$ . As formas de onda trifásicas no PAC são registradas e exportadas para processamento estatístico subsequente conforme descrito adiante.

Inicialmente, apresenta-se a modelagem do sistema de potência, especificando a arquitetura da microrrede, os modelos de geração distribuída e os parâmetros das cargas conectadas ao PAC. Em seguida, são discutidos os cenários de simulação que contemplam as condições de operação nominal, eventos de falta e situações de ilhamento não intencional, gerados por meio de variações de topologia e distúrbios na rede, garantindo diversidade estatística ao conjunto. Na sequência, discute-se a estratégia de particionamento dos dados, diferenciando os conjuntos destinados à escolha dos hiperparâmetros (conjuntos de otimização) daqueles usados para avaliar o desempenho final (conjuntos de treino/teste). Por fim, são formalizadas as métricas de avaliação empregadas, assegurando uma análise quantitativa robusta e compatível com estudos correlatos.

#### 3.1 MODELAGEM DO SISTEMA DE POTÊNCIA

Esta seção detalha o ambiente experimental utilizado para a simulação do sistema de potência. O sistema modelado representa uma planta Fotovoltaico (do inglês, *Photovoltaic*) (PV) de 250 kW conectada à rede elétrica, desenvolvida no ambiente *MATLAB/Simulink* R2022b, utilizando a biblioteca *Simscape Electrical* em regime discreto, com passo

$$T_{s,\text{Power}} = \frac{1}{33 \times f_{\text{nom}} \times 100} = \frac{1}{33 \times 60 \text{ Hz} \times 100} \approx 5,05 \mu\text{s}.$$

O passo de discretização  $T_{s,\text{Power}}$  corresponde a 100 amostras por período da portadora Modulação por Largura de Pulso (do inglês, *Pulse Width Modulation*) (PWM). Para a composição do banco de dados utilizado nas etapas de processamento estatístico, as tensões e correntes trifásicas no PAC são exportadas a uma frequência de amostragem  $F_s = 3300 f_{\text{nom}} = 198 \text{ kHz}$ , equivalente a 198 kSa/s (198 mil amostras por

segundo), mantendo coerência com o regime discreto do modelo e preservando os transitórios associados ao ilhamento e às faltas.

O modelo é baseado e adaptado do exemplo *250 kW Grid-Connected PV Array* da MathWorks<sup>1</sup> e nos ajustes reportados em [24].

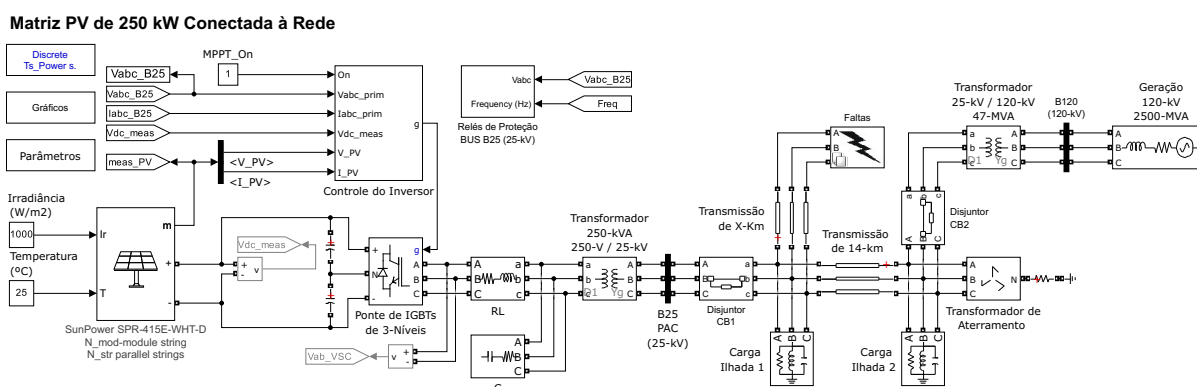
### 3.1.1 Arquitetura geral do sistema

O sistema de potência modelado, conforme ilustrado no diagrama da Figura 5, é composto pelos seguintes blocos principais:

- **Matriz Fotovoltaica (PV Array):** fonte GD primária.
- **Inversor de Potência:** converte a energia CC da matriz PV em CA para injeção na rede.
- **Transformadores Elevadores:** ajustam os níveis de tensão para a transmissão e conexão à rede.
- **Linhas de Transmissão:** representam a infraestrutura de conexão entre os componentes e a rede.
- **Cargas Locais:** consumidores de energia conectados ao sistema.
- **Ponto de Conexão à Rede Elétrica:** representa a rede de grande porte à qual o sistema PV está interligado.

O modelo ainda inclui disjuntores CB1/CB2 para controle de fluxo de potência e um transformador de aterramento para segurança e estabilidade do sistema. A microrrede opera no regime conectado à rede.

Figura 5 – Diagrama da matriz PV de 250 kW conectada à rede.



Legenda: Diagrama utilizado para representar a conexão da matriz PV de 250 kW ao sistema elétrico, indicando os principais blocos e pontos de interligação do arranjo.

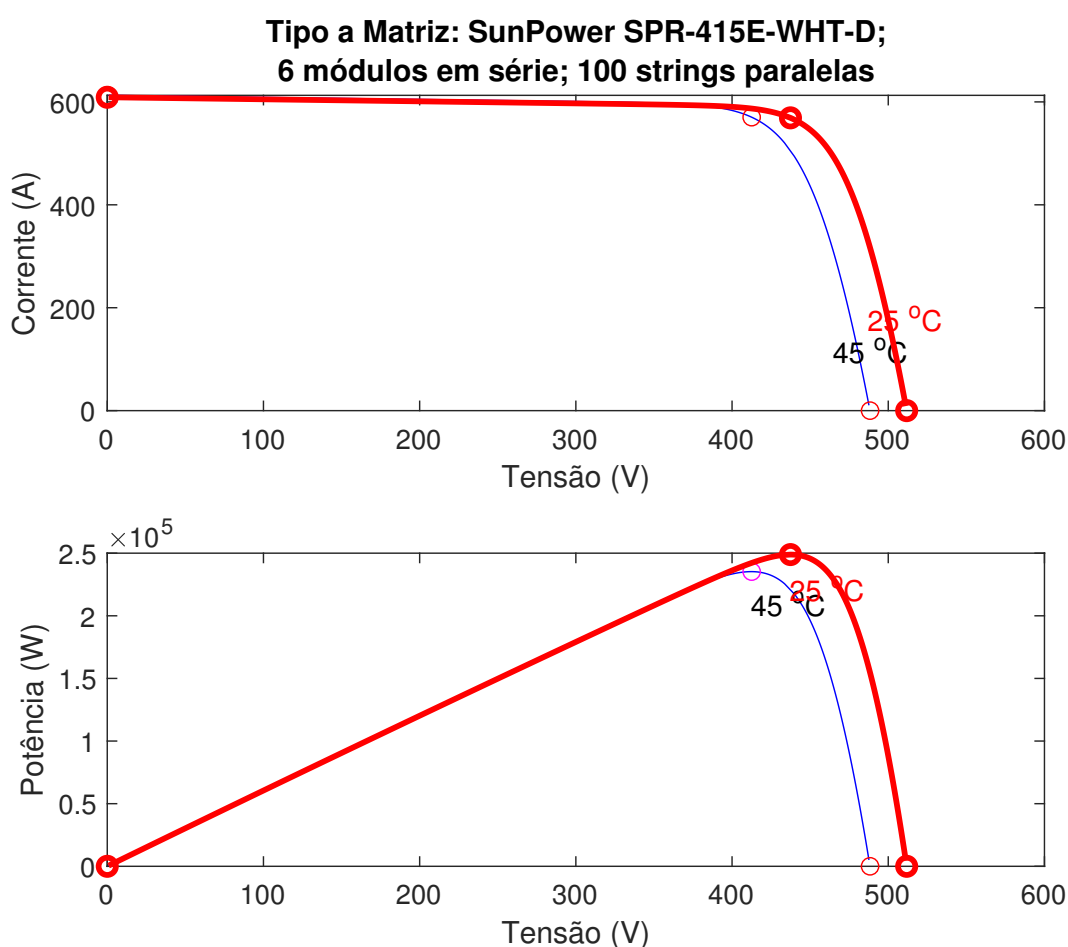
<sup>1</sup> Exemplo “*250 kW Grid-Connected PV Array*” disponível em: <https://www.mathworks.com/help/sps/ug/250-kw-grid-connected-pv-array.html>

### 3.1.2 Geração fotovoltaica

A matriz PV considerada possui 100 *strings* em paralelo ( $N_{str} = 100$ ), cada uma com 6 módulos *SunPower SPR-415E-WHT-D* em série ( $N_{mod} = 6$ ). Este arranjo produz potência nominal de  $P_{nom} = 250$  kW sob irradiância  $E_e = 1000$  W m<sup>-2</sup> e temperatura  $T_{cell} = 25$  °C. As curvas características *I-V* e *P-V* são apresentadas na Figura 6.

A interface de potência inclui um capacitor de enlace  $C_{link} = 50$  mF (25 mF por barramento) dimensionado para sustentar energia equivalente a  $\frac{3}{4}$  de um ciclo fundamental em  $P_{nom}$ .

Figura 6 – Curvas *I-V* e *P-V* para  $T_{cell} \in \{25, 45\}$  °C e  $E_e = 1000$  W m<sup>-2</sup>.



Legenda: Curvas características de corrente–tensão (*I-V*) e potência–tensão (*P-V*) do módulo, comparando o efeito da temperatura de célula para irradiância fixa  $E_e = 1000$  W m<sup>-2</sup>; a elevação de  $T_{cell}$  altera principalmente o nível de tensão e, conseqüentemente, o ponto de máxima potência.

### 3.1.3 Conversor CC-CA e Filtro de Saída

O conversor utiliza uma ponte multinível de Transistor Bipolar de Porta Isolada (do inglês, *Insulated-gate bipolar transistor*) (IGBT), na topologia *Neutral Point Clamped* (NPC), com modulação PWM senoidal de portadora triangular operando em  $f_c = 33 f_{nom} = 33 \times 60 \text{ Hz} = 1,98 \text{ kHz}$ . A interface com a rede é realizada por um reator  $RL$  com  $R = 0,406 \text{ m}\Omega$  e  $L = 108 \text{ }\mu\text{H}$ , dimensionado para suprimir harmônicas de comutação e limitar a taxa de variação de corrente  $\Delta i / \Delta t$ . Adicionalmente, utiliza-se um capacitor *shunt* dedicado à atenuação do *ripple* em alta frequência na tensão no PAC.

Em base 250 kV A, os parâmetros do reator correspondem aproximadamente a  $R \approx 0,15\%$  e  $L \approx 0,15 \text{ pu}$ , coerentes com  $R = 0,406 \text{ m}\Omega$  e  $L = 108 \text{ }\mu\text{H}$  na base física adotada.

### 3.1.4 Transformadores e Sistema de Distribuição

O acoplamento entre a geração fotovoltaica e a rede de média tensão é realizado por um transformador de 250 kV A (0,25/25 kV,  $Yg, D_1$ ), modelado a partir de três transformadores monofásicos equivalentes, cuja resistência e reatância de dispersão totalizam cerca de 6% na base de potência. A etapa seguinte de interligação com a rede de alta tensão é feita por um transformador de 47 MV A (25/120 kV,  $Yg, D_1$ ), cuja impedância percentual é de aproximadamente 8%. O barramento de 120 kV é modelado como uma fonte síncrona equivalente de 2500 MV A, representando um sistema de potência de elevado Razão de Curto-Circuito (do inglês, *Short-Circuit Ratio*) (SCR). Essa estrutura garante que a microrrede opere predominantemente em regime conectado à rede, permitindo o estudo do comportamento transitório durante a desconexão simulada.

O sistema de distribuição é representado por duas seções de linha de transmissão trifásica do tipo  $\pi$  transposto. A primeira seção, parametrizada pela variável `Fault_len` e fixada em 5 km, conecta o barramento B25 ao PAC e é utilizada para a inserção de faltas simétricas e assimétricas. A segunda seção, de 14 km, conecta a Carga 1 à Carga 2, reproduzindo o trecho de rede que permanece energizado durante a condição de ilhamento. Os disjuntores CB1 e CB2 permitem a configuração dos cenários de falta e a abertura controlada da interligação para a simulação do ilhamento não intencional.

### 3.1.5 Cargas

Foram consideradas duas cargas, sendo uma localizada ao lado do PAC e a outra a 14 km deste. Cada carga absorve um par de potências  $(P, Q)$  tal que  $P \in [206,6 \text{ kW}, 322,8 \text{ kW}]$  e  $Q \in [-5,9 \text{ kvar}, 4,1 \text{ kvar}]^2$ , conforme descrito na Seção 3.2.2.

<sup>2</sup> A faixa negativa de  $Q$  corresponde a potência capacitiva.

Durante os experimentos, os valores de potência ativa e reativa consumidos pelas cargas são ajustados de forma a cobrir uma faixa de operação crítica associada à zona de não detecção da IEEE 1547 no entorno do ponto de operação da GD. Essa variação é implementada por meio de uma grade  $\Delta P \times \Delta Q$  que combina  $P \in \{206,6, 250, 322,8\}$  kW e  $Q \in \{-5,9, 0, 4,1\}$  kvar, aplicada a cada carga e cenário de simulação, conforme detalhado na Seção 3.2.2.

### 3.1.6 Estrutura de Controle Digital

As malhas de controle são executadas em um processador digital equivalente, com período de amostragem  $T_{s,Control} = 10 T_{s,Power} \approx 50,5 \mu s$ . A estratégia de controle inclui:

- **Circuito de Sincronismo (do inglês, *Phase-Locked Loop*) (PLL) discreta**, baseada em Circuito PLL Baseado no Sistema de Coordenadas Síncrono (do inglês, *Synchronous Reference Frame Phase-Locked Loop*) (SRF-PLL) trifásica, implementada pelo bloco PLL (3PH) do *Simscape Electrical*, para extração da fase  $\theta$  e da frequência instantânea da rede;
- **Rastreador de Ponto de Máxima Potência (do inglês, *Maximum Power Point Tracking*) (MPPT)** do tipo Perturbe e Observe, com incremento  $\Delta D = 0,01$  e saturação da tensão do elo CC em [357 V, 583 V];
- **Regulador de tensão CC** do tipo PI, com  $K_p = 2$  e  $K_i = 400$ ;
- **Regulador de corrente** síncrono em  $dq$  (PI+FF), com  $K_p = 0,3$ ,  $K_i = 20$  e compensação *feedforward* por  $R_{ff}$  e  $L_{ff}$ , derivados dos parâmetros do reator e do transformador;
- **Modulador PWM** de portadora triangular, sincronizado à frequência de comutação  $f_c = 1,98$  kHz.

O conversor opera em modo *grid-following*, com sincronização de fase e frequência fornecida pela PLL. O sinal lógico produzido pelo MDI é utilizado, em arranjos *grid-following*, para comandar a desconexão conforme os requisitos de proteção. Já em arranjos *grid-forming*, o mesmo sinal pode atuar como gatilho para a transição de modo sob condição de ilhamento, mantendo-se a exigência de detecção rápida e alterando-se apenas a ação subsequente.

### 3.1.7 Cenários de Falta e Ilhamento

Para garantir diversidade estatística, são consideradas doze variantes de faltas simétricas e assimétricas, habilitadas via SIMULINK VARIANT, além de eventos de ilhamento provocados pela abertura do disjuntor CB2. As variantes de falta incluem configurações simétricas (LLL, LLL-G) e assimétricas (L, LL, LL-G).

Os tempos de ocorrência e duração dos eventos de falta e de ilhamento são

parametrizados por  $t_{fit}$  e  $t_{ilh}$ , respectivamente, enquanto a distância entre o PAC e o ponto de falta é fixada em  $Fault\_len = 5$  km. A janela total de simulação adotada é de  $t_{sim} = 0,5$  s.

Nos cenários analisados, os eventos de falta e deilhamento são aplicados em  $t = 0,35$  s, mantendo-se a distância do PAC ao ponto de falta em 5 km. Para diversificar as condições iniciais do sistema, a fase da tensão da fase A é aleatorizada em cada ensaio, com  $\theta_A \sim \mathcal{U}(-180^\circ, 180^\circ)$ .

### 3.2 CENÁRIOS DE SIMULAÇÃO

Este trabalho avalia o desempenho do MDI proposto em três grandes condições de operação — nominal (N), ilhamento não intencional (I) e faltas (F) — geradas a partir da microrrede de referência de 250 kW descrita na Seção 3.1. Para cada condição são sintetizados diversos perfis de carga (potências ativa e reativa) de modo a assegurar ampla variabilidade estatística e representar de forma realista as margens de potência da ZND recomendadas em normas (IEEE Std. 1547–2003 e a ABNT NBR–IEC 62116:2012). A Figura 7 resume o fluxo de geração dos cenários e o Tabela 4 apresenta uma visão geral dos mesmos.

Tabela 4 – Visão geral dos cenários de simulação.

Condição	Casos	Repetições	Total de execuções
Nominal (N)	9 (grade)	100 (grade)	900 (grade)
	ou 100 (aleatório)	ou 10 (aleatório)	ou 1000 (aleatório)
Ilhamento (I)	9 (grade)	100 (grade)	900 (grade)
	ou 100 (aleatório)	ou 10 (aleatório)	ou 1000 (aleatório)
Falta (F)	9 (grade)	100 (grade)	900 (grade)
	ou 100 (aleatório)	ou 10 (aleatório)	ou 1000 (aleatório)
<b>Total</b>	<b>2700 (grade) ou 3000 (aleatório)</b>		

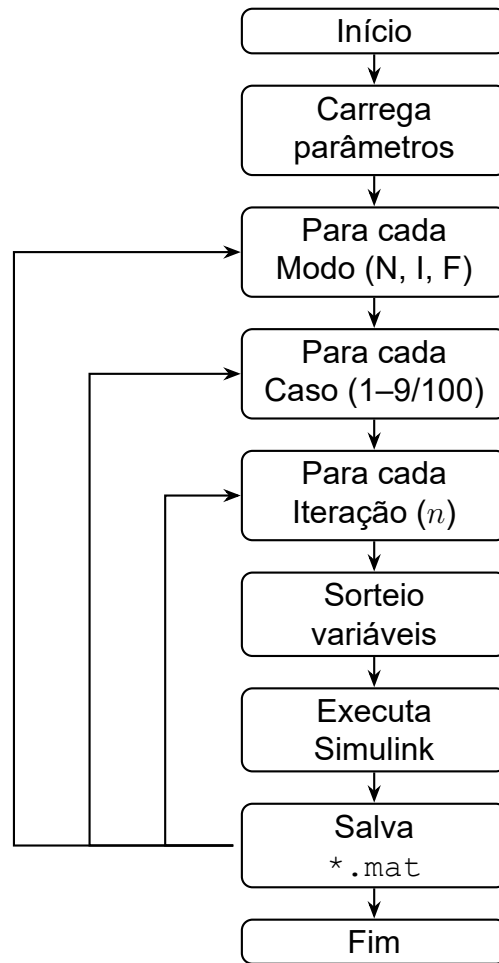
#### 3.2.1 Estratégia geral

Os experimentos são conduzidos em lotes, cada execução simula  $T_{sim} = 0,5$  s com frequência de amostragem  $f_{s,Power} = 3300 \times f_{nom} = 198$  kHz. As tensões trifásicas no PAC são registradas em matriz  $\mathbf{V} \in \mathbb{R}^{N \times 3}$  com  $N = f_{s,Power} T_{sim} + 1 = 99001$  amostras. Para aumentar a diversidade temporal, o ângulo inicial da fase A é sorteado uniformemente em  $[-180^\circ, 180^\circ]$ .

#### 3.2.2 Variabilidade de potência ativa e reativa

Sob condição de ilhamento, a frequência e a tensão no PAC são estabelecidas pelos desequilíbrios de potência ativa ( $\Delta P$ ) e reativa ( $\Delta Q$ ) entre a GD e a carga local

Figura 7 – Fluxo de geração dos cenários de simulação.



Legenda: Fluxograma do procedimento de geração e organização dos cenários: após o carregamento de parâmetros, executam-se laços aninhados por modo (N, I e F), por caso (1–9/100) e por iteração  $n$ . Em cada iteração, as variáveis são sorteadas, o modelo no Simulink é executado e os resultados são armazenados em arquivos \*.mat.

[8]. Utilizando o critério do Espaço de Desbalanços de Potências (EDP), os limites de operação de um MDI podem ser representados por [72]:

$$\left(\frac{V}{V_{\max}}\right)^2 - 1 \leq \frac{\Delta P}{P_{\text{DG}}} \leq \left(\frac{V}{V_{\min}}\right)^2 - 1, \quad (3.1)$$

$$Q_f(1 - (f/f_{\min})^2) \leq \frac{\Delta Q}{P_{\text{DG}}} \leq Q_f(1 - (f/f_{\max})^2), \quad (3.2)$$

onde  $V$  é a tensão nominal,  $P_{\text{DG}}$  é a potência ativa da GD e  $Q_f$  o fator de qualidade. A IEEE Std. 1547–2003 impõe  $V_{\max} = 110\%$ ,  $V_{\min} = 88\%$ ,  $f_{\max} = 60,5\text{Hz}$ ,  $f_{\min} = 59,3\text{Hz}$  e, para este estudo, adotou-se  $Q_f = 1$ . Substituindo esses valores nas Equações (3.1)–(3.2) obtém-se os limites da ZND:

$$-17,36\% \leq \frac{\Delta P}{P_{DG}} \leq 29,13\% \quad (3.3)$$

$$-2,37\% \leq \frac{\Delta Q}{P_{DG}} \leq 1,65\% \quad (3.4)$$

Considerando  $P_{DG} = 250 \text{ kW}$ , as faixas absolutas de potência tornam-se  $P \in [206,6 \text{ kW}, 322,8 \text{ kW}]$  e  $Q \in [-5,9 \text{ kvar}, 4,1 \text{ kvar}]$ .

Dois esquemas de amostragem foram empregados para mapear a ZND:

- (a) **Grade regular  $3 \times 3$ :** oito pontos nos limites da ZND mais o ponto central formam uma grade regular de nove casos (Tabela 5). Cada caso foi simulado  $n = 100$  vezes.
- (b) **Amostragem aleatória:** 100 pares  $(\Delta P, \Delta Q)$  foram sorteados com distribuição uniforme dentro da ZND; cada perfil foi repetido  $n = 10$  vezes.

Em ambos os esquemas, os valores de potência ativa e reativa de cada perfil foram divididos igualmente entre as duas cargas ilhadas descritas na Seção 3.1.5, de modo que a soma das potências consumidas por essas cargas reproduz exatamente o ponto correspondente no espaço  $(P, Q)$ . O ponto 5 representa o pior caso para a detecção de ilhamento, pois não há desequilíbrio entre a potência fornecida pela GD e a exigida pela carga [8]. Por fim, a Figura 8 ilustra esses nove perfis numerados de acordo com a Tabela 5 dentro da ZND.

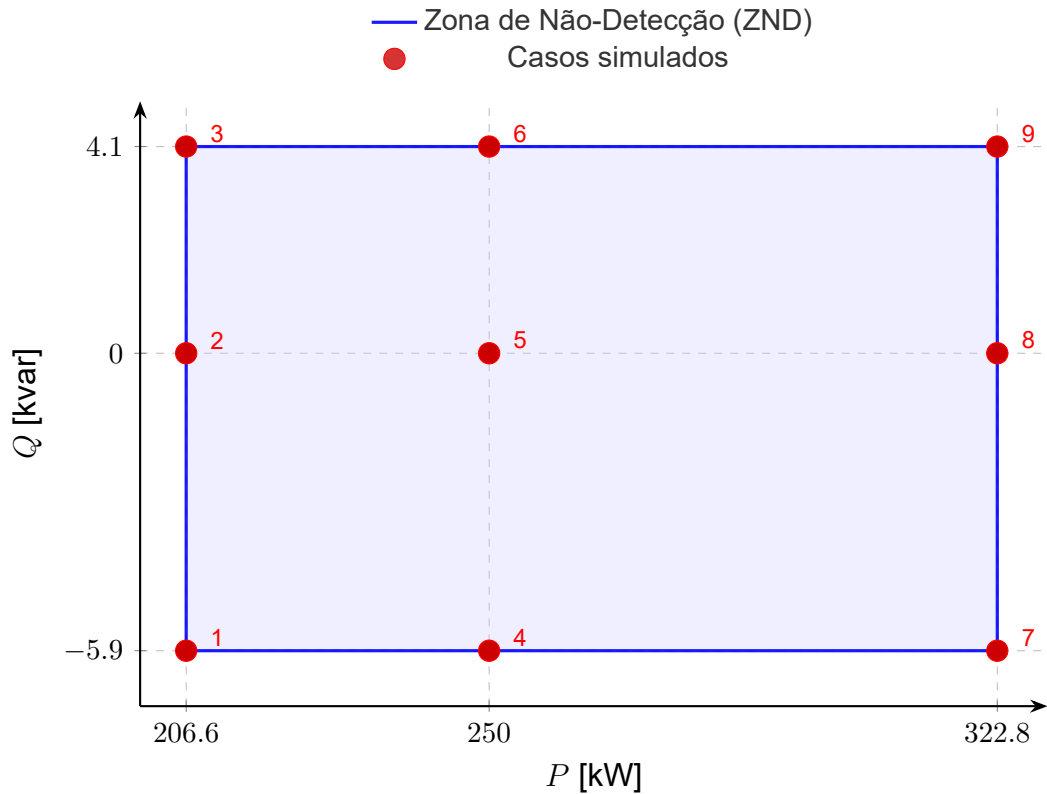
Tabela 5 – Perfis de potência usados na análise determinística (grade  $3 \times 3$ ).

Ponto	Descrição	$P$ [kW]	$Q_L$ [kvar]	$Q_C$ [kvar]
1	min $\Delta P$ , min $\Delta Q$	206,6	0,0	5,9
2	min $\Delta P$ , 0 $\Delta Q$	206,6	0,0	0,0
3	min $\Delta P$ , max $\Delta Q$	206,6	4,1	0,0
4	0 $\Delta P$ , min $\Delta Q$	250,0	0,0	5,9
5	0 $\Delta P$ , 0 $\Delta Q$	250,0	0,0	0,0
6	0 $\Delta P$ , max $\Delta Q$	250,0	4,1	0,0
7	max $\Delta P$ , min $\Delta Q$	322,8	0,0	5,9
8	max $\Delta P$ , 0 $\Delta Q$	322,8	0,0	0,0
9	max $\Delta P$ , max $\Delta Q$	322,8	4,1	0,0

### 3.2.2.1 Formalização do domínio de operação na ZND

Para referência, denotamos por

$$\delta = \begin{bmatrix} \Delta P/P_{DG} \\ \Delta Q/P_{DG} \end{bmatrix} \in \mathbb{R}^2 \quad (3.5)$$

Figura 8 – Espaço  $P$ - $Q$  e casos simulados na ZND.

Legenda: Diagrama no plano  $P$ - $Q$ , com  $Q = Q_L - Q_C$ , destacando a ZND (região retangular em azul) conforme a IEEE Std. 1547 e os nove perfis discretos da grade  $3 \times 3$  (pontos vermelhos 1–9), referenciados na Tabela 5. Os valores de  $P$  e  $Q$  nos eixos correspondem aos níveis mínimo, nominal e máximo adotados na varredura determinística.

o vetor formado pelos desbalanços relativos de potência ativa e reativa na barra do PAC, de forma consistente com as Equações (3.1)–(3.4).

O conjunto de operação explorado neste trabalho dentro da região de não detecção da IEEE Std. 1547 é então

$$\mathcal{D}_{\text{ZND}} = \{ \delta \in \mathbb{R}^2 : \delta \text{ satisfaz (3.3) e (3.4)} \}. \quad (3.6)$$

Os nove casos determinísticos da grade  $3 \times 3$  ilustrados na Figura 8 podem ser escritos como um subconjunto finito

$$\mathcal{D}_{\text{grade}} = \{ \delta^{(1)}, \delta^{(2)}, \dots, \delta^{(9)} \} \subset \mathcal{D}_{\text{ZND}}, \quad (3.7)$$

onde cada  $\delta^{(i)}$  corresponde a um dos pontos extremos ou central da região retangular definida por (3.3)–(3.4) e resumida na Tabela 5.

De forma complementar, o conjunto utilizado na validação fora da grade é obtido por amostragem uniforme de desbalanços no interior da ZND:

$$\mathcal{D}_{\text{rand}} = \{ \tilde{\delta}^{(1)}, \tilde{\delta}^{(2)}, \dots, \tilde{\delta}^{(100)} \} \subset \mathcal{D}_{\text{ZND}}, \quad (3.8)$$

em que cada  $\tilde{\delta}^{(j)}$  gera um perfil de carga distinto, mantendo os níveis de potência ativa e reativa dentro dos limites estabelecidos em (3.3)–(3.4).

### 3.2.3 Modos de operação

**Nominal (N):** A topologia permanece inalterada durante toda a janela de observação  $T_{\text{sim}}$ , mantendo a microrrede conectada à rede principal.

**Ilhamento não intencional (I):** O disjuntor de interconexão CB2 é aberto em  $t_{\text{ilh}} = 0,35 \text{ s}$ , isolando a microrrede. Neste ensaio, nenhuma falta é aplicada e  $\text{FAULT\_MODE} = 1$ .

**Faltas (F):** Aplica-se uma falta à montante em  $t_{\text{flt}} = 0,35 \text{ s}$ . O modo  $\text{FAULT\_MODE} \in \{2, \dots, 12\}$  é selecionado por amostragem uniforme a cada repetição, contemplando:

- *Monofásicas:* A, B, C (2–4);
- *Bifásicas:* AB, BC, CA (5–7);
- *Bifásicas-terra:* AB–g, BC–g, CA–g (8–10);
- *Trifásica:* ABC (11);
- *Trifásica-terra:* ABC–g (12).

A distância entre o PAC e o ponto de aplicação da falta é mantida em 5 km.

### 3.2.4 Tamanho e armazenamento do conjunto de dados

Cada execução do SIMULINK grava uma matriz tridimensional  $\mathbf{V} \in \mathbb{R}^{N \times 3 \times n}$  contendo as tensões trifásicas no PAC, onde  $N = f_{\text{s,Power}} T_{\text{sim}} + 1 = 99001$  amostras, os três canais correspondem às fases  $\{A, B, C\}$  e  $n$  é o número de repetições daquele caso.

**Regular**  $3 \times 3$ . Para  $n = 100$  repetições, 9 casos de carga e 3 modos de operação (N, I, F), geram-se

$$N \times 3_{\text{fases}} \times n \times (\text{casos}) \times 3_{\text{modos}} = 99\,001 \times 3 \times 100 \times 9 \times 3 \approx 8,02 \times 10^8 \text{ pontos.}$$

**Amostragem aleatória.** Tomando  $n = 10$  repetições, 100 casos e os mesmos 3 modos, obtém-se

$$99\,001 \times 3 \times 10 \times 100 \times 3 \approx 8,91 \times 10^8 \text{ pontos.}$$

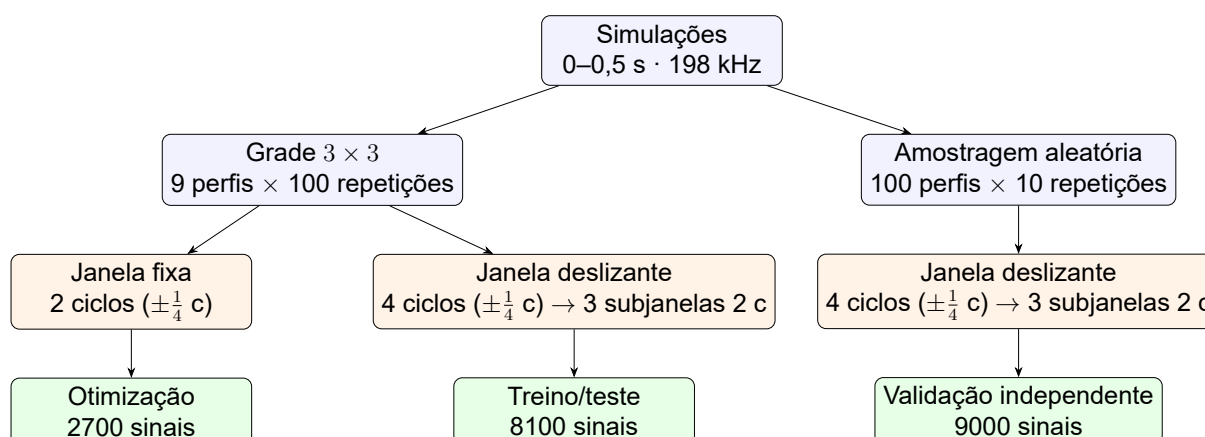
A Tabela 6 discrimina as dimensões e quantidade de arquivos gerados para o banco de dados.

Os dados são armazenados como `data_<mode>_case_<k>.mat`, usando o formato `-v7.3`

Tabela 6 – Dimensões dos arquivos \*.mat gerados.

Conjunto	Casos	Repetições $n$	Pontos por arquivo	Arquivos (3× modos)
Grade $3 \times 3$	9	100	$99\,001 \times 3 \times 100$	27
Aleatório	100	10	$99\,001 \times 3 \times 10$	300

Figura 9 – Particionamento dos dados simulados.



Legenda: Fluxograma do particionamento dos sinais simulados em dois esquemas: (i) grade  $3 \times 3$  (9 perfis repetidos 100 vezes), subdividida em um conjunto com janela fixa de 2 ciclos (otimização) e outro com janela deslizante de 4 ciclos segmentada em 3 subjanelas de 2 ciclos (treino/teste); e (ii) amostragem aleatória (100 perfis repetidos 10 vezes), utilizada com janela deslizante de 4 ciclos segmentada em 3 subjanelas de 2 ciclos para validação independente (Seção 3.2.2).

Além da matriz  $v$ , os arquivos da amostragem aleatória incluem a estrutura `powers_data` com as potências sorteadas ( $P, Q_L, Q_C$ ) e o identificador `FAULT_MODE`, o que facilita filtragens posteriores por perfil de carga ou tipo de falta.

### 3.2.4.1 Reprodutibilidade

Todos os scripts utilizam `rng default` para inicializar o gerador de números pseudo-aleatórios do MATLAB e, assim, garantir que os mesmos casos possam ser reproduzidos em execuções futuras.

## 3.3 DIVISÃO DOS DADOS

Conforme a estratégia delineada nas Seções 3.1-3.2, o acervo de sinais gerados nas simulações foi particionado em três bancos independentes, cada qual com papel bem definido no ciclo de *tuning*-treinamento-validação do MDI. A Figura 9 ilustra o fluxo de derivação dos dados e a Tabela 7 resume as dimensões de cada partição.

Tabela 7 – Visão geral dos conjuntos de dados.

Conjunto	Origem	Janelamento	Sinais
Otimização	Grade $3 \times 3$	Fixa (2 ciclos)	2 700
Treino/Teste	Grade $3 \times 3$	4 ciclos $\rightarrow$ $3 \times 2$ ciclos	8 100
Validação independente	Amostragem aleatória	4 ciclos $\rightarrow$ $3 \times 2$ ciclos	9 000

Essa separação, combinada com diferentes estratégias de janelamento, permite otimizar os hiperparâmetros de forma eficiente (janela fixa) e, subsequentemente, avaliar o desempenho em um cenário que emula o monitoramento em tempo real (janela deslizante), conforme detalhado no Capítulo 4.

### 3.3.1 Conjuntos para otimização

O banco de otimização foi pensado para ser pequeno o suficiente para permitir varreduras exaustivas de hiper-parâmetros e, ao mesmo tempo, suficientemente rico para capturar a variabilidade estatística dos três modos de operação — nominal (N), ilhamento (I) e falta (F).

Primeiro, cada sinal trifásico bruto<sup>3</sup> foi recortado em uma janela fixa de 2 ciclos (33,33 ms), posicionada com  $\frac{1}{2}$  ciclo antes e 1 ciclo  $+\frac{1}{2}$  ciclo depois de  $t = 0,35s$ , instante em que o evento ocorre. Um *jitter* uniforme de  $\pm\frac{1}{4}$  ciclo evita sincronismo artificial entre a fase-A e o rótulo.

Foram simulados nove pontos na grade  $3 \times 3$  da ZND (Tabela 5), cada qual repetido 100 vezes por modo de operação. O resultado final perfaz 2700 janelas, balanceadas nas três classes.

O banco é embaralhado com a configuração padrão do gerador pseudoaleatório do MATLAB e submetido a validação cruzada (Sec. 4.6.4.1). Em seguida, ele alimenta a otimização descrita na Sec. 4.6, que define as arquiteturas finais dos classificadores.

### 3.3.2 Conjuntos para treinamento e teste

Após fixada a topologia, é necessário expô-la a um cenário que reproduza o fluxo de dados online. Para isso, as mesmas 2700 instâncias da grade  $3 \times 3$  foram recortadas novamente, agora em 4 ciclos, posicionada com 1 ciclo  $+\frac{1}{2}$  ciclo antes e 2 ciclo  $+\frac{1}{2}$  ciclo depois de  $t = 0,35s$ , e transformadas em três sub-janelas de 2 ciclos com sobreposição de 50 %. O processo triplica o volume, gerando 8100 amostras.

Com permutação estratificada, 70% das amostras formam o conjunto de treinamento; as 30% restantes compõem o conjunto de teste. Somente o bloco de treino é usado para re-treinar os classificadores.

<sup>3</sup> Amostragem de 198 kHz, duração de 0,5 s.

### 3.3.3 Validação independente

A verdadeira prova de generalização do MDI é realizada sobre um banco fora da grade, formado por 100 perfis de carga inéditos, cada qual simulado dez vezes nos três modos operacionais. Após o janelamento deslizante ( $4 \rightarrow 3 \times 2$  ciclos, análogo ao conjunto de treino/teste), obtém-se 9000 sub-janelas avaliadas sob todas as condições de ruído consideradas. Nenhum hiper-parâmetro ou peso de rede é ajustado neste estágio.

## 4 METODOLOGIA

A presente etapa metodológica descreve, em detalhe, o fluxo completo de processamento de sinais empregado para transformar as tensões trifásicas brutas, simuladas no ambiente MATLAB/SIMULINK e caracterizadas no Capítulo 3, em vetores de atributos discriminantes capazes de alimentar os classificadores propostos. Esse fluxo é organizado em duas abordagens complementares de janelamento, cada qual direcionada a um objetivo específico dentro do ciclo de desenvolvimento do MDI:

1. **Janela fixa de 2 ciclos:** otimizada para experimentação rápida e seleção de hiperparâmetros, pois oferece um recorte temporal compacto, ancorado no instante do distúrbio, com *jitter* controlado para evitar viés<sup>1</sup>.
2. **Janela deslizante 4 → 3×2 ciclos:** configurada para simular monitoramento quase em tempo real; ao varrer a forma de onda com sobreposição de 50 %, fornece redundância temporal e maior robustez a atrasos de detecção.

Nas seções seguintes, são apresentados os passos executados em cada abordagem, desde a eliminação de artefatos até a geração do vetor final de cumulantes de ordem superior, evidenciando como as duas estratégias se articulam para compor o sistema de detecção proposto.

### 4.1 ABORDAGENS DE PRÉ-PROCESSAMENTO E ANÁLISE DE SINAIS

Esta seção descreve os fluxos de processamento responsáveis por converter as tensões trifásicas brutas, geradas conforme o Capítulo 3, em vetores de atributos estatísticos adequados à etapa de classificação. Adotam-se duas estratégias complementares de janelamento:

- (i) uma janela fixa de dois ciclos, empregada nas rotinas de otimização de hiperparâmetros e seleção de características, e
- (ii) uma janela deslizante de quatro ciclos, de onde se extraem três sub-janelas sobrepostas de dois ciclos, configurada para a operação contínua.

Ambos os fluxos compartilham um núcleo de pré-processamento comum, mas diferem no recorte temporal e na posterior extração de atributos, como detalhado nas Subsecs. 4.1.1 e 4.1.2.

---

<sup>1</sup> Os detalhes de geração desse conjunto encontram-se na Seção 3.3.1.

#### 4.1.1 Fluxo de processamento para janela fixa

O presente fluxo opera sobre a janela curta de dois ciclos definida na Sec. 3 e prepara os sinais para as rotinas de otimização de hiperparâmetros. Até a etapa imediatamente anterior à transformação PCA (Sec. 4.2), o processamento compreende três fases:

##### (1) Subamostragem

- **Motivação:** Os sinais brutos são obtidos no SIMULINK a  $f_{s,Power} = 3300 f_{nom} = 198 \text{ kHz}$ . Essa taxa é necessária para a correta implementação da simulação, porém impõe elevado custo computacional na extração de atributos; logo, opta-se por uma redução da resolução temporal de forma a não comprometer informação útil concentrada em banda fundamental e suas harmônicas de ordem baixa.
- **Taxa alvo:** adotou-se 256 amostras por ciclo fundamental ( $f_s^* = 256 f_{nom}$ ), valor clássico em análises.
- **Implementação:** escolheu-se o bloco `downsample` do MATLAB, que implica em uma taxa de decimação  $M$  inteira:

$$M = \left\lceil \frac{f_{s,Power}}{f_s^*} \right\rceil = 13 \quad \Rightarrow \quad f_{s,proc} = \frac{198 \text{ kHz}}{13} \approx 15,23 \text{ kHz}.$$

A pequena diferença de 0,8% em relação a  $f_s^*$  é irrelevante para as estatísticas estudadas e, em contrapartida, evita o atraso de grupo introduzido pelo método `resample` (que inclui filtro anti-aliasing de Resposta ao Impulso Finita (do inglês, *Finite Impulse Response*) (FIR)). O fator de subamostragem inteiro torna o processo determinístico e de baixa complexidade.

##### (2) Injeção controlada de ruído

- **Objetivo:** avaliar a robustez do método a interferências de instrumentação e quantização.
- **Procedimento:** cada fase ( $v_a, v_b, v_c$ ) recebe Ruído Branco Gaussiano Aditivo (do inglês, *Additive White Gaussian Noise*) (AWGN) gerado pelo bloco `awgn(v,SNR,'measured')` do MATLAB. O parâmetro `'measured'` calibra a potência do ruído com base na variância do próprio sinal, assegurando que  $\text{SNR} = 70 \text{ dB}$  seja respeitado em cada traço individual. A variância do ruído branco  $n(t)$  é calculada, para cada traço individual, por

$$\sigma_n^2 = \frac{\sigma_v^2}{10^{\text{SNR}/10}},$$

onde  $\sigma_v^2$  é a variância medida do sinal original e  $\text{SNR} = 70 \text{ dB}$ .

- **Cenário de otimização:** fixa-se  $\text{SNR} = 70 \text{ dB}$  por ser o nível típico. Relações entre 30–60dB são exploradas na validação final.

## (3) Normalização por fator de escala

- **Motivação:** padronizar as amostras para uma mesma faixa dinâmica antes de aplicar PCA ou calcular cumulantes produz matrizes de covariância mais bem condicionadas [73, 74, 75], o que torna a decomposição em componentes principais numericamente mais estável [76] e costuma facilitar a etapa posterior de aprendizado de máquina [77, 78].
- **Método** – Escalonamento simples:

$$\tilde{v}(t) = \frac{v(t)}{V_{\text{pico}}^{\text{nom}}}, \quad V_{\text{pico}}^{\text{nom}} = \frac{V_{\text{LL}}^{\text{nom}}}{\sqrt{3}} \sqrt{2} = 25 \text{ kV} \times \frac{\sqrt{2}}{\sqrt{3}} \approx 20,4 \text{ kV}.$$

O valor 25 kV corresponde à tensão fase–fase do barramento de média tensão onde se mede o PAC. A divisão converte as tensões para grandezas adimensionais, preservando relações de fase e amplitude relativas.

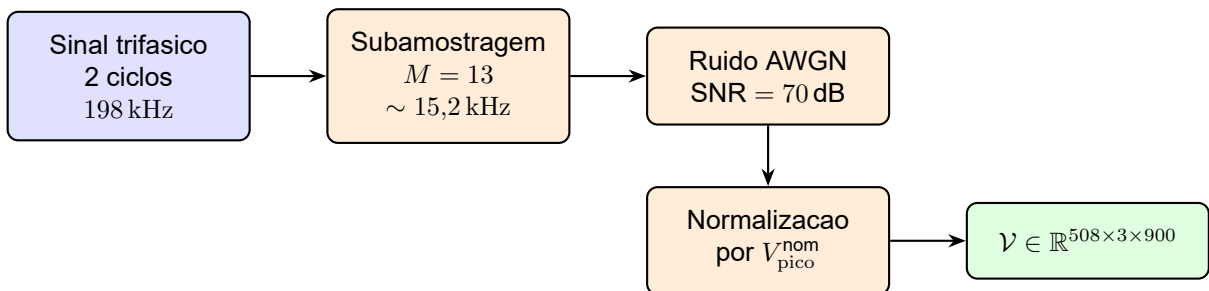
Após as três etapas obtêm-se três tensores, um por condição de operação,

$$\mathcal{V}^{(N)}, \mathcal{V}^{(I)}, \mathcal{V}^{(F)} \in \mathbb{R}^{508 \times 3 \times 900},$$

onde os eixos correspondem, respectivamente, a 508 amostras ( $\approx 2$  ciclos), três fases  $\{a,b,c\}$  e 900 instâncias (9 perfis  $\times$  100 repetições). Esses tensores alimentam a transformação PCA descrita na Sec. 4.2.

