



THALITA KELY PEREIRA

**LOCALIZAÇÃO DE FALTAS EM LINHAS DE
TRANSMISSÃO UTILIZANDO PROCESSAMENTO
ESTATÍSTICO DE SINAIS E APRENDIZADO DE MÁQUINA**

LAVRAS – MG

2025

THALITA KELY PEREIRA

**LOCALIZAÇÃO DE FALTAS EM LINHAS DE TRANSMISSÃO UTILIZANDO
PROCESSAMENTO ESTATÍSTICO DE SINAIS E APRENDIZADO DE MÁQUINA**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Engenharia de Sistemas e Automação, área de concentração Engenharia de Sistemas e Automação, para a obtenção do título de Mestre.

Orientador

Prof. Dr. Danton Diego Ferreira

Coorientador

Prof. Dr. Henrique Luis Moreira Monteiro

LAVRAS – MG

2025

Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).

Kely Pereira, Thalita

Localização de faltas em linhas de transmissão utilizando processamento estatístico de sinais e aprendizado de máquina / Thalita Kely Pereira. 2025

51 p. : il.

Orientador: Danton Diego Ferreira

Coorientador: Henrique Luis Moreira Monteiro

Dissertação (Mestrado Acadêmico) - UFLA, 2025.
Bibliografia.

1. Localização de faltas em linhas de transmissão. 2. Processamento estatístico de sinais. 3. Aprendizado de máquina. I. Diego Ferreira, Danton. II. Luis Moreira Monteiro, Henrique. III. Universidade Federal de Lavras. IV. Título.

THALITA KELY PEREIRA

**LOCALIZAÇÃO DE FALTAS EM LINHAS DE TRANSMISSÃO UTILIZANDO
PROCESSAMENTO ESTATÍSTICO DE SINAIS E APRENDIZADO DE MÁQUINA**

**FAULT LOCATION IN TRANSMISSION LINES USING STATISTICAL SIGNAL
PROCESSING AND MACHINE LEARNING**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Engenharia de Sistemas e Automação, área de concentração Engenharia de Sistemas e Automação, para a obtenção do título de Mestre.

APROVADA em 25 de fevereiro de 2025.

Dr. Bruno Henrique Groenner Barbosa

UFLA

Dr. Carlos Augusto Duque

UFJF

Orientador

Prof. Dr. Danton Diego Ferreira

Coorientador

Prof. Dr. Henrique Luis Moreira Monteiro

LAVRAS – MG

2025

*A Deus.
Aos meus pais, Gislene e João.*

AGRADECIMENTOS

Agradeço à Universidade Federal de Lavras – UFLA, pela excelência no ensino e pela estrutura ofertada ao longo de toda a minha formação. Ao Programa de Pós-Graduação em Engenharia de Sistemas e Automação (PPGESISA) e ao grupo de pesquisa AIA – Artificial Intelligence and Automation, deixo registrada minha gratidão pelo espaço e suporte acadêmico.

Aos meus pais, Gislene e João, meu mais profundo agradecimento por todo amor, apoio e encorajamento incondicional, fundamentais em cada etapa desta caminhada.

Aos professores Danton e Henrique, agradeço pela orientação dedicada, pela confiança depositada em meu trabalho e pelas contribuições valiosas para o desenvolvimento desta pesquisa.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

“A mente que se abre a uma nova ideia jamais voltará ao seu tamanho original.”

Albert Einstein

RESUMO

Este trabalho propõe uma abordagem para localização de faltas em linhas de transmissão, utilizando sinais simulados no ATP, o algoritmo SOBI para extração de componentes independentes e três técnicas de denoising: PCA, Wavelet e Filtro de Média Móvel. A terceira componente separada pelo SOBI foi usada como base para o pré-processamento. Após a remoção de ruído, foram aplicados algoritmos de regressão supervisionada: KNN, MLP (com e sem otimização por AG), Decision Tree, Random Forest e XGBoost. A base de dados foi composta por 7920 janelas com 700 amostras, reduzidas a 350 via subamostragem. Os modelos foram avaliados com validação cruzada K-Fold (10 folds), e as métricas de desempenho utilizadas foram o Erro Relativo percentual (ER%) e o Coeficiente de Determinação (R^2). Os experimentos foram realizados via simulação computacional, com medições de tempo de treinamento e inferência. O KNN apresentou os melhores resultados gerais, principalmente com SNR acima de 50 dB. A Wavelet destacou-se com baixo ruído e o Filtro de Média Móvel com ruído elevado. Os resultados demonstram a viabilidade da proposta com foco em baixo custo computacional e alta precisão.