As três operações são executadas em sequência: subamostragem, adição de ruído e normalização — antes de qualquer extração de atributos, conforme ilustrado na Figura 10.

Figura 10 – Pré-processamento da janela fixa de dois ciclos.



Legenda: Fluxo de pré-processamento aplicado aos sinais de tensão trifásica extraídos em janela fixa de 2 ciclos: subamostragem por fator  $M = 13$ , adição de ruído AWGN com  $\text{SNR} = 70$  dB e normalização pelo valor de pico nominal  $V_{\text{pico}}^{\text{nom}}$ . Ao final, cada amostra é organizada no tensor  $\mathcal{V} \in \mathbb{R}^{508 \times 3 \times 900}$ , compatível com a etapa de aprendizado.

#### 4.1.2 Fluxo de processamento para janela deslizante

O segundo fluxo reaproveita integralmente as três operações de pré-processamento, subamostragem, adição de ruído  $\text{SNR} = 70 \text{ dB}$  e normalização por  $V_{\text{pico}}^{\text{nom}}$ , já detalhadas na Subseção 4.1.1.

A diferença reside no recorte temporal e na estratégia de aquisição em modo contínuo, desenvolvida para emular a dinâmica de coleta dos conjuntos de treino/teste (Cap. 3).

##### (1) Janela-mãe de quatro ciclos

Para cada forma de onda bruta de  $0,5 \text{ s}$  recorta-se um trecho de quatro ciclos ( $\approx 66,7 \text{ ms}$ ) ancorado em  $t = 0,35 \text{ s}$ , com  $\pm \frac{1}{4}$  ciclo de *jitter* aleatório. A subamostragem  $M = 13$  produz um número  $L$  de amostras por fase igual a:

$$L = \left\lceil \frac{4 \text{ ciclos} \times 3300 \text{ pts./ciclo}}{13} \right\rceil = 1016 \text{ pts.}$$

##### (2) Janela deslizante de dois ciclos

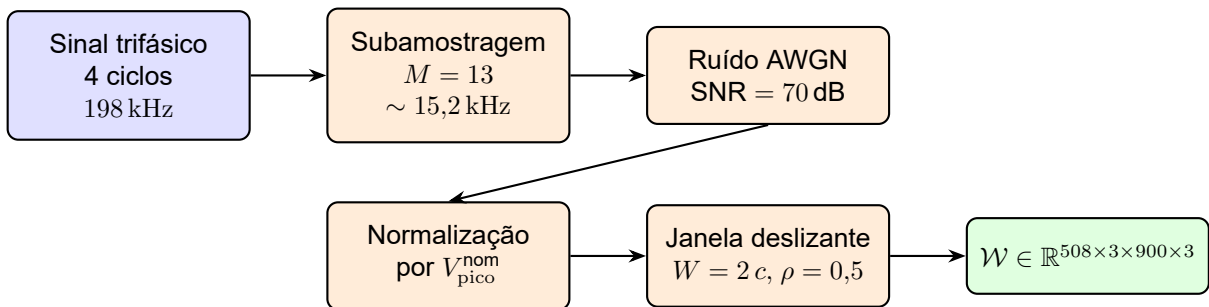
Sobre a janela-mãe aplica-se um deslocamento tipo “ $\frac{1}{2}$  overlap”:

$$W = 2 \text{ ciclos}, \quad \rho = 0,5, \quad s = W(1 - \rho) = 1 \text{ ciclo},$$

$$N_{\text{win}} = \left\lfloor \frac{L - W}{s} \right\rfloor + 1 = 3.$$

Geram-se, portanto, três sub-janelas  $\mathcal{J}_1, \mathcal{J}_2, \mathcal{J}_3$  de  $W = 508$  amostras, deslocadas de um ciclo entre si (Figura 11).

Figura 11 – Pré-processamento da janela deslizante de quatro ciclos.



Legenda: Fluxo de pré-processamento aplicado aos sinais de tensão trifásica extraídos em janela de 4 ciclos: subamostragem por fator  $M = 13$ , adição de ruído AWGN com  $\text{SNR} = 70 \text{ dB}$  e normalização por  $V_{\text{pico}}^{\text{nom}}$ . Em seguida, aplica-se janelamento deslizante com subjanelas de  $W = 2c$  e sobreposição  $\rho = 0,5$ , organizando-se as amostras no tensor  $\mathcal{W} \in \mathbb{R}^{508 \times 3 \times 900 \times 3}$ , onde o último eixo representa as três subjanelas por evento.

### (3) Estrutura de dados

Cada instância temporal original do conjunto base (2700 instâncias por classe no conjunto grade  $3 \times 3$ ) origina três amostras independentes, resultando em um total de  $2700 \times 3 = 8100$  tensores por condição de operação. A estrutura final dos dados para cada classe (Nominal, Ilhamento e Falta) é, portanto, expressa como:

$$\mathcal{W}^{(N)}, \mathcal{W}^{(I)}, \mathcal{W}^{(F)} \in \mathbb{R}^{508 \times 3 \times 900 \times 3}.$$

em que o primeiro eixo corresponde às amostras temporais da janela (508 pontos), o segundo às três fases  $\{a, b, c\}$ , o terceiro às 900 instâncias temporais originais, e o quarto às três subjanelas extraídas de cada uma delas.

Para garantir compatibilidade entre os dados extraídos com janela fixa e os obtidos por janela deslizante, reorganiza-se o quarto eixo (subjanelas) no eixo das instâncias. Realiza-se uma concatenação ao longo deste eixo, para se obter um tensor uniforme  $\widetilde{\mathcal{W}}^{(\cdot)} \in \mathbb{R}^{508 \times 3 \times 2700}$  para cada condição  $\{N, I, F\}$ .

Essa transformação permite unificar a representação dos dados, de modo que, tanto os sinais obtidos com janela fixa quanto os obtidos com janela deslizante passam a ter a mesma estrutura tridimensional. A partir deste ponto, todas as etapas de processamento (extração de características, treinamento e validação) podem ser aplicadas de forma idêntica sobre ambos os conjuntos.

A seguir, tais tensores são submetidos ao fluxo PCA (Sec. 4.2) e à rotina de extração de atributos descritos na Seção 4.3.

## 4.2 TRANSFORMAÇÃO PCA

Esta seção descreve a redução de dimensionalidade aplicada às tensões trifásicas normalizadas, visando concentrar a informação estatística relevante à detecção de ilhamento com baixo custo computacional. Emprega-se Análise de Componentes Principais (do inglês, *Principal Component Analysis*) (PCA) com uma estratégia dirigida ao evento-alvo: construímos uma rotação fixa  $\bar{\mathbf{R}}$  a partir de janelas de ilhamento e, em operação, projetamos cada janela em um subespaço de baixa dimensão.

### 4.2.1 Fundamentação

Considere, para cada janela de  $L$  amostras já normalizadas, a matriz de observações

$$\mathbf{X} = \begin{bmatrix} v_a(1) & v_b(1) & v_c(1) \\ v_a(2) & v_b(2) & v_c(2) \\ \vdots & \vdots & \vdots \\ v_a(L) & v_b(L) & v_c(L) \end{bmatrix} \in \mathbb{R}^{L \times 3},$$

onde cada coluna corresponde a uma fase trifásica (Sec. 4.1.1). Realiza-se a centralização por média (*mean-centering*). Seja

$$\boldsymbol{\mu} = \frac{1}{L} \sum_{t=1}^L \mathbf{X}(t, :) \in \mathbb{R}^3 \quad \text{e} \quad \tilde{\mathbf{X}} = \mathbf{X} - \mathbf{1}_L \boldsymbol{\mu}^\top,$$

em que  $\mathbf{1}_L \in \mathbb{R}^L$  denota um vetor coluna de uns.

O PCA clássico procura uma rotação ortonormal  $\mathbf{R} \in \mathbb{R}^{3 \times 3}$  tal que as colunas de

$$\mathbf{C} = \tilde{\mathbf{X}} \mathbf{R}$$

maximizem, sucessivamente, a variância projetada e sejam mutuamente não correlacionadas. Matematicamente,  $\mathbf{R}$  é constituída pelos autovetores da matriz de covariância amostral centrada

$$\boldsymbol{\Sigma} = \frac{1}{L-1} \tilde{\mathbf{X}}^\top \tilde{\mathbf{X}}, \quad \boldsymbol{\Sigma} \mathbf{R} = \mathbf{R} \boldsymbol{\Lambda},$$

com  $\boldsymbol{\Lambda} = \text{diag}(\lambda_1 \geq \lambda_2 \geq \lambda_3)$ . A fração de variância do  $j$ -ésimo componente é

$$\eta_j = \frac{\lambda_j}{\lambda_1 + \lambda_2 + \lambda_3}, \quad j \in \{1, 2, 3\}.$$

### 4.2.2 Estratégia centrada no ilhamento

Para privilegiar o evento de interesse (ilhamento) e, ao mesmo tempo, evitar o custo de estimar PCA a cada janela em tempo de execução, propõe-se o seguinte procedimento.

#### 1. Estimativa orientada ao evento-alvo

Para cada perfil de carga  $i$  e cada repetição  $j$  dos eventos de ilhamento no conjunto de otimização (Sec. 3.3.1), computa-se o PCA para cada janela de dados centralizada  $\tilde{\mathbf{X}}_{ij} \in \mathbb{R}^{L \times 3}$ .

Conservam-se apenas as direções principais, organizadas por colunas, formando assim a matriz de rotação  $\mathbf{R}_{ij} \in O(3)$ , onde:

$$O(3) = \{ \mathbf{R} \in \mathbb{R}^{3 \times 3} \mid \mathbf{R}^\top \mathbf{R} = \mathbf{I}, \det(\mathbf{R}) = \pm 1 \}.$$

## 2. Média de rotações

Define-se a rotação fixa como média elemento-a-elemento:

$$\bar{\mathbf{R}} = \frac{1}{N} \sum_{i,j} \mathbf{R}_{ij} \in \mathbb{R}^{3 \times 3}. \quad (4.1)$$

$\bar{\mathbf{R}}$  é fixada e passa a ser a única matriz de rotação empregada em todo o fluxo de processamento, independentemente do conjunto utilizado ou da classe do sinal.

Observa-se que, embora  $\bar{\mathbf{R}}$  preserve aproximadamente a ortogonalidade entre colunas, média euclidiana não garante ortonormalidade estrita ( $\bar{\mathbf{R}} \notin SO(3)$ )[79], mas preserva a direção média associada ao ilhamento, suficiente para projeção fixa. A validação é apresentada no Capítulo 5, Seção 5.3.

Enfatiza-se que o objetivo desta projeção não é a reconstrução do sinal ou a decorrelação perfeita, mas sim a criação de um transformador fixo e computacionalmente eficiente que maximize a separabilidade entre as classes de interesse.

## 3. Projeção em execução

Para qualquer nova janela  $\mathbf{X}$ , seja nominal, ilhamento ou falta, calcula-se  $\mathbf{C} = \mathbf{X}\bar{\mathbf{R}} \in \mathbb{R}^{L \times 3}$ . Em operação, utiliza-se apenas a primeira coluna  $\bar{\mathbf{r}}_1 = \bar{\mathbf{R}}_{:,1}$ , formando o traço unidimensional  $\mathbf{z} = \mathbf{X}\bar{\mathbf{r}}_1$ , que alimenta as Estatísticas de Ordem Superior (do inglês, *High-Order Statistics*) (HOS) (Sec. 4.3).

## 4. Intuição e vantagens

A matriz  $\bar{\mathbf{R}}$  foi estimada exclusivamente a partir de sinais de ilhamento, de modo que sua primeira coluna se alinha, em média, à direção de maior energia desse evento.

Embora  $\bar{\mathbf{R}} \notin SO(3)$  em sentido estrito, essa aproximação elimina o cálculo online de uma PCA completa e, quando aplicada aos modos nominal e falta, gera projeções “sub-ótimas” que, na prática, acentuam diferenças entre as classes, o que é exatamente o que se deseja para fins de detecção/classificação.

O custo computacional cai de uma PCA completa por janela para três produtos escalares por amostra, contribuindo significativamente para o requisito de execução em tempo real.

### 4.2.3 Critério para escolha da dimensionalidade

O MDI proposto tem dois requisitos centrais:

- (i) precisão na detecção e
- (ii) baixa complexidade computacional.

Para atender ao segundo requisito, investigamos se seria possível representar cada janela de sinais trifásicos por apenas uma componente principal ( $d = 1$ ), o que reduziria o custo computacional em tempo real. Assim, antes de fixar  $d = 1$  é preciso saber:

- (a) Quanta variância cada componente realmente contém?
- (b) Com essa redução, perde-se poder de discriminação entre classes?

Para responder a estas questões, sugerimos o seguinte protocolo para decisão:

- (a) *Energia*: estimar  $\eta_1, \eta_2, \eta_3$  por PCA clássico por janela (todas as classes) para aferir quanta variância cada componente retém.
- (b) *Separabilidade*: para cada componente, extrair cumulantes ( $2^{\text{a}}$ – $4^{\text{a}}$  ordens) ao longo dos atrasos, ranqueá-las e selecionar os melhores atributos, então avaliar os classificadores via validação cruzada.
- (c) *Escolha*: adotar o menor  $d$  que preserve separabilidade com baixa variância de desempenho.

#### 4.3 EXTRAÇÃO DE ATRIBUTOS VIA ESTATÍSTICAS DE ORDEM SUPERIOR

Esta seção descreve a extração de atributos baseada em Estatísticas de Ordem Superior (do inglês, *High-Order Statistics*) (HOS). As HOS são calculadas diretamente sobre as componentes principais obtidas na Sec. 4.2. Em [80], é demonstrado que os cumulantes são apropriados para uso em sinais elétricos e capazes de trazer uma boa caracterização de distúrbios elétricos. Neste trabalho, utilizam-se as aproximações estocásticas para os cumulantes conforme [81]. O objetivo é capturar assimetrias e não-gaussianidades introduzidas pelos eventos de interesse, preservando, por construção, a insensibilidade das cumulantes de ordem  $> 2$  ao ruído branco gaussiano aditivo [82, 83].

##### 4.3.1 Sinal de entrada e convenções

Seja  $\mathbf{X} \in \mathbb{R}^{L \times 3}$  a janela de tensões trifásicas (uma por coluna) e  $\bar{\mathbf{R}} \in \mathbb{R}^{3 \times 3}$  a rotação fixa estimada conforme Seção 4.2. Define-se

$$\mathbf{C} = \mathbf{X}\bar{\mathbf{R}}, \quad \bar{\mathbf{r}}_m \triangleq \bar{\mathbf{R}}_{:,m}, \quad \mathbf{c}_m \triangleq \mathbf{C}_{:,m} = \mathbf{X}\bar{\mathbf{r}}_m.$$

Considere, para cada janela temporal ( $L$  amostras) e para cada componente principal  $m \in \{1,2,3\}$ , o traço unidimensional

$$\mathbf{c}_m \triangleq [c_m[0], c_m[1], \dots, c_m[L-1]]^\top \in \mathbb{R}^L.$$

Remove-se a média amostral  $\mu_m = \frac{1}{L} \sum_{n=0}^{L-1} c_m[n]$  e define-se o traço centralizado

$$\tilde{c}_m[n] = c_m[n] - \mu_m.$$

Todas as cumulantes são estimadas com atraso circular (indexação modular) e normalização por  $1/L$ . Denotamos  $\tilde{c}_m[n + \tau] \equiv \tilde{c}_m[[n + \tau] \bmod L]$ .

#### 4.3.2 Definições e estimadores adotados

Sejam  $\tau \in \{0, 1, \dots, L-1\}$  os atrasos considerados. Para cada componente  $m$ , adotam-se os seguintes estimadores amostrais de cumulantes de 2ª, 3ª e 4ª ordens (fatias unidimensionais do tensor de cumulantes (atrasos 1-D)) [82, 83]:

$$\hat{\kappa}_2^{(m)}[\tau] = \frac{1}{L} \sum_{n=0}^{L-1} \tilde{c}_m[n] \tilde{c}_m[n + \tau], \quad (4.2)$$

$$\hat{\kappa}_3^{(m)}[\tau] = \frac{1}{L} \sum_{n=0}^{L-1} \tilde{c}_m[n] [\tilde{c}_m[n + \tau]]^2, \quad (4.3)$$

$$\hat{\kappa}_4^{(m)}[\tau] = \frac{1}{L} \sum_{n=0}^{L-1} \tilde{c}_m[n] [\tilde{c}_m[n + \tau]]^3 - 3\hat{\kappa}_2^{(m)}[\tau] \hat{\kappa}_2^{(m)}[0]. \quad (4.4)$$

Algumas observações importantes:

- (i) Em (4.4) foi utilizada a identidade de cumulantes  $\kappa\{x, y, y, y\} = \mathbb{E}[xy^3] - 3\mathbb{E}[xy]\mathbb{E}[y^2]$  sob centralização ( $\mathbb{E}[x] = \mathbb{E}[y] = 0$ ); no estimador amostral,  $\hat{\kappa}_2^{(m)}[0] = \frac{1}{L} \sum_n \tilde{c}_m[n]^2$ .
- (ii) A adoção de atraso circular torna os somatórios invariantes a bordas de janela.

#### 4.3.3 Arranjo de saída e notação

Para cada janela e modo de operação, as três famílias  $\{\hat{\kappa}_2^{(m)}[\tau]\}_{\tau=0}^{L-1}$ ,  $\{\hat{\kappa}_3^{(m)}[\tau]\}_{\tau=0}^{L-1}$  e  $\{\hat{\kappa}_4^{(m)}[\tau]\}_{\tau=0}^{L-1}$  são empilhadas por colunas na forma de vetores  $\mathbf{\Gamma}_r^{(m)} \in \mathbb{R}^{L \times 1}$ , com  $r \in \{2, 3, 4\}$ . Concretamente, o procedimento implementado gera, para cada janela,

$$\mathbf{\Gamma}_2^{(m)} = [\hat{\kappa}_2^{(m)}[0], \dots, \hat{\kappa}_2^{(m)}[L-1]]^\top, \quad \mathbf{\Gamma}_3^{(m)} = [\hat{\kappa}_3^{(m)}[0], \dots, \hat{\kappa}_3^{(m)}[L-1]]^\top,$$

$$\mathbf{\Gamma}_4^{(m)} = [\hat{\kappa}_4^{(m)}[0], \dots, \hat{\kappa}_4^{(m)}[L-1]]^\top,$$

para todos os  $m$  disponíveis (componentes principais como colunas do traço projetado). Essas séries por atraso são, então, a base para a etapa subsequente de ranqueamento/seleção de atributos (Sec. 4.4).

#### 4.3.4 Propriedades úteis ao problema

As cumulantes de ordem superior apresentam características particularmente alinhadas à detecção de eventos de faltas e ilhamento:

- **Potência e estrutura temporal.**

Enquanto as cumulantes de ordem superior capturam a não-gaussianidade, a cumulante de segunda ordem,  $\widehat{\kappa}_2^{(m)}[\tau]$ , que corresponde à função de autocovariância amostral sob centralização, captura as alterações mais fundamentais do sinal.

Em  $\tau = 0$ ,  $\widehat{\kappa}_2^{(m)}[0]$  representa a variância, sendo um indicador direto da potência média  $((RMS)^2)$  da janela e, para janelas de mesmo comprimento e sinal centrado, é proporcional à energia, sendo sensível a afundamentos e elevações de tensão.

Para  $\tau > 0$ , a autocovariância descreve a estrutura de correlação temporal do sinal; desvios de frequência fundamental e alterações no conteúdo harmônico, comuns durante o ilhamento, modificam sua forma. Assim,  $\widehat{\kappa}_2$  atua de forma complementar às HOS, garantindo a detecção de eventos que alteram a potência e a estrutura periódica do sinal, mesmo que estes não introduzam assimetrias ou caudas pesadas significativas em sua distribuição.

- **Robustez a ruído branco e gaussiano.**

Seja  $x[n] = s[n] + w[n]$ , com  $w$  AWGN independente e centrado. Pela aditividade das cumulantes sob independência,  $\kappa_r\{x[\cdot]\} = \kappa_r\{s[\cdot]\} + \kappa_r\{w[\cdot]\}$ . Para  $r \geq 3$ , como processos Gaussianos têm  $\kappa_r \equiv 0$  (teorema de Isserlis/Wick), segue que, em valor esperado,  $w$  não contribui para  $\kappa_3$  e  $\kappa_4$  [82, 83].

Além disso, para a cumulante de segunda ordem (autocovariância), tem-se  $\kappa_2^{(x)}[\tau] = \kappa_2^{(s)}[\tau] + \kappa_2^{(w)}[\tau]$ , e, por brancura,  $\kappa_2^{(w)}[\tau] = \sigma_w^2 \delta[\tau]$ . Logo, para  $\tau > 0$ ,  $\kappa_2^{(w)}[\tau] = 0$  e, portanto,  $\kappa_2^{(x)}[\tau] = \kappa_2^{(s)}[\tau]$ , isto é, não há viés sistemático nos atrasos não nulos.

Na prática, o AWGN aumenta a variância dos estimadores, mas não desloca sistematicamente as cumulantes de ordens  $r \geq 3$  nem as cumulantes de 2ª ordem em  $\tau > 0$ , o que sugere alta robustez do método a degradações por ruído branco. Para ruído colorido,  $\kappa_2^{(w)}[\tau]$  pode ser não nula em  $\tau > 0$ .

- **Assimetria e impulsividade.**

Eventos transitórios e não linearidades alteram a forma da distribuição: a 3ª ordem responde à assimetria (*skewness*) e a 4ª ordem à impulsividade/caudas pesadas (*kurtosis*). Na fatia utilizada (atrasos repetidos, 1-D), vale  $\kappa\{x,y,y,y\} =$

$\mathbb{E}[xy^3] - 3\mathbb{E}[xy]\mathbb{E}[y^2]$  (variáveis centradas), elevando a sensibilidade a desvios do regime nominal.

#### 4.3.5 Complexidade computacional

Seja  $L$  o número de amostras por janela. Para cada componente  $m$ , o cálculo direto de (4.2)–(4.4) para todos os  $\tau \in [0, L - 1]$  envolve  $\mathcal{O}(L)$  operações por atraso, totalizando  $\mathcal{O}(L^2)$  por cumulante e por componente. Como o processamento é independente por  $\{m, \tau\}$ , ele se paraleliza naturalmente por janelas, componentes e modos de operação.

#### 4.3.6 Resumo operacional

Em síntese, para cada janela e para cada componente principal  $m$ :

- (i) subtrai-se a média amostral;
- (ii) computam-se  $\hat{\kappa}_2^{(m)}[\tau]$ ,  $\hat{\kappa}_3^{(m)}[\tau]$  e  $\hat{\kappa}_4^{(m)}[\tau]$  com atraso circular e normalização  $1/L$  segundo (4.2)–(4.4);
- (iii) as séries resultantes por atraso alimentam o ranqueamento e a seleção de atributos descritos na Sec. 4.4.

### 4.4 RANQUEAMENTO E SELEÇÃO DE CARACTERÍSTICAS

Esta seção descreve o ranqueamento e a seleção dos atributos obtidos a partir das séries de cumulantes (Sec. 4.3) usando a Razão Discriminante de Fisher (do inglês, *Fisher's Discriminant Ratio*) (FDR). O procedimento implementado consiste em:

- (i) rearranjar os vetores de cumulantes para uma representação vetorial por janela;
- (ii) calcular a FDR multiclasse [84] por atributo;
- (iii) explorar simetria para reduzir redundância em  $\hat{\kappa}_2$ ; e
- (iv) selecionar os picos de FDR por ordem (cinco por ordem), concatenando-os no vetor final de características.

#### 4.4.1 Rearranjo dos cumulantes

Para cada ordem  $r \in \{2, 3, 4\}$ , componente  $m \in \{1, \dots, M\}$  e atraso  $\tau \in \{0, \dots, L - 1\}$ , denote por  $\hat{\kappa}_r^{(m)}(\tau)$  a amostra do cumulante correspondente (Sec. 4.3). Seja  $\mathcal{G} = \{N, I, F\}$  o conjunto de rótulos de classe e denote por  $g \in \{1, \dots, C\}$  o índice

inteiro da classe, com  $C = |\mathcal{G}| = 3$ . Para cada classe existem  $S$  janelas (“Casos  $\times$  Reps”).

Define-se a indexação linear de atributos

$$p = (m - 1)L + (\tau + 1), \quad p \in \{1, \dots, P\}, \quad P = LM,$$

obtendo-se, para cada ordem  $r$ , o tensor rearranjado

$$\tilde{\Gamma}^{(r)} \in \mathbb{R}^{C \times P \times S},$$

em que  $\tilde{\Gamma}^{(r)}(g, p, s)$  armazena a realização do atributo  $p$  (isto é, o par  $(m, \tau)$ ) na amostra  $s$  da classe  $g$ . Na prática, a indexação concatena os  $M$  componentes ao longo do eixo de atributos, preservando o número de janelas como eixo de amostras.

Assim, para cada ordem  $r$  tem-se  $\tilde{\Gamma}^{(r)} \in \mathbb{R}^{C \times P \times S}$  com  $P = L$  (um atributo por atraso).

#### 4.4.2 FDR multiclasse

Sejam, para cada classe  $g$  e atributo  $p$ ,

$$\mu_{g,p} = \frac{1}{S} \sum_{s=1}^S \tilde{\Gamma}^{(r)}(g, p, s), \quad \sigma_{g,p}^2 = \frac{1}{S-1} \sum_{s=1}^S (\tilde{\Gamma}^{(r)}(g, p, s) - \mu_{g,p})^2,$$

as estatísticas intra-classe (média e variância) do atributo  $p$ , computadas ao longo do eixo das janelas  $s$ . A FDR[84] é a soma sobre pares ordenados de classes:

$$\text{FDR}^{(r)}(p) = \sum_{g=1}^C \sum_{\substack{h=1 \\ h \neq g}}^C \frac{(\mu_{g,p} - \mu_{h,p})^2}{\sigma_{g,p}^2 + \sigma_{h,p}^2}. \quad (4.5)$$

#### 4.4.3 Exploração de simetria em $\hat{\kappa}_2$

Como  $\hat{\kappa}_2^{(m)}(\tau)$  (autocovariância centralizada) é par em  $\tau$  para sinais reais sob estacionariedade de segunda ordem, aplica-se um corte consistente por componente retendo apenas a “metade” de atrasos:

$$\tau \in \{0, 1, \dots, \lceil L/2 \rceil - 1\} \implies p \in \mathcal{K} \subset \{1, \dots, P\},$$

onde  $\mathcal{K}$  é construído por componente (mesmo subconjunto de atrasos para cada  $m$ ). Esse corte é aplicado apenas a  $r = 2$ .

#### 4.4.4 Seleção por picos locais e concatenação final

As curvas  $\text{FDR}^{(r)}(p)$  (Eq. 4.5), por serem calculadas sobre atrasos consecutivos de um sinal temporalmente correlacionado, exibem forte correlação serial. Isso faz

com que atributos vizinhos (e.g., para atrasos  $\tau$  e  $\tau + 1$ ) possuam poder discriminativo muito similar, formando “cordilheiras” ou aglomerados de valores elevados de FDR. Uma seleção dos  $k$  maiores valores absolutos poderia resultar na escolha de múltiplos pontos de uma mesma cordilheira, um fenômeno conhecido como colinearidade. Tal abordagem introduziria alta redundância no vetor de características final, aumentando a dimensionalidade do problema sem agregar nova informação e podendo degradar a performance e a estabilidade de classificadores sensíveis a matrizes de covariância mal condicionadas. Para mitigar este problema, opta-se pela seleção de picos locais de FDR. Esta estratégia visa identificar os representantes de cada aglomerado de alta relevância, garantindo a diversidade informacional e reduzindo a correlação interna do vetor de características final.

Para detectar picos em  $\text{FDR}^{(r)}(p)$  utilizou-se a função `findpeaks` do MATLAB. Uma amostra é considerada pico local se for estritamente maior do que as suas duas vizinhas imediatas ou se for igual a  $+\infty$ . Os pontos de extremidade do sinal que não sejam  $+\infty$  são excluídos da detecção. Em regiões planas (valores iguais consecutivos), a função retorna apenas o ponto de menor índice. Os picos são emitidos na ordem de ocorrência; neste trabalho, após a detecção, ordenam-se por altura (`SortStr = descend`) e selecionam-se os cinco maiores (`NPeaks = 5`). Não se impuseram restrições adicionais (e.g., distância ou proeminência mínimas), de modo que a seleção reflete exatamente as cinco cristas locais de maior magnitude de  $\text{FDR}^{(r)}(p)$  após os cortes definidos nesta seção.

Adota-se uma dimensionalidade por janela de  $B = 15$  atributos, estratificada por ordem ( $B_r = 5$  para cada  $r \in \{2, 3, 4\}$ ), escolhida a priori para controlar complexidade. Esse desenho: (i) mantém a razão  $B/N_{\text{eff}}$  baixa, mitigando aumento de variância e o fenômeno de *peaking* em amostra finita [85, 86]; (ii) reduz redundância entre atrasos vizinhos mesmo após a detecção local; e (iii) assegura paridade entre ordens, preservando a complementaridade entre cumulantes de 2ª, 3ª e 4ª ordens. No conjunto empregado,  $N_{\text{eff}} = C \times S = 3 \times 900 = 2700$ , de modo que  $B/N_{\text{eff}} = 15/2700 \approx 5,6 \times 10^{-3} \ll 0,2$ , mantendo  $B \ll N_{\text{eff}}$ .

Para cada ordem  $r$ , aplica-se `findpeaks` sobre  $\text{FDR}^{(r)}(p)$  e retêm-se os índices dos cinco picos de maior altura:

$$\mathcal{P}_2 = \{p_1^{(2)}, \dots, p_5^{(2)}\}, \quad \mathcal{P}_3 = \{p_1^{(3)}, \dots, p_5^{(3)}\}, \quad \mathcal{P}_4 = \{p_1^{(4)}, \dots, p_5^{(4)}\}.$$

Os atributos escolhidos são extraídos do tensor rearranjado:

$$\tilde{\Gamma}^{(2)}(:, \mathcal{P}_2, :), \quad \tilde{\Gamma}^{(3)}(:, \mathcal{P}_3, :), \quad \tilde{\Gamma}^{(4)}(:, \mathcal{P}_4, :),$$

e concatenados no eixo de atributos, perfazendo  $B = 15$  características por janela.

#### 4.4.5 Viabilidade em Tempo Real e Complexidade Online

É fundamental distinguir entre a fase de projeto e treinamento do método (descrita até aqui) e sua fase de operação ou inferência online. O procedimento de ranqueamento via FDR e seleção de picos, que envolve o cálculo de todas as cumulantes para todos os atrasos, é uma etapa de calibração realizada offline uma única vez sobre o conjunto de dados.

O resultado desta etapa é um conjunto fixo e reduzido de índices de atraso,  $\mathcal{P}_2, \mathcal{P}_3, \mathcal{P}_4$ . Uma vez que estes atrasos são identificados, o processamento para cada nova janela de dados em um sistema de monitoramento online torna-se computacionalmente muito mais leve. O fluxo de trabalho online consiste em:

- (a) Projetar as tensões a partir de  $\bar{\mathbf{R}}$ .
- (b) Calcular apenas os valores de cumulantes correspondentes aos atrasos pré-selecionados nos conjuntos  $\mathcal{P}_r$ .

Dessa forma, a etapa de cálculo da FDR é inteiramente suprimida na operação online. A complexidade computacional para a extração de atributos, por sua vez, é drasticamente reduzida. Em vez de calcular  $L$  valores para cada uma das três ordens de cumulantes (uma tarefa de complexidade  $\mathcal{O}(L^2)$  por ordem, conforme discutido na Sec. 4.3), a implementação online calcula apenas  $B_r = 5$  valores por ordem.

Cabe ressaltar, por fim, que o conjunto de  $B = 15$  atributos, embora computacionalmente eficiente, foi projetado para garantir a diversidade informacional ao incluir as três ordens de cumulantes. Este conjunto servirá como base para uma etapa final de otimização (Sec. 4.6.4.4), a ser detalhada posteriormente. Como será demonstrado, uma análise de relevância sobre este vetor permitirá refinar a seleção para um subconjunto ainda mais parcimonioso e robusto.

#### 4.5 MAPA DE ATRIBUTOS NO INTERIOR DA ZND

Seja  $\delta \in \mathcal{D}_{\text{ZND}}$  um ponto de operação definido no Capítulo 3. Para cada  $\delta$ , a simulação do sistema fornece um traço de tensões trifásicas  $v_\delta(t)$  no barramento do PAC. Após amostragem e seleção da janela de interesse, o sinal correspondente é organizado em um vetor  $x_\delta \in \mathbb{R}^L$ .

O fluxo de extração de atributos descrito nas Seções 4.2–4.4 pode ser representado, de forma compacta, por um mapeamento

$$S : \mathcal{D}_{\text{ZND}} \rightarrow \mathbb{R}^B, \quad S(\delta) = \Phi(x_\delta), \quad (4.6)$$

em que  $\Phi(\cdot)$  agrega as etapas de projeção em subespaço por PCA, estimação das cumulantes de interesse e normalização do vetor de características. O inteiro  $B$  denota,

em cada experimento, o número de atributos retidos após o ranqueamento e a seleção conduzidos na Seção 4.4.

Ao longo desta tese, assumiremos que  $S(\cdot)$  é localmente contínua no interior da ZND, no sentido de que pequenas variações nos desbalanços  $(\Delta P, \Delta Q)$  produzem pequenas variações em  $S(\delta)$ . De forma mais específica, admitimos que existem constantes  $K > 0$  e  $\rho > 0$  tais que

$$\|S(\delta_1) - S(\delta_2)\| \leq K \|\delta_1 - \delta_2\|, \quad \forall \delta_1, \delta_2 \in \mathcal{D}_{\text{ZND}} \text{ com } \|\delta_1 - \delta_2\| \leq \rho. \quad (4.7)$$

Essa hipótese de regularidade é compatível com o fato de que o modelo de rede e o processo de estimação de cumulantes dependem de forma contínua do ponto de operação [87], e será utilizada posteriormente para interpretar o comportamento do MDI quando  $\delta$  varia no interior da ZND.

## 4.6 CLASSIFICADORES E OTIMIZAÇÃO

Esta seção descreve as arquiteturas de classificação e o protocolo de otimização adotados para o MDI, conectando o vetor de atributos obtido nas Seções 4.2–4.4 ao treinamento e seleção de modelos. O desenho metodológico foi organizado em duas fases complementares:

- (i) **Fase de otimização em janela fixa** (Subseção 4.6.4): todos os experimentos de busca e seleção são conduzidos sobre janelas fixas de 2 ciclos (Subseção 4.1.1), com  $\text{SNR} = 70$  dB, partindo de um conjunto inicial de  $B = 15$  atributos (cumulantes de 2<sup>a</sup>–4<sup>a</sup> ordens). Nesta fase definem-se: (a) o número de partições  $K$  da validação cruzada; (b) o subconjunto ótimo de atributos para cada cardinalidade  $B \in \{1, \dots, 15\}$ ; (c) o número de neurônios na camada escondida da MLP; e (d) a combinação final arquitetura  $\times$  vetor de atributos.
- (ii) **Fase de treino/avaliação em janela deslizante** (Subseção 4.6.5): com a topologia e os atributos já fixados na Fase (i), a rede é re-treinada sobre janelas deslizantes  $4 \rightarrow 3 \times 2$  ciclos (Subseção 4.1.2) e avaliada por *hold-out* estratificado (30% teste), respeitando agrupamento por instância temporal. A decisão por evento considera a fusão das três sub-janelas.

Salvo menção em contrário, as classes são tri-rótulo  $\{N, I, F\}$ , os atributos são normalizados em  $[-1, 1]$ , e o vetor de entrada possui dimensionalidade  $B \in \{1, \dots, 15\}$ . As métricas reportadas são acurácia média  $\pm$  desvio-padrão sobre as partições.

### 4.6.1 Preparação dos dados e alvos

O vetor de características por janela é obtido a partir das séries de cumulantes descritas na Seção 4.3 e do ranqueamento/seleção por FDR na Seção 4.4. Adota-se

$B = 15$  atributos por janela, estratificados por ordem ( $B_r = 5$  para cada  $r \in \{2, 3, 4\}$ ). Sejam  $\mathcal{P}_2, \mathcal{P}_3, \mathcal{P}_4$  os índices de atraso selecionados na Seção 4.4 para cada ordem; o vetor final concatena, para cada janela, os elementos correspondentes  $\{\widehat{k}_2^{(1)}(\tau)\}_{\tau \in \mathcal{P}_2}$ ,  $\{\widehat{k}_3^{(1)}(\tau)\}_{\tau \in \mathcal{P}_3}$  e  $\{\widehat{k}_4^{(1)}(\tau)\}_{\tau \in \mathcal{P}_4}$ .

Seja  $\mathcal{G} = \{N, I, F\}$  o conjunto de classes e  $S$  o número de janelas por classe. O conjunto total tem  $N = C \cdot S$  amostras, com  $C = |\mathcal{G}| = 3$ . As amostras são organizadas em blocos contíguos por classe na ordem  $\{N, I, F\}$ . A matriz de dados e a de alvos são então

$$\mathbf{Z} \in \mathbb{R}^{N \times B}, \quad \mathbf{Y} \in \mathbb{R}^{N \times 3},$$

em que cada linha de  $\mathbf{Z}$  contém um vetor de  $B$  atributos de uma janela, e cada linha de  $\mathbf{Y}$  é a codificação *one-of-K* consistente com a ordem de classes adotada:  $(1, 0, 0)$  para N;  $(0, 1, 0)$  para I;  $(0, 0, 1)$  para F.

**Normalização.** Antes da partição em *folds*, aplica-se uma reescala linear *por atributo* para o intervalo  $[-1, 1]$ . Seja  $z_{i,b}$  o valor do atributo  $b$  na amostra  $i$ , e definam-se os limites globais

$$z_b^{\min} = \min_i z_{i,b}, \quad z_b^{\max} = \max_i z_{i,b}.$$

O valor normalizado é

$$\tilde{z}_{i,b} = 2 \frac{z_{i,b} - z_b^{\min}}{z_b^{\max} - z_b^{\min}} - 1,$$

equivalente à função do MATLAB “`mapminmax`”. Os limites  $(z_b^{\min}, z_b^{\max})$  são estimados *uma única vez* no conjunto disponível e aplicados de forma consistente a todos os *folds*.

**Validação cruzada.** Salvo menção em contrário, emprega-se validação cruzada *K-fold* com  $K = 10$  e partição aleatória uniforme de amostras em  $K$  subconjuntos (MATLAB: `crossvalind('Kfold', N, 10)`). Em cada *fold*  $i$ , define-se o par treino–teste por índice e constrói-se  $\mathbf{Z}_{\text{train}}, \mathbf{Y}_{\text{train}}$  e  $\mathbf{Z}_{\text{test}}, \mathbf{Y}_{\text{test}}$  por seleção de linhas. Denotamos por  $n_{\text{train}}$  e  $n_{\text{test}}$  o número de amostras de treino e teste em cada *fold*, com  $n_{\text{train}} = \frac{K-1}{K}N$  e  $n_{\text{test}} = \frac{1}{K}N$ .

#### 4.6.2 Arquiteturas consideradas

Foram avaliados dois classificadores complementares em custo e capacidade de modelagem: (i) um linear por Mínimos Quadrados (do inglês, *Least Squares*) (LS) multiclasse e (ii) um Perceptron Multicamadas (do inglês, *Multilayer Perceptron*) (MLP) com uma única camada escondida. A escolha equilibra baixo custo computacional em tempo de execução e capacidade de capturar não-linearidades [84].

#### 4.6.2.1 LS multiclasse via pseudoinversa.

O classificador LS é obtido por mínimos quadrados multiclasse, com termo de viés explícito. Dado  $\tilde{\mathbf{Z}}_{\text{train}} = [\mathbf{Z}_{\text{train}} \ \mathbf{1}] \in \mathbb{R}^{n_{\text{train}} \times (B+1)}$  e  $\mathbf{Y}_{\text{train}} \in \mathbb{R}^{n_{\text{train}} \times 3}$ , estima-se  $\mathbf{W} \in \mathbb{R}^{(B+1) \times 3}$  resolvendo

$$\mathbf{W}^* = \arg \min_{\mathbf{W}} \|\tilde{\mathbf{Z}}_{\text{train}} \mathbf{W} - \mathbf{Y}_{\text{train}}\|_F^2.$$

A solução é computada pela pseudoinversa de Moore–Penrose (MATLAB `PINV`, baseada em Decomposição em Valores Singulares (do inglês, *Singular Value Decomposition*) (SVD)):

$$\mathbf{W}^* = \tilde{\mathbf{Z}}_{\text{train}}^+ \mathbf{Y}_{\text{train}}.$$

Se  $\tilde{\mathbf{Z}}_{\text{train}}$  tem posto completo em colunas, então  $\tilde{\mathbf{Z}}_{\text{train}}^+ = (\tilde{\mathbf{Z}}_{\text{train}}^T \tilde{\mathbf{Z}}_{\text{train}})^{-1} \tilde{\mathbf{Z}}_{\text{train}}^T$  e recupera-se a solução das equações normais. Se for posto-deficiente, `PINV` entrega a solução de menor norma entre as infinitas soluções que minimizam o erro quadrático, de acordo com o truncamento de singulares definido internamente pela tolerância padrão. Seja  $\tilde{\mathbf{Z}}_{\text{test}} = [\mathbf{Z}_{\text{test}} \ \mathbf{1}] \in \mathbb{R}^{n_{\text{test}} \times (B+1)}$ , a predição no teste é

$$\mathbf{S} = \tilde{\mathbf{Z}}_{\text{test}} \mathbf{W}^* \in \mathbb{R}^{n_{\text{test}} \times 3}, \quad \hat{g}_j = \arg \max_{c \in \{1,2,3\}} S_{j,c}.$$

#### 4.6.2.2 MLP com uma camada escondida.

Considera-se uma rede MLP com uma única camada escondida e camada de saída com três neurônios (um por classe). Seja  $B$  o número de atributos ( $B \leq 15$ ),  $C = 3$  o número de classes e, em um dado *fold*,  $n_{\text{train}}$  e  $n_{\text{test}}$  as contagens de treino e teste, respectivamente. As matrizes empregadas são

$$\mathbf{Z}_{\text{train}} \in \mathbb{R}^{n_{\text{train}} \times B}, \quad \mathbf{Y}_{\text{train}} \in \mathbb{R}^{n_{\text{train}} \times C}, \quad \mathbf{Z}_{\text{test}} \in \mathbb{R}^{n_{\text{test}} \times B},$$

em que cada linha representa uma amostra;  $\mathbf{Y}$  utiliza codificação *one-of-K* com a ordem de classes  $\{N, I, F\}$ .

##### 4.6.2.2.1 Arquitetura e ativações.

Seja  $n_H$  o número de neurônios na camada escondida. Os parâmetros são

$$\mathbf{W}_1 \in \mathbb{R}^{B \times n_H}, \quad \mathbf{b}_1 \in \mathbb{R}^{n_H}, \quad \mathbf{W}_2 \in \mathbb{R}^{n_H \times C}, \quad \mathbf{b}_2 \in \mathbb{R}^C.$$

A propagação direta é dada por

$$\begin{aligned} \mathbf{H} &= \phi(\mathbf{Z}_{\text{train}} \mathbf{W}_1 + \mathbf{1}_{n_{\text{train}}} \mathbf{b}_1^T) \in \mathbb{R}^{n_{\text{train}} \times n_H}, \\ \hat{\mathbf{Y}}_{\text{train}} &= \phi(\mathbf{H} \mathbf{W}_2 + \mathbf{1}_{n_{\text{train}}} \mathbf{b}_2^T) \in \mathbb{R}^{n_{\text{train}} \times C}, \end{aligned}$$

onde  $\mathbf{1}_n$  denota um vetor coluna de 1's de tamanho  $n$ . Adota-se  $\phi(u) = \tanh(u)$ , aplicada elemento a elemento, com definição

$$\tanh(u) = \frac{\sinh(u)}{\cosh(u)} = \frac{e^{2u} - 1}{e^{2u} + 1} \in (-1,1), \quad \frac{d}{du} \tanh(u) = 1 - \tanh^2(u).$$

A  $\tanh$  é simétrica em torno de zero e saturante; a normalização dos atributos em  $[-1,1]$  (Seção 4.6.1) é compatível com essa faixa.

#### 4.6.2.2.2 Treinamento por Levenberg–Marquardt.

A MLP é treinada minimizando o Erro Médio Quadrático (do inglês, *Mean Squared Error*) (MSE) entre saídas e alvos. Escrevendo  $\Theta = \{\mathbf{W}_1, \mathbf{b}_1, \mathbf{W}_2, \mathbf{b}_2\}$  e  $\hat{\mathbf{Y}}_{\text{train}}(\Theta)$  como a saída da rede no treino, define-se o vetor de resíduos

$$\mathbf{e}(\Theta) = \text{vec}(\hat{\mathbf{Y}}_{\text{train}}(\Theta) - \mathbf{Y}_{\text{train}}) \in \mathbb{R}^{n_{\text{train}}C},$$

e a função de custo

$$\mathcal{L}(\Theta) = \frac{1}{n_{\text{train}}C} \|\mathbf{e}(\Theta)\|_2^2 = \frac{1}{n_{\text{train}}C} \sum_{i=1}^{n_{\text{train}}C} e_i(\Theta)^2.$$

O método LM utiliza a aproximação Gauss–Newton do Hessiano,  $\mathcal{H}(\Theta) \equiv \nabla^2 \mathcal{L}(\Theta) \approx \mathbf{J}^\top \mathbf{J}$ , e o gradiente  $\mathbf{g} = \mathbf{J}^\top \mathbf{e}$ , onde  $\mathbf{J} = \frac{\partial \mathbf{e}}{\partial \theta} \in \mathbb{R}^{(n_{\text{train}}C) \times P}$  é a jacobiana dos resíduos em relação ao vetor de parâmetros empilhados  $\theta = \text{vec}(\Theta) \in \mathbb{R}^P$ , com  $P = B n_H + n_H + n_H C + C$ . Em cada iteração, resolve-se

$$(\mathbf{J}^\top \mathbf{J} + \mu \mathbf{I}) \Delta \theta = -\mathbf{J}^\top \mathbf{e}, \quad \theta \leftarrow \theta + \Delta \theta,$$

com parâmetro de amortecimento  $\mu > 0$ . Para  $\mu$  pequeno, o passo se aproxima de Gauss–Newton; para  $\mu$  grande, aproxima um passo de descida íngreme. O parâmetro  $\mu$  é ajustado iterativamente: se  $\mathcal{L}$  reduz,  $\mu$  é diminuído; caso contrário,  $\mu$  é aumentado e o passo é reavaliado. O treinamento é interrompido ao atingir um limite de épocas (até 100) ou um alvo de erro ( $10^{-3}$ ); não se utiliza validação interna (o controle de generalização é feito pelo protocolo *K-fold* externo).

#### 4.6.2.2.3 Predição e decisão.

Para o conjunto de teste,

$$\mathbf{H}_{\text{test}} = \tanh(\mathbf{Z}_{\text{test}} \mathbf{W}_1 + \mathbf{1}_{n_{\text{test}}} \mathbf{b}_1^\top), \quad \hat{\mathbf{Y}}_{\text{test}} = \tanh(\mathbf{H}_{\text{test}} \mathbf{W}_2 + \mathbf{1}_{n_{\text{test}}} \mathbf{b}_2^\top).$$

O rótulo estimado para a amostra  $j$  é

$$\hat{g}_j = \arg \max_{c \in \{1,2,3\}} (\hat{\mathbf{Y}}_{\text{test}})_{j,c}.$$

Como a saída final é  $\tanh$ , os valores são escores em  $(-1,1)$ , sendo a decisão feita por *argmax* entre as três saídas.

### 4.6.3 Modelo de decisão no espaço de atributos

Do ponto de vista estatístico, o MDI implementa um classificador

$$h : \mathbb{R}^B \rightarrow \{N, I, F\}, \quad \hat{y} = h(S(\delta)), \quad (4.8)$$

treinado a partir de amostras rotuladas  $(S(\delta^{(i)}), y^{(i)})$ , obtidas para diferentes pontos de operação e modos de funcionamento do sistema.

Para fins de análise conceitual, é conveniente considerar, no espaço de atributos  $\mathbb{R}^B$ , os subconjuntos

$$\mathcal{X}_k = \{S(\delta^{(i)}) : y^{(i)} = k\}, \quad k \in \{N, I, F\}, \quad (4.9)$$

que reúnem os vetores de características associados a cada classe.

Uma forma simples de quantificar o quão separadas essas classes podem estar é por meio de uma margem geométrica

$$\gamma = \min_{k \neq \ell} \min_{s \in \mathcal{X}_k} \text{dist}(s, \mathcal{X}_\ell), \quad \text{dist}(s, \mathcal{X}_\ell) = \min_{t \in \mathcal{X}_\ell} \|s - t\|. \quad (4.10)$$

Quando os aglomerados  $\mathcal{X}_k$  são bem separados no espaço de atributos, o valor de  $\gamma$  tende a ser maior; quando há sobreposição significativa entre classes,  $\gamma$  pode se aproximar de zero. Essa noção será utilizada adiante na interpretação das fronteiras de decisão aprendidas pelo classificador e de sua capacidade de generalização.

### 4.6.4 Protocolo de Otimização

A otimização foi organizada em quatro rodadas encadeadas, mantendo fixas as convenções de preparação dos dados. Em todas as comparações, os classificadores LS e MLP operam sobre o mesmo vetor de características (mesma cardinalidade  $B$ ), de modo que ambos recebem o mesmo número de parâmetros de entrada. As rodadas de otimização e seu objetivo geral foram:

- (i) **Rodada 1 — validação cruzada.** Definir o número de partições  $K$  da validação cruzada por  $K$ -fold, comparando LS (linear, com viés) e MLP (uma camada escondida,  $n_h = 20$ , ativações tanh).
- (ii) **Rodada 2 — atributos.** Para cada cardinalidade  $B \in \{1, \dots, 15\}$ , selecionar a melhor combinação de atributos dentre os 15 candidatos (Seção 4.4), avaliando LS e MLP sob o  $K$  definido na Rodada 1.
- (iii) **Rodada 3 — tamanho da camada escondida.** Para cada cardinalidade  $B$  e seu melhor subconjunto (Rodada 2), escolher  $n_h$  (número de neurônios na camada escondida da MLP) por varredura discreta. Nesta rodada apenas a MLP é considerada.

- (iv) **Rodada 4 — arquitetura final.** Fixados  $B$  (subconjunto ótimo) e  $n_h$  (Rodadas 2–3), consolidar a MLP final e compará-la ao LS como referência linear para compor o par arquitetura  $\times$  vetor de características adotado nas avaliações subsequentes.

A seguir cada rodada é detalhada.

#### 4.6.4.1 Validação Cruzada

Esta rodada teve como objetivo escolher o número de partições  $K$  do protocolo  $K$ -fold que produza estimativas de desempenho estáveis e com baixo viés.

Foram avaliados  $K \in \{5, 10, 15, 20\}$  para ambas as arquiteturas, utilizando todas as  $B = 15$  características e as três classes  $\{N, I, F\}$ . Em cada *fold*  $i \in \{1, \dots, K\}$ , treina-se no subconjunto de treino e avalia-se no subconjunto de teste, obtendo a acurácia por *fold*:

$$A_K^{(i)} \in [0, 1], \quad \bar{A}_K = \frac{1}{K} \sum_{i=1}^K A_K^{(i)}, \quad s_K = \sqrt{\frac{1}{K-1} \sum_{i=1}^K (A_K^{(i)} - \bar{A}_K)^2}.$$

Reporta-se  $\bar{A}_K \pm s_K$  e registram-se as matrizes de confusão por *fold*.

Para cada arquitetura, testou-se a hipótese nula de igualdade das médias entre os níveis de  $K$  por ANOVA de um fator (níveis  $K$ ), tomando como observações os  $\{A_K^{(i)}\}$ . Em caso de significância ( $\alpha = 0,05$ ), aplicou-se a Diferença Significativa Honesta (do inglês, *Honestly Significant Difference*) (HSD) de Tukey entre pares de níveis.

O critério de escolha adotado foi

- (a) Se a ANOVA indicou diferença significativa, escolheu-se um nível  $K$  pertencente ao grupo de maior média (HSD de Tukey); entre empates, preferiu-se o menor desvio-padrão  $s_K$ .
- (b) Se a ANOVA não indicou diferença significativa, adotou-se o  $K$  com menor  $s_K$  (estimativa mais estável); persistindo empate, preferiu-se o menor  $K$  (menor custo).

O valor selecionado é denotado por  $K^*$  e é utilizado nas Rodadas 2–4.

#### 4.6.4.2 Atributos

Esta rodada tem por objetivo, para cada cardinalidade  $B$ , selecionar a melhor combinação de atributos a partir de um conjunto candidato (cumulantes previamente

selecionados; ver Sec. 4.4), avaliando LS e MLP sob o protocolo de validação cruzada com  $K = K^*$  definido na Seção 4.6.4.1.

Seja  $\mathcal{I} = \{1, \dots, M\}$ ,  $M = 15$ , o índice de atributos candidatos. Para cada cardinalidade  $B \in \{1, \dots, 15\}$ , considera-se o conjunto de combinações

$$\mathcal{C}_B = \{\mathcal{S} \subset \mathcal{I} : |\mathcal{S}| = B\}, \quad |\mathcal{C}_B| = \binom{M}{B}.$$

Para cada  $\mathcal{S} \in \mathcal{C}_B$ , formam-se as matrizes de dados restringindo  $\mathbf{Z}$  às colunas (atributos) em  $\mathcal{S}$ .

Em cada combinação  $\mathcal{S}$ , ambos os classificadores são avaliados por validação cruzada  $K$ -fold, produzindo, para o modelo  $m \in \{\text{LS}, \text{MLP}\}$ , as acurácias por fold  $\{A_{B,\mathcal{S},m}^{(i)}\}_{i=1}^K$ , a média

$$\bar{A}_{B,\mathcal{S},m} = \frac{1}{K} \sum_{i=1}^K A_{B,\mathcal{S},m}^{(i)}$$

e o desvio-padrão

$$s_{B,\mathcal{S},m} = \sqrt{\frac{1}{K-1} \sum_{i=1}^K (A_{B,\mathcal{S},m}^{(i)} - \bar{A}_{B,\mathcal{S},m})^2}.$$

Para cada cardinalidade  $B$  e para cada modelo  $m$ , aplica-se ANOVA de um fator sobre as acurácias por fold  $\{A_{B,\mathcal{S},m}^{(i)}\}$  ao longo de todas as combinações  $\mathcal{S} \in \mathcal{C}_B$ . Havendo significância global ( $\alpha = 0,05$ ), usa-se o teste HSD de Tukey para identificar o(s) grupo(s) de maior média. Mantêm-se, então, as combinações  $\mathcal{G}_{B,m}$  cuja média não difere de forma significativa desses grupos para o modelo  $m$ . Caso a ANOVA não indique significância, toma-se  $\mathcal{G}_{B,m} \equiv \mathcal{C}_B$ . A concordância entre modelos retém a interseção

$$\mathcal{G}_B = \mathcal{G}_{B,\text{LS}} \cap \mathcal{G}_{B,\text{MLP}}.$$

Entre as combinações remanescentes  $\mathcal{S} \in \mathcal{G}_B$ , define-se um escore que privilegia alta acurácia média e baixa variabilidade, agregando os dois modelos:

$$F_B(\mathcal{S}) = \underbrace{\frac{1}{2}(\bar{A}_{B,\mathcal{S},\text{LS}} + \bar{A}_{B,\mathcal{S},\text{MLP}})}_{\text{acurácia média}} - \underbrace{\frac{1}{2}(s_{B,\mathcal{S},\text{LS}} + s_{B,\mathcal{S},\text{MLP}})}_{\text{penalização por instabilidade}}.$$

Escolhe-se, para cada  $B$ , a combinação ótima

$$\mathcal{S}_B^* = \arg \max_{\mathcal{S} \in \mathcal{G}_B} F_B(\mathcal{S}).$$

Esses subconjuntos  $\mathcal{S}_B^*$  são usados na rodada seguinte, dedicada ao ajuste do tamanho da camada escondida da MLP.

#### 4.6.4.3 Camada escondida

Nesta rodada ajusta-se o número de neurônios na camada escondida da MLP, mantendo fixos, para cada cardinalidade  $B$ , o subconjunto ótimo de atributos  $S_B^*$  obtido na Rodada 2. Os demais hiperparâmetros da rede (ativação  $\tanh$ , função de treino LM, épocas e meta de erro) permanecem conforme a Seção 4.6.2.

Para cada  $B$ , considerou-se o conjunto de contagens de neurônios

$$\mathcal{N} = \{2, 3, \dots, 20\}.$$

Para cada  $n_h \in \mathcal{N}$ , avaliou-se a MLP por validação cruzada  $K$ -fold com  $K = K^*$  (Seção 4.6.4.1), usando somente as colunas de  $Z$  correspondentes a  $S_B^*$ . Em cada  $fold$   $i \in \{1, \dots, K^*\}$ , a rede é treinada no subconjunto de treino (sem partição interna) e avaliada no de teste, produzindo as acurácias

$$A_{B,n_h}^{(i)} \in [0,1], \quad \bar{A}_{B,n_h} = \frac{1}{K^*} \sum_{i=1}^{K^*} A_{B,n_h}^{(i)}, \quad s_{B,n_h} = \sqrt{\frac{1}{K^* - 1} \sum_{i=1}^{K^*} (A_{B,n_h}^{(i)} - \bar{A}_{B,n_h})^2}.$$

As partições  $K$ -fold são amostradas aleatoriamente para cada  $n_h$  sob semente fixa, garantindo reprodutibilidade.

Aplicou-se ANOVA de um fator sobre as acurácias por  $fold$   $\{A_{B,n_h}^{(i)}\}$  ao longo de todos os  $n_h \in \mathcal{N}$ . Havendo significância global ( $\alpha = 0,05$ ), utilizou-se o teste HSD de Tukey para identificar os grupos de maior média; mantêm-se, então, os valores de  $n_h$  cuja média não difere de forma significativa desses grupos. Na ausência de significância, todos os  $n_h \in \mathcal{N}$  são mantidos.

Entre os  $n_h$  remanescentes, a escolha é feita por um escore que privilegia alta acurácia e baixa variabilidade, com penalização leve por complexidade:

$$J_B(n_h) = \bar{A}_{B,n_h} - s_{B,n_h} - \lambda \frac{n_h}{\max(\mathcal{N})}, \quad \lambda = 0,01.$$

Seleciona-se, para cada  $B$ , o tamanho ótimo da camada escondida

$$n_h^*(B) = \arg \max_{n_h} J_B(n_h).$$

Em caso de empate em  $J_B$ , adota-se o menor  $s_{B,n_h}$ ; persistindo, prefere-se o menor  $n_h$ . O par  $(S_B^*, n_h^*(B))$  alimenta a Rodada 4.

#### 4.6.4.4 Arquiteturas e vetor de características finais

Com os melhores subconjuntos de atributos por cardinalidade  $S_B^*$  (Rodada 2) e os tamanhos ótimos de camada escondida  $n_h^*(B)$  (Rodada 3) já definidos, esta rodada consolida a arquitetura final da MLP por comparação direta entre as candidatas indexadas por  $B$ .

Para cada  $B$  considerado, forma-se uma rede candidata

$$\mathcal{M}_B = (B, \mathcal{S}_B^*, n_h^*(B)),$$

mantendo fixos os demais hiperparâmetros (ativação  $\tanh$ , treino LM, número de épocas e meta de erro), e avalia-se cada  $\mathcal{M}_B$  por validação cruzada  $K$ -fold com  $K = K^*$  (Seção 4.6.4.1). Em cada *fold*  $i$  obtém-se a acurácia  $A_B^{(i)}$ ; reportam-se, por candidata,

$$\bar{A}_B = \frac{1}{K^*} \sum_{i=1}^{K^*} A_B^{(i)}, \quad s_B = \sqrt{\frac{1}{K^* - 1} \sum_{i=1}^{K^*} (A_B^{(i)} - \bar{A}_B)^2}.$$

Realiza-se ANOVA de um fator sobre as acurácias por *fold* das candidatas  $\{\mathcal{M}_B\}$ ; havendo significância ( $\alpha = 0,05$ ), aplica-se o teste HSD de Tukey para identificar os grupos de maior média e retêm-se os  $B$  cuja média não difere significativamente desses grupos. Na ausência de significância, todas as candidatas seguem adiante.

Entre as candidatas remanescentes, a seleção final usa um escore que combina desempenho e estabilidade, com leve penalização por complexidade da camada escondida:

$$J(B) = \bar{A}_B - s_B - \lambda \frac{n_h^*(B)}{\max(\mathcal{N})}, \quad \lambda = 0,01.$$

Define-se então

$$B^* = \arg \max_B J(B), \quad \text{MLP}^* \equiv \mathcal{M}_{B^*} = (B^*, \mathcal{S}_{B^*}^*, n_h^*(B^*)).$$

Critérios de desempate são, nesta ordem: menor  $s_B$  e menor  $n_h^*(B)$ . Para referência linear, o classificador LS é também treinado e avaliado com o mesmo vetor de características de dimensão  $B^*$ .

#### 4.6.5 Protocolo de Validação e Controle de Viés

Uma vez definida a arquitetura ótima do classificador na fase de otimização (Seção 4.6.4), a etapa subsequente é dedicada a treinar e validar o modelo final de forma a garantir sua validade externa e capacidade de generalização. Para tal, implementa-se um protocolo projetado para emular a operação em tempo real (usando janelas deslizantes) e, crucialmente, para impedir qualquer forma de vazamento de informação entre os estágios de desenvolvimento e de teste final.

Este protocolo se fundamenta em três pilares essenciais:

1. **Congelamento do Fluxo de Extração de Atributos:** Todos os parâmetros e hiperparâmetros que definem a transformação do sinal bruto em um vetor de características são herdados da fase de otimização e permanecem imutáveis. Isso inclui:

- A matriz de rotação PCA fixa ( $\bar{R}$ ), estimada exclusivamente sobre os eventos de ilhamento do conjunto de otimização (Seção 4.2.2).
- A dimensionalidade da projeção.
- O conjunto exato de  $B^*$  atrasos das cumulantes ( $\mathcal{S}_{B^*}^*$ ), selecionado na Seção 4.6.4.4.

Esta medida é fundamental para evitar o viés de seleção, garantindo que a performance final não seja um artefato de um ajuste otimista sobre os dados de teste. O fluxo de atributos, uma vez calibrado, é tratado como um sistema fixo e não adaptativo.

2. **Partição Estratificada por Instância Temporal:** A geração de dados via janela deslizante ( $4 \rightarrow 3 \times 2$  ciclos) introduz uma forte correlação entre as três subjanelas ( $\mathcal{J}_1, \mathcal{J}_2, \mathcal{J}_3$ ) extraídas da mesma forma de onda original. Uma partição aleatória simples (a nível de subjanela) levaria a um vazamento de informação, pois o modelo poderia ser treinado em  $\mathcal{J}_1$  e testado em sua “irmã” quase idêntica  $\mathcal{J}_2$ . Para impedir isso:

- As três subjanelas geradas a partir da mesma instância temporal são tratadas como um grupo atômico e indivisível.
- A divisão do conjunto de dados em subconjuntos de treino e teste é sempre realizada a nível de instância temporal, garantindo que todas as três subjanelas de um evento pertençam exclusivamente a um único subconjunto.

3. **Avaliação Final em Conjunto de Dados Independente:** A prova definitiva de generalização é obtida através da avaliação do modelo treinado em um conjunto de dados completamente inédito, o banco fora da grade (Seção 3.3.3). Este conjunto:

- Contém 100 perfis de carga e geração que não foram vistos em nenhuma etapa anterior de otimização ou treinamento.
- É processado exatamente com o mesmo fluxo (Pilar 1).
- Serve como um teste cego, onde o desempenho medido reflete a capacidade real do MDI de operar em condições novas e desconhecidas.

#### 4.6.5.1 *Treinamento Final e Avaliação de Desempenho*

O procedimento de treinamento e validação segue, então, os fluxos descritos:

1. **Treinamento em Janela Deslizante:** O modelo final (MLP<sup>\*</sup> e o LS de referência) é treinado utilizando os dados do conjunto grade  $3 \times 3$  (Seção 3.3.2), processados com o mecanismo de janela deslizante. Adota-se uma partição *hold-out* (70% treino, 30% teste), respeitando o agrupamento por instância temporal (Pilar 2).

O treinamento de todas as três subjanelas ( $\mathcal{J}_1, \mathcal{J}_2, \mathcal{J}_3$ ) assegura que o modelo aprenda a lidar com as pequenas variações temporais inerentes ao processo de varredura, o que é essencial para a robustez de decisão.

2. **Validação Independente:** Após o treinamento, os classificadores (com seus pesos e parâmetros congelados) são aplicados ao conjunto de validação fora da grade. Para cada evento neste conjunto, o sistema gera três previsões independentes, uma para cada subjanela. As matrizes de confusão e métricas de desempenho são calculadas primeiramente a nível de subjanela, para então serem consolidadas em uma decisão final por evento, conforme detalhado na seção seguinte.

Este protocolo metodológico garante que os resultados reportados sejam uma estimativa conservadora e confiável do desempenho do MDI em um cenário operacional realista.

#### 4.7 FLUXO EM EXECUÇÃO

As seções anteriores detalharam, de forma modular, os componentes de projeto, otimização e validação do MDI proposto. Esta seção final consolida essas etapas em um único fluxograma operacional, ilustrando a sequência de processamento que seria executada em tempo real para classificar um novo evento a partir das tensões trifásicas monitoradas.

É fundamental reiterar a distinção entre a fase de projeto (*offline*), que envolveu a busca exaustiva por hiperparâmetros e a seleção de atributos com custo computacional elevado, e a fase de execução (*online*), descrita aqui. O fluxo em execução herda os resultados da fase de projeto, isto é, a matriz de projeção fixa  $\bar{\mathbf{R}}$ , o subconjunto ótimo de  $B^*$  atrasos de cumulantes e a arquitetura do classificador treinado, para operar com latência mínima e baixo custo computacional.

A Figura 12 apresenta este fluxo de trabalho otimizado, que pode ser decomposto em duas fases interligadas:

1. **Calibração Offline (Etapa Única):** Executada uma única vez, esta fase utiliza o conjunto de dados de otimização para calcular e “congelar” os componentes-chave do sistema. Conforme ilustrado no lado esquerdo da figura, esta etapa produz:
  - A matriz de rotação média  $\bar{\mathbf{R}}$ , que encapsula a direção de máxima variância dos eventos de ilhamento.
  - O conjunto fixo de atrasos  $\{\mathcal{P}_2, \mathcal{P}_3, \mathcal{P}_4\}$ , que representa as características mais discriminantes.

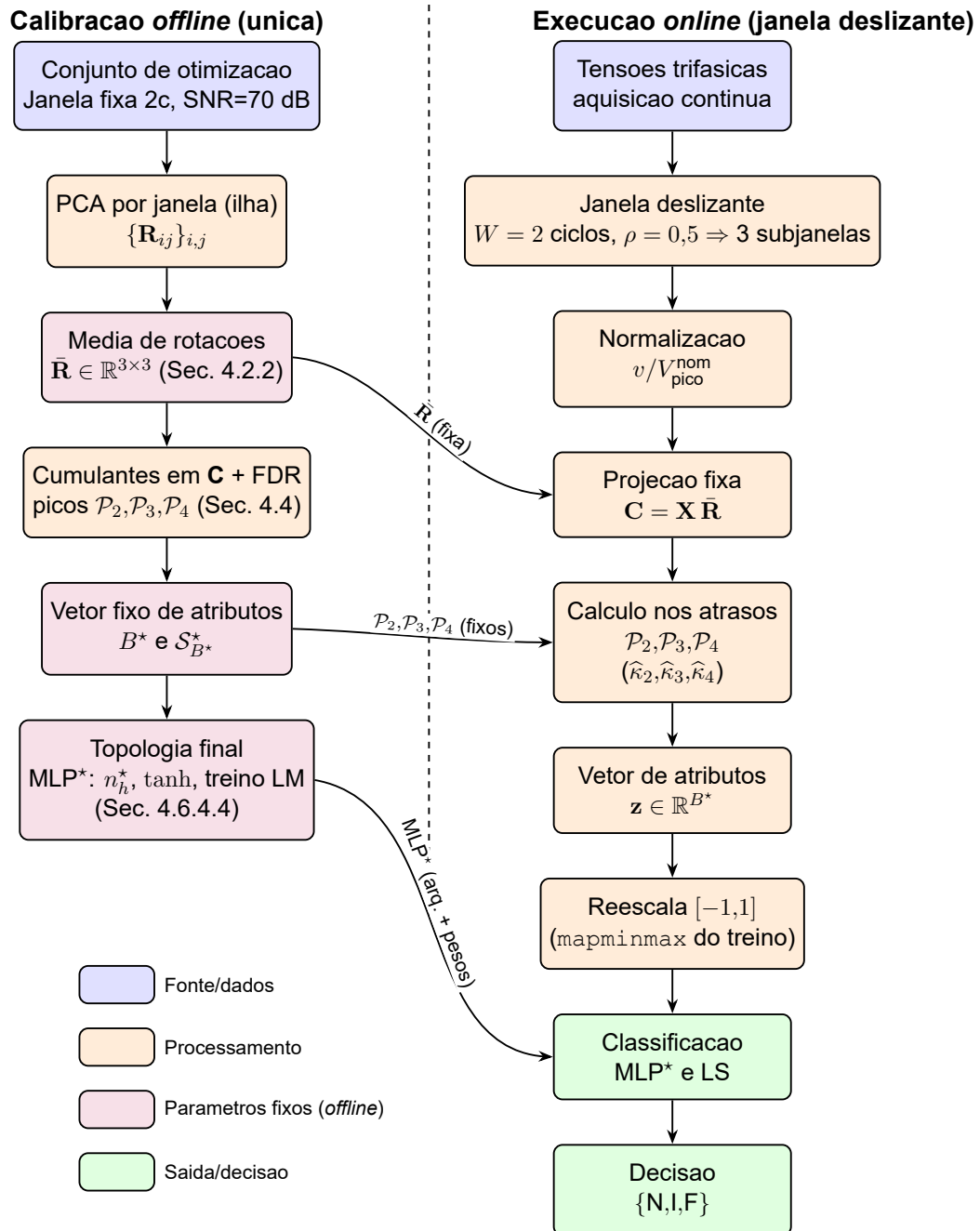
- A arquitetura e os pesos da rede MLP\* final, já treinada e otimizada.

Estes três resultados são os únicos artefatos transferidos para a fase de execução.

2. **Execução Online (Tempo Real):** Para cada nova forma de onda adquirida continuamente, o sistema executa o fluxo de baixo custo computacional ilustrado no lado direito da figura:

- (i) O sinal é recortado, normalizado e segmentado em três subjanelas sobrepostas.
- (ii) Cada subjanela é projetada em componentes principais usando a matriz fixa  $\bar{\mathbf{R}}$ .
- (iii) O sistema calcula as cumulantes apenas nos atrasos pré-selecionados  $\{\mathcal{P}_k\}$ , formando o vetor final de atributos.
- (iv) O vetor de atributos é reescalado e classificado pela MLP\* já treinada.
- (v) As três previsões (uma por subjanela) são fundidas para gerar a decisão final sobre o estado do sistema.

Este design desacoplado é a chave para a viabilidade do método: a complexidade computacional é confinada à fase de projeto, permitindo que o fluxo em tempo real opere com máxima eficiência, equilibrando alta precisão com os requisitos de baixa latência indispensáveis para aplicações de proteção e controle em sistemas de energia elétrica.

Figura 12 – Fluxo de calibração *offline* e execução *online*.

Legenda: Fluxo operacional em duas fases. Na calibração *offline*, fixam-se  $\bar{\mathbf{R}}$ , os atrasos selecionados por FDR e a arquitetura MLP\*. Na execução *online*, cada nova forma de onda é janelada, normalizada, projetada, descrita por cumulantes nos atrasos selecionados e classificada em  $\{N, I, F\}$ .

## 5 RESULTADOS

Este capítulo apresenta os resultados obtidos com a execução do fluxo proposto no Cap. 4, aplicado aos sinais sintetizados no Cap. 3. Segue-se a filosofia de projeto adotada:

- (i) a etapa de otimização baseada em janelas fixas de 2 ciclos (Sec. 4.1.1) para selecionar atributos e hiperparâmetros;
- (ii) a etapa operacional sobre janelas deslizantes  $4 \rightarrow 3 \times 2$  ciclos (Sec. 4.1.2), que emula o monitoramento quase em tempo real.

É reportado primeiro a configuração efetiva usada na execução. Em seguida, apresentamos os achados das rodadas de otimização (escolha de  $K$ , subconjuntos de atributos e  $n_h$  da MLP), os resultados em treino/teste com janela deslizante e, por fim, a validação independente no banco fora da grade. Ao longo das subseções, mantemos os vínculos com o desenho metodológico (Caps. 3–4) e utilizamos as mesmas métricas e convenções.

### 5.1 CONFIGURAÇÃO EXPERIMENTAL E REPRODUTIBILIDADE

#### 5.1.1 Dados e janelamento

Todos os ensaios foram gerados a partir da microrrede de referência de 250 kW (Cap. 3). As formas de onda brutas têm duração  $T_{\text{sim}} = 0,5\text{ s}$  e evento ancorado em  $t_0 = 0,35\text{ s}$  (Sec. 3.1.7). Na etapa de otimização (Sec. 4.1.1) usou-se uma janela fixa de 2 ciclos centrada em  $t_0$ , com *jitter* uniforme de  $\pm \frac{1}{4}$  de ciclo.

Para treino/teste e validação independente aplicou-se um recorte de 4 ciclos com o mesmo *jitter*, seguido de janela deslizante  $4 \rightarrow 3 \times 2$  ciclos com sobreposição de 50% (Sec. 4.1.2).

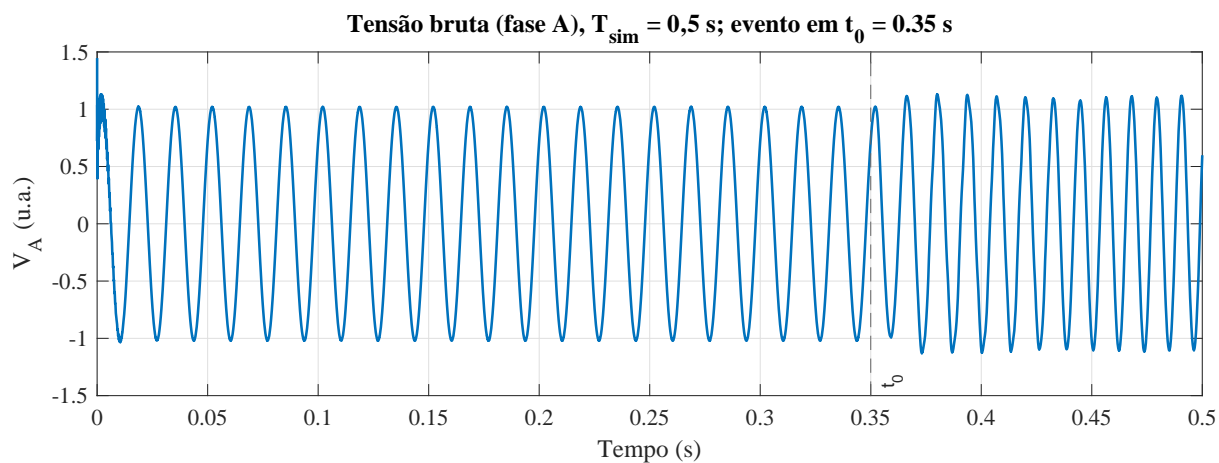
As Figuras 13 e 14 ilustram, para a mesma realização da fase A: (i) o traço bruto de 0,5 s com o instante  $t_0$ ; (ii) a janela fixa de 2 ciclos da otimização; e (iii) a janela-mãe de 4 ciclos e as três sub-janelas de 2 ciclos (50% de sobreposição) da operação em janela deslizante. Os parâmetros estão sintetizados na Tabela 8.

#### 5.1.2 Reamostragem

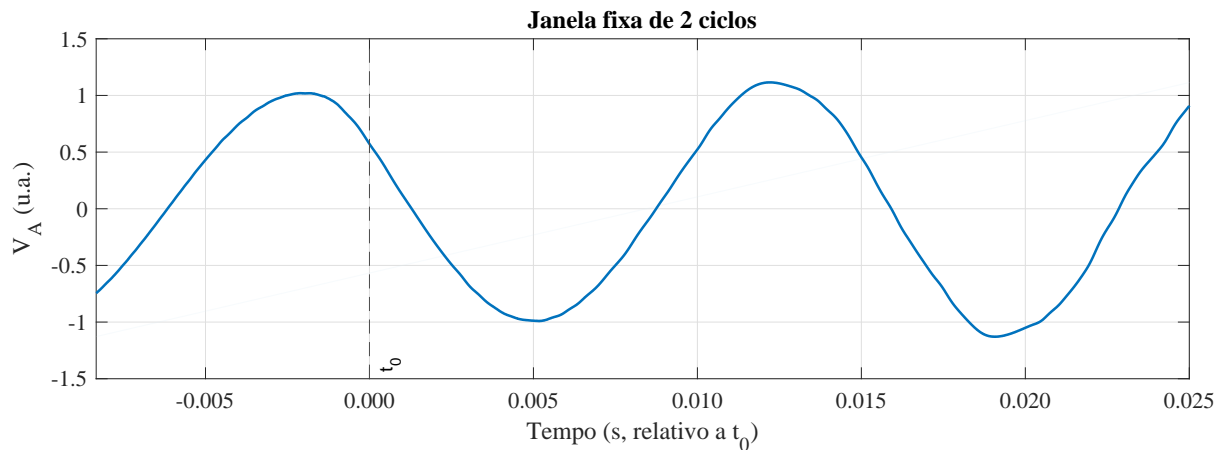
O alvo foi 256 amostras por ciclo fundamental ( $f_{\text{nom}} = 60\text{ Hz}$ ), ou seja,

$$f_s^* = 256 f_{\text{nom}} = 15\,360\text{ Hz}.$$

Figura 13 – Exemplo do janelamento temporal: sinal bruto e janela fixa.



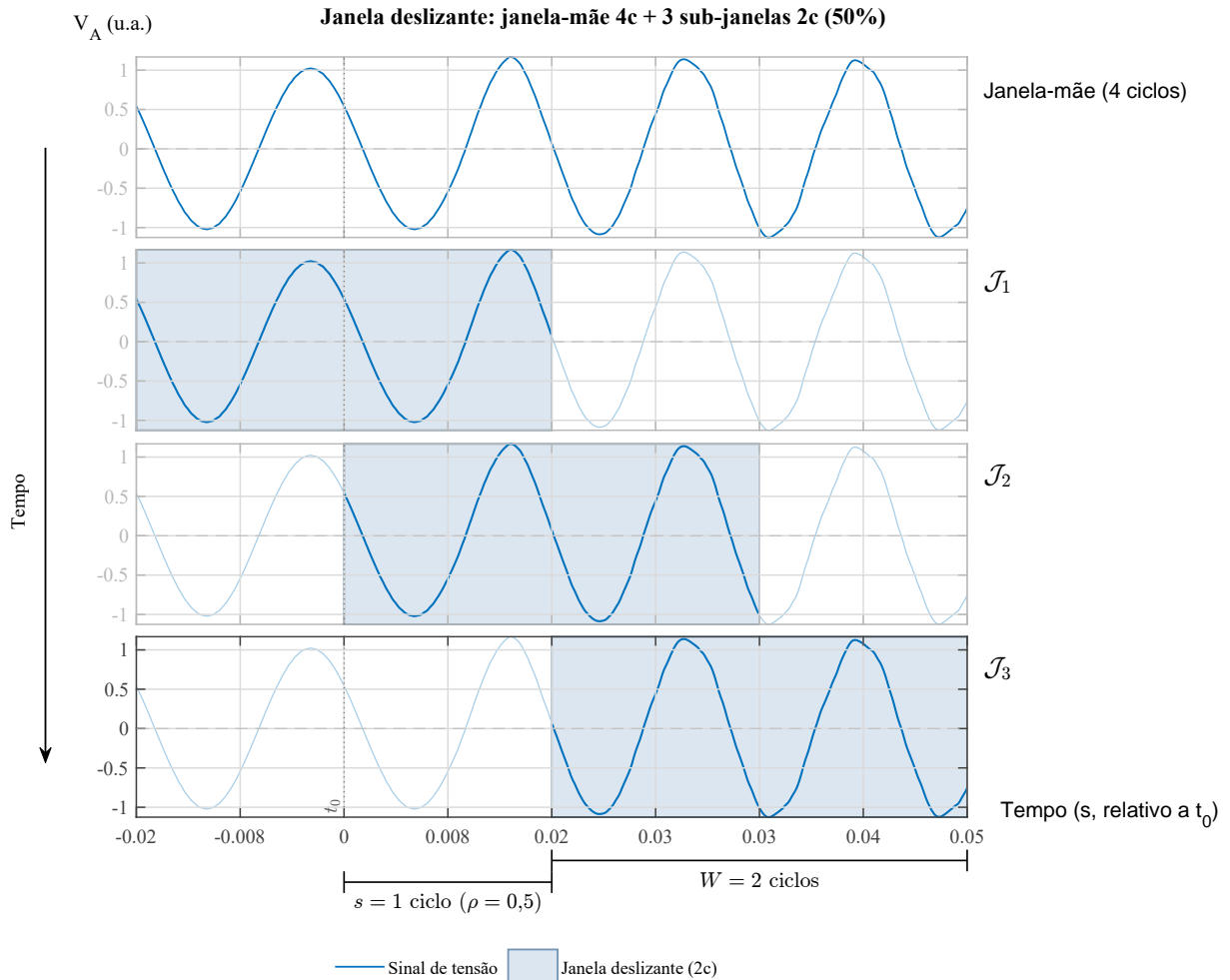
(a) Tensão bruta no PAC (fase A),  $T_{\text{sim}} = 0,5$  s; evento em  $t_0 = 0,35$  s.



(b) Recorte fixo de 2 ciclos centrado em  $t_0$  (com *jitter*  $\pm \frac{1}{4}$  ciclo).

Legenda: Exemplo ilustrativo do janelamento aplicado aos sinais no PAC, mantendo-se a mesma fase e a mesma realização: (a) sinal bruto com duração  $T_{\text{sim}} = 0,5$  s e instante do evento  $t_0$ ; (b) recorte fixo de 2 ciclos centrado em  $t_0$ , com *jitter* de  $\pm \frac{1}{4}$  ciclo.

Figura 14 – Exemplo do janelamento temporal: janela deslizante.



(a) Janela deslizante: janela-mãe de 4 ciclos ( $\pm \frac{1}{4}$  ciclo de *jitter*) e três sub-janelas de 2 ciclos com sobreposição de 50%.

Legenda: Detalhe da etapa operacional: janela-mãe de 4 ciclos e três sub-janelas de 2 ciclos com 50% de sobreposição, usadas para compor múltiplas observações por evento.

Implementou-se decimação por fator inteiro  $M = \lceil 198,0 \text{ kHz} / 15,36 \text{ kHz} \rceil = 13$ , resultando em  $f_{s,proc} \approx 15,23 \text{ kHz}$  ( $\approx 254$  pontos/ciclo). Essa diferença de  $\sim 0,8\%$  em relação ao alvo preserva as estatísticas de interesse e evita atrasos de grupo associados a métodos que utilizam filtros (Sec. 4.1.1).

### 5.1.3 Ruído e normalização

Em todas as fases de projeto/treino aplicou-se AWGN com  $\text{SNR} = 70 \text{ dB}$  traço a traço (Sec. 4.1.1). As tensões foram normalizadas por  $V_{pico}^{nom} = V_{LL}^{nom} \sqrt{2}/\sqrt{3}$ , com  $V_{LL}^{nom} = 25 \text{ kV}$  no PAC, produzindo grandezas adimensionais.

#### 5.1.4 Projeção PCA

Usou-se a rotação fixa  $\bar{\mathbf{R}}$  estimada exclusivamente com janelas de ilhamento do conjunto de otimização (Sec. 4.2.2). A projeção opera com  $d = 1$ , retendo apenas a primeira componente  $c_1$  (Sec. 5.3). Cada janela é mapeada para um traço unidimensional  $z$  antes do cálculo das cumulantes.

#### 5.1.5 Atributos

Os atributos são cumulantes de 2<sup>a</sup>, 3<sup>a</sup> e 4<sup>a</sup> ordens calculados sobre  $z$ . Eles são avaliados por FDR multiclasse e reduzidos por picos locais (cinco por ordem) na fase de projeto (Sec. 4.4). Na operação, apenas os atrasos pré-selecionados são computados (Sec. 4.4.5).

#### 5.1.6 Partição e decisão

No fluxo de janela deslizante, as três sub-janelas de cada instância temporal são tratadas como um grupo atômico na partição 70/30 (Sec. 3.3.2); a decisão por evento é obtida por fusão das três previsões (Sec. 4.6.5). As métricas são reportadas em nível de sub-janela e de evento.

#### 5.1.7 Implementação e organização dos scripts

As rotinas MATLAB/Simulink foram organizadas em três blocos: (i) geração de dados; (ii) pré-processamento da janela fixa; e (iii) janela deslizante e avaliação. Todos os códigos utilizaram gerador de números aleatórios com semente fixa (`rng default`) e foram executados na mesma versão do MATLAB/Simulink do Cap. 3. A Tabela 8 sintetiza os parâmetros efetivamente adotados, em conformidade com o Cap. 4 (Secs. 4.1.1–4.4.5) e o Cap. 3.

Tabela 8 – Parâmetros efetivamente usados na execução.

Item	Configuração
Duração da simulação	$T_{\text{sim}} = 0,5 \text{ s}$ ; evento em $t_0 = 0,35 \text{ s}$ .
Janelas (otimização)	Fixa $2c$ , <i>jitter</i> $\pm \frac{1}{4}c$ .
Janelas (operação)	$4c \rightarrow 3 \times 2c$ , sobreposição 50%, <i>jitter</i> $\pm \frac{1}{4}c$ .
Reamostragem	<code>downsample</code> $M = 13 \Rightarrow 15,23 \text{ kHz}$ .
Ruído	AWGN 70 dB por traço.
Normalização	$v/V_{\text{pico}}^{\text{nom}}$ ; $V_{\text{LL}}^{\text{nom}} = 25 \text{ kV}$ .
PCA	$\bar{\mathbf{R}}$ fixa (estimada em ilhamento), $d = 1$ , uso de $c_1$ .
Atributos	Cumulantes (2 <sup>a</sup> –4 <sup>a</sup> ) em atrasos pré-selecionados por FDR.
Partições	Otimização: <i>K-fold</i> (Seção 4.6.4.1); operação: 70/30 por instância.
RNG	<code>rng default</code> (todas as etapas MATLAB).

## 5.2 CARACTERIZAÇÃO DO CONJUNTO

Esta seção documenta o conjunto efetivamente utilizado nas etapas de otimização (janela fixa, 2 ciclos) e operação (janela deslizante  $4 \rightarrow 3 \times 2$  ciclos). O objetivo é: (i) explicitar as contagens e o equilíbrio por classe, caso e repetição; (ii) consolidar os níveis de SNR empregados em cada bloco de análise; (iii) evidenciar a distribuição dos modos de falta; e (iv) descrever a validação independente fora da grade. Adotamos a convenção de classes {N,I,F} para normal, ilhamento e falta, respectivamente.

### 5.2.1 Composição e Estratificação

A Tabela 9 resume as contagens por janelamento e por classe, distinguindo instâncias (janelas centrais) e sub-janelas (no regime deslizante). No conjunto de otimização (2 ciclos), cada classe possui  $9 \times 100 = 900$  janelas (9 casos  $\times$  100 repetições), perfazendo 2700 amostras balanceadas. No conjunto de treino/teste ( $4 \rightarrow 3 \times 2$  ciclos), cada instância temporal gera três sub-janelas: são 2700 sub-janelas por classe (900 instâncias  $\times$  3) e 8100 no total. O conjunto fora da grade contém 10 perfis inéditos, cada um com 10 repetições nas três classes {N,I,F}, totalizando 3000 instâncias e 9000 sub-janelas (três por instância).

Tabela 9 – Resumo de contagens por janelamento e classe.

Conjunto	N	I	F	Total
Otimização (2 ciclos; instâncias)	900	900	900	2700
Treino/Teste ( $4 \rightarrow 3 \times 2$ ciclos; sub-janelas)	2700	2700	2700	8100
Validação ( <i>fora da grade</i> ; instâncias)	1000	1000	1000	3000
Validação ( <i>fora da grade</i> ; sub-janelas)	3000	3000	3000	9000

Para explicitar o equilíbrio interno, a Tabela 10 agrega a estratificação por caso (grade  $3 \times 3$ ) e por perfil (fora da grade). No janelamento fixo (2 ciclos), cada caso da grade contribui com 100 instâncias por classe; no fora da grade, cada perfil contribui com 10 instâncias por classe. No regime em janela deslizante ( $4 \rightarrow 3 \times 2$  ciclos), cada instância gera três sub-janelas por classe.

Tabela 10 – Estratificação e balanceamento por conjunto.