Palavras-chave: localização de faltas em linhas de transmissão; identificação cega de segunda ordem; análise de componentes independentes; K-Nearest Neighbors; Extreme Gradient Boosting; Multilayer Perceptron, Árvore de Decisão.

ABSTRACT

This work proposes an approach for fault location in transmission lines using signals simulated in ATP, the SOBI algorithm for extracting independent components, and three denoising techniques: Principal Component Analysis (PCA), Wavelet Transform, and Moving Average Filter. The third component extracted by SOBI was used as the basis for preprocessing. After noise removal, supervised regression algorithms were applied: K-Nearest Neighbors (KNN), Multilayer Perceptron (MLP), MLP optimized by Genetic Algorithm (GA), Decision Tree (DT), Random Forest (RF), and Extreme Gradient Boosting (XGBoost). The dataset consisted of 7,920 windows with 700 samples, reduced to 350 through subsampling. The models were evaluated using 10-fold cross-validation, and the performance metrics used were Relative Error percentage (ER%) and the Coefficient of Determination (R^2). Experiments were conducted via computational simulation, with training and inference times recorded. KNN achieved the best overall performance, especially for SNR levels above 50 dB. The Wavelet technique performed better under low noise, while the Moving Average Filter was more effective under high noise conditions. The results demonstrate the feasibility of the proposed method, with a focus on low computational cost and high accuracy.

Keywords: fault location in transmission lines; second-order blind identification (SOBI); independent component analysis; K-Nearest Neighbors; Multilayer Perceptron; Decision Tree; Random Forest; Extreme Gradient Boosting.

INDICADORES DE IMPACTO

O presente trabalho contribui de forma significativa para o avanço tecnológico no setor elétrico, propondo uma abordagem eficiente para a localização de faltas em linhas de transmissão, com potencial de aplicação direta em sistemas reais de monitoramento e diagnóstico. A pesquisa apresenta impacto tecnológico concreto, ao utilizar sinais simulados e algoritmos de inteligência artificial que demonstraram elevada precisão na identificação da localização da falha, mesmo em condições de ruído, promovendo maior confiabilidade e segurança na operação do sistema elétrico. Em potencial, os resultados podem ser incorporados por concessionárias de energia elétrica, impactando positivamente a gestão de redes de transmissão e reduzindo os tempos de interrupção e os custos associados às manutenções emergenciais. Os impactos se concentram nos territórios atendidos por sistemas de transmissão, podendo beneficiar populações urbanas e rurais de forma indireta, ao garantir maior estabilidade e continuidade do fornecimento de energia elétrica. O estudo alinha-se à área temática 7 - Tecnologia e Produção, da Política Nacional de Extensão, e está vinculado aos Objetivos de Desenvolvimento Sustentável (ODS) da ONU, especialmente o ODS 7 (Energia limpa e acessível), ODS 9 (Indústria, inovação e infraestrutura) e ODS 11 (Cidades e comunidades sustentáveis), ao contribuir para soluções tecnológicas mais inteligentes, sustentáveis e com menor custo operacional. Embora o estudo não tenha caráter extensionista direto, seu potencial de aplicação prática pode ser explorado por empresas e instituições de ensino, impactando a formação de estudantes e profissionais das áreas de engenharia elétrica, sistemas inteligentes e automação. A pesquisa contou com o envolvimento de docentes e discentes do Programa de Pós-Graduação em Engenharia de Sistemas e Automação da UFLA, promovendo também o fortalecimento da formação científica e acadêmica da universidade.

IMPACT INDICATORS

This work significantly contributes to technological advancement in the electric power sector by proposing an efficient method for fault location in transmission lines, with direct potential for implementation in real-world monitoring and diagnostic systems. The research demonstrates concrete technological impact by employing simulated signals and artificial intelligence algorithms that showed high accuracy in locating faults, even under noisy conditions, thereby enhancing reliability and safety in power system operations. The results have the potential to be adopted by electric power utilities, positively influencing transmission grid management while reducing outage durations and associated emergency maintenance costs. The impacts primarily affect territories covered by transmission networks and can indirectly benefit both urban and rural populations by ensuring better power supply continuity and system stability. The study aligns with Theme Area 7 – Technology and Production of Brazil’s National Extension Policy and supports several United Nations Sustainable Development Goals (SDGs), notably SDG 7 (Affordable and Clean Energy), SDG 9 (Industry, Innovation, and Infrastructure), and SDG 11 (Sustainable Cities and Communities), by advancing smarter, more sustainable, and cost-effective technological solutions. Although the study does not have a direct extension component, its practical application potential can be leveraged by companies and educational institutions, benefiting the training of students and professionals in electrical engineering, intelligent systems, and automation. The research involved faculty and students from the Graduate Program in Systems and Automation Engineering at UFLA, contributing to the academic and scientific development of the university.