Estrato (tipo)	Número estratos	Repetições/ estrato	Janela fixa	Janela deslizante
Casos 1–9 (grade $3 \times 3$ )	9	100	900	2700
Perfis 1–100 ( <i>fora da grade</i> )	100	10	1000	3000

### 5.2.2 Níveis de ruído (SNR) avaliados

No projeto/treino aplicou-se AWGN com  $\text{SNR} = 70$  dB traço a traço (Sec. 4.1.1). Para robustez, replica-se o processamento para  $\text{SNR} \in \{60, 50, 40, 30\}$  dB. A Tabela 11 consolida os cenários usados neste capítulo.

Tabela 11 – Cenários de SNR por bloco de análise.

Bloco de análise	Conjunto	SNR aplicado
Otimização (janela fixa)	Grade $3 \times 3$	70 dB
Treino/teste (janela deslizante)	Grade $3 \times 3$	70 dB
Robustez (variação de ruído)	Grade $3 \times 3$	$\{60, 50, 40, 30\}$ dB
Validação independente	Fora da grade (aleatória)	70 dB
Robustez na validação	Fora da grade (aleatória)	$\{60, 50, 40, 30\}$ dB

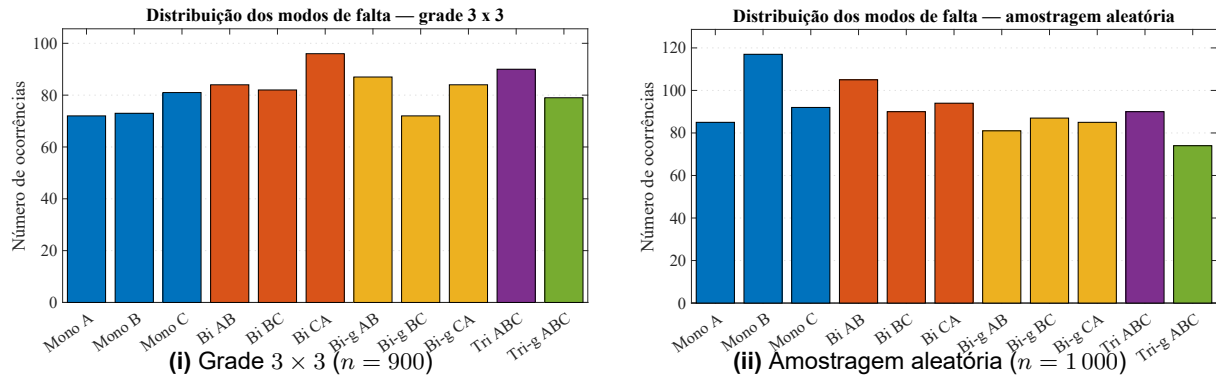
### 5.2.3 Modos de falta e checagens de aleatoriedade

A seleção do modo de falta ( $\text{FAULT\_MODE} \in \{2, \dots, 12\}$ ) é aleatória a cada repetição e a distância do PAC ao ponto de aplicação é fixa em 5 km (Cap. 3). A Figura 15 resume as frequências relativas por modo de falta nos dois esquemas de geração (Sec. 3.2.2): (i) grade  $3 \times 3$  ( $n = 900$ ) e (ii) amostragem aleatória na ZND ( $n = 1000$ ). Em ambos os casos observa-se distribuição aproximadamente uniforme entre faltas monofásicas, bifásicas, bifásicas-terra e trifásicas.

## 5.3 VALIDAÇÃO DA PROJEÇÃO FIXA ( $\bar{\mathbf{R}}$ )

Esta seção tem como objetivo é verificar empiricamente que a rotação fixa  $\bar{\mathbf{R}}$  (Sec. 4.2.2) é adequada e que a projeção unidimensional ( $d = 1$ ; uso de  $c_1$ ) é suficiente. Reporta-se: (A) o erro de ortogonalidade de  $\bar{\mathbf{R}}$ ; (B) a energia explicada média (e percentis) por componente em cada classe; e (C) a distribuição de  $\eta_1$ .

Figura 15 – Distribuição dos modos de falta por esquema de amostragem.



Legenda: Histogramas das classes de falta adotadas nos cenários (Seção correspondente), comparando a distribuição de ocorrências entre (i) o conjunto determinístico em grade 3 x 3 e (ii) o conjunto por amostragem aleatória fora da grade. O valor  $n$  indica o número total de instâncias consideradas em cada esquema.

### 5.3.1 Erro de ortogonalidade de $\bar{\mathbf{R}}$

Calcula-se  $\bar{\mathbf{R}}$  de acordo com a Equação 4.1 sobre as matrizes ortonormais  $\mathbf{R}$  calculadas por PCA sobre as janelas do conjunto de otimização, perfazendo 900 matrizes:  $i = 9$  perfis de carga e  $j = 100$  repetições.

O afastamento ou erro de ortonormalidade pode ser calculado como

$$\varepsilon_{\text{orth}} = \|\bar{\mathbf{R}}^T \bar{\mathbf{R}} - \mathbf{I}\|_F. \quad (5.1)$$

Onde  $\|\cdot\|_F$  denota a norma de Frobenius [88]. Para o conjunto estudado obteve-se:

$$\varepsilon_{\text{orth}} \approx 1,41.$$

Observa-se que, embora preserve aproximadamente a ortogonalidade entre colunas, a matriz obtida não é ortonormal exata, com o desvio resultando principalmente da diferença nas normas das colunas, não unitárias. Trata-se de uma consequência natural da média direta no espaço euclidiano.

Esse comportamento é esperado e aceitável dentro da proposta adotada. Como discutido em [79], a média euclidiana de matrizes em  $O(3)$  pode resultar em uma matriz fora do grupo, sendo necessário aplicar projeções específicas para reestabelecer a ortogonalidade, o que não foi adotado aqui por priorizar simplicidade computacional.

Enfatiza-se que o objetivo desta projeção não é a reconstrução do sinal ou a decorrelação perfeita, mas sim a criação de um transformador fixo e computacionalmente eficiente que maximize a separabilidade entre as classes de interesse.

Use-se, portanto, apenas a primeira coluna  $\bar{\mathbf{r}}_1$  de  $\bar{\mathbf{R}}$  para projetar ( $\mathbf{z} = \mathbf{X}\bar{\mathbf{r}}_1$ ), uma vez que esse desvio não prejudica a separabilidade, como confirmam as próximas

análises.

### 5.3.2 Energia explicada por componente

Seja  $\mathbf{X} \in \mathbb{R}^{508 \times 3}$  a janela de observações já normalizada (ver Seção 4.2.1) e  $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \lambda_3)$  os autovalores da matriz de covariância  $\boldsymbol{\Sigma} = \frac{1}{507} \tilde{\mathbf{X}}^T \tilde{\mathbf{X}}$ . Definimos a fração de variância (ou energia) preservada pelo  $j$ -ésimo componente principal como

$$\eta_j = \frac{\lambda_j}{\lambda_1 + \lambda_2 + \lambda_3}, \quad j \in 1, 2, 3, \quad 0 \leq \eta_j \leq 1. \quad (5.2)$$

A Tabela 12 apresenta as estatísticas descritivas da fração de variância explicada ( $\eta_j$ ) pelos três primeiros componentes principais para cada classe de operação: nominal, ilhamento e falta. Foram computadas a média e os percentis 5% ( $p_5$ ) e 95% ( $p_{95}$ ) com base em  $N = 900$  janelas por classe, todas extraídas do conjunto de otimização.

Tabela 12 – Energia explicada pelos componentes principais ( $\eta_j$ ) em porcentagem.

Classe	$\eta_1$			$\eta_2$			$\eta_3$		
	Média	$p_5$	$p_{95}$	Média	$p_5$	$p_{95}$	Média	$p_5$	$p_{95}$
Nominal	50,03	50,03	50,03	49,97	49,97	49,97	< 0,01	< 0,01	< 0,01
Ilhamento	54,01	50,49	56,17	45,99	43,83	49,51	< 0,01	< 0,01	< 0,01
Falta	68,07	52,14	84,79	29,87	14,11	44,75	2,05	< 0,01	5,96

Observa-se que, na classe nominal, a variância é praticamente dividida igualmente entre os dois primeiros componentes principais ( $\eta_1 \approx 50,03\%$ ,  $\eta_2 \approx 49,97\%$ ), com variações nulas, enquanto o terceiro componente não contribui significativamente ( $\eta_3 \approx 0,00\%$  em todos os casos). Isso indica que, sob operação normal, os sinais trifásicos são altamente correlacionados e linearmente redundantes, sendo essencialmente bidimensionais.

Na classe de ilhamento, nota-se uma leve assimetria: o primeiro componente concentra em média 54,01% da variância (com mínimo de 50,49% e máximo de 56,17%), enquanto o segundo retém os 45,99% restantes. O terceiro componente segue com contribuição próxima de nula. A pequena diferença entre os dois primeiros componentes sugere que, embora ainda haja correlação entre as fases, o ilhamento introduz uma ligeira ruptura nessa estrutura, o que pode ser útil para discriminação.

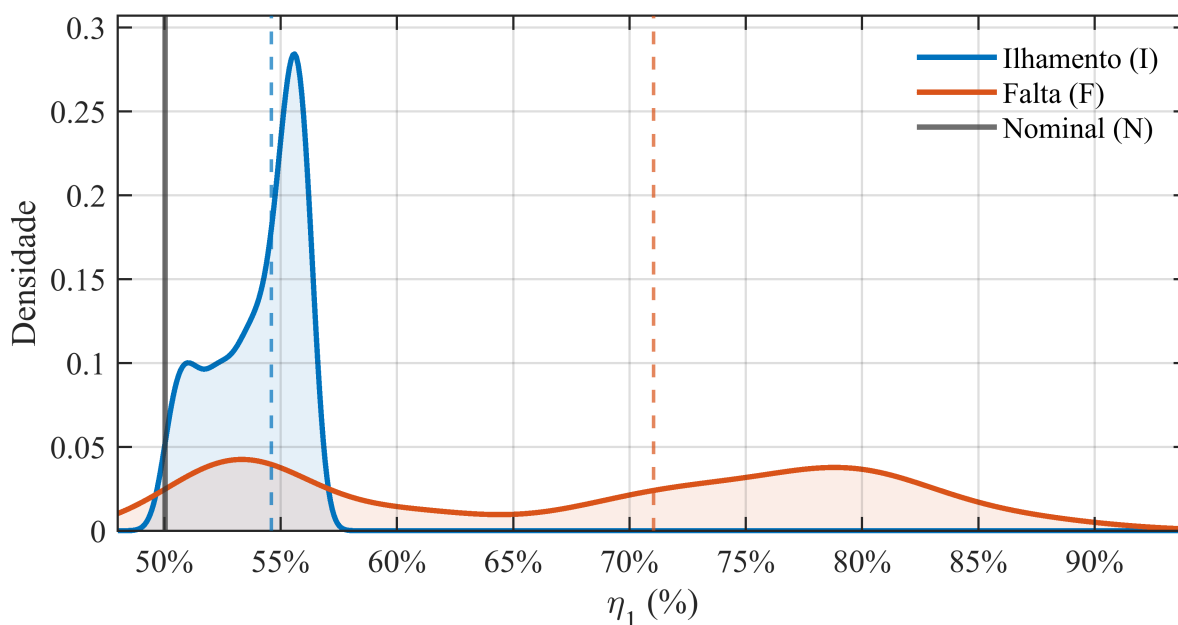
Já na classe de falta, o comportamento muda substancialmente. O primeiro componente explica, em média, 68,07% da variância, podendo chegar a 84,79%, enquanto o segundo cai para 29,87% e o terceiro passa a ter participação não desprezível (até 5,96%). Essa dispersão maior entre os componentes reflete a forte assimetria introduzida pelas faltas, o que reduz a redundância entre as fases.

Em conjunto, os resultados indicam que os dois primeiros componentes capturam  $> 97\%$  da variância para todos os cenários, sugerindo que  $d = 2$  é dimensionalidade suficiente para os classificadores. Entretanto, a subseção seguinte demonstra que, mesmo com  $d = 1$ , obtém-se desempenho competitivo em virtude da grande separabilidade proporcionada por  $\eta_1$ . Isto confirma a robustez da projeção de ordem reduzida sem sacrificar a discriminação entre classes.

### 5.3.3 Distribuição de $\eta_1$ por classe

A Figura 16 mostra as densidades de  $\eta_1$  para as três classes ( $N$ ,  $I$ ,  $F$ ) calculadas sobre as 900 janelas de otimização por classe. Para  $N$ , a distribuição é praticamente degenerada em torno de 50% (linha vertical preta), refletindo a simetria trifásica e a dominância da sequência positiva. Em  $I$ , o pico desloca-se para a direita, tipicamente no intervalo 50,5–56%, com mediana indicada pela linha tracejada azul. Já em  $F$ , a densidade é mais espalhada e deslocada para valores maiores (mediana laranja  $\approx 70\%$ ), com cauda longa até  $\sim 90\%$ . Essa separação  $F \gg \{N, I\}$  em  $\eta_1$  antecipa a boa discriminabilidade quando trabalhamos com  $d = 1$  (Sec. 5.3.2) e explica por que o uso exclusivo de  $c_1$  já oferece alto desempenho (Tab. 13).

Figura 16 – Densidade de  $\eta_1$  por classe no conjunto de otimização.



Legenda: Estimativa de densidade (KDE) da energia explicada pelo primeiro componente principal ( $\eta_1$ ) para cada classe ( $N$ ,  $I$  e  $F$ ) no conjunto de otimização.

Em síntese,  $\eta_1$  concentra metade da variância em  $N$  ( $\approx 50\%$ ), cresce moderadamente em  $I$  ( $\approx 54\%$ ) e torna-se dominante em  $F$  ( $\approx 68\%$ , com ampla variabilidade). Essa hierarquia é consistente com a física do problema (quebra de simetria e surgi-

mento de sequência negativa em faltas) e fundamenta a adoção sistemática de  $d = 1$  com a primeira coluna de  $\bar{\mathbf{R}}$  nas etapas seguintes.

#### 5.3.4 Separabilidade por componente

A fração de variância explicada oferece um primeiro indício de quão informativa pode ser cada componente principal, mas não garante que a projeção unidimensional preserve a separabilidade entre as classes  $\{N, I, F\}$ . Portanto, conduziu-se um experimento empírico para avaliar diretamente essa capacidade de discriminação quando se utiliza cada componente de forma isolada ( $d = 1$ ). O procedimento segue quatro etapas:

- (1) Cálculo das cumulantes de 2<sup>a</sup>, 3<sup>a</sup> e 4<sup>a</sup> ordens em todos os 508 atrasos, gerando matrizes  $\Gamma_2, \Gamma_3, \Gamma_4 \in \mathbb{R}^{508 \times 900}$ .
- (2) Cálculo do FDR multiclasse para ranquear atributos.
- (3) Seleção dos cinco máximos locais de cada ordem, resultando em um vetor de  $5 \times 3 = 15$  atributos por janela. A escolha de picos locais, em vez dos maiores valores absolutos de FDR, mitiga correlações entre cumulantes vizinhas.
- (4) Treinamento, em validação cruzada estratificada de 10 *folds*, de dois classificadores:
  - (i) LS multiclasse;
  - (ii) MLP com uma camada oculta (20 neurônios, `tansig`, `trainlm`).

Os resultados desta subseção têm caráter preliminar. Adotou-se validação cruzada estratificada com  $K = 10$  apenas como sonda para verificar a informatividade de cada componente ( $c_1, c_2, c_3$ ). A escolha definitiva de  $K$  será realizada adiante, na Seção 5.5. As conclusões aqui reportadas não antecipam essa escolha, servindo apenas para motivar o uso de  $d = 1$  nas etapas seguintes.

Os desempenhos médios obtidos em cada cenário estão sumarizados na Tabela 13. Os valores apresentados correspondem à média de acurácia e ao respectivo desvio-padrão ao longo das 10 partições.

Os resultados de  $c_1$  revelam que a MLP quase atinge 100% de acurácia, com um desvio-padrão inferior a 0,4%, enquanto o classificador LS permanece em torno de 93%. Essa combinação de alto desempenho e baixa variabilidade sugere que os 15 atributos derivados de  $c_1$  concentram a maior parte da informação relevante para a tarefa, mesmo sem qualquer etapa adicional de otimização.

No caso da segunda componente, observa-se que o LS mantém praticamente o mesmo patamar de acurácia verificado em  $c_1$ ; entretanto, a MLP apresenta uma

Tabela 13 – Acurácia média (%)  $\pm$  desvio-padrão em validação cruzada 10-fold por componente principal (ensaio preliminar).

Componente	dim(Z)	LS (%)	MLP (%)
$c_1$	15	93,48 $\pm$ 1,25	<b>99,74 <math>\pm</math> 0,36</b>
$c_2$	15	93,63 $\pm$ 1,94	97,33 $\pm$ 9,41
$c_3$	15	49,48 $\pm$ 2,01	48,78 $\pm$ 2,24

variabilidade muito superior (desvio-padrão de 9,4%). Esse comportamento indica que, embora haja informação discriminativa em  $c_2$ , ela se encontra mesclada a ruído não gaussiano ou a atributos redundantes que tornam o processo de treinamento da rede sensível às partições dos dados.

Por fim, a terceira componente produz desempenhos que não ultrapassam 50% de acurácia em nenhum dos classificadores. Tal resultado reforça a conclusão de que  $c_3$  carrega essencialmente residual pouco informativo, corroborando a análise apresentada na Seção 5.3.2.

Em síntese, a projeção unidimensional sobre  $c_1$  preserva, e de fato maximiza, a separabilidade entre as classes enquanto cumpre o requisito de baixa complexidade computacional: apenas três multiplicações escalares por amostra no estágio de projeção. Por essa razão, adota-se  $d = 1$  como dimensionalidade de trabalho nas etapas subsequentes do MDI. Questões relativas à seleção fina de cumulantes e ao ajuste dos classificadores serão exploradas nas Seções 4.4–4.6.

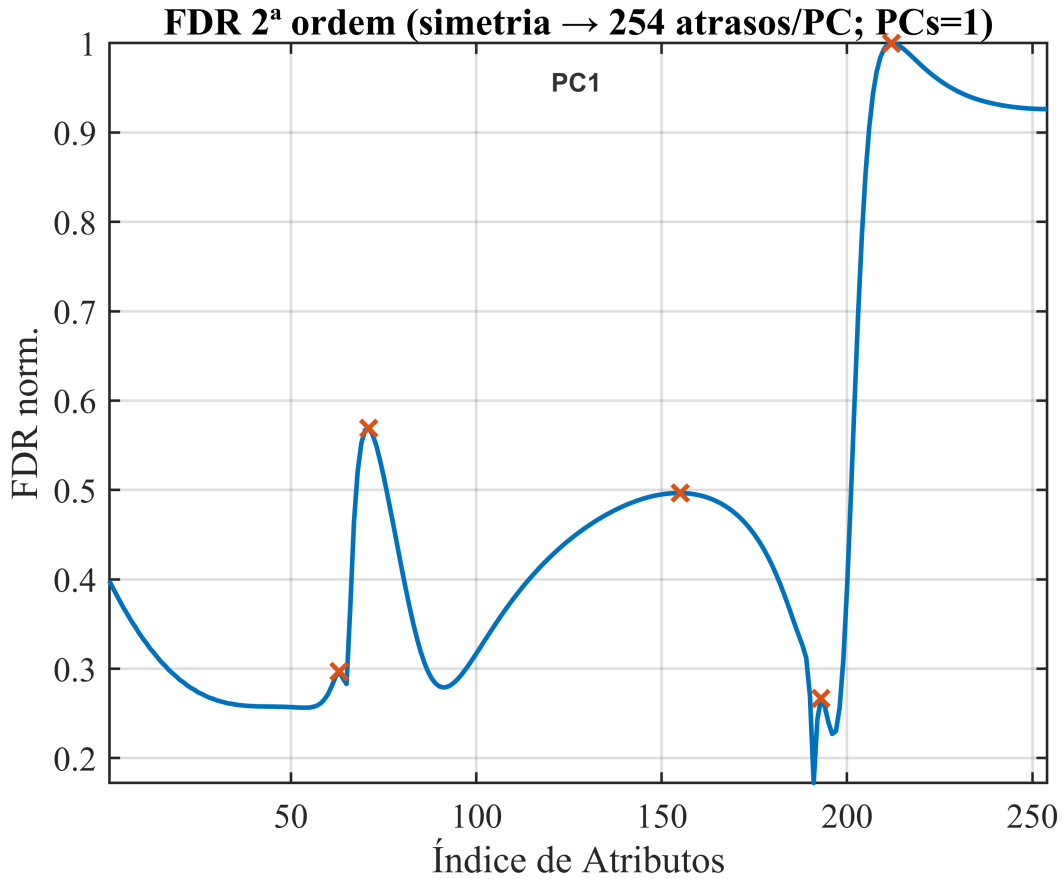
## 5.4 SELEÇÃO DE ATRIBUTOS POR FDR

Após fixar  $d = 1$ , estimaram-se as cumulantes conforme a Seção 4.3 e calculou-se a FDR multiclasse por atraso  $\tau$  (Seção 4.4). Para  $r = 2$  explorou-se a simetria par da autocovariância, retendo apenas  $\tau \in [0, \lceil L/2 \rceil - 1]$  (Seção 4.4.3). Em seguida, aplicou-se a seleção por picos locais (cinco por ordem), evitando colinearidade entre atrasos vizinhos e preservando diversidade informacional.

### 5.4.1 Mapas de FDR e leitura inicial

As Figuras 17–19 exibem os mapas 1-D de FDR para  $r \in \{2,3,4\}$  em  $c_1$ , com marcadores nas cinco cristas selecionadas  $\mathcal{P}_r$ . Para facilitar a leitura visual da forma dos mapas (e não apenas das alturas absolutas), as curvas foram normalizadas por ordem dividindo-se pela FDR máxima de cada ordem. Assim, valores “1” indicam o pico daquela ordem e os demais aparecem em escala relativa a ele.

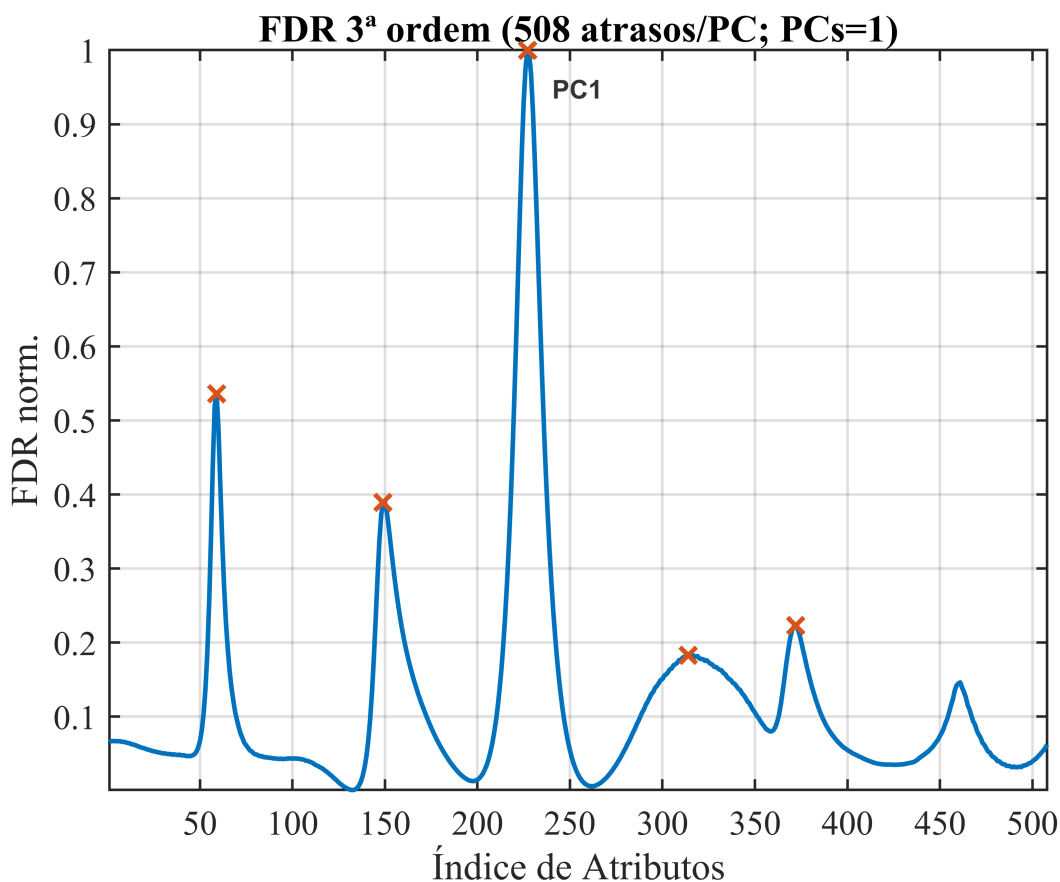
Observa-se a recorrência de cristas em torno das quadraturas ( $\sim 90^\circ$ ) e entre

Figura 17 – Mapa 1-D de FDR de 2ª ordem em  $c_1$ .

Legenda: Mapa unidimensional da FDR de 2ª ordem calculada sobre  $c_1$ , com normalização pela própria ordem. Os marcadores  $\times$  em vermelho indicam os cinco picos selecionados, definidos como  $\mathcal{P}_2$ , que representam atrasos candidatos para extração de atributos.

$260^\circ$ – $320^\circ$ <sup>1</sup>. Em 2ª ordem (energia/defasagem) e 4ª ordem (impulsividade) essas regiões concentram os maiores picos; em 3ª ordem (assimetria) há cristas menores, o que sugere que a separação entre  $\{N, I, F\}$  se dá mais por nível/energia e impulsividade do que por *skewness*.

<sup>1</sup> Conversões de  $\tau$  para tempo e ângulo elétrico na Seção 5.4.1.4.

Figura 18 – Mapa 1-D de FDR de 3ª ordem em  $c_1$ .

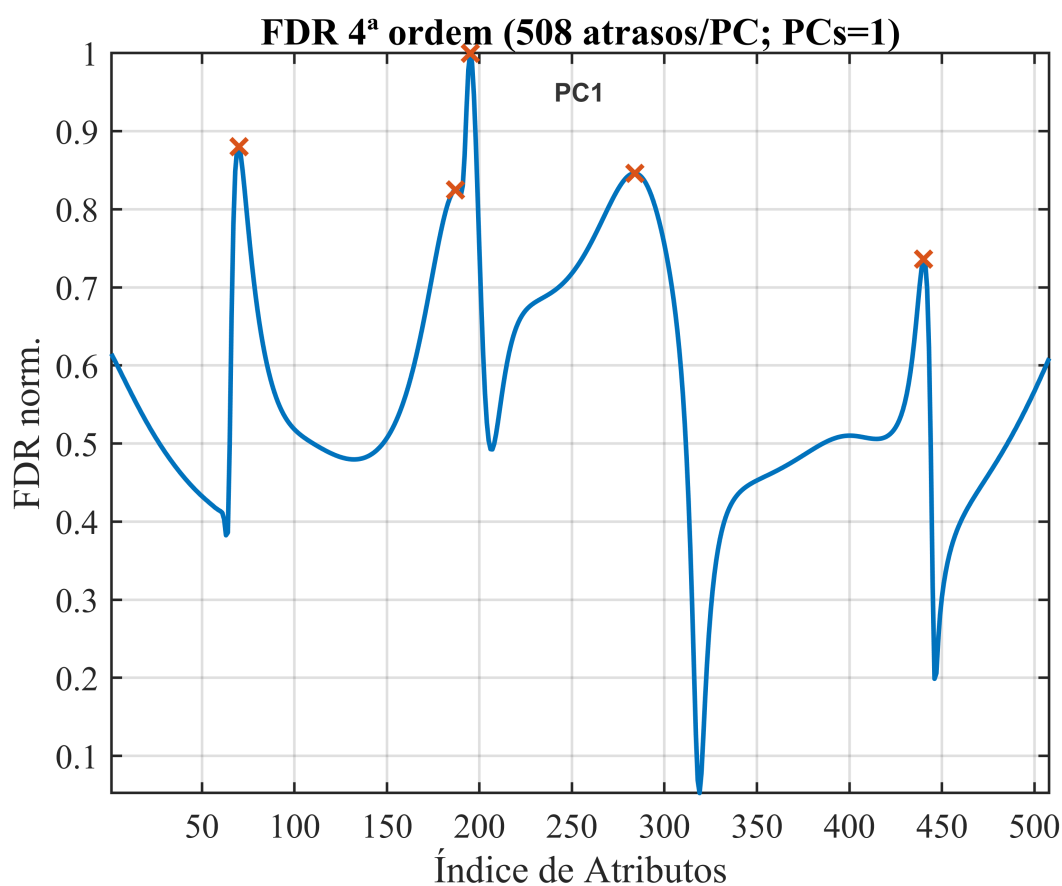
Legenda: Mapa unidimensional da FDR de 3ª ordem calculada sobre  $c_1$ , com normalização pela própria ordem. Os marcadores  $\times$  em vermelho indicam os cinco picos selecionados, definidos como  $\mathcal{P}_3$ , que representam atrasos candidatos para extração de atributos.

#### 5.4.1.1 Poder discriminativo geral

A Tabela 14 resume, por ordem, os cinco picos selecionados com suas FDR absolutas e normalizadas (divisão pelo pico máximo de cada ordem). Os valores confirmam que:

- (i) em 2ª e 4ª ordens há separação forte, com valores de FDR na casa das dezenas;
- (ii) em 3ª ordem os valores são muito menores (centésimos). O seu pico máximo de FDR é de  $\sim 0,03$ , um valor três ordens de magnitude inferior aos picos das outras ordens.

Na Tabela 14, a FDR normalizada é obtida pela divisão pela FDR máxima dentro de cada ordem. Essa diferença drástica indica que a contribuição da 3ª ordem

Figura 19 – Mapa 1-D de FDR de 4ª ordem em  $c_1$ .

Legenda: Mapa unidimensional da FDR de 4ª ordem calculada sobre  $c_1$ , com normalização pela própria ordem. Os marcadores  $\times$  em vermelho indicam os cinco picos selecionados, definidos como  $\mathcal{P}_4$ , que representam atrasos candidatos para extração de atributos.

Tabela 14 – Poder discriminativo dos picos por ordem (FDR bruta e normalizada).

Ranking	Ordem $r = 2$			Ordem $r = 3$			Ordem $r = 4$		
	$\tau$	FDR	FDR norm.	$\tau$	FDR	FDR norm.	$\tau$	FDR	FDR norm.
1	211	29,8473	1,000	226	0,0274	1,000	194	20,4921	1,000
2	70	16,9824	0,569	58	0,0147	0,537	69	18,0404	0,881
3	154	14,8217	0,497	148	0,0107	0,391	283	17,3394	0,846
4	62	8,8591	0,297	371	0,0061	0,223	186	16,9018	0,825
5	192	7,9583	0,267	313	0,0050	0,182	439	15,0897	0,736

para a classificação é marginal e potencialmente prejudicial, podendo adicionar ruído e complexidade computacional desnecessária ao modelo. Diante dessa evidência quantitativa, optou-se por descartar os cumulantes de 3ª ordem e construir o vetor de atributo final exclusivamente com as características de 2ª e 4ª ordem. Assim, o conjunto final de atributos é composto por  $B^* = 10$  características, os cinco picos de  $r = 2$  e os cinco de  $r = 4$ .

A baixa relevância da 3ª ordem não é uma particularidade deste conjunto de dados, mas sim um reflexo da física do problema. A cumulante de 3ª ordem quantifica a assimetria (*skewness*) da distribuição de probabilidade do sinal. Sinais de CA, como os analisados, são intrinsecamente simétricos em regime permanente, possuindo cumulantes de ordem ímpar teoricamente nulas. Os fenômenos de interesse, faltas e ilhamentos, alteram drasticamente a energia (capturada pela 2ª ordem) e a impulsividade (capturada pela 4ª ordem) do sinal, mas preservam em grande parte essa simetria. Uma falta, por exemplo, manifesta-se como um aumento de amplitude simétrico, enquanto o ilhamento acarreta desvios de frequência e amplitude que também não introduzem assimetria estatística relevante. Portanto, os resultados empíricos da Tabela 14 corroboram a expectativa teórica de que a assimetria não é uma característica distintiva para este problema de classificação.

#### 5.4.1.2 Pares 2D mais discriminativos

Para inspecionar a separabilidade bivariada, avaliou-se todo par  $(x_p, x_q)$  formado pelos  $B^* = 10$  atributos selecionados via FDR. Após padronização  $z$ -score, cada par foi ranqueado pelo critério de Fisher multiclasse em duas dimensões,

$$J_{2D}(p, q) = \text{tr}(S_w^{-1} S_b),$$

onde  $S_w$  é a covariância intra-classes e  $S_b$  a covariância entre-classes restritas ao par [84]. Selecionaram-se os seis maiores  $J_{2D}$ , a Figura 20 mostra os *scatter plots* desses pares, onde se observam fronteiras quase lineares entre  $\{N, I, F\}$ . Essa etapa é meramente visual.

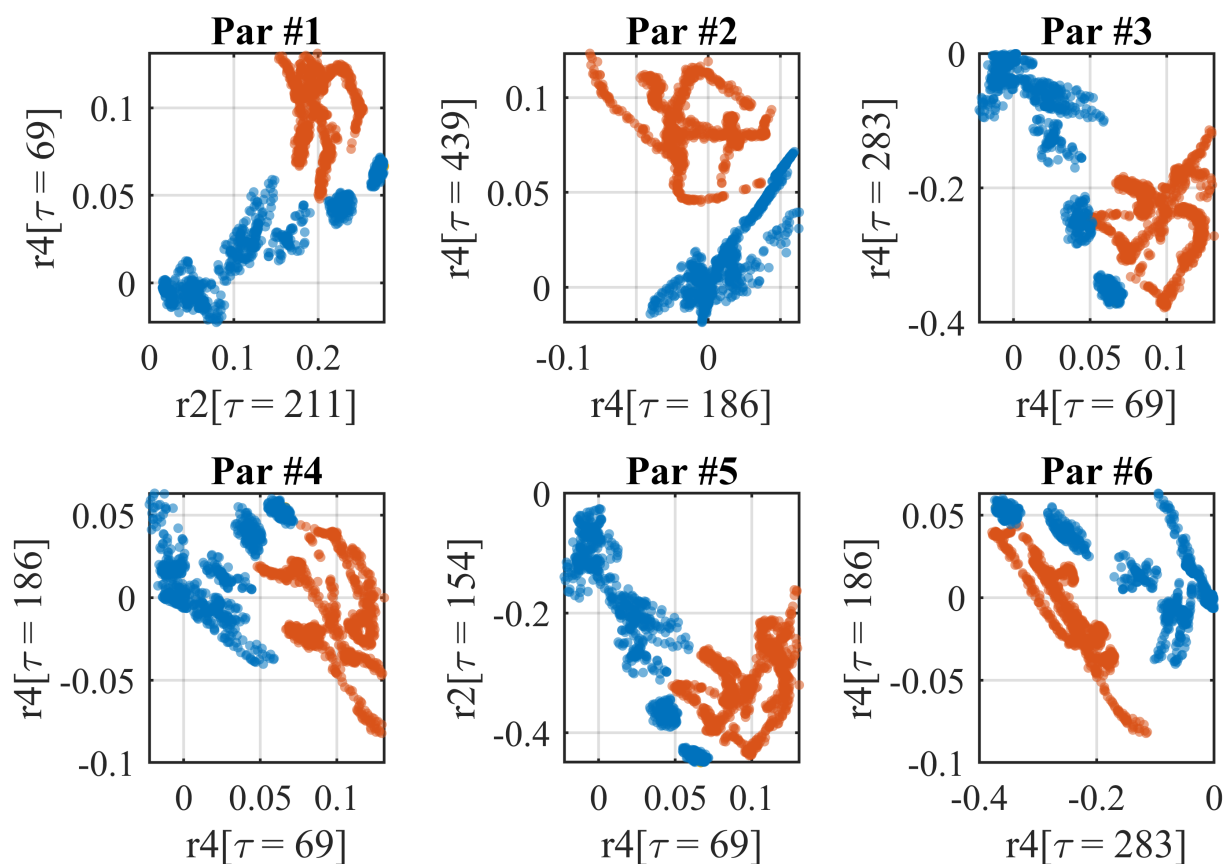
#### 5.4.1.3 Projeção linear e contribuições individuais

Como síntese global do espaço de atributos, aplicou-se a Análise Discriminante Linear (do inglês, *Linear Discriminant Analysis*) (LDA) ao vetor de 15 atributos, projetando-o nas duas primeiras Discriminantes Lineares (LD1 e LD2), solução do problema

$$S_b \mathbf{w} = \lambda S_w \mathbf{w},$$

com  $S_b$  e  $S_w$  computadas no espaço completo após  $z$ -score [84]. Como  $C = 3$  classes, o posto de  $S_b$  é no máximo  $C - 1$ , logo a LDA produz até duas discriminantes (LD1

Figura 20 – Seis pares com maior separabilidade segundo  $J_{2D}$ .

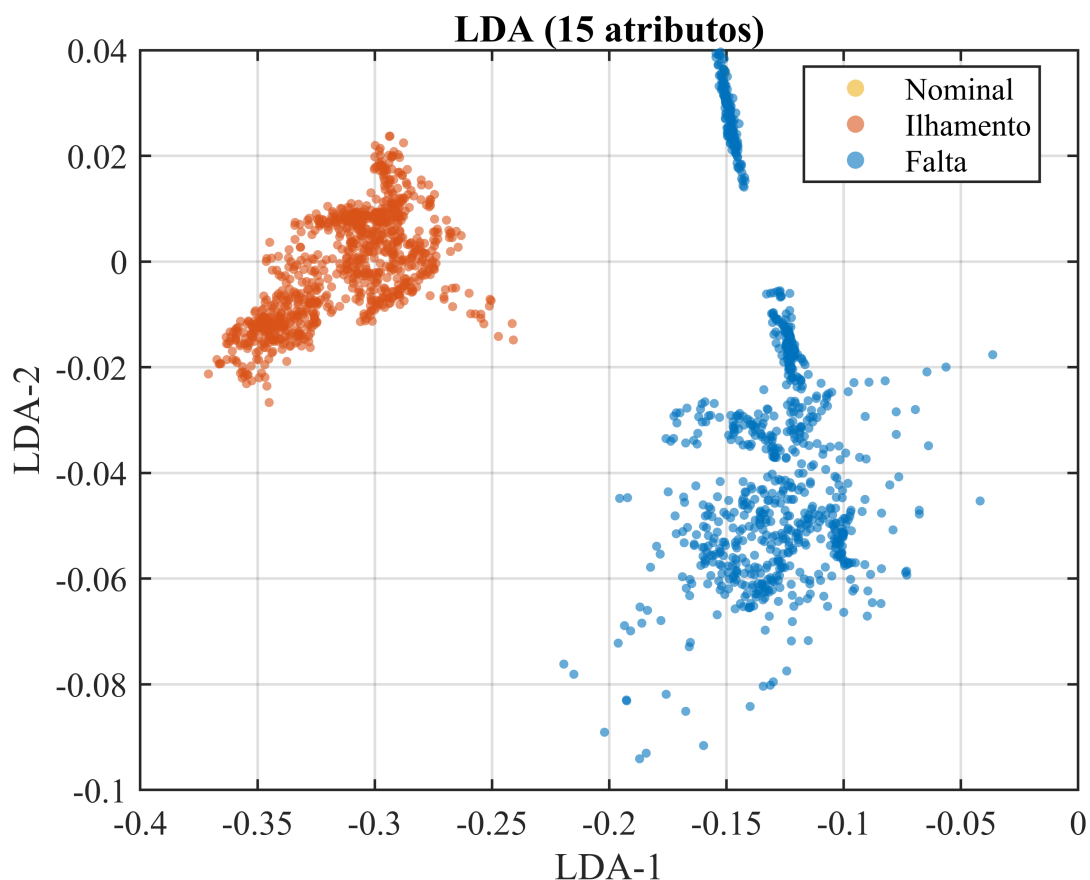


● Nominal ● Ilhamento ● Falta

Legenda: Seis pares de atributos com maior separabilidade entre classes segundo o critério  $J_{2D} = \text{tr}(S_w^{-1}S_b)$ , calculado a partir das matrizes de dispersão intra-classe ( $S_w$ ) e entre-classes ( $S_b$ ). Os dados foram previamente padronizados em  $z$ -score para tornar comparáveis escalas distintas entre atributos.

e LD2). A Figura 21 exibe a projeção em (LD1, LD2), evidenciando separação clara entre {I} e {F}. A nuvem {N} é muito compacta e sobreposta à de falta, ficando pouco visível no plano.

Figura 21 – Projeção LDA com 15 atributos padronizados.



Legenda: Projeção bidimensional obtida por LDA a partir do conjunto completo de 15 atributos. Antes da projeção, os atributos foram padronizados em  $z$ -score, de modo que a separação reflita predominantemente diferenças estruturais entre classes e não apenas diferenças de escala entre variáveis.

Para validar a decisão de excluir a 3ª ordem, analisou-se a importância dos atributos na projeção LDA. A Tabela 15, que aqui ainda exhibe os 15 atributos para fins de comparação, demonstra que as cinco características de 3ª ordem estão nas últimas posições, com contribuições praticamente nulas. Isso corrobora que sua remoção não acarreta perda de informação relevante para a discriminação linear. A análise final, portanto, foca-se nos atributos de 2ª e 4ª ordem, que dominam completamente a formação das discriminantes LD1 e LD2.

Na Tabela 15,  $|LD1|$  e  $|LD2|$  representam contribuições relativas normalizadas, com soma 1 em cada discriminante, enquanto *combinado* corresponde à média aritmética dessas importâncias.

Tabela 15 – Importância normalizada dos atributos na LDA.

Atributo	Ordem	LD1	LD2	Combinado
r2[ $\tau = 154$ ]	2	0,3941	0,3214	0,3578
r2[ $\tau = 211$ ]	2	0,3256	0,1972	0,2614
r4[ $\tau = 283$ ]	4	0,0064	0,2252	0,1158
r2[ $\tau = 62$ ]	2	0,1210	0,0477	0,0843
r2[ $\tau = 70$ ]	2	0,0757	0,0331	0,0544
r4[ $\tau = 186$ ]	4	0,0158	0,0828	0,0493
r2[ $\tau = 192$ ]	2	0,0401	0,0419	0,0410
r4[ $\tau = 69$ ]	4	0,0008	0,0332	0,0170
r4[ $\tau = 194$ ]	4	0,0139	0,0124	0,0131
r4[ $\tau = 439$ ]	4	0,0047	0,0016	0,0031
r3[ $\tau = 148$ ]	3	0,0004	0,0013	0,0009
r3[ $\tau = 58$ ]	3	0,0005	0,0009	0,0007
r3[ $\tau = 226$ ]	3	0,0004	0,0008	0,0006
r3[ $\tau = 313$ ]	3	0,0005	0,0003	0,0004
r3[ $\tau = 371$ ]	3	0,0002	0,0002	0,0002

#### 5.4.1.4 Atrasos em amostras, tempo e ângulo

Após a decimação, a taxa de amostragem é

$$f_s = \frac{198000}{13} \approx 15\,230,77 \text{ Hz}, \quad N_{\text{ciclo}} = \frac{f_s}{60} \approx 253,85 \text{ amostras/ciclo.}$$

Dado um atraso  $\tau$  (em amostras), o tempo e o ângulo correspondentes são

$$t(\tau) = \frac{\tau}{f_s} \text{ [s]} \quad \Rightarrow \quad t_{\text{ms}}(\tau) = 10^3 \frac{\tau}{f_s} \text{ [ms]},$$

$$\phi(\tau) = \left( \frac{\tau}{f_s} \cdot 360 \cdot 60 \right) \bmod 360^\circ = \left( \tau \cdot \frac{21600}{f_s} \right) \bmod 360^\circ.$$

Reporta-se  $\phi(\tau)$  no intervalo angular de  $0^\circ$  a  $360^\circ$ , pois picos separados de mais de um ciclo podem cair em fases equivalentes.

Os ângulos  $\phi(\tau)$  são relativos ao início da janela. Como a fase da Fase A é aleatória e o instante do evento possui *jitter* de  $\pm \frac{1}{4}$  de ciclo, o ângulo efetivo do evento na janela é  $\phi_{\text{evt}} \equiv (\phi_A + 360 f_0 \delta t) \bmod 360^\circ$ , onde  $\phi_A$  é o ângulo da Fase A no início da janela e  $\delta t \in [-\frac{1}{4}T_0, \frac{1}{4}T_0]$  é o *jitter* do ancoramento temporal do evento ( $T_0 = 1/f_0$ ).

A Tabela 16 lista os mesmos picos em *ordem crescente de*  $\tau$ , convertendo para tempo e fase elétrica com  $f_s = \frac{198000}{13}$  e  $f_0 = 60$  Hz. Note que a janela de análise tem dois ciclos e o evento é ancorado após meio ciclo com *jitter* de  $\pm \frac{1}{4}$  de ciclo; isto é, na prática o distúrbio entra entre  $\approx \frac{1}{4}$  e  $\frac{3}{4}$  de ciclo após o início. A recorrência de picos perto de  $\sim 90^\circ$  e  $260^\circ$ – $320^\circ$  é compatível com esse posicionamento relativo do evento e com a sensibilidade de  $2^a/4^a$  ordens nessas quadraturas.

Tomando os picos dominantes, observamos  $\tau = 211$  ( $2^a$  ordem),  $\tau = 226$  ( $3^a$ ) e  $\tau = 194$  ( $4^a$ ), que caem em  $299^\circ$ ,  $320,5^\circ$  e  $275,1^\circ$ , respectivamente, todos na região de

Tabela 16 – Atrasos selecionados por ordem (picos de FDR) em  $c_1$ .

Ordem $r = 2$			Ordem $r = 3$			Ordem $r = 4$		
$\tau$	$t$ (ms)	$\phi$ (graus)	$\tau$	$t$ (ms)	$\phi$ (graus)	$\tau$	$t$ (ms)	$\phi$ (graus)
62	4,07	87,9	58	3,81	82,3	69	4,53	97,9
70	4,60	99,3	148	9,72	209,9	186	12,21	263,8
154	10,11	218,4	226	14,84	320,5	194	12,74	275,1
192	12,61	272,3	313	20,55	83,9	283	18,58	41,3
211	13,85	299,2	371	24,36	166,1	439	28,82	262,6

três-quartos de ciclo, onde a combinação entre potência/defasagem e impulsividade maximiza diferenças entre  $\{N, I, F\}$ . As “segundas melhores” cumulantes de cada ordem ( $\tau = 70, 58, 69$ ) aparecem próximas de um quarto de ciclo ( $\sim 90^\circ$ ), formando, junto com os picos em  $\sim 270^\circ\text{--}320^\circ$ , dois aglomerados angulares estáveis. Alguns atrasos acima de um ciclo (p.ex.,  $\tau = 313, 371$  em 3ª ordem) retornam a fases equivalentes, refletindo cordilheiras de FDR úteis quando o transiente se estende por mais de um período.

Em termos de magnitude absoluta, conforme demonstrado pela Tabela 14, a 3ª ordem apresenta FDRs insignificantes em comparação com as ordens 2 e 4. Isso valida a conclusão de que, para este problema, a assimetria do sinal (*skewness*) não é uma característica relevante para a separação das classes, que são dominadas pela estrutura de energia e pela impulsividade. Por essa razão, optou-se por um modelo mais enxuto e robusto, descartando os atributos de 3ª ordem. A coerência da seleção dos atributos de 2ª e 4ª ordens é confirmada pela análise dos pares 2D (Figura 20) e pela projeção LDA (Tabela 15).

## 5.5 OTIMIZAÇÃO EM JANELA FIXA

Com o conjunto de  $B^* = 10$  atributos candidatos definido pela etapa de FDR (Seção 5.4) e mantendo a janela de dois ciclos, procede-se à otimização do classificador no regime janela fixa. O objetivo aqui é seguir o protocolo experimental da Seção 4.6.4 para escolher, de forma reprodutível:

- (i) o particionamento de validação cruzada,
- (ii) o melhor subconjunto de atributos para cada cardinalidade  $B^*$ ,
- (iii) o tamanho da camada escondida da MLP e, por fim,
- (iv) consolidar a arquitetura final, comparando-a a uma referência de LS.

Em todas as rodadas, os dados são previamente escalonados para  $[-1,1]$  e as comparações entre configurações usam a mesma família de partições, evitando vieses entre métodos. A apresentação segue quatro blocos:

- a Rodada 1 define o número de partições  $K^*$  via estabilidade de acurácia;
- a Rodada 2 seleciona, para cada  $B^*$ , o subconjunto  $S_{B^*}^*$  que maximiza o critério adotado;
- a Rodada 3, restrita à MLP, varre o número de neurônios escondidos  $n_h$  no melhor  $B^*$ ;
- e a Rodada 4 consolida a arquitetura escolhida e reporta o desempenho (média  $\pm$  desvio padrão) e a Matriz de Confusão (MC).

Quando pertinente, são incluídas análises de significância para comparações entre configurações.

### 5.5.1 Rodada 1 — escolha do tamanho do particionamento

Avaliou-se o impacto do número de partições na validação cruzada, testando  $K \in \{5, 10, 15, 20\}$  com os mesmos índices para ambos os classificadores (LS e MLP). A MLP utilizou camada escondida com 20 neurônios e *trainlm*. A Tabela 17 reporta a acurácia média geral e o desvio padrão por  $K$  e por classificador.

Tabela 17 – Acurácia (%) na validação cruzada em função de  $K$  ( $B = 10$  atributos).

$K$	Least Squares (LS)		MLP	
	média	dp	média	dp
5	93,48	0,58	99,63	0,19
10	93,48	1,15	99,56	0,34
15	93,48	1,99	97,37	8,65
20	93,48	1,68	99,63	0,56

Para o LS, a acurácia média permaneceu essencialmente inalterada em 93,48% entre os valores de  $K$ , com leve aumento do desvio padrão à medida que o conjunto de teste por *fold* diminui. Notavelmente,  $K = 5$  apresentou o menor desvio padrão (0,58%), indicando maior estabilidade do modelo nessa configuração. A ANOVA de uma via confirmou a ausência de diferenças entre os grupos ( $p = 1,00$ ); o teste Tukey–HSD não identificou pares significativamente distintos (todas as comparações com  $p_{\text{ajust}} = 1,00$ ).

Para a MLP, observa-se um desempenho consistentemente alto e estável, com acurácias próximas a 99,6% e desvios padrão muito baixos para a maioria dos valores

de  $K$ , alcançando em  $K = 5$  a maior acurácia média (99,63%) e o menor desvio padrão (0,19%). Uma pequena queda de desempenho foi observada em  $K = 15$ , mas a ANOVA não indicou diferença estatística global entre os grupos ( $p = 0,520$ ), e o teste Tukey–HSD subsequente não encontrou pares significativamente distintos. Isso demonstra a robustez do modelo frente a diferentes estratégias de particionamento.

Considerando que os grupos não apresentaram diferença estatística global, a escolha de  $K^*$  passa a ser guiada exclusivamente por critérios de estabilidade e custo computacional, assim, considerando que  $K = 5$  proporcionou o desempenho mais estável (menor variância) para ambos os classificadores e o melhor desempenho médio para a MLP, além de implicar um menor custo computacional, ele é selecionado como o valor ótimo. Adota-se, portanto,  $K^* = 5$  nas rodadas subsequentes.

### 5.5.1.1 Matrizes de confusão e métricas por classe

A Figura 22 apresenta as matrizes de confusão agregadas para  $K^* = 5$  folds, representando a soma das predições sobre todo o conjunto de dados.<sup>2</sup> Cada célula exibe a contagem absoluta de amostras (negrito) e seu percentual relativo ao total de dados.

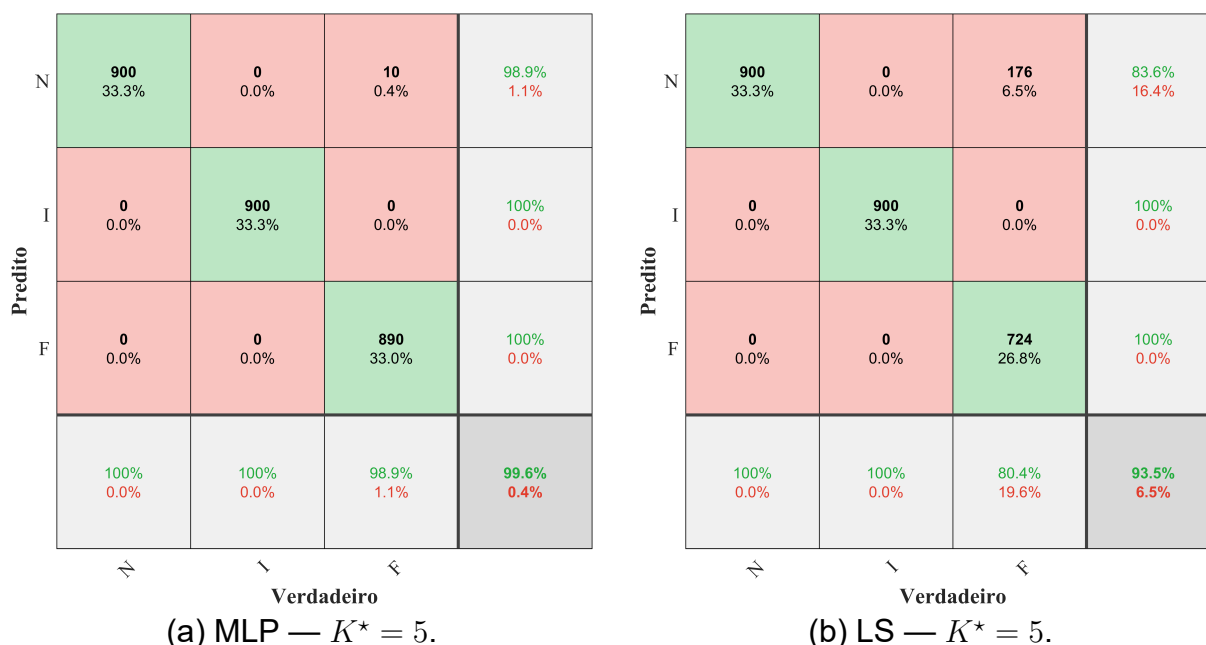
A matriz do LS (Fig. 22b) demonstra precisão igual a 100% para as classes Falta (F) e Ilhamento (I), mas seu principal erro consiste na classificação incorreta de 176 amostras de Falta (F) como Nominal (N), resultando em um *recall* de apenas 80,4% para a classe F.

Em contrapartida, a MLP (Fig. 22a) exibe um desempenho superior e com um padrão de erro mínimo. As classes Ilhamento (I) e Nominal (N) são perfeitamente identificadas (*recall* de 100%). O único erro residual ocorre na classe Falta (F), onde apenas 10 amostras (1,1%) são confundidas com a classe Nominal (N). Essa drástica redução na confusão F→N, em comparação com o LS, justifica a acurácia global superior da MLP.

Com  $N_{\text{tot}} = 2700$  janelas, a acurácia é reportada com duas casas decimais e intervalo de confiança (IC95%): LS 93,48% (IC95% 92,49–94,35) e MLP 99,63% (IC95% 99,32–99,80). As métricas por classe na Tabela 18, calculadas a partir das matrizes agregadas, quantificam essas observações.

O LS alcança precisão perfeita para F e I, mas ao custo de um baixo *recall* para a classe Falta (80,44%). A MLP, por sua vez, apresenta um perfil de desempenho excepcional, com scores F1 superiores a 99,4% para todas as classes, o que a qualifica como uma solução extremamente robusta e confiável.

<sup>2</sup> A matriz de confusão *média* por *fold* (média das matrizes normalizadas de cada *fold*) também foi calculada para avaliar a estabilidade entre partições. Seus resultados são consistentes com a análise agregada aqui apresentada, não sendo exibida por brevidade.

Figura 22 – Matrizes de confusão agregadas ( $K^* = 5$ ).

Legenda: Matrizes de confusão agregadas para os classificadores MLP e LS, com  $K^* = 5$ . Os painéis laterais resumem o desempenho por classe predita (aprox. precisão por linha) e os painéis inferiores por classe verdadeira (aprox. *recall* por coluna). O canto inferior direito apresenta a acurácia global do modelo.

Tabela 18 – Métricas por classe em  $K = 5$  (valores em %).

Classe	LS			MLP		
	precision	recall	F1	precision	recall	F1
N	83,64	100,00	91,09	98,90	100,00	99,45
I	100,00	100,00	100,00	100,00	100,00	100,00
F	100,00	80,44	89,16	100,00	98,89	99,44

Em síntese, a escolha de  $K^* = 5$  é justificada pela obtenção do desempenho mais estável e preciso, confirmando o alto poder discriminativo do conjunto de atributos selecionado e estabelecendo uma base sólida para as próximas rodadas de otimização.

### 5.5.2 Rodada 2 — melhor subconjunto para cada $B^*$

Com o particionamento fixado em  $K^*=5$  (Seção 5.5) e o conjunto-base de 10 atributos selecionado por FDR (cinco cumulantes de 2ª ordem e cinco de 4ª, Seção 5.4), conduz-se nesta rodada uma busca exaustiva para, em cada cardinalidade  $B^* \in \{1, \dots, 10\}$ , determinar o subconjunto ótimo  $S_{B^*}^*$  que maximize simultaneamente o desempenho dos dois classificadores avaliados (LS e MLP). Para manter a notação inequívoca, adota-se:  $B$  para o conjunto-base de 10 atributos;  $B^*$  para a cardinalidade

do subconjunto; e  $S_{B^*}^* \subset \mathcal{B}$  para o subconjunto ótimo de tamanho  $B^*$ .

A motivação desta rodada é que a seleção via FDR, por ser univariada e linear, ordena bem os atrasos mais informativos, mas não modela relações de complementaridade, nem elimina redundâncias entre atributos de ordens distintas. Assim, embora o conjunto completo com 10 atributos seja informativo, não há garantia de que ele seja o mais adequado para os classificadores considerados; frequentemente, subconjuntos menores combinam melhor atributos de 2ª e 4ª ordens e levam a soluções mais estáveis em validação cruzada. Por isso, avaliam-se exaustivamente todos os subconjuntos de cada cardinalidade, sempre sob o mesmo particionamento  $K^*$ , buscando o melhor compromisso entre LS e MLP.

### 5.5.2.1 Resultados por cardinalidade

Para cada cardinalidade  $B^*$ , enumeraram-se todas as combinações possíveis de  $B^*$  atributos em  $\mathcal{B}$  e aplicou-se validação cruzada estratificada com dobras fixas e compartilhadas ( $K^* = 5$ ). O classificador LS foi ajustado por mínimos quadrados com viés (pseudoinversa). A MLP empregou uma camada escondida com 20 neurônios, função de ativação `tansig` e algoritmo de treinamento `trainlm` (100 épocas, `goal = 10-3`), sem partição de validação interna.

O critério de escolha de  $S_{B^*}^*$  foi o seguinte: (i) maximização mútua da média das acurácias de LS e MLP em validação cruzada; e (ii) em caso de empate nessa média, seleção do subconjunto com menor desvio-padrão médio.

As posições 1–10 apresentadas na Tabela 19 fixam o mapeamento entre índice e atributo efetivo: cada número que aparecer nos subconjuntos adiante refere-se exatamente a um desses cumulantes de 2ª ou 4ª ordens. Com esse mapeamento estabelecido, a próxima tabela passa a listar, para cada cardinalidade  $B^*$ , qual combinação de posições obteve o melhor desempenho em validação cruzada.

Tabela 19 – Mapeamento das posições do conjunto de atributos  $r[\tau]$ .

Posição	Atributo ( $r[\tau]$ )
1	r2 [211]
2	r4 [194]
3	r4 [69]
4	r4 [283]
5	r2 [70]
6	r4 [186]
7	r4 [439]
8	r2 [154]
9	r2 [62]
10	r2 [192]

A Tabela 20 resume então, para cada  $B^*$ , o subconjunto selecionado  $\mathcal{S}_{B^*}^*$  (em termos das posições da Tabela 19) e as acurácias de validação cruzada ( $K^* = 5$ ) dos classificadores LS e MLP, com média e desvio-padrão.

Tabela 20 – Melhor subconjunto  $\mathcal{S}_{B^*}^*$  por cardinalidade e desempenho em validação cruzada ( $K^* = 5$ ).

$B^*$	Subconjunto $\mathcal{S}_{B^*}^*$ posições	LS (Acurácia %)		MLP (Acurácia %)	
		média	dp	média	dp
1	[5]	85,00	18,77	99,33	0,53
2	[2, 5]	94,52	1,56	99,56	0,21
3	[1, 2, 4]	92,37	0,98	99,56	0,36
4	[1, 4, 5, 9]	93,33	0,81	99,74	0,31
5	[1, 2, 4, 5, 8]	93,48	0,50	99,89	0,17
6	[1, 3, 4, 5, 8, 9]	93,48	0,61	99,78	0,24
7	[3, 4, 5, 6, 7, 8, 9]	93,41	0,21	99,59	0,15
8	[1, 2, 3, 5, 6, 7, 8, 9]	93,48	0,30	99,59	0,15
9	[1, 2, 3, 5, 6, 7, 8, 9, 10]	93,48	0,80	99,48	0,56
10	[1, 2, 3, 4, 5, 6, 7, 8, 9, 10]	93,48	1,29	99,63	0,26

Os resultados permitem algumas observações relevantes. Primeiro, mesmo uma única característica de 2ª ordem ( $B^*=1$ , posição 5,  $r_2 [70]$ ) já fornece à MLP desempenho muito alto ( $99,33\% \pm 0,53\%$ ), mas o LS apresenta maior variabilidade ( $85,00\% \pm 18,77\%$ ), o que indica que uma única projeção linear não é suficiente para separar de modo homogêneo todas as dobras. Segundo, o salto para  $B^*=2$ , com o par  $\{r_4 [194], r_2 [70]\}$ , estabiliza o classificador linear e o leva a  $94,52\% \pm 1,56\%$ , enquanto a MLP mantém desempenho acima de 99,5%. Isso mostra que a combinação de uma informação de 4ª ordem com uma de 2ª ordem já captura boa parte da variabilidade discriminativa do problema.

Para cardinalidades intermediárias, notadamente  $B^* \geq 4$ , o LS passa a oscilar em torno de 93,3–93,5% com desvios-padrão pequenos, sugerindo que, a partir desse ponto, o espaço de atributos é suficientemente rico para uma fronteira quase linear sob o particionamento adotado. A MLP, por sua vez, alcança seu melhor resultado em  $B^* = 5$  ( $99,89\% \pm 0,17\%$ ), com leve decréscimo, ainda dentro da faixa de 99,5–99,8%, quando mais atributos são adicionados. Esse comportamento é coerente com a ideia de que um conjunto moderado de atributos bem escolhidos explora as interações entre cumulantes de ordens distintas sem introduzir ruído ou redundância excessivos.

Assim, a Rodada 2 estabelece, para cada cardinalidade  $B^*$ , o subconjunto  $\mathcal{S}_{B^*}^*$  que melhor combina atributos de 2ª e 4ª ordens sob o mesmo particionamento  $K^*$ . Conjuntos moderados ( $B^* \in [5,8]$ ) oferecem desempenho praticamente perfeito para

a MLP e desempenho estável para o LS, o que os torna candidatos naturais para as etapas seguintes. A escolha final da cardinalidade ótima será refinada na Rodada 3, na qual a MLP será reavaliada com arquiteturas alternativas para cada  $\mathcal{S}_{B^*}^*$ , empregando a estrutura aqui definida como referência de atributos. O LS, por ser determinístico dado  $B^*$  e  $K^*$ , não requer ajuste adicional nessa etapa.

### 5.5.3 Rodada 3 — ajuste de arquitetura da MLP

Nesta rodada a escolha do número de neurônios da MLP foi feita com base em: (i) no particionamento e nos subconjuntos ótimos de atributos encontrados nas Rodadas anteriores; (ii) na avaliação sistemática do desempenho da rede para diferentes tamanhos de camada escondida; e (iii) em uma etapa estatística (ANOVA e teste de Tukey), seguida de um critério composto de decisão para selecionar a arquitetura mais adequada. O objetivo foi evitar escolher uma rede apenas porque, por acaso, obteve a maior média em uma única execução de validação cruzada, priorizando em vez disso arquiteturas estatisticamente indistinguíveis da melhor e, dentro desse grupo, as mais estáveis e parcimoniosas.

Para cada cardinalidade de atributos  $B^* \in \{1, \dots, 10\}$  aplicou-se o seguinte procedimento:

- (i) Selecionou-se o subconjunto  $\mathcal{S}_{B^*}^*$  identificado na Tabela 20;
- (ii) Manteve-se o mesmo particionamento estratificado de validação cruzada com  $K^* = 5$  dobras;
- (iii) Varreu-se o número de neurônios na camada escondida na faixa

$$n_h \in \{2, 3, \dots, 20\},$$

treinando em cada caso uma MLP com a mesma configuração de treinamento (`tansig`, `trainlm`, 100 épocas, `goal = 10-3`) e sem partição de validação interna adicional;

- (iv) Para cada par  $(B^*, n_h)$  foram armazenadas as acurácias por dobra, a acurácia média e o desvio padrão em validação cruzada;
- (v) Sobre o conjunto de acurácias por dobra foi aplicada uma ANOVA de uma via, tendo o número de neurônios como fator;
- (vi) Quando a ANOVA não indicou diferença global significativa, todas as arquiteturas foram consideradas equivalentes; quando indicou diferença, aplicou-se o teste de comparações múltiplas de Tukey–HSD para identificar o subconjunto de arquiteturas que não eram significativamente piores que as melhores (isto é, com  $p \geq 0,05$  em relação ao grupo de maior acurácia);

(vii) Dentro desse subconjunto estatisticamente indiferenciável, calculou-se um índice composto

$$J(n_h) = \bar{a}(n_h) - s(n_h) - \lambda \frac{n_h}{n_h^{\max}}, \quad (5.3)$$

em que  $\bar{a}(n_h)$  é a acurácia média na validação cruzada,  $s(n_h)$  é o desvio padrão associado e o último termo penaliza arquiteturas mais complexas ( $\lambda = 10^{-2}$  e  $n_h^{\max} = 20$ ). Assim, duas redes com mesma média caem no critério de desempate pela menor variabilidade; persistindo o empate, vence a de menor número de neurônios.

Aplicando esse protocolo para cada subconjunto  $\mathcal{S}_{B^*}^*$ , chegaram-se às arquiteturas resumidas na Tabela 21.

Tabela 21 – Arquiteturas MLP selecionadas por cardinalidade  $B^*$ .

$B^*$	Subconjunto $\mathcal{S}_{B^*}^*$	$n_h^*$	Acurácia (%)	dp (%)
1	[5]	2	99,33	0,10
2	[2, 5]	5	99,59	0,15
3	[1, 2, 4]	7	100,00	0,00
4	[1, 4, 5, 9]	2	99,96	0,08
5	[1, 2, 4, 5, 8]	3	99,96	0,08
6	[1, 3, 4, 5, 8, 9]	2	99,93	0,10
7	[3, 4, 5, 6, 7, 8, 9]	2	99,93	0,17
8	[1, 2, 3, 5, 6, 7, 8, 9]	2	99,81	0,19
9	[1, 2, 3, 5, 6, 7, 8, 9, 10]	2	99,89	0,10
10	[1, 2, 3, 4, 5, 6, 7, 8, 9, 10]	3	99,96	0,08

Alguns pontos merecem destaque:

- Já com  $B^* = 1$  (um único cumulante de 2ª ordem) uma MLP extremamente compacta, com apenas dois neurônios escondidos, atinge 99,33% de acurácia média com desvio padrão de apenas 0,10%. Isso indica que o problema já apresenta forte separabilidade mesmo em uma dimensão.
- A partir de  $B^* = 2-3$  observa-se um claro platô de desempenho. Em particular, para  $B^* = 3$  o subconjunto [1,2,4] leva a 100,00% de acurácia média com desvio padrão nulo, ao custo de uma camada escondida um pouco maior ( $n_h^* = 7$ ).
- Para cardinalidades intermediárias ( $B^* \in \{4, \dots, 7\}$ ) surgem combinações que estabilizam o desempenho em torno de 99,9%, agora com arquiteturas mais enxutas ( $n_h^* = 2$  ou 3). Isso mostra que, uma vez combinados cumulantes de 2ª e 4ª ordens, poucos neurônios são suficientes para saturar a acurácia.
- Mesmo com quase todos os 10 atributos disponíveis, a rede não cresce significativamente. Para  $B^* = 10$  a melhor solução segundo a Equação 5.3 utiliza apenas

três neurônios escondidos e atinge  $99,96\% \pm 0,08\%$ , evidenciando que o problema, na janela de dois ciclos e com os cumulantes selecionados, não exige arquiteturas profundas nem largas.

Nos experimentos conduzidos, os testes de Tukey–HSD não indicaram diferenças estatisticamente significativas entre as arquiteturas avaliadas em cada cardinalidade  $B^*$ , de modo que as redes selecionadas via  $J(n_h)$  devem ser interpretadas como escolhas parcimoniosas dentro de um conjunto de soluções estatisticamente equivalentes.

Em resumo, a Rodada 3 mostra que a alta acurácia observada na Rodada 2 não dependia de uma MLP grande (como a configuração inicial de 20 neurônios utilizada apenas como referência). Ao controlar estatisticamente as diferenças entre arquiteturas e penalizar a complexidade, chega-se a um conjunto de redes muito pequenas (tipicamente 2–3 neurônios, com até 7 no caso mais extremo) que preservam o desempenho de 99,8–100% e apresentam variabilidade mínima entre as dobras. Essas arquiteturas serão, portanto, as candidatas naturais a serem consolidadas na Rodada 4, quando serão comparadas entre si para definição do melhor modelo não linear estudado.

#### 5.5.4 Rodada 4 — consolidação da arquitetura

Na Rodada 3 foram obtidas, para cada cardinalidade  $B^*$ , as arquiteturas candidatas  $MLP(B^*, n_h^*)$  que maximizam o  $J(n_h)$  sob o particionamento fixo  $K^* = 5$ . Em todos os casos não se indicou diferenças estatisticamente significativas entre as arquiteturas avaliadas para um mesmo  $B^*$ , isto é, dentro de cada cardinalidade as redes com diferentes números de neurônios não puderam ser distinguidas em termos de acurácia média. Assim, as arquiteturas selecionadas pela Equação 5.3 devem ser entendidas como representantes parcimoniosas de um conjunto de soluções estatisticamente equivalentes.

Nesta rodada, essas arquiteturas  $MLP(B^*, n_h^*)$  são reavaliadas de forma comparativa, mantendo-se:

- o mesmo conjunto de atributos  $\mathcal{S}_{B^*}^*$  definido na Rodada 2;
- o mesmo particionamento estratificado de validação cruzada ( $K^* = 5$ );
- o mesmo número de neurônios escondidos  $n_h^*$  selecionado na Rodada 3 para cada  $B^*$ .

O objetivo desta etapa é consolidar a escolha da arquitetura final à luz de um critério de compromisso entre desempenho médio, estabilidade (desvio padrão) e com-

plexidade (número de atributos e neurônios), comparando ainda a MLP “ótima” com um classificador linear de referência (LS) que utiliza o mesmo vetor de atributos.

A Tabela 22 resume as acurácias em validação cruzada para cada par  $(B^*, n_h^*)$  nesta rodada.

Tabela 22 – Desempenho em validação cruzada das arquiteturas MLP $(B^*, n_h^*)$  reavaliadas na Rodada 4.

$B^*$	$n_h^*$	Acurácia (%)	dp (%)
1	2	99,33	0,10
2	5	99,44	0,29
3	7	99,81	0,32
4	2	99,93	0,10
5	3	99,85	0,33
6	2	99,96	0,08
7	2	99,96	0,08
8	2	93,89	12,53
9	2	99,93	0,17
10	3	99,85	0,15

Algumas observações podem ser feitas:

- Para  $B^* = 1$  e  $B^* = 2$  a MLP já apresenta desempenho elevado (entre 99,3% e 99,4%) com variabilidade pequena, mas o modelo trabalha com um ou dois atributos apenas e não explora completamente a complementaridade entre cumulantes de 2<sup>a</sup> e 4<sup>a</sup> ordens.
- A partir de  $B^* = 3$ , com exceção de  $B^* = 8$ , forma-se um platô de desempenho em torno de 99,8%–99,96%, com desvios padrão inferiores a 0,4%. As cardinalidades  $B^* \in \{3,4,5,6,7,9,10\}$  pertencem claramente a esse platô.
- A cardinalidade  $B^* = 8$  constitui um ponto fora da curva, com acurácia média de 93,89% e desvio padrão elevado (12,53%), indicando forte sensibilidade a partições e inicializações. Tal combinação é, portanto, descartada como candidata à arquitetura final.
- As melhores médias globais são obtidas para  $B^* = 6$  e  $B^* = 7$ , ambas com 99,96% de acurácia e desvio padrão de 0,08%, usando apenas dois neurônios na camada escondida. Em termos de complexidade,  $B^* = 6$  utiliza seis atributos, enquanto  $B^* = 7$  emprega sete atributos com o mesmo número de neurônios.

Como os testes na Rodada 3 não apontaram diferenças estatisticamente significativas entre as arquiteturas candidatas dentro de cada  $B^*$ , e as médias da Tabela 22

se encontram, em sua maioria, em um patamar praticamente saturado, as arquiteturas devem ser consideradas equivalentes do ponto de vista estatístico. Diante desse cenário de empate, adotou-se explicitamente um critério de consolidação baseado em:

- (a) selecionar apenas as arquiteturas com acurácia média  $\geq 99,8\%$  e desvio padrão  $\leq 0,2\%$ ;
- (b) dentro desse grupo, preferir a menor cardinalidade  $B^*$ ; em caso de empate, preferir o menor número de neurônios  $n_h^*$ .

Aplicando esse critério à Tabela 22, chega-se à escolha de  $B^* = 6$  atributos e  $n_h^* = 2$  neurônios como arquitetura final da MLP. Essa configuração corresponde ao subconjunto  $\mathcal{S}_6^* = [1, 3, 4, 5, 8, 9]$ , corresponde aos cumulantes de 2ª ordem nos atrasos  $\tau = \{211, 70, 154, 62\}$  e de 4ª ordem em  $\tau = \{69, 283\}$  amostras, isto é, às posições  $r_2[211]$ ,  $r_2[70]$ ,  $r_2[154]$ ,  $r_2[62]$ ,  $r_4[69]$  e  $r_4[283]$ . Em termos de topologia, o MDI proposto em janela fixa de dois ciclos é implementado por uma rede do tipo

$$\text{MLP}(6,2,3),$$

com seis entradas (atributos), dois neurônios na camada escondida e três saídas (classes  $\{N, I, F\}$ ), treinada com função de ativação `tansig` na camada escondida e de saída e algoritmo `trainlm`.

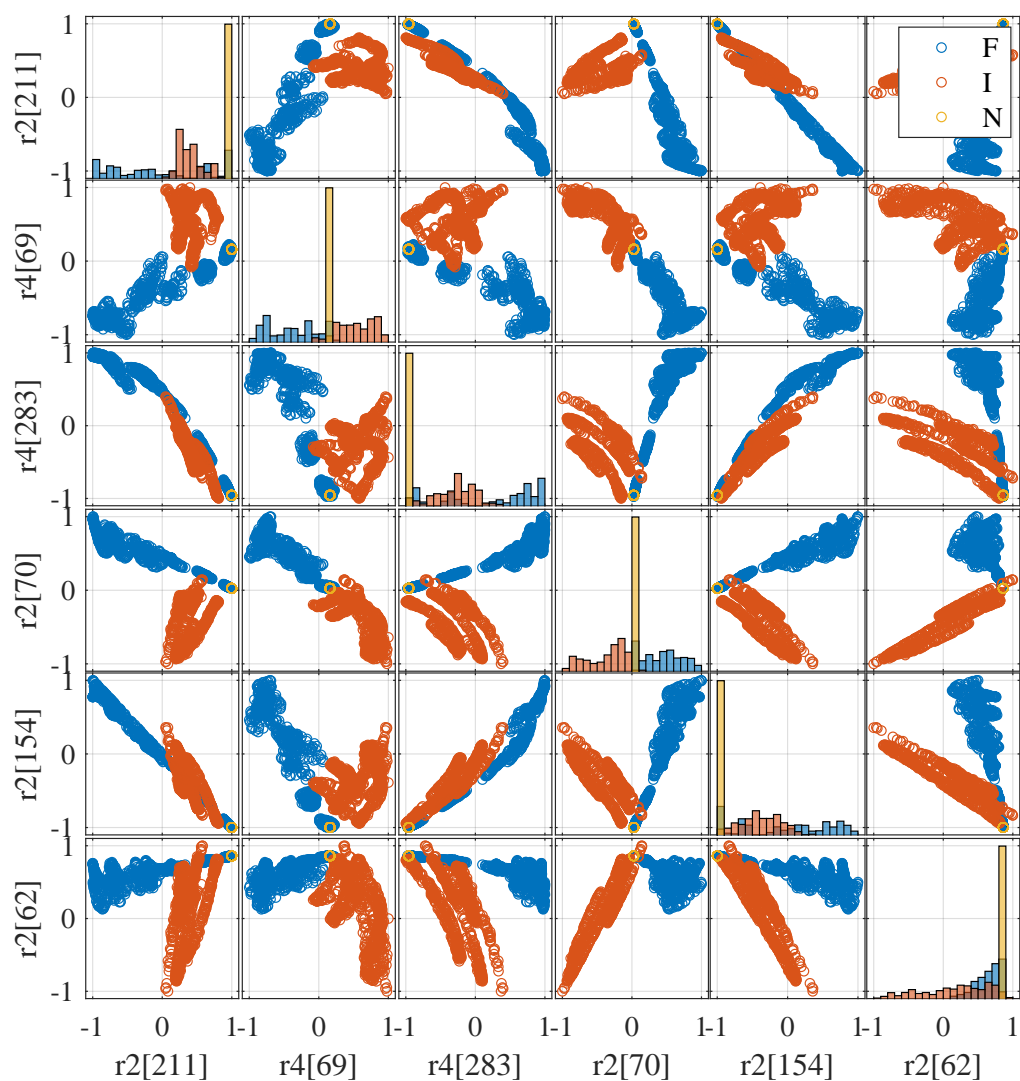
Para efeito de comparação linear, considera-se o classificador LS ajustado sobre o mesmo conjunto de atributos  $\mathcal{S}_6^*$ . Como mostrado na Tabela 20, o LS atinge acurácia média de aproximadamente 93,5% neste cenário, com desvio padrão inferior a 1%. Assim, a MLP final de baixa complexidade ( $B^* = 6$ ,  $n_h^* = 2$ ) fornece um ganho absoluto de cerca de 6 p.p. em acurácia global em relação ao classificador linear de referência, mantendo ao mesmo tempo número reduzido de parâmetros e custo computacional compatível com aplicações quase em tempo real.

#### 5.5.4.1 Visualização do espaço de atributos da arquitetura final

Para complementar a análise numérica da Rodada 4, a Figura 23 apresenta a matriz de dispersão de todas as combinações bidimensionais dos seis cumulantes que compõem o vetor de atributos final em janela fixa,  $\mathcal{S}_6^*$ . Cada painel mostra a projeção das janelas de dois ciclos em um par de atributos, com as cores identificando as classes Nominal (N), Ilhamento (I) e Falta (F). Ainda que a figura seja densa, é possível notar que as nuvens de pontos tendem a se organizar em faixas bem definidas ao longo de curvas não lineares, com sobreposição relativamente limitada em boa parte das projeções.

Para tornar mais evidente a estrutura de separação induzida por  $\mathcal{S}_6^*$ , a Figura 24 detalha a projeção no plano formado pelos cumulantes  $r_2[70]$  e  $r_4[69]$ , pertencentes ao

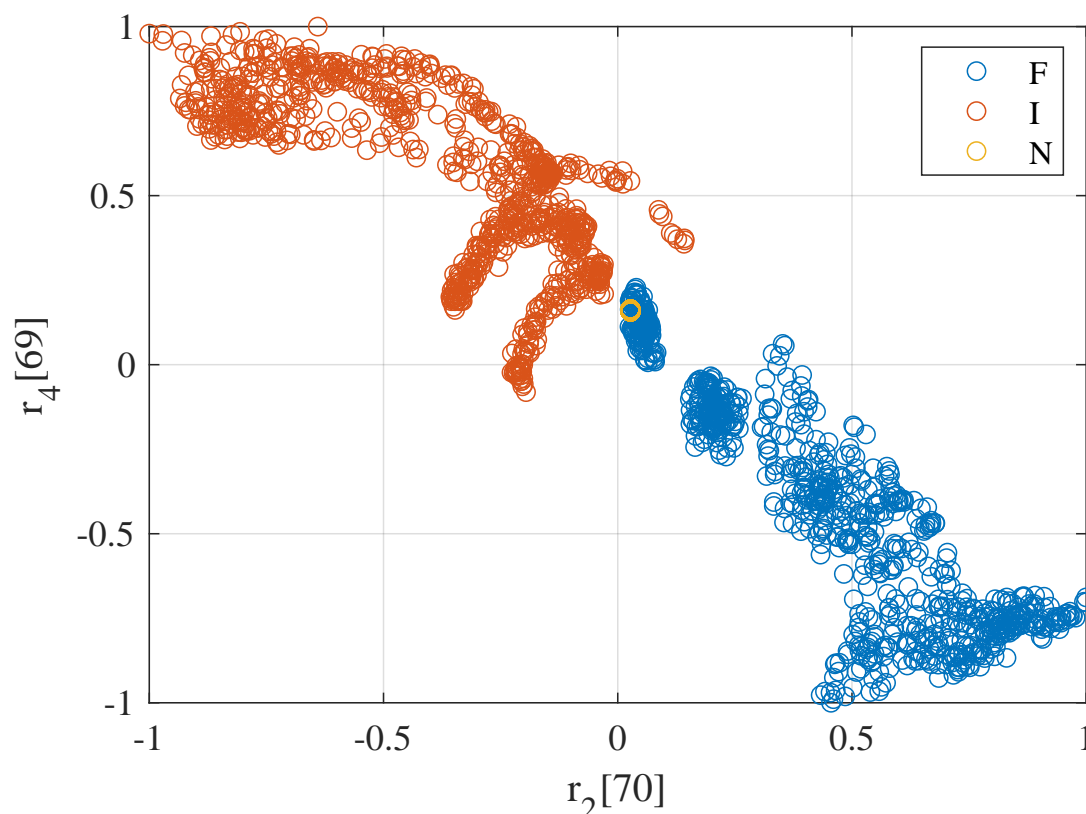
Figura 23 – Matriz de dispersão dos seis atributos cumulantes selecionados (janela fixa).



Legenda: Matriz de dispersão (pares bidimensionais) construída a partir do subconjunto final de 6 atributos cumulantes,  $S_6^*$ , para o caso de janela fixa. As cores codificam as classes Nominal (N), Ilhamento (I) e Falta (F) em cada projeção 2D.

vetor final. Nesta vista, as três classes formam regiões com geometria característica: a classe Nominal concentra-se em uma faixa relativamente compacta, enquanto as classes Ilhamento e Falta se distribuem em ramos distintos ao longo das diagonais. Embora a separação perfeita só seja alcançada pela combinação multivariada dos seis atributos na MLP(6,2,3), essa projeção ilustra de maneira mais clara como a combinação de cumulantes de 2ª e 4ª ordens contribui para a discriminação entre os regimes operativos.

Figura 24 – Projeção no plano  $r_2[70] \times r_4[69]$  (janela fixa).



Legenda: Projeção bidimensional do conjunto de dados no plano formado pelos cumulantes  $r_2[70]$  e  $r_4[69]$ , pertencentes ao vetor final de atributos  $S_6^*$ . A separação entre as nuvens de pontos indica a discriminação geométrica entre os regimes Nominal (N), Ilhamento (I) e Falta (F) para este par representativo.

Essa arquitetura consolidada será utilizada nas seções seguintes como base para a avaliação em janela deslizante e para a validação independente no conjunto fora da grade.

## 5.6 TREINAMENTO EM JANELA DESLIZANTE

Nesta seção, a arquitetura final obtida na janela fixa, MLP(6,2,3) com vetor de atributos  $S_6^*$ , é treinada no regime de janela deslizante descrito na Seção 4.6. O obje-

tivo é adaptar o MDI ao cenário em que o instante de detecção não está rigidamente ancorado ao início do distúrbio, mantendo, contudo, a mesma representação estatística de ordem superior consolidada na etapa de otimização.

### 5.6.1 Formação do conjunto em janela deslizante

O ponto de partida desta etapa são exatamente as tensões trifásicas simuladas na grade  $3 \times 3$  apresentada no Capítulo 3, isto é, os mesmos perfis de operação que deram origem às janelas de dois ciclos utilizadas na etapa de otimização em janela fixa. Na etapa anterior, para cada evento (Nominal, Ilhamento, Falta) e para cada combinação de parâmetros da grade, era extraído um único recorte de dois ciclos, rigidamente ancorado em torno do instante do distúrbio.

No regime de janela deslizante, o procedimento é reorganizado da seguinte forma. Para cada realização de tensão trifásica  $\mathbf{v}(n)$  pertencente à grade  $3 \times 3$ , seleciona-se inicialmente uma janela ampla de quatro ciclos em torno do instante de ocorrência do evento (ou, no caso nominal, do ponto central de referência). Sobre essa janela de quatro ciclos aplica-se então um janelamento deslizante de comprimento fixo de dois ciclos e sobreposição de 50%. Isso resulta, para cada realização, em três janelas menores de dois ciclos, denotadas por  $\mathcal{W}_1$ ,  $\mathcal{W}_2$  e  $\mathcal{W}_3$ :

- $\mathcal{W}_1$ : contém majoritariamente o regime nominal, com apenas uma pequena fração do distúrbio ao final do recorte;
- $\mathcal{W}_2$ : é centrada aproximadamente no mesmo ponto que a janela de dois ciclos utilizada na otimização em janela fixa, representando uma condição intermediária, em que o transitório nominal–distúrbio é bem capturado;
- $\mathcal{W}_3$ : contém predominantemente o regime perturbado (ilhamento ou falta), com pouca ou nenhuma contribuição do regime puramente nominal.

Do ponto de vista estatístico, cada sinal da grade  $3 \times 3$  que anteriormente gerava um único vetor de atributos em janela fixa passa a gerar três vetores associados às janelas  $\mathcal{W}_1$ ,  $\mathcal{W}_2$  e  $\mathcal{W}_3$ . Mantendo-se o mesmo número de casos e repetições utilizado na etapa de janela fixa, obtêm-se, para cada classe, três vezes mais padrões de treino no regime deslizante. No experimento considerado, isso resulta em  $N_{\text{jan}} = 2700$  janelas por classe (Nominal, Ilhamento e Falta), totalizando 8100 janelas rotuladas, todas derivadas do mesmo conjunto de simulações em grade  $3 \times 3$  com SNR de 70 dB.

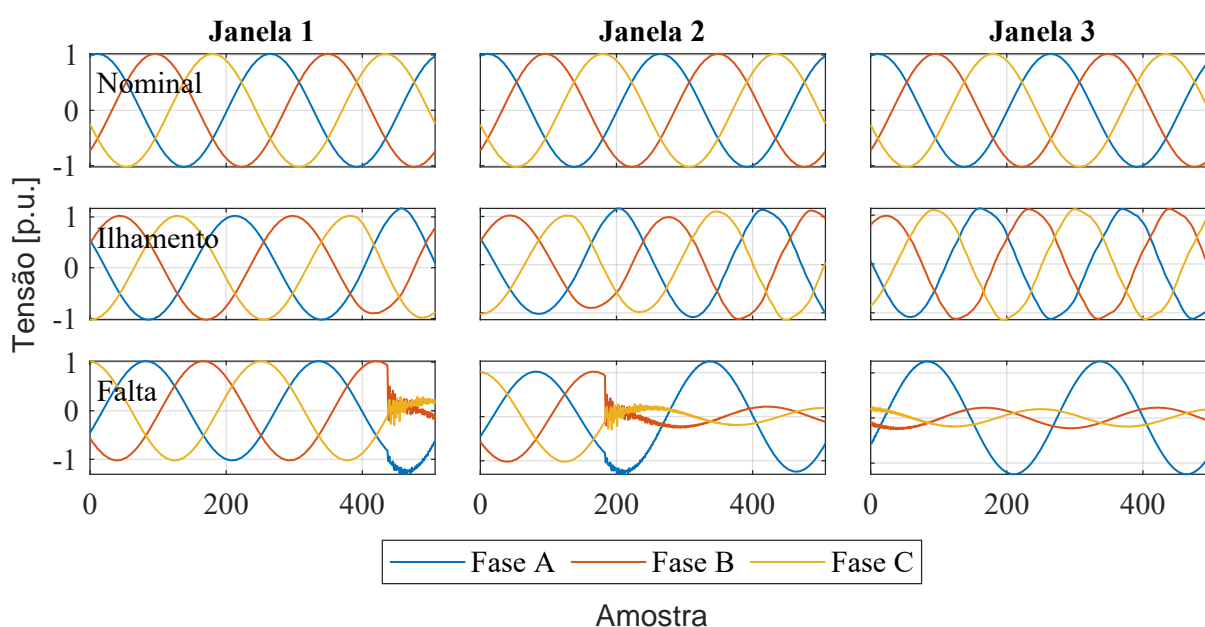
Esse arranjo permite avaliar se o conjunto de atributos ótimos extraídos em janela fixa ( $S_6^*$ ) permanece discriminante quando a decisão é tomada a partir de recortes que capturam diferentes frações do transitório nominal–distúrbio, desde janelas ainda

predominantemente nominais até janelas quase totalmente dominadas pela perturbação.

Para tornar mais concreta essa organização, a Figura 25 ilustra, para um caso representativo da grade  $3 \times 3$ , as três janelas  $\mathcal{W}_1$ ,  $\mathcal{W}_2$  e  $\mathcal{W}_3$  extraídas de uma mesma realização de tensão trifásica. As linhas superior, intermediária e inferior correspondem, respectivamente, às classes Nominal, Ilhamento e Falta.

Observa-se que a primeira janela capta majoritariamente o regime nominal. A segunda recorta o transitório nominal–distúrbio em torno do ponto de referência utilizado na janela fixa, enquanto a terceira concentra quase exclusivamente o regime perturbado (ilhamento ou falta).

Figura 25 – Exemplo de janelamento deslizante em uma realização trifásica.



Legenda: Exemplo do esquema de janela deslizante aplicado a uma realização de tensão trifásica da grade  $3 \times 3$ . Cada painel corresponde a uma sub-janela de dois ciclos:  $\mathcal{W}_1$  predominantemente nominal,  $\mathcal{W}_2$  centrada na transição nominal–distúrbio e  $\mathcal{W}_3$  contendo majoritariamente o regime perturbado. As linhas superior, intermediária e inferior representam, respectivamente, os casos Nominal, Ilhamento e Falta.

### 5.6.2 Projeção PCA global e cálculo de cumulantes por janela

Seguindo a metodologia descrita na Seção 4.2, o processamento em janela deslizante não introduz uma nova etapa de estimação de componentes, mas reutiliza a mesma base obtida na fase de otimização em janela fixa. Em particular, adota-se um vetor de rotação global  $\bar{\mathbf{r}}_1 \in \mathbb{R}^3$ , correspondente à primeira componente média cal-

culada a partir das tensões trifásicas do evento de ilhamento na grade  $3 \times 3$ , e mantido fixo para todos os cenários considerados nesta seção.

Cada janela de tensão trifásica  $\mathbf{v}[n] \in \mathbb{R}^3$  é projetada sobre essa direção principal por meio de

$$c_1[n] = \mathbf{v}[n] \bar{\mathbf{r}}_1,$$

produzindo, para cada janela  $\mathcal{W}_1$ ,  $\mathcal{W}_2$  e  $\mathcal{W}_3$ , uma série escalar  $c_1[n]$ . Em conformidade com a etapa de otimização, nenhuma outra componente é utilizada, uma vez que a primeira componente já se mostrou suficiente para concentrar a maior parcela da variância e discriminar adequadamente as classes na janela fixa. Esse procedimento é aplicado a todos os eventos, casos e repetições, resultando em uma coleção de séries escalares, uma para cada janela deslizante.

Sobre cada uma dessas séries calcula-se, em seguida, um conjunto reduzido de estatísticas de ordem superior. Avaliam-se diretamente as cumulantes de segunda e de quarta ordens apenas nos atrasos  $\tau$  associados ao vetor ótimo  $\mathcal{S}_6^*$ , conforme definido na Seção 5.5. No caso em estudo, utilizam-se as cumulantes de segunda ordem nos atrasos  $\tau = \{211, 70, 154, 62\}$  amostras e as cumulantes de quarta ordem em  $\tau = \{69, 283\}$  amostras. Assim, para cada janela deslizante obtém-se diretamente o vetor de atributos

$$\mathbf{s} = [r_2[211], r_2[70], r_2[154], r_2[62], r_4[69], r_4[283]]^\top.$$

Desse modo, o mapeamento de “tensão trifásica em janela de dois ciclos  $\rightarrow$  vetor de atributos” permanece idêntico ao utilizado na fase de otimização, sendo agora apenas aplicado a múltiplos recortes ao longo do transitório nominal–distúrbio. Isso garante que os resultados apresentados a seguir reflitam exclusivamente o impacto do janelamento deslizante sobre o desempenho do MDI, sem a introdução de novos graus de liberdade na etapa de extração de características.

### 5.6.3 Vetor de atributos e treinamento *hold-out* 70/30

Como detalhado na subseção anterior, cada janela deslizante  $\mathcal{W}_k$  é mapeada em um vetor de atributos  $\mathbf{s} \in \mathbb{R}^6$ , construído a partir das cumulantes de segunda e quarta ordens associadas ao conjunto ótimo  $\mathcal{S}_6^*$ . No regime de janela deslizante, cada realização da grade  $3 \times 3$  passa a contribuir com três vetores  $\mathbf{s}$  (um por janela), de modo que, ao final, são obtidas  $N_{\text{jan}} = 2700$  janelas por classe (Nominal, Ilhamento e Falta), totalizando 8100 vetores de atributos rotulados.

Os vetores de atributos são então organizados em uma matriz de dados

$$\mathbf{X} = [\mathbf{S}_N \ \mathbf{S}_I \ \mathbf{S}_F] \in \mathbb{R}^{6 \times 8100},$$

em que cada bloco  $S_N, S_I, S_F \in \mathbb{R}^{6 \times 2700}$  reúne os vetores de atributos de uma mesma classe. Os rótulos correspondentes são codificados em formato *one-hot*, resultando em uma matriz de alvos  $Y \in \{0,1\}^{3 \times 8100}$  com uma linha por classe (N, I, F) e uma coluna por janela.

Antes do treinamento, cada componente do vetor de atributos é escalonada para o intervalo  $[-1,1]$  por meio de uma normalização linear aplicada linha a linha em  $X$ , em conformidade com o protocolo descrito na Seção 4.6.1. Essa etapa garante que todas as características contribuam em uma mesma escala numérica para o ajuste dos classificadores.

A avaliação de generalização em janela deslizante é conduzida por meio de um esquema de partição *hold-out*, com aproximadamente 70% dos padrões destinados ao treino e 30% reservados para teste interno. As colunas de  $X$  e  $Y$  são inicialmente embaralhadas de forma aleatória; em seguida, o conjunto é dividido em dois subconjuntos disjuntos,  $D_{tr}^{slide}$  (treino) e  $D_{ts}^{slide}$  (teste), contendo  $N_{tr} = 5670$  e  $N_{ts} = 2430$  amostras, respectivamente. Como as classes são originalmente balanceadas, a amostragem aleatória simples resulta, neste experimento, em contagens de janelas por classe bastante próximas em cada subconjunto, conforme resumido na Tabela 23. As colunas de  $\%_{tr}$  e  $\%_{ts}$  evidenciam que, mesmo com a aleatorização imposta pelo particionamento *hold-out*, as três classes permanecem aproximadamente balanceadas em ambos os subconjuntos, com proporções em torno de um terço em treino e teste. A coluna  $\%_{total}$  confirma que cada classe responde por cerca de 33,3% do conjunto completo de janelas.

Tabela 23 – Contagem de janelas por classe nos conjuntos de treino e teste (janela deslizante).

Classe	$N_{tr}$	$\%_{tr}$	$N_{ts}$	$\%_{ts}$	$\%_{total}$
N	1922	33,9	778	32,0	33,3
I	1894	33,4	806	33,2	33,3
F	1854	32,7	846	34,8	33,3

#### 5.6.4 Desempenho da MLP(6,2,3) em janela deslizante

A rede neural adotada nesta etapa é exatamente a arquitetura consolidada na Rodada 4: uma MLP com seis entradas, dois neurônios na camada escondida e três saídas, MLP(6,2,3). A camada escondida utiliza função de ativação  $tansig$ , assim como a camada de saída, e o treinamento é realizado com o algoritmo de retropropagação  $trainlm$ . O número máximo de épocas é fixado em  $10^3$ , sem critério explícito de validação interna, uma vez que o papel de conjunto de teste é cumprido diretamente por  $D_{ts}^{slide}$ .

A rede é treinada apenas com os padrões de  $\mathcal{D}_{tr}^{slide}$ , produzindo um conjunto de pesos específico para o regime de janela deslizante. Em seguida, o modelo é avaliado tanto no conjunto de treino quanto no de teste, e as predições são convertidas em rótulos de classe a partir da posição do maior valor na saída (regra do máximo).

#### 5.6.4.1 Matrizes de confusão e métricas por classe

A Tabela 24 resume as acurácias globais obtidas pela MLP(6,2,3) em treino e teste, enquanto a Figura 26 apresenta a matriz de confusão agregada no conjunto de teste em janela deslizante. A partir dessa matriz derivam-se as métricas de desempenho utilizadas em todo este capítulo.

Seja  $c \in \{N, I, F\}$  uma das três classes (Nominal, Ilhamento e Falta) e denote-se por  $TP_c$ ,  $FP_c$  e  $FN_c$  os números de verdadeiros positivos, falsos positivos e falsos negativos relativos a  $c$ . A acurácia global é definida como

$$Acc = \frac{\sum_c TP_c}{N_{tot}}, \quad (5.4)$$

em que  $N_{tot}$  é o número total de amostras avaliadas. A precisão, o *recall* e o F1-score por classe são dados por

$$Prec(c) = \frac{TP_c}{TP_c + FP_c}, \quad (5.5)$$

$$Rec(c) = \frac{TP_c}{TP_c + FN_c}, \quad (5.6)$$

$$F1(c) = \frac{2 Prec(c) Rec(c)}{Prec(c) + Rec(c)}. \quad (5.7)$$

As métricas por classe extraídas a partir da matriz de confusão de teste são reunidas na Tabela 25.

Tabela 24 – Acurácia global da MLP(6,2,3) em treino e teste (janela deslizante).

Conjunto	Acurácia (%)
Treino	97,76
Teste	98,52

Tabela 25 – Métricas por classe da MLP(6,2,3) no teste (janela deslizante).

Classe	Precisão (%)	Sensibilidade (%)	F1 (%)
N	99,62	100,00	99,81
I	98,49	97,15	97,81
F	97,54	98,46	98,00

Uma inspeção direta da matriz de confusão da Figura 26 mostra que todas as 778 janelas nominais do conjunto de teste foram corretamente identificadas como Nominal, sem qualquer falso alarme associado à classe de operação normal. Os erros

Figura 26 – Matriz de confusão da MLP(6,2,3) no teste (janela deslizante).

MLP - Teste			
Prev.	Alvo		
	N	I	F
N	778 32,0%	2 0,1%	1 0,0%
I	0 0,0%	783 32,2%	12 0,5%
F	0 0,0%	21 0,9%	833 34,3%

Legenda: As linhas indicam as classes *previstas* e as colunas as classes *alvo*. Em cada célula, apresenta-se a contagem e sua porcentagem no conjunto de teste; a diagonal principal representa as classificações corretas.

concentram-se exclusivamente na fronteira entre os regimes de Ilhamento e Falta: entre as 806 janelas de ilhamento, apenas 23 ( $\approx 2,9\%$ ) foram mal classificadas, sendo 2 como Nominal e 21 como Falta, ao passo que, das 846 janelas de falta, apenas 13 ( $\approx 1,5\%$ ) foram confundidas com outras classes, 1 como Nominal e 12 como Ilhamento. Em termos de desempenho para a função de proteção, isso se reflete em uma precisão de 98,49% e *recall* de 97,15% para a classe de ilhamento, valores compatíveis com um esquema altamente seletivo e pouco propenso a disparos indevidos.

Os resultados numéricos destas tabelas mostram que a MLP treinada em janela deslizante mantém desempenho elevado no conjunto de teste, com acurácia global próxima à observada na etapa de validação cruzada em janela fixa e sem discrepâncias pronunciadas entre as métricas das três classes. Isso indica que o vetor de atributos  $S_6^*$ , originalmente otimizado em dois ciclos ancorados no distúrbio, continua fornecendo um espaço de representação favorável mesmo quando a decisão é tomada a partir de janelas deslizantes ao longo da forma de onda.

### 5.6.5 Comparação com o classificador LS

Para fins de comparação linear, ajusta-se também um classificador por mínimos quadrados (LS) sobre o mesmo conjunto de atributos  $S_6^*$  e a mesma partição hold-out 70/30. O modelo é obtido pela solução de pseudoinversa da relação  $\mathbf{X}_{tr}^T \mathbf{W} \approx \mathbf{Y}_{tr}^T$ , resultando em uma matriz de pesos  $\mathbf{W}$  que implementa uma fronteira de decisão linear no espaço de atributos.

De forma análoga à MLP, o LS é avaliado nos conjuntos de treino e teste, e suas acurácias são reunidas na Tabela 26. A Figura 27 apresenta a matriz de confusão agregada no conjunto de teste, enquanto a Tabela 27 resume as métricas por classe.

Tabela 26 – Acurácia global do classificador LS em treino e teste (janela deslizante).

Conjunto	Acurácia (%)
Treino	80,26
Teste	79,59

Figura 27 – Matriz de confusão do classificador LS no teste (janela deslizante).

LS - Teste			
Prev.	Alvo		
	N	I	F
N	<b>778</b> 32,0%	<b>261</b> 10,7%	<b>221</b> 9,1%
I	<b>0</b> 0,0%	<b>535</b> 22,0%	<b>4</b> 0,2%
F	<b>0</b> 0,0%	<b>10</b> 0,4%	<b>621</b> 25,6%

Legenda: As linhas indicam as classes *previstas* e as colunas as classes *alvo*. Em cada célula, apresenta-se a contagem e sua porcentagem no conjunto de teste; a diagonal principal representa as classificações corretas.

Tabela 27 – Métricas por classe do classificador LS no teste (janela deslizante).

Classe	precision (%)	recall (%)	F1 (%)
N	61,75	100,00	76,35
I	99,26	66,38	79,55
F	98,42	73,40	84,09

No conjunto de teste, o classificador LS atinge acurácia global de aproximadamente 79,6 %, com F1-score entre 76,4 % e 84,1 % para as três classes (Tabela 27). A matriz de confusão da Figura 27 revela, contudo, um padrão bem distinto daquele observado para a MLP(6,2,3): todas as janelas nominais são corretamente reconhecidas ( $\text{recall}_N = 100\%$ ), mas à custa de um número expressivo de falsos nominais. Entre as 806 janelas de ilhamento em teste, 261 ( $\approx 33,6\%$ ) são classificadas como Nominal e 10 como Falta, enquanto, dentre as 846 janelas de falta, 221 ( $\approx 26,6\%$ ) são igualmente absorvidas pela classe Nominal e apenas quatro são confundidas com Ilhamento. Assim, as classes de distúrbio apresentam elevadas precisões ( $> 98\%$ ), mas sensibilidades mais modestas, sobretudo para o ilhamento ( $\text{recall}_I \approx 66,4\%$ ).

De maneira geral, o desempenho do LS em janela deslizante repete o padrão observado na etapa de janela fixa: embora a acurácia global seja elevada, ela permanece bem abaixo daquela obtida pela MLP(6,2,3), e o padrão de erros concentra-se em confusões entre os estados Nominal e de distúrbio. Esse comportamento é típico

de uma fronteira linear que privilegia o aglomerado nominal no espaço de atributos, levando a um classificador mais conservador e propenso a subestimar a ocorrência de ilhamento e faltas em regiões limítrofes. Em contraste, a MLP(6,2,3) consegue manter simultaneamente alta precisão e alto *recall* para as três classes, explorando de forma mais eficiente a separabilidade proporcionada pelas cumulantes de ordem superior e justificando a adoção de um modelo não linear para a decisão final.

### 5.6.6 Síntese e preparação para a validação

Os resultados desta seção mostram que a arquitetura MLP(6,2,3), otimizada em janela fixa de dois ciclos, mantém desempenho elevado quando treinada e testada em janelas deslizantes extraídas do mesmo conjunto de simulações (SNR de 70 dB). A combinação de cumulantes de segunda e quarta ordens em  $\mathcal{S}_6^*$  continua fornecendo um espaço de atributos altamente discriminante, no qual uma rede neural de baixa complexidade apresenta desempenho consistentemente superior ao classificador linear por mínimos quadrados, tomado como referência. Ainda assim, o LS permanece como um *linha de base* linear relevante, representando uma solução de menor complexidade para fins de comparação sistemática na etapa de validação.

A partir deste ponto, considera-se encerrada a fase de treinamento: os pesos da rede MLP(6,2,3) ajustada sobre  $\mathcal{D}_{tr}^{slide}$  e a matriz de pesos  $\mathbf{W}$  do classificador LS calculada a partir do mesmo conjunto são mantidos fixos e passam a definir, respectivamente, o *modelo principal* do MDI e o modelo linear de referência. Na próxima seção, esses modelos são apenas avaliados — sem qualquer novo ajuste de parâmetros — em cenários de validação independentes, incluindo condições fora da grade e diferentes níveis de SNR, de modo a investigar sua robustez e capacidade de generalização em situações mais próximas da operação real do sistema.

## 5.7 VALIDAÇÃO INDEPENDENTE

Concluída a etapa de treinamento e seleção de modelos em janela deslizante, resta verificar se o desempenho obtido se mantém quando o MDI é exposto a condições de operação não utilizadas no processo de otimização. Esta seção é dedicada justamente a essa validação independente, na qual o par formado pelo vetor de atributos  $\mathcal{S}_6^*$  e pelos classificadores treinados na seção anterior, a rede MLP(6,2,3) e o classificador linear por mínimos quadrados (LS), é avaliado em novos conjuntos de simulações.

A validação é conduzida sobre perfis de carga sorteados aleatoriamente no interior da região de não detecção definida na etapa de planejamento, bem como sob diferentes níveis de SNR, variando de 70 dB até 30 dB. Em todos os casos, o fluxo de processamento de sinais é rigorosamente o mesmo adotado na fase de treinamento:

aplicação da matriz de projeção obtida por PCA, cálculo das cumulantes de ordem superior e extração do vetor  $\mathcal{S}_6^*$ , seguido pela normalização e pela classificação. Nenhum parâmetro interno dos modelos é reajustado nesta etapa; isto é, tanto os pesos da MLP(6,2,3) quanto a matriz de ganho do LS são mantidos fixos.

Ao longo da seção, são inicialmente caracterizados os conjuntos de validação e o protocolo experimental empregado. Em seguida, analisam-se em detalhe os resultados obtidos para o cenário de referência com SNR = 70 dB, para então investigar a influência sistemática do ruído sobre o desempenho dos classificadores. Por fim, é apresentada uma síntese comparativa entre a MLP(6,2,3) e o LS, destacando os compromissos entre robustez, complexidade do modelo e capacidade de detecção em condições mais próximas da operação real.

### 5.7.1 Conjuntos de validação e protocolo experimental

A validação independente do MDI é conduzida sobre o banco fora da grade descrito na Seção 3.3.3, construído a partir de 100 perfis de carga inéditos sorteados uniformemente dentro da ZND definida nas Equações (3.3)-(3.4). Para cada par  $(\Delta P, \Delta Q)$  são simulados os três modos de operação, nominal (N), ilhamento não intencional (I) e faltas (F), conforme os cenários da Seção 3.2.3, de modo que, antes do janelamento, cada classe disponha de 100 perfis distintos, cada um repetido  $n = 10$  vezes.

Na etapa de validação, esse mesmo conjunto é avaliado sob cinco condições de ruído distintas, com SNR igual a 70 dB, 60 dB, 50 dB, 40 dB e 30 dB. Para cada nível de SNR é gerado um conjunto de sinais ruidosos contendo, em cada classe, os mesmos 100 perfis e 10 repetições por perfil. Em seguida, as tensões trifásicas no PAC são recortadas em janelas deslizantes de quatro ciclos e, tal como no conjunto de treino em janela deslizante (Seção 3.3.2), cada janela de quatro ciclos é decomposta em três sub-janelas de dois ciclos com sobreposição de 50%. Assim, para cada nível de SNR obtém-se, por classe, 100 perfis  $\times$  10 repetições  $\times$  3 sub-janelas, totalizando 3000 sub-janelas por classe e 9000 sub-janelas por conjunto de validação.

O processamento de cada conjunto reutiliza exatamente o mapeamento de características consolidado na etapa de treinamento em janela deslizante (Seção 5.6), conforme formalizado no Capítulo 4. Todas as sub-janelas de tensão trifásica são projetadas na mesma base global  $\bar{\mathbf{R}}$  obtida a partir da grade  $3 \times 3$ , utilizando, em particular, a sua primeira direção principal  $\bar{\mathbf{r}}_1$  para gerar, em cada janela, uma série escalar  $c_1[n]$ . Sobre essa série, calculam-se diretamente apenas as cumulantes de segunda e quarta ordens nos atrasos fixados na etapa de otimização em janela fixa, reconstruindo o vetor de atributos ótimo

$$\mathbf{s} = [r_2[211], r_2[70], r_2[154], r_2[62], r_4[69], r_4[283]]^T,$$

associado ao conjunto  $S_6^*$ . Em seguida, cada componente de  $s$  é escalonada pelo mesmo mapeamento afim adotado na fase de treinamento em janela deslizante, garantindo que os bancos de treino, teste e validação independente sejam processados de forma estritamente consistente, sem introdução de novos hiperparâmetros ou reotimização de características.

A etapa de classificação é inteiramente congelada: os parâmetros da rede MLP(6,2,3) e os pesos do classificador por mínimos quadrados (LS) são exatamente aqueles ajustados na fase de treino e teste em janela deslizante (Seção 5.6), sem qualquer retreinamento ou ajuste adicional. Para cada sub-janela, os dois classificadores produzem um vetor de saídas contínuas associado às três classes (N, I, F), e a decisão por janela é obtida pela posição do maior valor na saída. Os rótulos verdadeiros são definidos ao nível de perfil de carga e modo de operação, sendo replicados para as três sub-janelas correspondentes, de forma que cada janela derivada de um mesmo sinal compartilhe a mesma classe de referência.

Nas subseções seguintes, esse processo é aplicado a todos os níveis de SNR e a ambos os classificadores. Primeiro, analisam-se as matrizes de confusão e as métricas por classe em nível de sub-janela; em seguida, avaliam-se estratégias de agregação das três janelas de cada sinal, explorando o compromisso entre robustez à degradação da SNR e tempo de detecção do ilhamento.

### 5.7.2 Resultados para SNR de 70 dB na ZND

O cenário com SNR de 70 dB é adotado como caso de referência na validação independente porque reproduz o mesmo nível de ruído empregado na etapa de treinamento em janela deslizante, porém agora aplicado a perfis de carga totalmente fora da grade  $3 \times 3$ . Dessa forma, isola-se o efeito da mudança de perfis dentro da ZND, mantendo constantes as demais condições de simulação.

Em todas as análises desta seção, o desempenho dos classificadores é quantificado a partir das matrizes de confusão associadas às três classes (Nominal, Ilhamento e Falta), utilizando as métricas de acurácia global, precisão, *recall* e F1-score definidas em (5.4)–(5.7). As tabelas e figuras a seguir reportam esses indicadores para cada classificador, sub-janela deslizante e cenário de validação, e as mesmas definições permanecem válidas na análise de diferentes níveis de SNR apresentada na subseção seguinte.

As matrizes de confusão obtidas com a MLP em cada sub-janela de dois ciclos são apresentadas na Fig. 28, enquanto a Tab. 28 resume a acurácia global e as métricas por classe. Na primeira sub-janela (Fig. 28a), a rede já atinge acurácia global em torno de 90,7%. A classe nominal é identificada sem erros (*recall* de aproximadamente 100%), ao passo que ilhamento e faltas apresentam *recall* ainda moderado, da

ordem de 85%–90%. Os erros concentram-se quase exclusivamente em trocas entre ilhamento e faltas: praticamente não há confusões entre regimes perturbados e operação nominal, indicando que, mesmo em uma janela inicial, as cumulantes ao longo de  $\bar{r}_1$  já separam bem a presença de distúrbios.

Figura 28 – Matrizes de confusão da MLP na ZND para as três sub-janelas de dois ciclos (SNR=70 dB).

Janela 1				Janela 2				Janela 3			
Prev.	Alvo			Prev.	Alvo			Prev.	Alvo		
	N	I	F		N	I	F		N	I	F
N	1000 33,3%	7 0,2%	52 1,7%	N	1000 33,3%	0 0,0%	22 0,7%	N	1000 33,3%	0 0,0%	0 0,0%
I	0 0,0%	849 28,3%	77 2,6%	I	0 0,0%	1000 33,3%	19 0,6%	I	0 0,0%	1000 33,3%	0 0,0%
F	0 0,0%	144 4,8%	871 29,0%	F	0 0,0%	0 0,0%	959 32,0%	F	0 0,0%	0 0,0%	1000 33,3%

(a) Janela 1                      (b) Janela 2                      (c) Janela 3

Legenda: Matrizes de confusão agregadas por sub-janela (dois ciclos) no cenário em ZND com SNR=70 dB. Linhas indicam a classe prevista (*Prev.*) e colunas a classe alvo (*Alvo*); cada célula reporta contagem e porcentagem no respectivo conjunto. Abreviações: N = nominal, I = ilhamento, F = falta.

Na segunda sub-janela (Fig. 28b), o desempenho da MLP se aproxima do limite ideal. A acurácia global sobe para cerca de 98,6%, mantendo *recall* de 100% para a classe nominal e acima de 95% para as classes de ilhamento e falta, com escores F1 elevados para todas as classes (Tab. 28). Na terceira sub-janela (Fig. 28c), obtém-se classificação perfeita sobre o conjunto de validação (100% de acurácia global e F1 igual a 100% em todas as classes), o que sugere que, após aproximadamente quatro ciclos de observação, o mapeamento de características consolidado na grade generaliza de forma praticamente irrestrita para perfis inéditos dentro da ZND.

Tabela 28 – Desempenho da MLP na validação independente na ZND (SNR=70 dB).

Janela	Acurácia	Nominal		Ilhamento		Falta	
	[%]	Rec. [%]	F1 [%]	Rec. [%]	F1 [%]	Rec. [%]	F1 [%]
1	90,67	100,00	84,90	87,10	97,13	88,16	86,45
2	98,63	100,00	100,00	95,90	98,91	99,06	97,91
3	100,00	100,00	100,00	100,00	100,00	100,00	100,00

Para o classificador linear por mínimos quadrados, as matrizes de confusão são mostradas na Fig. 29 e os indicadores numéricos correspondentes na Tab. 29. Na

primeira sub-janela (Fig. 29a), a acurácia global é de apenas 54,70%. Embora todos os perfis nominais sejam corretamente rotulados (*recall* de 100%), nenhum perfil de ilhamento é identificado como tal (*recall* de 0%), sendo majoritariamente absorvido pela classe nominal (e, em menor grau, confundido com faltas). Além disso, uma fração significativa das faltas é confundida com operação normal, resultando em *recall* de apenas 64,10% para a classe de faltas. O padrão de confusão indica que, tão próximo ao instante de perturbação, a fronteira linear em  $\mathbb{R}^6$  induzida pelo conjunto  $\mathcal{S}_6^*$  não é suficientemente flexível para separar ilhamento das demais condições.

Figura 29 – Matrizes de confusão do classificador LS na ZND (SNR=70 dB).

Janela 1				Janela 2				Janela 3			
Prev.	Alvo			Prev.	Alvo			Prev.	Alvo		
	N	I	F		N	I	F		N	I	F
N	1000 33,3%	988 32,9%	359 12,0%	N	1000 33,3%	0 0,0%	178 5,9%	N	1000 33,3%	0 0,0%	178 5,9%
I	0 0,0%	0 0,0%	0 0,0%	I	0 0,0%	929 31,0%	0 0,0%	I	0 0,0%	1000 33,3%	0 0,0%
F	0 0,0%	12 0,4%	641 21,4%	F	0 0,0%	71 2,4%	822 27,4%	F	0 0,0%	0 0,0%	822 27,4%

(a) Janela 1

(b) Janela 2

(c) Janela 3

Legenda: Matrizes de confusão agregadas por sub-janela de dois ciclos na validação independente, restrita à ZND e com SNR de 70 dB. Linhas correspondem às classes preditas (*Prev.*) e colunas às classes verdadeiras (*Alvo*). Classes: N = nominal, I = ilhamento, F = falta.

Nas sub-janelas seguintes, o classificador LS se beneficia da evolução temporal das cumulantes e recupera um desempenho mais competitivo. Na segunda janela (Fig. 29b), a acurácia global sobe para 91,70%, com *recall* de 100% para nominal e de 92,90% e 82,20% para ilhamento e faltas, respectivamente. Na terceira sub-janela (Fig. 29c), a acurácia global atinge 94,07%, com *recall* de 100% para nominal e ilhamento e de 82,20% para a classe de faltas (Tab. 29). Permanece, portanto, um resíduo de confusão entre faltas e regime nominal, mesmo após quatro ciclos de observação.

Considerando exclusivamente o cenário de 70 dB e perfis fora da grade, os resultados indicam que o MDI baseado em cumulantes projetados ao longo de  $\bar{r}_1$  apresenta forte capacidade de generalização quando associado à MLP, que atinge desempenho praticamente perfeito a partir da segunda/terceira sub-janela. O classificador por mínimos quadrados também alcança acurácia elevada nas janelas posteriores, mas sofre degradação marcante na detecção precoce do ilhamento, especialmente na primeira sub-janela. Em termos de compromisso entre tempo de detecção e acurácia,

Tabela 29 – Desempenho do classificador por mínimos quadrados (LS) na validação independente na ZND (SNR=70 dB).

Janela	Acurácia	Nominal		Ilhamento		Falta	
	[%]	Rec. [%]	F1 [%]	Rec. [%]	F1 [%]	Rec. [%]	F1 [%]
1	54,70	100,00	59,76	0,00	0,00	64,10	77,56
2	91,70	100,00	91,83	92,90	96,32	82,20	86,85
3	94,07	100,00	91,83	100,00	100,00	82,20	90,23

o uso de janelas mais precoces favorece claramente a MLP, enquanto janelas mais tardias reduzem o risco de decisão equivocada à custa de alguns ciclos adicionais de observação.

### 5.7.3 Influência do nível de ruído (30–70 dB)

A robustez do MDI frente a ruídos de medição foi avaliada por meio de uma varredura sistemática do nível de SNR, variando-o de 70 dB até 30 dB. Em todos os casos, o conjunto de perfis fora da grade descrito na Seção 5.7.1 foi reaproveitado, preservando-se integralmente o fluxo de processamento estabelecido anteriormente: projeção das tensões trifásicas no subespaço principal, extração de cumulantes em atrasos pré-selecionados, normalização e classificação pelos modelos MLP e LS ao longo de três sub-janelas deslizantes de dois ciclos. Assim, apenas a potência do ruído gaussiano branco aditivo foi modificada entre os cenários, o que permite isolar o impacto da SNR sobre o desempenho de detecção.

Como discutido na Seção 5.5, o vetor de características que alimenta os classificadores é composto por cumulantes de segunda e quarta ordens avaliados em atrasos não nulos. Em presença de ruído gaussiano branco, as cumulantes de ordem superior do próprio ruído são teoricamente nulas, e as cumulantes de segunda ordem em atrasos distintos de zero também se anulam. Como as cumulantes são aditivas para componentes estatisticamente independentes, a contribuição do ruído nessas estatísticas tende a ser nula, de modo que o valor observado das cumulantes reflete essencialmente a dinâmica do sinal útil. Espera-se, portanto, que as características utilizadas apresentem baixa sensibilidade a esse tipo de ruído, e que a varredura em SNR funcione como uma validação empírica dessa propriedade de robustez.

Para cada combinação de classificador, sub-janela e nível de SNR, são calculadas as métricas de acurácia global, precisão, *recall* e F1-score definidas em (5.4)–(5.7), a partir das matrizes de confusão associadas às três classes (Nominal, Ilhamento e Falta). Além dessas métricas, duas quantidades derivadas são particularmente relevantes do ponto de vista da segurança da operação: a probabilidade de não detecção de ilhamento ( $P_{miss}^I$ ) e a probabilidade de falso alarme de ilhamento ( $P_{FA}^I$ ). Denotando

a classe de ilhamento por I, define-se

$$P_{\text{miss}}^I = 1 - \text{Rec}(I), \quad (5.8)$$

que quantifica a fração de eventos de ilhamento classificados incorretamente como Nominal ou Falta, e

$$P_{\text{FA}}^I = \frac{N_{\text{N} \rightarrow \text{I}} + N_{\text{F} \rightarrow \text{I}}}{N_{\text{N}} + N_{\text{F}}}, \quad (5.9)$$

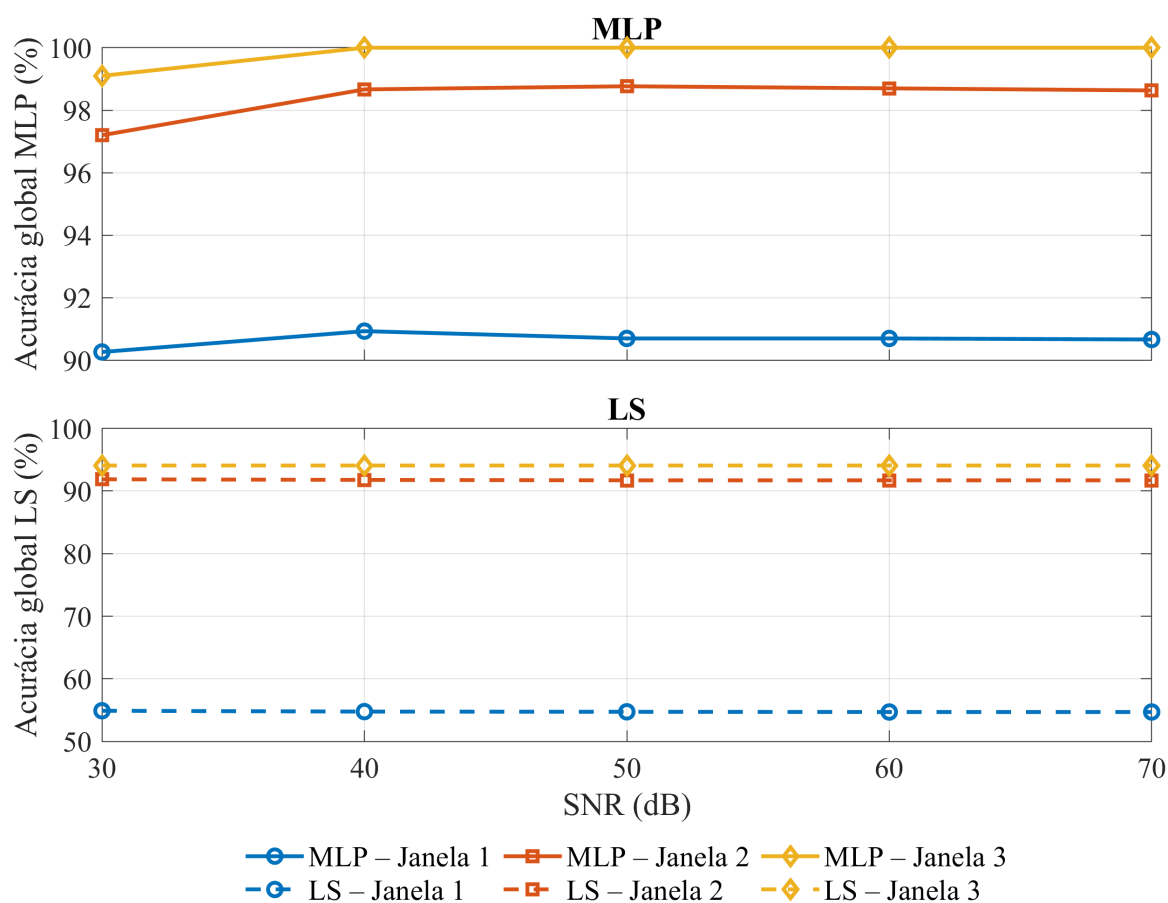
em que  $N_{\text{N} \rightarrow \text{I}}$  e  $N_{\text{F} \rightarrow \text{I}}$  representam, respectivamente, o número de amostras nominais e de falta que foram classificadas como ilhamento, enquanto  $N_{\text{N}}$  e  $N_{\text{F}}$  são os totais de amostras dessas classes. Essa decomposição por tipo de erro permite avaliar não apenas o desempenho médio, mas a natureza dos erros cometidos em cada nível de SNR.

A Figura 30 resume a acurácia global dos classificadores MLP e LS em função da SNR, para as três sub-janelas deslizantes. Observa-se, em primeiro lugar, que a acurácia permanece elevada em todo o intervalo de SNR analisado, com variações pouco expressivas mesmo na presença de ruído mais intenso. Em segundo lugar, nota-se que o uso de janelas mais tardias tende a consolidar a decisão, resultando em desempenho superior quando comparado à primeira sub-janela, especialmente para o classificador LS. Para a MLP, as três sub-janelas já se encontram praticamente saturadas em altos níveis de SNR, e a degradação com a redução de SNR é marginal, em torno de poucos décimos de ponto percentual.

Apenas a acurácia global, entretanto, não é suficiente para caracterizar o comportamento do MDI em situações críticas. A Figura 31 detalha o desempenho específico na detecção de ilhamento por meio das curvas de *recall* e F1-score da classe de ilhamento em função da SNR. Para a MLP, as métricas associadas às sub-janelas intermediária e final mantêm-se próximas de seu valor máximo em todo o intervalo de SNR, indicando que praticamente todos os eventos de ilhamento são corretamente identificados mesmo em níveis de ruído mais severos. Para o LS, a primeira sub-janela apresenta desempenho insatisfatório, com *recall* de ilhamento próximo de zero e  $P_{\text{miss}}^I \approx 100\%$  independentemente da SNR, o que evidencia uma limitação estrutural da combinação entre janela com pouca contribuição do evento e classificador linear. As sub-janelas subsequentes, por sua vez, compensam essa limitação e alcançam desempenho comparável ao da MLP, sobretudo na sub-janela final, em que o ilhamento passa a ser corretamente identificado em praticamente todas as realizações.

A mesma análise pode ser refinada ao se considerar explicitamente as probabilidades de não detecção e de alarme falso de ilhamento, apresentadas na Figura 32. Nota-se que, para a MLP,  $P_{\text{miss}}^I$  é praticamente nulo a partir da segunda sub-janela em todo o intervalo de SNR, refletindo uma taxa de erro para ilhamento desprezível. Já o LS apresenta  $P_{\text{miss}}^I$  elevado na primeira sub-janela, mas converge para valores reduzidos (da ordem de poucos por cento) nas sub-janelas subsequentes. Em relação ao

Figura 30 – Acurácia global de MLP e LS em função da SNR.

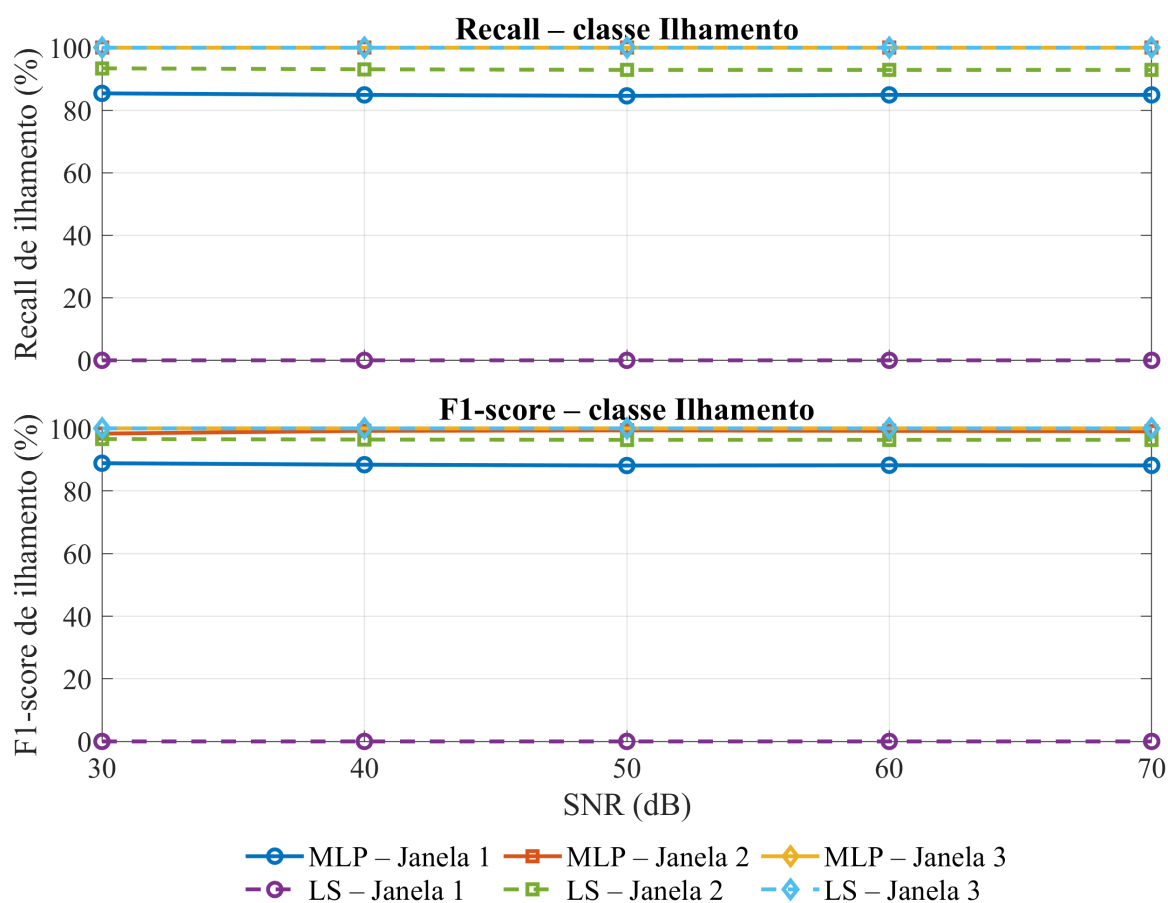


Legenda: A figura compara a acurácia global dos classificadores MLP e LS para diferentes níveis de SNR, considerando as três sub-janelas deslizantes de dois ciclos. Cada curva representa o desempenho agregado no respectivo regime de ruído gaussiano branco aditivo, permitindo avaliar a sensibilidade do desempenho às variações de ruído.

alarme falso,  $P_{FA}^I$  permanece baixo para ambos os classificadores, com ligeira elevação apenas nas janelas mais precoces da MLP e, ainda assim, dentro de um patamar operacionalmente aceitável.

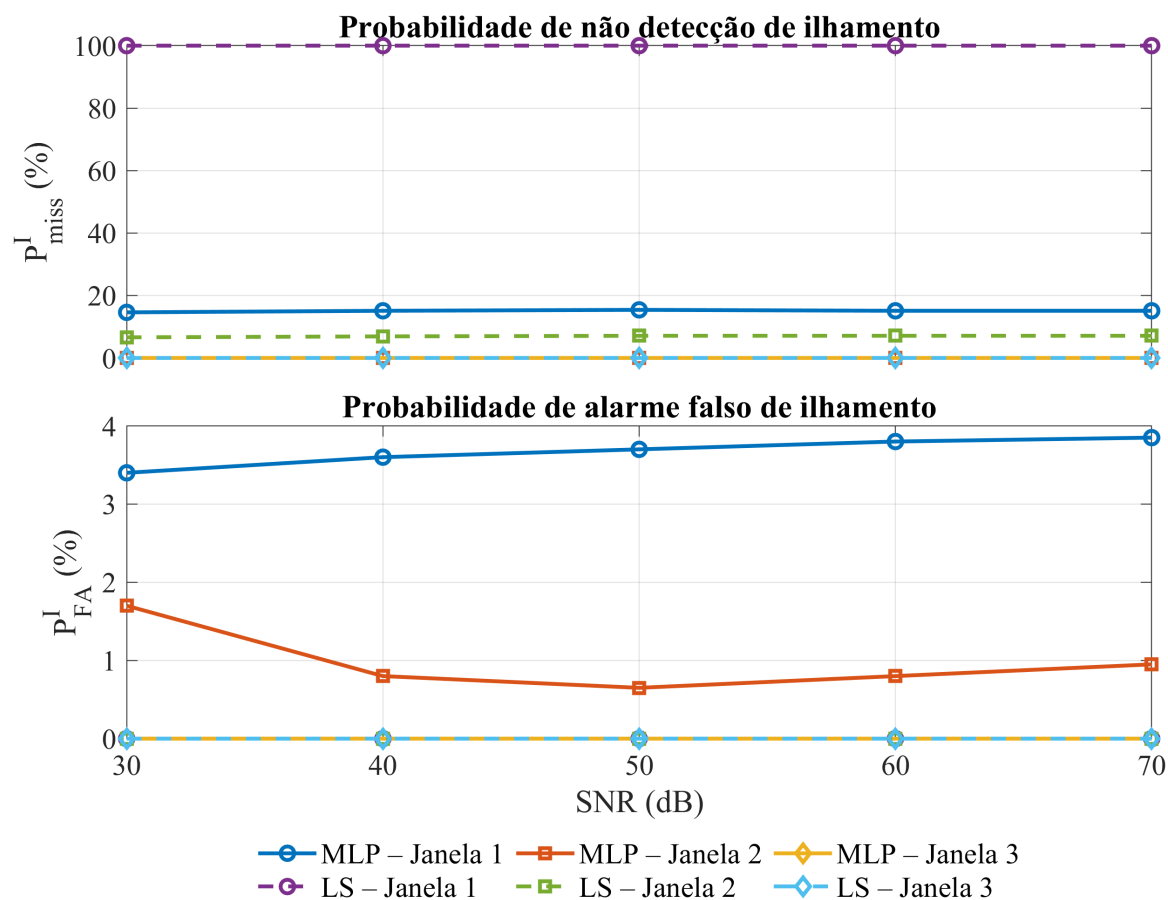
Embora a detecção de ilhamento seja o foco principal do MDI, é igualmente importante assegurar que o sistema preserve boa capacidade de identificar faltas. A Figura 33 apresenta as curvas de *recall* e F1-score da classe Falta em função da SNR. De forma semelhante ao observado para o ilhamento, as métricas associadas às sub-janelas intermediária e final mostram-se pouco sensíveis à redução da SNR, indicando que a introdução do ruído não induz a uma confusão sistemática entre faltas e condições nominais ou de ilhamento. Dessa forma, o método mantém simultaneamente boa seletividade para faltas e alta sensibilidade para ilhamento ao longo de toda a faixa de SNR considerada.

Figura 31 – Recall e F1-score da classe de ilhamento em função da SNR.



Legenda: A figura apresenta as métricas de *recall* e F1-score para a classe de ilhamento, avaliadas em diferentes níveis de SNR, considerando as três sub-janelas deslizantes de dois ciclos. Para cada nível de ruído, as curvas comparam o desempenho dos classificadores MLP e LS, permitindo analisar a estabilidade da detecção de ilhamento sob a adição de ruído gaussiano branco aditivo.

Para facilitar a comparação entre diferentes níveis de ruído, a Tabela 30 resume o desempenho dos classificadores MLP e LS em três cenários representativos: 70 dB, 50 dB e 30 dB. Nesta síntese, considera-se a terceira sub-janela deslizante, que agrega a informação correspondente aos ciclos 3 e 4 da componente fundamental desde o início do monitoramento. Embora represente o pior caso em termos de atraso de decisão, o tempo máximo associado a quatro ciclos permanece da ordem de algumas dezenas de milissegundos, muito inferior ao limite de 2 s estabelecido em normas de interconexão. Em contrapartida, essa sub-janela apresenta os melhores resultados em todas as métricas de interesse, o que a torna a opção mais conservadora do ponto de vista da segurança da proteção. A tabela apresenta, portanto, a acurácia global, o F1-score de ilhamento e as probabilidades de não detecção e de alarme falso de ilhamento para essa sub-janela final.

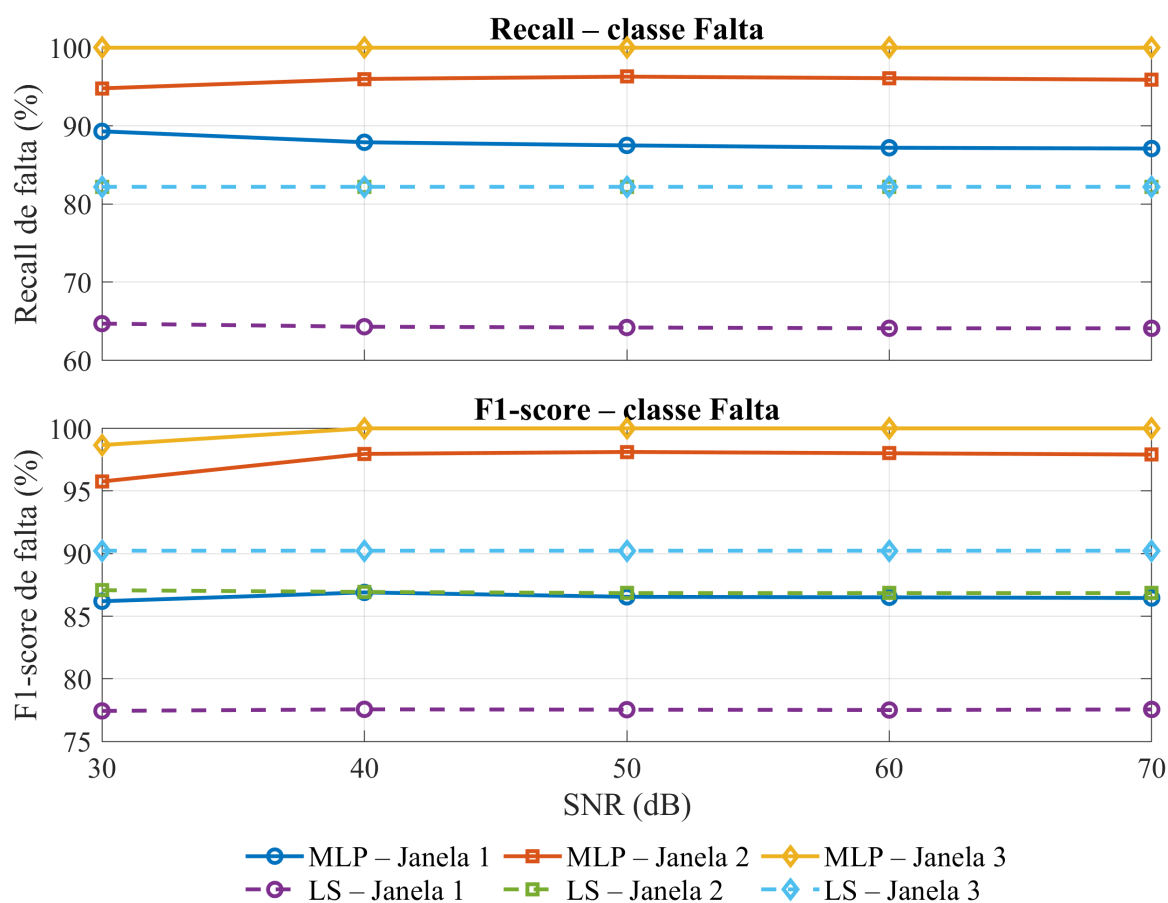
Figura 32 –  $P_{miss}^I$  e  $P_{FA}^I$  em função da SNR.

Legenda: A figura apresenta a probabilidade de não detecção de ilhamento ( $P_{miss}^I$ ) e a probabilidade de alarme falso de ilhamento ( $P_{FA}^I$ ) em função do nível de SNR, comparando os classificadores MLP e LS no regime de janela deslizante (três sub-janelas de dois ciclos). As curvas permitem avaliar como o ruído afeta, separadamente, erros do tipo não detecção e alarme falso, bem como a influência da posição temporal da sub-janela na ocorrência desses erros.

Tabela 30 – Desempenho dos classificadores MLP e LS em três níveis de SNR (sub-janela 3).

SNR (dB)	MLP – Sub-janela 3				LS – Sub-janela 3			
	Acc (%)	F1 <sub>I</sub> (%)	$P_{miss}^I$ (%)	$P_{FA}^I$ (%)	Acc (%)	F1 <sub>I</sub> (%)	$P_{miss}^I$ (%)	$P_{FA}^I$ (%)
70	100.00	100.00	0.00	0.00	94.07	100.00	0.00	0.00
50	100.00	100.00	0.00	0.00	94.07	100.00	0.00	0.00
30	99.10	100.00	0.00	0.00	94.07	100.00	0.00	0.00

Figura 33 – Recall e F1-score da classe Falta em função da SNR.



Legenda: A figura apresenta as métricas de *recall* e F1-score da classe Falta em função do nível de SNR, para as três sub-janelas deslizantes de dois ciclos, comparando os classificadores MLP e LS. As curvas permitem avaliar a estabilidade das métricas com a degradação de SNR e o efeito da posição temporal da sub-janela no desempenho de detecção de faltas.

No caso particular do LS na sub-janela final, é interessante notar que, apesar da acurácia global de aproximadamente 94%, todos os erros concentram-se na distinção entre condições nominais e de falta: nenhum evento de ilhamento é perdido e nenhum estado não-ilhado é indevidamente rotulado como ilhamento em toda a faixa de SNR considerada. Isso reforça que o vetor de características separa de forma consistente a classe de ilhamento das demais.

Em conjunto, os resultados indicam que a introdução de ruído gaussiano branco aditivo, mesmo em níveis relativamente baixos de SNR, não provoca degradação significativa nas métricas de interesse. A estabilidade das curvas de acurácia, *recall* e F1-score, bem como os valores praticamente nulos de  $P_{miss}^I$  e  $P_{FA}^I$  na sub-janela final, corroboram experimentalmente a expectativa teórica de que cumulantes de ordem superior e cumulantes de segunda ordem em atrasos não nulos são pouco sensíveis

a esse tipo de ruído. Além disso, ao se adotar a terceira sub-janela como referência, obtém-se um atraso de detecção amplamente compatível com os requisitos normativos e, ao mesmo tempo, um desempenho praticamente ideal na detecção de ilhamento. A comparação entre MLP e LS mostra ainda que as diferenças de desempenho estão mais relacionadas à capacidade de modelagem de cada classificador e ao uso de sub-janelas deslizantes do que à SNR propriamente dita, reforçando a robustez do vetor de características adotado para a detecção de ilhamento.

### 5.7.3.1 Interpretação geométrica da validação fora da grade

Os resultados desta seção mostram que, embora o classificador tenha sido treinado apenas com amostras associadas a uma discretização regular da ZND (grade  $3 \times 3$ ), a MLP(6,2,3) mantém elevado desempenho ao ser avaliada em perfis aleatórios  $\delta \in \mathcal{D}_{\text{ZND}}$  não contemplados na etapa de otimização. Essa generalização pode ser interpretada à luz da modelagem introduzida nos Capítulos 3 e 4.

Para cada  $\delta \in \mathcal{D}_{\text{ZND}}$ , o fluxo de processamento define um vetor de atributos  $S(\delta) \in \mathbb{R}^B$  (Equação 4.7), enquanto o classificador  $h$  particiona esse espaço em três regiões de decisão associadas às classes  $\{N, I, F\}$ . Seja  $\delta^{(\text{tr})}$  um dos pontos de operação usados na fase de treinamento, com vetor de características  $S(\delta^{(\text{tr})}) \in \mathcal{X}_k$  para algum  $k \in \{N, I, F\}$ . Assumindo que a margem geométrica  $\gamma$  definida em (4.10) é estritamente positiva, a hipótese de regularidade (4.7) garante que, para qualquer outro ponto  $\delta \in \mathcal{D}_{\text{ZND}}$  suficientemente próximo de  $\delta^{(\text{tr})}$ ,

$$\|S(\delta) - S(\delta^{(\text{tr})})\| \leq K \|\delta - \delta^{(\text{tr})}\|. \quad (5.10)$$

Enquanto o deslocamento induzido pela mudança de perfil satisfizer

$$K \|\delta - \delta^{(\text{tr})}\| < \gamma, \quad (5.11)$$

o vetor  $S(\delta)$  permanecerá dentro da mesma região de decisão que  $S(\delta^{(\text{tr})})$ , garantindo a consistência da classificação.

Como os perfis aleatórios utilizados na validação independente são sorteados no interior da ZND, e, portanto, no interior do “envelope” definido pela grade  $3 \times 3$ , é de se esperar que uma parcela significativa deles satisfaça uma condição do tipo (5.11) em relação a algum ponto de treinamento  $\delta^{(\text{tr})}$ . Os resultados empíricos desta seção corroboram essa interpretação geométrica, evidenciando que o MDI preserva a capacidade de discriminar ilhamentos, faltas e regime nominal em uma ampla variedade de perfis de carga não vistos na etapa de otimização.

### 5.7.3.2 Extensão qualitativa além da ZND e limitações

Importa enfatizar que todos os cenários de treinamento, teste e validação independente foram restritos à região de não detecção da IEEE Std. 1547, descrita pelas

Equações (3.1)–(3.4) no plano  $(\Delta P/P_{DG}, \Delta Q/P_{DG})$ . Essa escolha é deliberada: a ZND corresponde precisamente à faixa em que relés convencionais de sub/sobretensão e sub/sobrefrequência, ajustados conforme os limites normativos, não conseguem garantir detecção sem aumento indesejado da taxa de falsos acionamentos.

Por outro lado, as relações entre EDP, tensão e frequência discutidas no Capítulo 4 mostram que, para pontos de operação que ultrapassam a ZND, ao menos uma das grandezas  $V$  ou  $f$  sai dos intervalos nominais. Nessa situação, espera-se que as perturbações induzidas nas formas de onda sejam mais pronunciadas do que aquelas presentes no interior da ZND, produzindo vetores de atributos  $S(\delta)$  ainda mais afastados do aglomerado nominal no espaço de características.

Sob a mesma hipótese de continuidade do mapa  $S(\delta)$  introduzida na Equação (4.7), é razoável conjecturar que o afastamento em termos de EDP se traduza, em média, em um aumento da distância de  $S(\delta)$  às fronteiras de decisão aprendidas pela MLP, tornando a detecção de ilhamento potencialmente mais favorável do que no caso limite estudado dentro da ZND. No entanto, como o presente trabalho não inclui uma campanha sistemática de simulações fora da região de não detecção, o desempenho quantitativo do MDI nessas condições permanece como uma hipótese de trabalho, constituindo uma limitação deste estudo e um tópico natural para investigações futuras.

#### 5.7.4 Análise comparativa entre MLP e LS

As subseções anteriores apresentaram, separadamente, o desempenho do MDI para o cenário fora da grade em 70 dB e para a varredura em SNR de 70 dB a 30 dB. Nesta subseção, o foco passa a ser a comparação qualitativa entre os classificadores MLP e LS, enfatizando como a escolha do modelo de decisão se relaciona com o vetor de características e com o uso das sub-janelas deslizantes.

Um primeiro ponto importante é que, quando o problema é visto sob a ótica binária ilhamento versus não-ilhamento (isto é, agrupando as classes Nominal e Falta em uma única categoria), ambos os classificadores apresentam, na terceira sub-janela, um comportamento praticamente ideal para o conjunto de cenários avaliados. As matrizes de confusão indicam que, nessa sub-janela, nenhum evento de ilhamento é confundido com condições não ilhadas e tampouco há amostras nominais ou de falta classificadas como ilhamento. Em termos das métricas definidas em (5.8)–(5.9), obtém-se  $P_{\text{miss}}^I = 0$  e  $P_{\text{FA}}^I = 0$  para todas as combinações de SNR estudadas, tanto para a MLP quanto para o LS, enquanto o F1-score da classe de ilhamento atinge 100% em todos os casos. Assim, do ponto de vista estrito da detecção de ilhamento no conjunto de teste construído, até mesmo o classificador linear LS funciona como um detector perfeito quando alimentado pelas características extraídas na terceira sub-janela.

Esse resultado reforça a ideia central desta tese de que o desempenho é fortemente determinado pelo processamento de sinais e pelo desenho do vetor de características, mais do que pela sofisticação do classificador em si. A projeção das tensões no subespaço principal, seguida da extração de cumulantes de segunda e quarta ordens em atrasos não nulos, produz um espaço de características no qual a classe de ilhamento se torna quase linearmente separável das demais. Nessa situação, um estimador linear por mínimos quadrados já é suficiente para realizar a separação com desempenho ideal, desde que se disponha das amostras correspondentes à terceira sub-janela.

Do ponto de vista de implementação, em um cenário de monitoramento em tempo real, o relé pode operar com janelas deslizantes de dois ciclos sobre as tensões trifásicas, projetá-las no subespaço principal e extrair, a cada nova janela, o mesmo vetor de cumulantes utilizado nos experimentos. A estratégia sugerida pelos resultados é que a decisão efetiva de ilhamento seja tomada apenas quando a janela corrente estiver posicionada de forma equivalente à terceira sub-janela analisada neste estudo, isto é, quando o regime pós-distúrbio já estiver estabelecido. Sob essa política de decisão, o atraso máximo entre o início do distúrbio e a sinalização de ilhamento é da ordem de quatro ciclos da componente fundamental, ainda muito abaixo do limite de  $2s$  usualmente especificado em normas de interconexão, ao mesmo tempo em que se obtém, para o conjunto de validação analisado, detecção de ilhamento sem erros observados mesmo com um classificador linear simples.

A MLP, por sua vez, passa a oferecer vantagens claras quando se busca reduzir ligeiramente o tempo de detecção ou enriquecer o problema para além da simples detecção binária. Os resultados mostram que, na segunda sub-janela, a MLP já atinge F1-score de 100 % para a classe de ilhamento em toda a faixa de SNR, com probabilidades de não detecção e de alarme falso igualmente desprezíveis. Em outras palavras, a rede neural consegue antecipar uma decisão confiável de ilhamento utilizando apenas os ciclos 2 e 3 da componente fundamental, o que representa um compromisso interessante entre rapidez e robustez. Além disso, quando se considera a classificação completa em três classes (Nominal, Ilhamento e Falta), a MLP apresenta acurácias globais mais altas e menor confusão entre as classes Nominal e Falta, especialmente na terceira sub-janela, onde o desempenho se aproxima de um classificador ideal também para faltas.

Em síntese, a comparação entre MLP e LS sugere a seguinte leitura de projeto: o vetor de características baseado em cumulantes projetados no subespaço principal é suficientemente informativo para que um classificador linear simples atue como detector de ilhamento praticamente perfeito, desde que seja permitido utilizar a terceira sub-janela deslizante como referência para a decisão. A MLP, por outro lado, agrega valor principalmente em dois aspectos: (i) antecipação da decisão de ilhamento na

segunda sub-janela, com desempenho equivalente ao da terceira sub-janela do LS; e (ii) melhoria da classificação multiclasse, reduzindo erros entre condições nominais e de falta. As diferenças observadas entre os classificadores, portanto, refletem mais a capacidade de modelagem e o compromisso entre tempo de detecção e seletividade do que uma limitação intrínseca do vetor de características, que se mostrou robusto e eficaz em todos os cenários de SNR considerados.

### 5.7.5 Síntese da validação independente

A validação independente teve como objetivo principal verificar se o desempenho obtido na fase de otimização se sustentaria em cenários de operação não utilizados no ajuste dos classificadores. Para isso, todos os experimentos deste capítulo foram conduzidos com classificadores fixos: a arquitetura e pesos da MLP, os pesos da solução por mínimos quadrados (LS) e o vetor de características  $\mathcal{S}_6^*$  foram mantidos inalterados. Do ponto de vista dos dados, todo o conjunto de perfis fora da grade, bem como as variações de SNR consideradas, são estritamente independentes daqueles utilizados no processo de otimização descrito na Seção 5.5. Dessa forma, os resultados reportados nas Seções 5.7.1–5.7.4 refletem, de fato, a capacidade de generalização do MDI proposto.

Em termos de desempenho, os resultados convergem para um quadro bastante consistente. No cenário de referência com 70 dB e no conjunto fora da grade, o par  $\mathcal{S}_6^* + \text{MLP}$  apresenta acurácia global muito próxima de 100% já a partir da segunda sub-janela, alcançando desempenho praticamente ideal na terceira sub-janela, tanto para detecção de ilhamento quanto para a classificação multiclasse entre estados Nominal, Ilhamento e Falta. Quando se introduz ruído gaussiano branco aditivo e a SNR é reduzida até 30 dB, as curvas de acurácia, *recall* e F1-score mostram variações marginais, indicando forte robustez do vetor de características baseado em cumulantes projetados no subespaço principal.

O classificador LS, empregado como referência linear, reforça essa interpretação. Apesar de apresentar acurácias globais inferiores às da MLP, sobretudo em janelas iniciais, o LS é capaz de operar como um detector de ilhamento praticamente perfeito quando se considera a terceira sub-janela: para todos os níveis de SNR avaliados, as probabilidades de não detecção e de alarme falso de ilhamento são nulas no conjunto de validação, e o F1-score da classe de ilhamento atinge 100%. Esses resultados sugerem que o processamento de sinais proposto, projeção no subespaço principal, extração de cumulantes de segunda e quarta ordens em atrasos não nulos e seleção de  $\mathcal{S}_6^*$ , é o principal responsável por tornar o ilhamento quase linearmente separável das demais condições de operação.

A influência da posição da sub-janela também se mostrou decisiva. As primei-

ras janelas, que capturam predominantemente o início do transitório, representam um compromisso entre rapidez de decisão e estabilidade das métricas: a MLP consegue antecipar uma decisão segura de ilhamento na segunda sub-janela, enquanto o LS se beneficia de sub-janelas mais tardias, nas quais o regime pós-distúrbio está mais bem estabelecido. A terceira sub-janela, em particular, agrega a informação de quatro ciclos completos da componente fundamental, o que ainda resulta em atrasos de decisão da ordem de algumas dezenas de milissegundos na prática, largamente compatíveis com os requisitos normativos, mas com desempenho que, no conjunto testado, se aproxima de um classificador ideal.

Em conjunto, a validação independente permite formular algumas conclusões gerais sobre o MDI proposto. Em primeiro lugar, o vetor de características  $\mathcal{S}_6^*$ , baseado em cumulantes de ordem superior e cumulantes de segunda ordem em atrasos não nulos, mostrou-se robusto à presença de ruído gaussiano branco em uma faixa ampla de SNR. Em segundo lugar, a comparação entre MLP e LS indica que, uma vez estabelecido um espaço de características adequadamente informativo, até mesmo um classificador linear simples é suficiente para garantir detecção confiável de ilhamento, enquanto a MLP agrega valor ao antecipar a decisão e melhorar a classificação multiclasse. Por fim, embora os resultados sejam bastante promissores no contexto da microrrede e dos cenários simulados, eles também apontam direções claras para trabalhos futuros, como a avaliação do método em topologias distintas de sistema de distribuição, em condições de ruído mais complexas (por exemplo, ruído colorido e distorções harmônicas) e em ambientes com múltiplos pontos de medição, temas que serão retomados nas conclusões desta tese.

## 5.8 COMPLEXIDADE COMPUTACIONAL E CONSIDERAÇÕES DE TEMPO DE EXECUÇÃO

Além das métricas de desempenho apresentadas nas subseções anteriores, é importante verificar se o MDI proposto é compatível com uma implementação prática em tempo real. Nesta seção, é feita uma estimativa da complexidade computacional da fase operacional do método (Fig. 12), isto é, do custo para processar uma janela de dois ciclos e tomar uma decisão a partir do vetor de características e de um classificador previamente treinado, sem considerar o custo da etapa offline de otimização e treinamento.

Considere uma janela deslizante de comprimento  $L$  amostras, correspondente a dois ciclos da componente fundamental. Em um cenário operacional, assume-se que, para cada janela, estão disponíveis as tensões trifásicas no barramento do PAC, organizadas em uma matriz de dimensão  $3 \times L$  (três fases,  $L$  amostras por fase). O fluxo online utilizado pode ser resumido nas seguintes etapas:

- (i) **normalização:** as três fases são escalonadas pelo valor nominal de pico do barramento ( $V_{\text{pico}}^{\text{nom}} = 25 \text{ kV}$ ), produzindo tensões adimensionais;
- (ii) **projeção no subespaço principal:** as tensões normalizadas são projetadas ao longo da primeira componente  $\bar{r}_1$ , gerando uma série escalar  $c_1[n]$  ao longo da janela, por meio de uma combinação linear dos três canais de fase;
- (iii) **extração de cumulantes:** o sinal projetado  $c_1[n]$  é centralizado (subtração da média) e, em seguida, são calculadas cumulantes de segunda e quarta ordens  $\hat{\kappa}_2[\tau]$  e  $\hat{\kappa}_4[\tau]$  em atrasos pré-selecionados, de acordo com as definições em (4.2)–(4.4);
- (iv) **formação do vetor de atributos:** o vetor de características  $S_6^*$  é composto por quatro cumulantes de segunda ordem em atrasos não nulos e duas cumulantes de quarta ordem em atrasos não nulos, totalizando seis atributos;
- (v) **classificação:** o vetor  $S_6^*$  é aplicado a um classificador previamente treinado, que pode ser um estimador linear por mínimos quadrados (LS) ou uma rede neural MLP com arquitetura 6–2–3, utilizando a função de ativação tangente hiperbólica sigmoideal ( $tansig$ ) tanto na camada oculta quanto na camada de saída.

A Tabela 31 resume, em termos do comprimento  $L$  da janela, o número aproximado de operações aritméticas necessárias em cada bloco da fase operacional, considerando somas, multiplicações e funções não lineares. As contagens foram obtidas a partir das expressões de referência e da estrutura da MLP utilizada nesta tese, assumindo que: (i) a normalização consiste na divisão de cada amostra de tensão por  $V_{\text{pico}}^{\text{nom}}$ ; (ii) a projeção em  $\bar{r}_1$  é realizada por uma combinação linear dos três canais a cada amostra; (iii) as cumulantes de segunda ordem  $\hat{\kappa}_2[\tau]$  são calculadas a partir do sinal centralizado, segundo (4.2), para um conjunto de atrasos que inclui tanto os quatro atrasos utilizados como atributos quanto os atrasos auxiliares necessários às cumulantes de quarta ordem (incluindo  $\tau = 0$ ); (iv) as cumulantes de quarta ordem  $\hat{\kappa}_4[\tau]$  são obtidas pela expressão (4.4), reutilizando as cumulantes de segunda ordem já calculadas; e (v) os classificadores operam sobre o vetor de seis características já extraídas, de modo que o custo de decisão independe de  $L$ .

Na etapa de normalização, cada amostra de tensão de fase é multiplicada por uma constante de escala  $1/V_{\text{pico}}^{\text{nom}}$ . Como são três fases e  $L$  amostras por janela, o custo é de aproximadamente  $3L$  multiplicações e nenhuma soma. Em seguida, na projeção em  $\bar{r}_1$ , para cada amostra  $n$  calcula-se

$$c_1[n] = r_A v_A[n] + r_B v_B[n] + r_C v_C[n],$$

o que exige três multiplicações e duas somas por amostra. Para uma janela de comprimento  $L$ , isso totaliza  $3L$  multiplicações e  $2L$  somas.

Tabela 31 – Número aproximado de operações matemáticas na fase operacional do MDI por sub-janela de dois ciclos, em função do comprimento  $L$  da janela.

Operação	Norm.	Proj. em $\bar{r}_1$	Cum. 2 <sup>aa</sup>	Cum. 4 <sup>a</sup>	LS	MLP (6,2,3)
Somas	0	$2L$	$9L - 1$	$2L + 2$	18	18
Multiplicações	$3L$	$3L$	$7L + 8$	$4L + 6$	18	18
Funções não lineares	0	0	0	0	0	5

<sup>a</sup> Inclui as cumulantes de segunda ordem utilizadas como atributos e as cumulantes auxiliares necessárias ao cálculo das cumulantes de quarta ordem, incluindo  $\hat{\kappa}_2[0]$ .

O bloco de cumulantes de segunda ordem começa pela centralização do sinal projetado, com o cálculo da média de  $c_1[n]$  e a subtração desse valor de cada amostra, o que implica em  $2L$  somas e uma multiplicação para o fator  $1/L$ . Em seguida, são avaliadas cumulantes de segunda ordem  $\hat{\kappa}_2[\tau]$  em sete atrasos: quatro atrasos não nulos que compõem o vetor de atributos e três atrasos auxiliares (dois atrasos adicionais utilizados na correção das cumulantes de quarta ordem e o atraso  $\tau = 0$ ). Para cada atraso, a expressão

$$\hat{\kappa}_2[\tau] = \frac{1}{L} \sum_{n=0}^{L-1} \tilde{c}_1[n] \tilde{c}_1[n + \tau]$$

demanda, de forma conservadora,  $L$  multiplicações e  $L$  somas no somatório, além de uma multiplicação adicional pelo fator  $1/L$ . O custo total do bloco de cumulantes de segunda ordem é, portanto, da ordem de  $7L + 8$  multiplicações e  $9L - 1$  somas por janela.

As cumulantes de quarta ordem para dois atrasos não nulos são calculadas a partir da expressão

$$\hat{\kappa}_4[\tau] = \frac{1}{L} \sum_{n=0}^{L-1} \tilde{c}_1[n] [\tilde{c}_1[n + \tau]]^3 - 3\hat{\kappa}_2[\tau]\hat{\kappa}_2[0],$$

reutilizando o sinal centralizado  $\tilde{c}_1[n]$  e as cumulantes de segunda ordem já computadas. Considerando uma implementação direta, em que se pré-calcula  $\tilde{c}_1[n]^3$  ao longo da janela (cerca de  $2L$  multiplicações), cada atraso não nulo exige aproximadamente  $3L$  multiplicações e  $L$  somas no somatório, mais três multiplicações e uma soma na etapa de correção. Para os dois atrasos de interesse, o custo global do bloco de cumulantes de quarta ordem é da ordem de  $4L + 6$  multiplicações e  $2L + 2$  somas por janela.

Os classificadores operam sobre o vetor de seis atributos  $S_6^*$ , de modo que o custo independe do comprimento  $L$ . No caso do estimador linear por mínimos quadrados, admite-se uma matriz de pesos de dimensão  $6 \times 3$  e três termos de *bias*, resultando em cerca de 18 multiplicações e 18 somas para calcular as três saídas e aplicar a regra

de decisão (escolha da maior saída). Para a MLP com arquitetura 6–2–3 e função de ativação  $\tanh$  em ambas as camadas, a camada oculta realiza aproximadamente 12 multiplicações e 12 somas, seguidas de duas ativações não lineares; a camada de saída utiliza cerca de 6 multiplicações e 6 somas adicionais, além de três ativações  $\tanh$ , totalizando 18 multiplicações, 18 somas e cinco funções não lineares por decisão.

Agrupando os termos dominantes em  $L$ , o custo por janela de dois ciclos é da ordem de

$$(0 + 2 + 9 + 2)L \approx 13L \text{ somas} \quad \text{e} \quad (3 + 3 + 7 + 4)L \approx 17L \text{ multiplicações,}$$

aos quais se somam apenas algumas dezenas de operações adicionais associadas aos classificadores. Para um valor típico de  $L \approx 500$  amostras por janela (dois ciclos), isso corresponde a aproximadamente  $8,5 \times 10^3$  multiplicações e  $6,5 \times 10^3$  somas por sub-janela, isto é, cerca de  $1,5 \times 10^4$  operações aritméticas no total. Mesmo considerando a necessidade de acumular três janelas de dois ciclos para tomar a decisão final na janela mais tardia, o custo por evento permanece da ordem de poucas dezenas de milhares de operações, plenamente compatível com a capacidade de *DSPs* ou microcontroladores modernos operando na faixa de dezenas de megahertz.

Um ponto importante é que a etapa de extração de características domina o custo computacional, enquanto o classificador em si (seja LS ou MLP) contribui com uma fração muito pequena do total. Isso reforça a interpretação de que o desempenho alcançado decorre principalmente do desenho do vetor de atributos e da projeção no subespaço principal, mais do que da sofisticação do modelo de decisão. Além disso, a seleção de apenas seis cumulantes específicos, em vez de séries completas em grades densas de atrasos, reduz significativamente o número de operações necessárias na fase operacional, sem comprometer a capacidade de detecção de ilhamento demonstrada nos cenários de validação independente.

Por fim, ressalta-se que a análise apresentada fornece uma estimativa conservadora da ordem de grandeza do custo computacional do MDI proposto. Implementações otimizadas em linguagem de baixo nível ou em hardware dedicado podem explorar reutilização de termos intermediários e instruções vetoriais, reduzindo ainda mais o número efetivo de operações por decisão. A caracterização detalhada do tempo de execução em plataformas específicas é deixada como perspectiva de trabalhos futuros, mas os números obtidos aqui já indicam que o método é computacionalmente leve e compatível com aplicações de proteção em tempo real.

## 5.9 ANÁLISE DE ERROS, INCERTEZA ESTATÍSTICA E CASOS LIMITE

Os resultados de validação apresentados nas seções anteriores mostram que o MDI proposto atinge, para o conjunto de cenários simulados, desempenho extremamente elevado na detecção de ilhamento, em particular na terceira sub-janela deslizante, tanto para o classificador linear por mínimos quadrados (LS) quanto para a MLP. Ainda assim, mesmo diante de matrizes de confusão sem erros observados em determinados cenários, é importante discutir: (i) a incerteza estatística associada às métricas estimadas a partir de um conjunto finito de amostras, (ii) a estrutura qualitativa dos erros residuais que de fato ocorrem em janelas e classes específicas e (iii) casos limite nos quais o problema de decisão tende a se tornar mais desafiador.

### **Incerteza estatística nas métricas de desempenho**

Do ponto de vista estatístico, as quantidades reportadas (acurácia, *recall*, F1-score, probabilidades de não detecção e de alarme falso) são estimativas obtidas a partir de um número finito de amostras de validação. Suponha, por exemplo, que, para uma dada combinação de classificador, sub-janela e nível de SNR, a classe de ilhamento tenha sido avaliada em  $N_I$  amostras na validação independente, e que nenhuma delas tenha sido classificada incorretamente. Denotando por  $p_{\text{err}}^I$  a probabilidade verdadeira de erro de classificação para ilhamento nesse cenário (isto é, a probabilidade de um evento de ilhamento ser rotulado como não-ilhamento), o resultado empírico de zero erros em  $N_I$  ensaios não implica  $p_{\text{err}}^I = 0$ , mas apenas que nenhum erro foi observado nessa amostra finita.

Modelando cada decisão como uma variável Bernoulli (acerto/erro) com probabilidade de erro  $p_{\text{err}}^I$ , é possível quantificar a incerteza por meio de um intervalo de confiança binomial. No caso  $N_{\text{err}} = 0$ , obtém-se um limite superior unilateral. Na validação independente considerada, a classe de ilhamento possui  $N_I = 1000$  amostras por sub-janela; assim, quando  $P_{\text{miss}}^I = 0$  é observado, um limite superior unilateral com 95 % de confiança fornece

$$p_{\text{miss}}^I \leq 2,99 \times 10^{-3},$$

valor consistente com as estimativas empíricas reportadas. De modo análogo, para o conjunto não-I com  $N_{\text{nonI}} = 2000$  amostras, a observação de  $P_{\text{FA}}^I = 0$  implica um limite superior unilateral típico da ordem de  $10^{-3}$ . Logo, sempre que se afirmam resultados como  $P_{\text{miss}}^I = 0$  ou F1-score de 100 % para ilhamento, deve-se entender que tais valores indicam ausência de erros observados no conjunto de validação, permanecendo sujeitos a uma incerteza estatística inversamente proporcional ao número de amostras avaliadas.

### Estrutura qualitativa dos erros residuais (ênfase em SNR=30 dB)

A análise das matrizes de confusão e das métricas por sub-janela indica que os erros residuais se concentram, de modo geral, em dois aspectos:

- **sub-janelas mais precoces**, especialmente a primeira sub-janela, quando o distúrbio ainda não se consolidou por completo ao longo da janela de observação;
- **separação entre classes não ilhadas** (nominal e falta), com confusões pontuais em regiões de fronteira no espaço de atributos, mesmo quando a classe de ilhamento já se encontra bem separada.

Esses padrões tornam-se particularmente claros no pior caso de ruído analisado (SNR = 30 dB). Para a MLP, observa-se na primeira sub-janela um total de 292/3000 erros multiclasse, dominados pela confusão  $I \rightarrow F$  (140 ocorrências) e  $F \rightarrow I$  (68 ocorrências), refletindo a proximidade entre assinaturas transitórias de ilhamento e falta quando o distúrbio é parcialmente observado. Na segunda sub-janela, a detecção de ilhamento se estabiliza ( $P_{\text{miss}}^I = 0$ ), e os erros remanescentes passam a concentrar-se em confusões entre modos não ilhados (principalmente  $F \rightarrow I$ ,  $N \rightarrow F$  e  $F \rightarrow N$ ). Na terceira sub-janela, a MLP torna a classe I perfeitamente separável no conjunto avaliado e elimina também os erros de falta, restando apenas confusões residuais do tipo  $N \rightarrow F$  (27 ocorrências).

Para o LS, a primeira sub-janela em 30 dB revela um caso limite relevante: a classe de ilhamento é integralmente perdida ( $P_{\text{miss}}^I = 1$ ), com predominância de  $I \rightarrow N$  (976 ocorrências). Esse comportamento evidencia a sensibilidade do separador linear quando a janela ainda contém grande mistura temporal e quando a transformação de projeção, orientada ao ilhamento, ainda não produz atributos suficientemente estáveis para uma fronteira linear. Na segunda sub-janela, o LS recupera grande parte da separabilidade de ilhamento ( $P_{\text{miss}}^I = 0,066$ ), e na terceira sub-janela elimina os erros observados de ilhamento ( $P_{\text{miss}}^I = 0$ ). Contudo, mesmo nessa sub-janela tardia, persistem confusões entre falta e nominal ( $F \rightarrow N$ , 178 ocorrências), indicando que, em baixo SNR, a distinção F vs. N é estruturalmente mais desafiadora para o modelo linear e para o conjunto de atributos selecionados.

### Papel da projeção orientada ao ilhamento e implicações para erros N–F

A interpretação dos erros entre N e F deve ser feita à luz da transformação empregada. A projeção  $x[n] \mapsto c[n]$  foi construída a partir de uma matriz de rotação média obtida no treinamento utilizando apenas ilhamento, isto é, uma transformação orientada ao ilhamento. Nesse sentido, o objetivo da projeção não é preservar variância global, reconstruir o sinal, nem recuperar componentes principais no sentido clássico;

o objetivo é produzir uma mudança de coordenadas que favoreça a separabilidade das classes com ênfase no problema I vs. não-I.

Essa escolha metodológica ajuda a explicar dois fenômenos observados:

- em alguns casos de falta (sobretudo em baixo SNR), a componente projetada  $c_1[n]$  pode permanecer dominada por uma componente de natureza fundamental, produzindo estatísticas de ordem superior semelhantes às do regime nominal e, portanto, confusões do tipo  $F \rightarrow N$ ;
- a classe de falta, embora presente e reportada nas matrizes de confusão, não é o foco principal do MDI; sua classificação correta emerge como um efeito colateral interessante da estrutura discriminativa aprendida para o ilhamento. Assim, erros residuais  $N \leftrightarrow F$  em condições adversas de ruído não contradizem o objetivo central do método, desde que a separação I vs. não-I permaneça robusta.

Em termos práticos, a persistência de  $N \leftrightarrow F$  em 30 dB é compatível com um cenário em que: (i) a projeção prioriza direções relevantes ao ilhamento; e (ii) as cumulantes selecionadas em  $\mathcal{S}_6^*$  capturam assinaturas estatísticas altamente discriminativas para I, mas não necessariamente maximizam a separação entre todas as classes não ilhadas, em especial sob ruído mais elevado.

### **Janela de observação e mistura temporal nas sub-janelas iniciais**

A relação entre a posição temporal da sub-janela e a consolidação do distúrbio fornece uma explicação conceitual para a maior taxa de erros em janelas precoces. Quando o distúrbio (ilhamento ou falta) se inicia dentro da janela de dois ciclos, especialmente próximo de suas extremidades, o sinal projetado  $c_1[n]$  contém uma mistura entre o regime pré-distúrbio e o regime pós-distúrbio. Nessa situação, as cumulantes estimadas na janela representam, efetivamente, estatísticas de um processo não estacionário composto, o que tende a suavizar contrastes entre classes e a aproximar amostras das fronteiras de decisão.

Do ponto de vista do espaço de atributos, essa mistura temporal se reflete em valores de cumulantes ocupando regiões intermediárias, o que explica: (i) as confusões  $I \leftrightarrow F$  na MLP na sub-janela 1 em 30 dB; e (ii) a degradação severa do LS na sub-janela 1, quando os atributos ainda não se estabilizaram a ponto de permitirem uma separação linear robusta. À medida que a janela se desloca e passa a capturar predominantemente o regime pós-distúrbio (segunda e, sobretudo, terceira sub-janela), os atributos convergem para valores característicos do evento, e o desempenho na detecção de ilhamento melhora rapidamente, mesmo sob SNR adverso.

## Síntese

Em síntese, a análise de erros, incerteza estatística e casos limite permite destacar quatro pontos principais:

- (i) mesmo quando não são observados erros de ilhamento na validação independente, a probabilidade verdadeira de erro é limitada por uma incerteza estatística inversamente proporcional ao número de amostras (por exemplo, limites superiores unilaterais da ordem de  $10^{-3}$  para  $N_1 = 1000$ );
- (ii) os erros residuais concentram-se nas sub-janelas iniciais, onde a mistura temporal entre regimes reduz a separabilidade, e diminuem drasticamente na sub-janela mais tardia, na qual o regime pós-distúrbio domina a janela;
- (iii) em  $\text{SNR} = 30$  dB, a MLP estabiliza a detecção de ilhamento já na segunda sub-janela ( $P_{\text{miss}}^l = 0$  observado) e praticamente elimina erros na terceira, enquanto o LS apresenta um caso limite na sub-janela 1, mas recupera a separação de ilhamento nas janelas seguintes;
- (iv) confusões  $N \leftrightarrow F$  em baixo SNR são compatíveis com a escolha metodológica de uma projeção orientada ao ilhamento, cujo propósito é maximizar a separabilidade do evento de ilhamento, sendo a classificação de faltas um resultado secundário.

Esses elementos reforçam a estratégia de projeto adotada neste trabalho, em que a terceira sub-janela é privilegiada como referência para a decisão final de ilhamento, ao mesmo tempo em que se reconhece a possibilidade de decisões mais rápidas com a MLP em janelas intermediárias, em um compromisso controlado entre rapidez de detecção, seletividade e robustez estatística.

## 5.10 CONCLUSÕES PARCIAIS DO CAPÍTULO

Este capítulo consolidou, de forma experimental, os principais elementos do método de detecção e classificação proposto, desde as escolhas estruturais de pré-processamento até sua avaliação em cenários operacionais e de generalização. Em primeiro lugar, demonstrou-se que a projeção fixa construída a partir de janelas de ilhamento constitui um compromisso adequado entre simplicidade computacional e preservação de informação discriminativa. A análise de energia e os ensaios por componente sustentam a adoção de uma representação unidimensional baseada na primeira componente principal, a qual mantém elevada capacidade de separação entre os regimes Nominal, Ilhamento e Falta.

Na etapa de otimização, o procedimento de ranqueamento e seleção de atributos a partir de cumulantes projetados resultou em um vetor compacto de seis características, combinando cumulantes de segunda e de quarta ordens em atrasos específicos. Esse vetor mostrou-se suficiente para sustentar alto desempenho de classificação com uma arquitetura MLP mínima (6–2–3), preservando acurácia elevada e baixa variabilidade em validação cruzada, ao mesmo tempo em que estabelece um modelo com custo de decisão constante e adequado a aplicações em tempo real.

Ao migrar para o regime de janela deslizante, verificou-se que a MLP treinada nesse regime mantém desempenho elevado no conjunto de teste e apresenta erros residuais concentrados principalmente na fronteira entre ilhamento e falta, com alta seletividade para o regime nominal. Em contraste, o classificador linear por mínimos quadrados, embora apresente elevada precisão para classes de distúrbio, mostrou menor sensibilidade (recall) para ilhamento e maior tendência a absorver janelas perturbadas na classe nominal, evidenciando limitações estruturais de fronteiras lineares em janelas precoces.

A validação independente, com perfis fora da grade na zona de não detecção, confirmou a capacidade de generalização do método quando acoplado à MLP, atingindo desempenho praticamente perfeito a partir das sub-janelas mais informativas. O classificador LS também melhora de forma consistente em janelas posteriores, mas apresenta degradação marcante na detecção precoce do ilhamento, reforçando o papel do modelo não linear para reduzir o risco de não detecção em horizontes curtos.

Quanto à robustez a ruídos de medição, os experimentos com variação sistemática da SNR indicaram baixa sensibilidade do desempenho global ao ruído gaussiano branco aditivo, em especial nas sub-janelas intermediária e final. Tal comportamento é coerente com as propriedades estatísticas dos cumulantes utilizados e sustenta a aplicabilidade do método em cenários realistas de medição, nos quais a degradação por ruído é inevitável.

Por fim, a análise de complexidade computacional mostrou que o custo operacional do método é dominado por operações lineares no comprimento da janela, com um número reduzido de operações adicionais no estágio de classificação. Assim, o conjunto de resultados deste capítulo estabelece que o método proposto combina (i) alta acurácia, (ii) generalização fora da grade, (iii) robustez frente a ruído e (iv) viabilidade computacional, criando uma base sólida para as discussões finais e a síntese conclusiva do trabalho.

Este capítulo consolidou a etapa de validação do MDI proposto, discutindo não apenas métricas agregadas de desempenho, mas também a coerência do fluxo de processamento e sua adequação a uma implementação prática (Secs. 5.1–5.8). A arquitetura avaliada combina (i) uma projeção linear fixa das tensões trifásicas, construída

por uma estratégia dirigida ao evento-alvo (ilhamento), e (ii) a extração de cumulantes de segunda e quarta ordens em atrasos não nulos, formando um vetor compacto de atributos que alimenta classificadores supervisionados.

Do ponto de vista metodológico, é importante enfatizar que a transformação empregada não tem como finalidade reconstrução de sinal, nem deve ser interpretada como uma PCA clássica executada em tempo real. Em vez disso, utiliza-se uma rotação média  $\bar{R}$  estimada exclusivamente a partir de janelas de ilhamento (Sec. 5.3), de modo a induzir uma projeção orientada ao ilhamento. Essa escolha deliberada sacrifica propriedades estritas de componentes principais (e.g., ortonormalidade perfeita e decorrelação ótima para todas as classes) em favor de um mapeamento fixo, estável e computacionalmente eficiente, cujo propósito central é aumentar a separabilidade entre classes no espaço de atributos. Como consequência, é esperado que em parte dos cenários não-ilhados a componente projetada ainda seja dominada por comportamento de baixa ordem (incluindo traços fortemente associados à fundamental), pois o eixo de projeção foi ajustado para privilegiar o evento de ilhamento e não para caracterizar exaustivamente o universo de faltas. Assim, a capacidade de discriminar faltas deve ser interpretada como um efeito adicional útil, enquanto o objetivo primário permanece a detecção robusta de ilhamento. Essa leitura é consistente com (i) a caracterização do erro de ortogonalidade e estabilidade da projeção (Sec. 5.3) e (ii) a distribuição de energia por componente e por classe (Sec. 5.3.2, Tabela 12).

Os principais resultados e conclusões do capítulo podem ser resumidos como segue:

- **Projeção de baixa dimensão suficiente e coerente com o fenômeno:** a análise da energia explicada por componente (Tabela 12) e o ensaio preliminar por componente (Tabela 13) sustentam que uma projeção de baixa dimensão é compatível com a preservação de informação discriminativa relevante ao ilhamento. Em particular, a primeira componente projetada concentra a estrutura estatística útil à separação.
- **Vetor compacto e informativo:** a seleção por FDR (Sec. 5.4) conduz a um conjunto reduzido de atributos cumulantes em atrasos não nulos, permitindo estruturar um espaço de decisão separável com baixa dimensionalidade. A arquitetura consolidada utiliza seis características, com mapeamento explícito de posições e atrasos (Tabela 19), e a visualização do espaço de atributos reforça a existência de organização geométrica não linear entre as classes (p.ex., Figuras da Sec. 5.5).
- **Comparação entre classificadores e papel do modelo não linear:** uma vez estabelecido um espaço de atributos informativo, o classificador linear LS já for-

nece desempenho estável em validação cruzada, porém com acurácia global inferior à solução não linear, enquanto a MLP agrega ganho consistente (Tabela 17) sem exigir aumento substancial de complexidade na etapa de decisão. Esse ganho torna-se particularmente relevante quando a separação é mais difícil (e.g., nas transições temporais), pois a MLP ajusta fronteiras mais adequadas à geometria observada no espaço de atributos.

- **Consistência entre otimização (janela fixa) e operação (janela deslizante):** ao migrar para o regime de janela deslizante (Sec. 5.6), observa-se manutenção de desempenho elevado para a arquitetura final (Tabela 24), ao passo que o LS sofre degradação relevante no mesmo regime (Tabela 26). Esse resultado sustenta que o desenho do vetor de atributos é necessário, mas não suficiente, para garantir robustez quando há maior mistura temporal entre regimes dentro da janela.
- **Validação independente e generalização na ZND:** a validação com perfis não vistos (Sec. 5.7, Tabelas 28 e 29) confirma a capacidade de generalização do método, com melhora acentuada do desempenho nas sub-janelas mais tardias, quando o regime pós-distúrbio domina a janela. Em particular, a terceira sub-janela emerge como referência conservadora para a decisão final, ao mesmo tempo em que se reconhece a possibilidade de decisões mais rápidas em sub-janelas intermediárias, em um compromisso controlado entre rapidez de detecção, seletividade e robustez estatística (Sec. 5.7.5; ver também o exemplo de janelamento na Fig. 25).
- **Robustez ao ruído e leitura correta dos casos limite:** a variação sistemática da SNR indica baixa sensibilidade do desempenho do método no regime mais informativo (sub-janela 3), com métricas de ilhamento preservadas mesmo em  $SNR = 30$  dB (Sec. 5.7.3, Tabela 30). Em cenários adversos, confusões residuais envolvendo classes não prioritárias (p.ex.,  $N \leftrightarrow F$ ) são compatíveis com a escolha metodológica de uma projeção orientada ao ilhamento, na qual a classificação de faltas é um resultado secundário.
- **Compatibilidade com tempo real e dominância do custo de extração:** a análise de complexidade mostra que o custo operacional por sub-janela é dominado pela extração de características (normalização, projeção e cumulantes), enquanto o classificador contribui com uma fração pequena do total (Sec. 5.8, Tabela 31). Assim, o desempenho alcançado decorre principalmente do desenho da projeção e do vetor de atributos, e não de um modelo de decisão excessivamente complexo.

Além disso, o capítulo explicitou a necessidade de interpretar corretamente resultados de “erro zero” em conjuntos finitos: a ausência de erros observados não implica probabilidade verdadeira nula, sendo mais apropriado discutir limites superiores e incerteza estatística em paralelo às métricas pontuais (Sec. 5.9 e Sec. 5.7.5). Também se verificou que erros residuais tendem a concentrar-se em sub-janelas iniciais, nas quais a mistura temporal entre regimes reduz a separabilidade, e diminuem drasticamente na sub-janela mais tardia.

Por fim, embora os experimentos tenham sido conduzidos no contexto de uma microrrede e com ênfase em cenários restritos à ZND, a própria natureza do problema sugere uma perspectiva favorável de generalização: a ZND representa precisamente o caso mais desafiador para relés convencionais, e perturbações mais pronunciadas tendem a produzir deslocamentos maiores no espaço de atributos, afastando  $\mathcal{S}(\delta)$  das fronteiras de decisão. Dessa forma, o capítulo sustenta a tese de que a combinação projeção fixa orientada ao alinhamento + cumulantes em atrasos selecionados constitui um MDI robusto e promissor, com forte apelo prático: baixa dimensionalidade, custo computacional reduzido e desempenho consistente em validações independentes. Ao mesmo tempo, os resultados delimitam com clareza os próximos passos naturais, como a avaliação em topologias distintas de sistemas de distribuição, em condições de ruído mais complexas e com múltiplos pontos de medição, para quantificar a generalização em cenários progressivamente mais próximos do campo.

## 6 CONCLUSÃO

A integração crescente de Geração Distribuída (GD) tem ampliado o desafio de garantir seletividade, rapidez e robustez da proteção em redes de distribuição. Entre os eventos críticos, o ilhamento não intencional permanece especialmente delicado por (i) poder ocorrer com desequilíbrio mínimo entre geração e carga e (ii) expor riscos à segurança, à integridade de equipamentos e à coordenação da proteção. Esta tese enfrentou esse problema propondo e validando um método passivo de detecção de ilhamento (MDI) orientado por processamento estatístico de sinais e reconhecimento de padrões.

### 6.1 SÍNTESE DO TRABALHO E RESPOSTA AOS OBJETIVOS

O objetivo central do trabalho foi desenvolver um MDI passivo capaz de operar com elevada confiabilidade inclusive em condições críticas associadas à Zona de Não-Detecção (ZND), preservando simultaneamente as vantagens típicas de métodos passivos: simplicidade operacional e ausência de perturbação deliberada na rede.

Para atender a esse objetivo, a tese estruturou um fluxo com duas fases complementares:

- **Calibração e consolidação offline:** construção de uma transformação linear fixa para projeção das tensões trifásicas do PAC em baixa dimensão, seguida de extração de estatísticas de ordem superior (cumulantes) e seleção de atrasos/atributos mais discriminantes. Nessa fase, também se define e treina o decisor final (detector/classificador).
- **Execução online em janela deslizante:** aplicação da projeção fixa às janelas de tensão, cálculo de cumulantes apenas nos atrasos pré-selecionados (vetor compacto de atributos) e decisão em tempo curto, compatível com requisitos normativos.

Do ponto de vista conceitual, a escolha-chave foi tratar o problema como uma tarefa de classificação supervisionada no espaço de atributos estatísticos: a dinâmica transitória contida nas tensões trifásicas é comprimida por uma projeção linear de baixa dimensão e, em seguida, descrita por cumulantes ( $2^a$  e  $4^a$  ordens, em atrasos não nulos) que funcionam como assinaturas robustas para distinguir ilhamento de operação nominal e de faltas.

### 6.2 PRINCIPAIS CONTRIBUIÇÕES CIENTÍFICAS E TÉCNICAS

As contribuições desta tese podem ser sintetizadas em quatro eixos:

### 6.2.1 Contribuições metodológicas

1. **Projeção linear fixa orientada ao ilhamento:** em vez de executar uma PCA “clássica” em tempo real (com autovetores variando a cada janela), o método adota uma rotação média estimada offline e mantida fixa na operação, deliberadamente orientada ao evento-alvo. Essa escolha preserva eficiência computacional e estabilidade, ao mesmo tempo em que aumenta a separabilidade entre classes no espaço de atributos.
2. **Uso de cumulantes como descritores do transitório:** a extração de estatísticas de ordem superior em atrasos não nulos reforça aspectos não gaussianos e não lineares do sinal útil, contribuindo para discriminar eventos que, no domínio de grandezas elétricas tradicionais, podem ser pouco separáveis.
3. **Seleção parcimoniosa de atributos por FDR:** a combinação de cumulantes em atrasos específicos com ranqueamento por separabilidade permite um vetor final compacto, reduzindo risco de sobreajuste e favorecendo implementação prática.

### 6.2.2 Contribuições experimentais

1. **Geração e organização de um banco de dados abrangente:** a tese estruturou um protocolo de simulação e particionamento com cenários de operação nominal, faltas e ilhamentos, incluindo condições críticas de baixo desequilíbrio de potência e varredura de ruído.
2. **Validação independente e análise de generalização:** além de desempenho agregado, o trabalho avaliou consistência em perfis inéditos e investigou a influência da SNR, acompanhando métricas específicas de segurança (não detecção e falso alarme de ilhamento).

### 6.2.3 Contribuições em decisão, comparação de modelos e modos de operação

1. **Demonstração de suficiência do LS como detector de ilhamento:** nos conjuntos avaliados, um decisor LS operando sobre o vetor compacto de atributos mostrou-se suficiente para desempenhar a tarefa de detecção de ilhamento com desempenho elevado, destacando-se como alternativa simples, leve e atrativa quando o requisito principal é detectar.
2. **Superioridade e generalidade do MLP para classificação:** embora o LS seja suficiente como detector no escopo estudado, o MLP apresentou desempenho superior e, sobretudo, a vantagem de oferecer classificação (p. ex., ilhamento versus faltas versus operação nominal), permitindo ampliar a função do método de “detector” para “classificador” de eventos.

3. **Definição implícita de dois modos práticos:** a tese estabelece um caminho claro para duas configurações operacionais: (i) *modo detector* (mais simples) com LS e (ii) *modo classificador* (mais completo) com MLP, preservando o mesmo pré-processamento e os mesmos atributos.

#### 6.2.4 Contribuições em implementabilidade e transferibilidade

1. **Estratégia naturalmente implementável:** projeção fixa + vetor pequeno de cumulantes em atrasos selecionados + decisor leve resultam em um fluxo compatível com execução em tempo real com baixa demanda computacional.
2. **Arcabouço portátil para outras redes:** a metodologia proposta é modular e, portanto, promissora para adaptação a outras topologias/redes ao ajustar (i) o componente/projeção escolhida, (ii) os atrasos de cumulantes selecionados e (iii) a quantidade de atributos, com potencial de preservar as características atingidas nesta tese: representação compacta, atributos estatísticos robustos e alto desempenho inclusive sob condições críticas.

### 6.3 RESULTADOS CONSOLIDADOS E IMPLICAÇÕES PARA PROTEÇÃO

Os resultados obtidos sustentam que o método proposto atingiu o objetivo de combinar acurácia, robustez e tempo de decisão compatível com requisitos de interconexão. Os principais destaques são:

#### 6.3.1 Desempenho em classificação e seletividade

No regime de janela deslizante, o decisor MLP final apresentou desempenho global elevado em teste, com erros residuais concentrados sobretudo na fronteira entre ilhamento e falta (sem penalizar a classe nominal com falsos alarmes relevantes). Em termos de qualidade de decisão voltada à proteção, isso se traduz em alta precisão/recall para a classe de ilhamento, com desempenho equilibrado também para faltas.

Além disso, os resultados indicam que um decisor mais simples, baseado em LS, é suficiente para a detecção de ilhamento nos conjuntos avaliados, preservando a estrutura de atributos do método e reduzindo complexidade de implementação. Nesse contexto, a principal vantagem prática do MLP é expandir o escopo para classificação de eventos, o que pode ser determinante quando se busca aumentar a inteligência da proteção e reduzir ambiguidades operacionais.

### 6.3.2 Desempenho na ZND e tempo de detecção

Um ponto central foi a avaliação em condições associadas à ZND. A estratégia de decisão por sub-janelas mostrou que janelas mais tardias consolidam o regime pós-evento, elevando a confiabilidade. Em particular, a terceira sub-janela (pior caso em atraso) corresponde a aproximadamente quatro ciclos de observação e ainda assim permanece muito abaixo de limites normativos típicos, mantendo-se na ordem de dezenas de milissegundos. Essa configuração, embora conservadora, entrega os melhores resultados de segurança (minimizando tanto não detecção quanto falso alarme de ilhamento).

### 6.3.3 Robustez a ruído (SNR) e estabilidade operacional

A varredura de SNR evidenciou baixa sensibilidade do desempenho a ruído gaussiano branco aditivo, coerente com o uso de cumulantes em atrasos não nulos (nos quais a contribuição teórica do ruído tende a ser desprezível). Do ponto de vista operacional, isso é particularmente relevante em medições reais, nas quais a SNR pode variar com instrumentação, condições de rede e interferências.

### 6.3.4 Síntese: por que o método funciona bem

Em retrospecto, o desempenho observado decorre da combinação de três fatores:

- **Compressão informativa:** a projeção de baixa dimensão retém a dinâmica essencial do transitório em uma representação compacta.
- **Atributos estatísticos adequados ao fenômeno:** cumulantes em atrasos não nulos capturam estrutura temporal e não linearidade de forma mais discriminante do que métricas tradicionais em cenários críticos.
- **Decisor compatível com a geometria observada:** tanto LS quanto MLP exploram a separabilidade construída no espaço de atributos; o LS oferece simplicidade e suficiência para detecção no escopo avaliado, enquanto o MLP fornece maior poder de discriminação e capacidade de classificação multiclasse.

## 6.4 LIMITAÇÕES

Apesar dos resultados, algumas limitações devem ser explicitadas para interpretação correta e para orientar avanços:

1. **Base predominantemente simulada:** embora o modelo utilizado seja detalhado e o protocolo de validação tenha buscado generalização, medições reais (com

efeitos de instrumentação, harmônicos não ideais, distorções e eventos não modelados) podem introduzir novas fontes de variabilidade.

2. **Dependência do cenário e do tipo de GD:** o estudo concentrou-se em uma topologia e em um sistema fotovoltaico específico. Outras tecnologias (eólica, armazenamento, síncronos, microturbinas) e estratégias de controle podem alterar assinaturas transitórias.
3. **Compromissos entre simplicidade e generalidade do decisor:** embora o LS tenha se mostrado suficiente como detector no escopo avaliado, sua natureza linear pode limitar desempenho em cenários mais heterogêneos; já o MLP tende a generalizar melhor para fronteiras não lineares, ao custo de maior complexidade.
4. **Escolha de janelas e compromisso rapidez–segurança:** decisões mais precoces são possíveis, mas com maior mistura temporal entre regimes; a tese privilegiou a alternativa mais conservadora para garantir robustez.

## 6.5 TRABALHOS FUTUROS

Como desdobramentos diretos e viáveis, destacam-se:

1. **Validação com dados reais e/ou em ambiente *Hardware-in-the-Loop* (HIL):** incorporar registros oscilográficos de campo e experimentos *hardware-in-the-loop* para verificar robustez frente a fenômenos não modelados e restrições de aquisição.
2. **Generalização para múltiplas topologias e níveis de penetração de GD:** replicar o protocolo em alimentadores distintos, diferentes impedâncias equivalentes e cenários com múltiplas unidades de GD operando simultaneamente.
3. **Transferência metodológica orientada por adaptação modular:** investigar sistematicamente como a escolha do componente/projeção, os atrasos de cumulantes e a cardinalidade do vetor de atributos afetam desempenho em diferentes redes. Essa linha é especialmente promissora por permitir adaptar o método ao “contexto elétrico” da topologia alvo, mantendo o núcleo do arcabouço (projeção fixa + cumulantes + decisor leve) e buscando preservar as características alcançadas nesta tese.
4. **Extensão multimodal:** combinar tensões e correntes (quando disponíveis) e avaliar fusão de atributos para aumentar redundância e reduzir ambiguidades com faltas específicas.

5. **Adaptação a diferentes frequências e normas:** avaliar a transferência do método para redes de 50 Hz e para diferentes requisitos normativos, mantendo rastreabilidade do compromisso rapidez–segurança.
6. **Aprendizado contínuo e detecção de novidade:** investigar mecanismos de monitoramento de *drift* e técnicas para rejeição de eventos fora do domínio treinado, elevando confiabilidade em ambientes abertos.
7. **Explicabilidade e auditoria de decisão:** incorporar análises de contribuição de atributos (por evento) para tornar a decisão mais transparente a operadores e engenheiros de proteção.

## 6.6 CONSIDERAÇÕES FINAIS

Esta tese demonstrou que é possível construir um MDI passivo que combina representação compacta, atributos estatísticos robustos e decisão supervisionada eficiente, atingindo desempenho elevado inclusive em condições críticas associadas à ZND, sem degradar a qualidade de energia e sem recorrer à injeção ativa de perturbações.

Além disso, os resultados mostraram que o mesmo arcabouço suporta dois usos complementares: um detector de ilhamento simples e suficiente (baseado em LS) e um classificador mais geral e superior (baseado em MLP), capaz de discriminar eventos para além da detecção. Por fim, pela sua natureza modular, a metodologia proposta é promissora para adaptação a outras topologias/redes mediante ajustes na projeção, nos atrasos de cumulantes e na quantidade de atributos, com potencial de preservar as características centrais obtidas nesta tese: método passivo, compacto, robusto e com elevado desempenho mesmo sob condições críticas.

## REFERÊNCIAS

- 1 CALVIN, K.; DASGUPTA, D.; KRINNER, G.; MUKHERJI, A.; THORNE, P. W.; TRISOS, C.; ROMERO, J.; ALDUNCE, P.; BARRETT, K.; BLANCO, G.; CHEUNG, W. W.; CONNORS, S.; DENTON, F.; DIONGUE-NIANG, A.; DODMAN, D.; GARSCHAGEN, M.; GEDEN, O.; HAYWARD, B.; JONES, C.; JOTZO, F.; KRUG, T.; LASCO, R.; LEE, Y.-Y.; MASSON-DELMOTTE, V.; MEINSHAUSEN, M.; MINTENBECK, K.; MOKSSIT, A.; OTTO, F. E.; PATHAK, M.; PIRANI, A.; POLOCZANSKA, E.; PÖRTNER, H.-O.; REVI, A.; ROBERTS, D. C.; ROY, J.; RUANE, A. C.; SKEA, J.; SHUKLA, P. R.; SLADE, R.; SLANGEN, A.; SOKONA, Y.; SÖRENSON, A. A.; TIGNOR, M.; VUUREN, D. van; WEI, Y.-M.; WINKLER, H.; ZHAI, P.; ZOMMERS, Z.; HOURCADE, J.-C.; JOHNSON, F. X.; PACHAURI, S.; SIMPSON, N. P.; SINGH, C.; THOMAS, A.; TOTIN, E.; ALEGRÍA, A.; ARMOUR, K.; BEDNAR-FRIEDL, B.; BLOK, K.; CISSÉ, G.; DENTENER, F.; ERIKSEN, S.; FISCHER, E.; GARNER, G.; GUIVARCH, C.; HAASNOOT, M.; HANSEN, G.; HAUSER, M.; HAWKINS, E.; HERMANS, T.; KOPP, R.; LEPRINCE-RINGUET, N.; LEWIS, J.; LEY, D.; LUDDEN, C.; NIAMIR, L.; NICHOLLS, Z.; SOME, S.; SZOPA, S.; TREWIN, B.; WIJST, K.-I. van der; WINTER, G.; WITTING, M.; BIRT, A.; HA, M. **IPCC, 2023: Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, H. Lee and J. Romero (eds.)]. IPCC, Geneva, Switzerland.** [s.n.], 2023. Acesso em 14 dez. 2025. Disponível em: <https://www.ipcc.ch/report/ar6/syr/>.
- 2 Associação Brasileira de Energia Solar Fotovoltaica (ABSOLAR). **Energia Solar no Brasil (dados e estatísticas do setor)**. 2025. Dados atualizados em 11 nov. 2025. Acesso em 14 dez. 2025. Disponível em: <https://www.absolar.org.br/mercado/energia-solar-no-brasil/>.
- 3 Agência Nacional de Energia Elétrica (ANEEL). **Micro e Minigeração Distribuída**. 2025. Atualizado em 21 jan. 2025. Acesso em 14 dez. 2025. Disponível em: <https://www.gov.br/aneel/pt-br/assuntos/geracao-distribuida>.
- 4 Agência Nacional de Energia Elétrica (ANEEL). **Resolução Normativa ANEEL nº 482, de 17 de abril de 2012**. 2012. Acesso em 14 dez. 2025. Disponível em: <https://www.aneel.gov.br/cedoc/ren2012482.pdf>.
- 5 Agência Nacional de Energia Elétrica (ANEEL). **Resolução Normativa ANEEL nº 687, de 24 de novembro de 2015**. 2015. Acesso em 14 dez. 2025. Disponível em: <https://www.aneel.gov.br/cedoc/ren2015687.pdf>.
- 6 Brasil. **Lei nº 14.300, de 6 de janeiro de 2022**. 2022. Atualizado em 26 nov. 2025. Acesso em 14 dez. 2025. Disponível em: <https://www.gov.br/mme/pt-br/aceso-a-informacao/legislacao/leis/lei-n-14-300-2022.pdf/view>.
- 7 1547-2018 - IEEE Standard for Interconnection and Interoperability of Distributed Energy Resources with Associated Electric Power Systems Interfaces. [S.l.]: IEEE, 2018. Title from content provider. ISBN 1504446399.
- 8 DUTTA, S.; SADHU, P. K.; REDDY, M. J. B.; MOHANTA, D. K. Shifting of research trends in islanding detection method - a comprehensive survey. **Protection and**

**Control of Modern Power Systems**, Springer Science and Business Media LLC, v. 3, n. 1, jan 2018.

9 JALIL, M. W.; ISHTIAQUE, W.; ARIF, A. A review of islanding detection techniques for inverter-based distributed generation. In: **IEEC 2023**. [S.I.]: MDPI, 2023. (IEEC 2023), p. 40.

10 MOHAPATRA, S. S.; MAHARANA, M. K.; PRADHAN, A. Review study on recent advancements in islanding detection and diagnosis in microgrids using signal processing and machine learning techniques. **Electric Power Components and Systems**, Informa UK Limited, p. 1–20, maio 2024. ISSN 1532-5016.

11 ABNT. **NBR IEC 62116:2012. Procedimento de ensaio de anti-ilhamento para inversores de sistemas fotovoltaicos conectados à rede elétrica**. Rio de Janeiro, 2012.

12 IEEE. IEEE standard for interconnecting distributed resources with electric power systems. **IEEE Std 1547-2003**, IEEE, p. 1–28, 2003. Disponível em: <https://ieeexplore.ieee.org/document/1225051>.

13 ADMASIE, S.; BUKHARI, S. B. A.; GUSH, T.; HAIDER, R.; KIM, C. H. Intelligent islanding detection of multi-distributed generation using artificial neural network based on intrinsic mode function feature. **J. Mod Power Syst. Clean Energy**, Journal of Modern Power Systems and Clean Energy, v. 8, n. 3, p. 511–520, 2020.

14 MISHRA, M.; CHANDAK, S.; ROUT, P. K. Taxonomy of islanding detection techniques for distributed generation in microgrid. **Renewable Energy Focus**, Elsevier BV, v. 31, p. 9–30, dec 2019.

15 ROPP, M. E.; BEGOVIC, M.; ROHATGI, A. Prevention of islanding in grid-connected photovoltaic systems. **Prog. Photovoltaics Res. Appl.**, Wiley, v. 7, n. 1, p. 39–59, jan 1999.

16 CHAITANYA, B. K.; YADAV, A.; PAZOKI, M.; ABDELAZIZ, A. Y. A comprehensive review of islanding detection methods. In: **Uncertainties in Modern Power Systems**. [S.I.]: Elsevier, 2021. p. 211–256.

17 ABOKHALIL, A.; AWAN, A.; AL-QAWASMI, A.-R. Comparative study of passive and active islanding detection methods for PV grid-connected systems. **Sustainability**, MDPI AG, v. 10, n. 6, p. 1798, may 2018.

18 VIEIRA, J. C. M.; SALLES, D.; FREITAS, W. Power imbalance application region method for distributed synchronous generator anti-islanding protection design and evaluation. **Electr. Power Syst. Res.**, Elsevier BV, v. 81, n. 10, p. 1952–1960, oct 2011.

19 SARANGI, S.; SAHU, B. K.; ROUT, P. K. Detection and classification of islanding by using variational mode decomposition and adaptive multi-kernel based extreme learning machine technique. **Sustainable Energy, Grids and Networks**, Elsevier BV, v. 30, p. 100668, jun. 2022. ISSN 2352-4677.

- 20 EL-SOUSY, F. F. M.; LARIK, N. A.; LUE, W.; ALMUTAIRI, S. Z.; ALQAHTANI, M. H.; MOBAYEN, S. Robust modified passive islanding detection for microgrids using mathematical morphology based dual algorithm. **Scientific Reports**, Springer Science and Business Media LLC, v. 15, n. 1, fev. 2025. ISSN 2045-2322.
- 21 AMINI, H.; MEHRIZI-SANI, A.; NOROOZIAN, R. Passive islanding detection of inverter-based resources in a noisy environment. **Energies**, MDPI AG, v. 17, n. 17, p. 4405, set. 2024. ISSN 1996-1073.
- 22 HAIDER, R.; KIM, C. H.; GHANBARI, T.; BUKHARI, S. B. A. Harmonic-signature-based islanding detection in grid-connected distributed generation systems using kalman filter. **IET Renew. Power Gener.**, Institution of Engineering and Technology (IET), v. 12, n. 15, p. 1813–1822, sep 2018.
- 23 BANU, I. V.; ISTRATE, M. **Passive anti-islanding protection for three-phase grid-connected photovoltaic power systems**. [S.l.]: IEEE DataPort, 2021.
- 24 BANU, I. V.; BARKAT, F.; ISTRATE, M.; GUERRERO, J. M.; CULEA, G.; LIVINTI, P.; MOTAS, J. G.; NEAGU, B.; ANDRIOAIA, D. Passive anti-islanding protection for three-phase grid-connected photovoltaic power systems. **International Journal of Electrical Power & Energy Systems**, Elsevier BV, v. 148, p. 108946, jun 2023. ISSN 0142-0615.
- 25 KULKARNI, M. S.; MISHRA, S.; SUDABATTULA, S.; SHARMA, N. K.; JADOUN, V. K. Efficient islanding detection in hybrid microgrids: the hybrid approach integrating rocpad and ib-rpv. **Renewable Energy Focus**, Elsevier BV, v. 51, p. 100629, out. 2024. ISSN 1755-0084.
- 26 MERLIN, V. L.; SANTOS, R. C.; GRILO, A. P.; VIEIRA, J. C. M.; COURRY, D. V.; OLESKOVICZ, M. A new artificial neural network based method for islanding detection of distributed generators. **International Journal of Electrical Power & Systems**, Elsevier BV, v. 75, p. 139–151, feb 2016.
- 27 KONG, X.; XU, X.; YAN, Z.; CHEN, S.; YANG, H.; HAN, D. Deep learning hybrid method for islanding detection in distributed generation. **Applied Energy**, Elsevier BV, v. 210, p. 776–785, jan 2018.
- 28 MISHRA, P. P.; BHENDE, C. N. Islanding detection using sparse s-transform in distributed generation systems. **Electr. Eng.**, Springer Science and Business Media LLC, v. 100, n. 4, p. 2397–2406, jul 2018.
- 29 YU, S.; YIN, L. Islanding detection method based on s transform and ANFIS. **Journal of Renewable and Sustainable Energy**, AIP Publishing, v. 10, n. 5, p. 055503, sep 2018.
- 30 CHAITANYA, B. K.; YADAV, A. Hilbert–huang transform based islanding detection scheme for distributed generation. In: **2018 IEEE 8th Power India International Conference (PIICON)**. [S.l.]: IEEE, 2018.
- 31 ACHLERKAR, P. D.; SAMANTARAY, S. R.; MANIKANDAN, M. S. Variational mode decomposition and decision tree based detection and classification of power quality disturbances in grid-connected distributed generation system. **IEEE Trans.**

**Smart Grid**, Institute of Electrical and Electronics Engineers (IEEE), v. 9, n. 4, p. 3122–3132, jul 2018.

32 RADHAKRISHNAN, R. M.; SANKAR, A.; RAJAN, S. Synchrophasor based islanding detection for microgrids using moving window principal component analysis and extended mathematical morphology. **IET Renew. Power Gener.**, v. 14, n. 12, p. 2089–2099, 2020. Disponível em: <https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/iet-rpg.2019.1240>.

33 HARTMANN, N. B.; SANTOS, R. C. dos; GRILO, A. P.; VIEIRA, J. C. M. Hardware implementation and real-time evaluation of an ANN-based algorithm for anti-islanding protection of distributed generators. **IEEE Trans. Ind. Electron.**, Institute of Electrical and Electronics Engineers (IEEE), v. 65, n. 6, p. 5051–5059, jun 2018.

34 EZZAT, A.; ELNAGHI, B. E.; ABDELSALAM, A. A. Microgrids islanding detection using fourier transform and machine learning algorithm. **Electr. Power Syst. Res.**, Elsevier BV, v. 196, p. 107224, jul 2021.

35 WANG, Y.; RAZA, A.; MOHAMMED, F. P.; RAVISHANKAR, J.; PHUNG, T. Detection and classification of disturbances in the islanded micro-grid by using wavelet transformation and feature extraction algorithm. **The Journal of Engineering**, Institution of Engineering and Technology (IET), v. 2019, n. 18, p. 5284–5286, jun 2019.

36 PAIVA, S. C.; RIBEIRO, R. L. de A.; ALVES, D. K.; COSTA, F. B.; ROCHA, T. de O. A. A wavelet-based hybrid islanding detection system applied for distributed generators interconnected to AC microgrids. **International Journal of Electrical Power & Energy Systems**, Elsevier BV, v. 121, p. 106032, oct 2020.

37 SARANGI, S.; BISWAL, C.; ROUT, P. K.; SAHU, B. K. Comparative analysis of time-frequency transform-based differential protection strategy for distributed generation integrated microgrid. **e-Prime - Advances in Electrical Engineering, Electronics and Energy**, Elsevier BV, v. 9, p. 100676, set. 2024. ISSN 2772-6711.

38 XIE, D.; ZANG, D.; GAO, P. Research on a new method of islanding detection based on lifting wavelet and neural network. **Journal of Physics: Conference Series**, IOP Publishing, v. 2800, n. 1, p. 012010, jul. 2024. ISSN 1742-6596.

39 MISHRA, M.; ROUT, P. K. Fast discrete s-transform and extreme learning machine based approach to islanding detection in grid-connected distributed generation. **Energy Systems**, Springer Science and Business Media LLC, v. 10, n. 3, p. 757–789, abr. 2018. ISSN 1868-3975.

40 AGRAWAL, S.; PATRA, S.; MOHANTY, S. R.; AGARWAL, V.; BASU, M. Use of matrix-pencil method for efficient islanding detection in static DG and a parallel comparison with DWT method. **IEEE Trans. Ind. Electron.**, Institute of Electrical and Electronics Engineers (IEEE), v. 66, n. 11, p. 8937–8946, nov 2019.

41 KOLLI, A. T.; GHAFFARZADEH, N. A novel phaselet-based approach for islanding detection in inverter-based distributed generation systems. **Electr. Power Syst. Res.**, Elsevier BV, v. 182, p. 106226, may 2020.

- 42 ADMASIE, S.; BUKHARI, S. B. A.; HAIDER, R.; GUSH, T.; KIM, C.-H. A passive islanding detection scheme using variational mode decomposition-based mode singular entropy for integrated microgrids. **Electr. Power Syst. Res.**, Elsevier BV, v. 177, p. 105983, dec 2019.
- 43 SALIMI, S.; KOOCHAKI, A. An effective method for islanding detection based on variational mode decomposition. **Electrica**, AVES Publishing Co., v. 19, n. 2, p. 135–145, jul 2019.
- 44 NALE, R.; VENKATANAGARAJU, K.; BISWAL, S.; BISWAL, M.; KISHOR, N. Islanding detection in distributed generation system using intrinsic time decomposition. **IET Generation, Transmission & Distribution**, Institution of Engineering and Technology (IET), v. 13, n. 5, p. 626–633, feb 2019.
- 45 MISHRA, S.; MALLICK, R. K.; GADANAYAK, D. A. Islanding detection of microgrid using emd and random forest classifier. In: **2020 International Conference on Computational Intelligence for Smart Power System and Sustainable Energy (CISPSSE)**. [S.I.]: IEEE, 2020. p. 1–5.
- 46 THOMAS, S. R.; KURUPATH, V.; NAIR, U. A passive islanding detection method based on k-means clustering and emd of reactive power signal. **Sustainable Energy, Grids and Networks**, Elsevier BV, v. 23, p. 100377, set. 2020. ISSN 2352-4677.
- 47 OLIVEROS, O. F.; ROSA, J. J. G. de la; SIERRA-FERNÁNDEZ, J.-M.; ESPINOSA-GAVIRA, M.-J.; AGÜERA-PÉREZ, A. **Power quality measurement and analysis using higher-order statistics: understanding HOS contribution on the smart(er) grid: Understanding hos contribution on the smart grid**. [S.I.]: Wiley & Sons, Incorporated, John, 2022. ISBN 9781119747796.
- 48 DUBEY, R.; POPOV, M.; SAMANTARAY, S. R. Transient monitoring function-based islanding detection in power distribution network. **IET Generation, Transmission & Distribution**, Institution of Engineering and Technology (IET), v. 13, n. 6, p. 805–813, jun 2018.
- 49 NALE, R.; BISWAL, M.; KISHOR, N. A transient component based approach for islanding detection in distributed generation. **IEEE Trans. Sustainable Energy**, Institute of Electrical and Electronics Engineers (IEEE), v. 10, n. 3, p. 1129–1138, jul 2019.
- 50 GHALAVAND, F.; ALIZADE, B.; GABER, H.; KARIMIPOUR, H. Microgrid islanding detection based on mathematical morphology. **Energies**, MDPI AG, v. 11, n. 10, p. 2696, oct 2018.
- 51 DUTTA, S.; OLLA, S.; SADHU, P. K. A secured, reliable and accurate unplanned island detection method in a renewable energy based microgrid. **Engineering Science and Technology, an International Journal**, Elsevier BV, v. 24, n. 5, p. 1102–1115, oct 2021.
- 52 MICKY, R. R.; SANKAR, A.; SUNITHA, R. Synchrophasor data driven islanding detection, localization and prediction for microgrid using energy operator. **IEEE Trans. Power Syst.**, Institute of Electrical and Electronics Engineers (IEEE), v. 36, n. 5, p. 4052–4065, sep 2021.

- 53 AREFIN, A. A.; BAI, F.; CUI, Y. A distribution-pmu-based islanding detection approach for active distribution networks. **International Journal of Green Energy**, Informa UK Limited, v. 22, n. 16, p. 3745–3761, jul. 2025. ISSN 1543-5083.
- 54 MUDA, H.; JENA, P. Phase angle-based PC technique for islanding detection of distributed generations. **IET Renew. Power Gener.**, Institution of Engineering and Technology (IET), v. 12, n. 6, p. 735–746, mar 2018.
- 55 RABUZIN, T.; NORDSTRÖM, L. Data-driven islanding detection using a principal subspace of voltage angle differences. **IEEE Trans. Smart Grid**, Institute of Electrical and Electronics Engineers (IEEE), v. 12, n. 5, p. 4250–4258, sep 2021.
- 56 MLAKIC, D.; BAGHAEE, H. R.; NIKOLOVSKI, S. A novel ANFIS-based islanding detection for inverter-interfaced microgrids. **IEEE Trans. Smart Grid**, Institute of Electrical and Electronics Engineers (IEEE), v. 10, n. 4, p. 4411–4424, jul 2019.
- 57 BAGHAEE, H. R.; MLAKIC, D.; NIKOLOVSKI, S.; DRAGICEVIC, T. Anti-islanding protection of PV-based microgrids consisting of PHEVs using SVMs. **IEEE Trans. Smart Grid**, Institute of Electrical and Electronics Engineers (IEEE), v. 11, n. 1, p. 483–500, jan 2020.
- 58 BAGHAEE, H. R.; MLAKIC, D.; NIKOLOVSKI, S.; DRAGICEVIC, T. Support vector machine-based islanding and grid fault detection in active distribution networks. **IEEE Journal of Emerging and Selected Topics in Power Electronics**, Institute of Electrical and Electronics Engineers (IEEE), v. 8, n. 3, p. 2385–2403, sep 2020.
- 59 KE, J.; ZHENGXUAN, Z.; ZHE, Y.; YU, F.; TIANSHU, B.; JIANKANG, Z. Intelligent islanding detection method for photovoltaic power system based on adaboost algorithm. **IET Generation, Transmission & Distribution**, Institution of Engineering and Technology (IET), v. 14, n. 18, p. 3630–3640, jul 2020. Disponível em: 10.1049/iet-gtd.2018.6841.
- 60 PAL, K.; GOSWAMI, S.; BHUYAN, A.; KOLEY, S.; ROY, S.; BHAKAT, R. Islanding detection using pattern recognition by hog features with ensemble classifier. In: **2024 IEEE 3rd International Conference on Electrical Power and Energy Systems (ICEPES)**. [S.l.]: IEEE, 2024. v. 5, p. 1–6.
- 61 THOMAS, M. S.; TERANG, P. P. Islanding detection using decision tree approach. In: **2010 Joint International Conference on Power Electronics, Drives and Energy Systems & 2010 Power India**. [S.l.]: IEEE, 2010.
- 62 CHEN, R.; ZHOU, L.; XIONG, C.; XU, H.; ZHANG, Z.; HE, X.; DONG, Q.; WANG, C. Islanding detection method for microgrids based on catboost. **Frontiers in Energy Research**, Frontiers Media SA, v. 10, jan. 2023. ISSN 2296-598X.
- 63 MENEZES, T. S.; COURY, D. V.; FERNANDES, R. A. S. Islanding detection based on artificial neural network and s-transform for distributed generators. In: **2019 IEEE Milan PowerTech**. [S.l.]: IEEE, 2019.
- 64 MANIKONDA, S. K. G.; GAONKAR, D. N. IDM based on image classification with CNN. **The Journal of Engineering**, Institution of Engineering and Technology (IET), v. 2019, n. 10, p. 7256–7262, jun 2019.

- 65 ABDELSALAM, A. A.; SALEM, A. A.; ODA, E. S.; ELDESOUKY, A. A. Islanding detection of microgrid incorporating inverter based DGs using long short-term memory network. **IEEE Access**, Institute of Electrical and Electronics Engineers (IEEE), v. 8, p. 106471–106486, 2020.
- 66 OZCANLI, A. K.; BAYSAL, M. Islanding detection in microgrid using deep learning based on 1d cnn and cnn-lstm networks. **Sustainable Energy, Grids and Networks**, Elsevier BV, v. 32, p. 100839, dez. 2022. ISSN 2352-4677.
- 67 HUSSAIN, A.; MIRZA, S.; KIM, C.-H. Islanding detection and classification of non-islanding disturbance in multi-distributed generation power system using deep neural networks. **Electric Power Systems Research**, Elsevier BV, v. 224, p. 109807, nov. 2023. ISSN 0378-7796.
- 68 DAVID, E.; ORILLAZA, J. R.; PEDRASA, J. R. Islanding detection using transformer neural networks. In: **2024 IEEE 4th International Conference on Power, Electronics and Computer Applications (ICPECA)**. [S.l.]: IEEE, 2024. p. 592–596.
- 69 CHANDAK, S.; MISHRA, M.; NAYAK, S.; ROUT, P. K. Optimal feature selection for islanding detection in distributed generation. **IET Smart Grid**, Institution of Engineering and Technology (IET), v. 1, n. 3, p. 85–95, sep 2018.
- 70 MISHRA, M.; ROUT, P. K. Loss of main detection in distribution generation system based on hybrid signal processing and machine learning technique. **International Transactions on Electrical Energy Systems**, Hindawi Limited, v. 29, n. 1, p. e2676, aug 2018.
- 71 KIM, M.-S.; HAIDER, R.; CHO, G.-J.; KIM, C.-H.; WON, C.-Y.; CHAI, J.-S. Comprehensive review of islanding detection methods for distributed generation systems. **Energies**, MDPI AG, v. 12, n. 5, p. 837, mar 2019.
- 72 YE, Z.; KOLWALKAR, A.; ZHANG, Y.; DU, P.; WALLING, R. Evaluation of anti-islanding schemes based on nondetection zone concept. **IEEE Trans. Power Electron.**, Institute of Electrical and Electronics Engineers (IEEE), v. 19, n. 5, p. 1171–1176, sep 2004.
- 73 KESSY, A.; LEWIN, A.; STRIMMER, K. Optimal whitening and decorrelation. **The American Statistician**, Informa UK Limited, v. 72, n. 4, p. 309–314, jan. 2018. ISSN 1537-2731.
- 74 JOLLIFFE, I. T. **Principal component analysis**. 2. ed., [nachdr.]. ed. New York: Springer, 2004. (Springer series in statistics). ISBN 9780387954424.
- 75 HIGHAM, N. J. **Accuracy and stability of numerical algorithms**. 2nd ed. ed. Philadelphia, Pa.: Society for Industrial and Applied Mathematics, 2002. (Other titles in applied mathematics, 80). System requirements: Adobe Acrobat Reader. ISBN 9780898718027.
- 76 GOLUB, G. H. **Matrix computations**. Fourth edition. Baltimore: The Johns Hopkins University Press, 2013. (Johns Hopkins studies in the mathematical sciences). Literaturangaben und Index. ISBN 9781421407944.

- 77 BISHOP, C. M. **Pattern recognition and machine learning**. New York, NY: Springer Science+Business Media, LLC, 2019. (Information Science and Statistics).
- 78 HASTIE, T. **The elements of statistical learning**: Data mining, inference, and prediction. Second edition. New York, NY: Springer, 2017. (Springer Series in Statistics). Literaturverzeichnis: Seite 699-727. ISSN 2197-568X. ISBN 9780387848587.
- 79 MOAKHER, M. Means and averaging in the group of rotations. **SIAM Journal on Matrix Analysis and Applications**, Society for Industrial & Applied Mathematics (SIAM), v. 24, n. 1, p. 1–16, jan. 2002. ISSN 1095-7162.
- 80 FERREIRA, D. D.; MARQUES, C. A. G.; SEIXAS, J. M. e de; CERQUEIRA, A. S.; RIBEIRO, M. V.; DUQUE, C. A. Exploiting higher-order statistics information for power quality monitoring. In: **Power Quality**. InTech, 2011. cap. 17, p. 345–362. Disponível em: <https://www.intechopen.com/chapters/14973>.
- 81 RIBEIRO, M. V.; MARQUES, C. A. G.; DUQUE, C. A.; CERQUEIRA, A. S.; PEREIRA, J. L. R. Detection of disturbances in voltage signals for power quality analysis using HOS. **EURASIP Journal on Advances in Signal Processing**, Springer Science and Business Media LLC, v. 2007, n. 1, apr 2007.
- 82 MENDEL, J. M. Tutorial on higher-order statistics (spectra) in signal processing and system theory: theoretical results and some applications. **Proceedings of the IEEE**, Institute of Electrical and Electronics Engineers (IEEE), v. 79, n. 3, p. 278–305, mar 1991.
- 83 NIKIAS, C. L.; PETROPULU, A. P. **Higher-order spectra analysis**: A nonlinear signal processing framework. Englewood Cliffs, N.J [u.a.]: PTR Prentice Hall, 1993. (Prentice Hall signal processing series). Literaturangaben. ISBN 0136782108.
- 84 THEODORIDIS, S.; KOUTROUMBAS, K. Pattern recognition and neural networks. In: **Machine Learning and Its Applications**. [S.l.]: Springer Berlin Heidelberg, 2001. p. 169–195.
- 85 HUGHES, G. On the mean accuracy of statistical pattern recognizers. **IEEE Transactions on Information Theory**, Institute of Electrical and Electronics Engineers (IEEE), v. 14, n. 1, p. 55–63, jan. 1968. ISSN 1557-9654.
- 86 GUYON, I.; ELISSEEFF, A. An introduction to variable and feature selection. **J. Mach. Learn. Res.**, JMLR.org, v. 3, n. null, p. 1157–1182, mar. 2003. ISSN 1532-4435.
- 87 CHEN, Y.; CHI, Y.; FAN, J.; MA, C. Spectral methods for data science: a statistical perspective. **Foundations and Trends® in Machine Learning**, Emerald, v. 14, n. 5, p. 566–806, 2021. ISSN 1935-8245.
- 88 BOYD, S. P. **Convex optimization**. Version 29. Cambridge: Cambridge University Press, 2023. First published 2004. ISBN 0521833787.