LISTA DE ILUSTRAÇÕES

Figura 1- Denominações de faltas.	22
Figura 2- Separação de fontes.	23
Figura 3- RNA típica.	27
Figura 4- Fluxograma do Método Proposto.	34
Figura 5- Modelo da Linha de Transmissão de 500 kV.	36
Figura 6- Circuito para simulação das faltas.	37
Figura 7- Falta Elétrica a 48 km, 20 ohms e incidência em 90°.	38
Figura 8- Três componente do SOBI referentes a uma falta a 48 km.	40
Figura 9- Terceira componente do SOBI referente a uma falta a 48 km, 20 Ohms, incidência em 180° e SNR 50dB.	41
Figura 10- Região de interesse do sinal.	42
Figura 11- Sinal filtrado pelo PCA referente a uma falta a 48 km, 20 Ohms, incidência em 180° e SNR 50dB.	47
Figura 12- Sinal filtrado pela Wavelet referente a uma falta a 48 km, 20 Ohms, incidência em 180° e SNR 50dB.	48
Figura 13- Sinal filtrado pelo Filtro Média Móvel referente a uma falta a 48 km, 20 Ohms, incidência em 180° e SNR 50dB.	49
Figura 14- Comparação entre os métodos de denoising.	50
Figura 15- Comparação ao longo do comprimento da LT.	60

LISTA DE TABELAS

Tabela 1- Exemplos de causas que provocam faltas transitórias ou permanentes.	22
Tabela 2- Avaliação comparativa de métodos de localização de faltas em Linhas de Transmissão.	33
Tabela 3- Propriedades das Faltas.	37
Tabela 4- Métricas de avaliação para regressão.	44
Tabela 5- Comparação da Métrica ER (%).	55
Tabela 6- Comparação da Métrica R^2	55
Tabela 7- Vantagens e desvantagens de cada método.	56
Tabela 8- Desempenho computacional dos modelos.	57
Tabela 9- Comparação do ER (%) entre as técnicas de denoising.	59

LISTA DE SIGLAS

ATP	Alternative Transients Program
DT	Decision Tree (Árvore de Decisão)
EA	Erro Absoluto
ER%	Erro Relativo Percentual
KNN	K-Nearest Neighbors (K-Vizinhos Mais Próximos)
LT	Linha de Transmissão
MLP	Multilayer Perceptron (Perceptron de Múltiplas Camadas)
PCA	Principal Component Analysis (Análise de Componentes Principais)
RF	Random Forest (Floresta Aleatória)
RNA	Rede Neural Artificial
R ²	Coefficiente de Determinação
SOBI	Second-Order Blind Identification (Identificação Cega de Segunda Ordem)
SNR	Signal-to-Noise Ratio (Relação Sinal-Ruído)
XGB	Extreme Gradient Boosting

SUMÁRIO

1	INTRODUÇÃO	16
1.1	Justificativa da dissertação	16
1.2	Motivação da dissertação	16
1.3	Objetivos da dissertação	20
1.3.1	Objetivos gerais	20
1.3.2	Objetivos específicos	20
2	REFERENCIAL TEÓRICO	21
2.1	Faltas no Sistema Elétrico de Potência (SEP)	21
2.2	Análise de Componentes Independentes (ICA)	22
2.3	Second Order Blind Identification (SOBI)	24
2.4	K-Nearest Neighbors (KNN)	25
2.5	Redes Neurais Artificiais (RNA)	26
2.6	Decision Tree (DT)	27
2.7	Random Forest (RF)	28
2.8	Extreme Gradient Boosting (XGBoost)	28
3	ESTUDOS RELACIONADOS	29
4	METODOLOGIA	34
4.1	Ambiente de implementação e simulação dos dados	34
4.2	Simulação da Linha de Transmissão	35
4.3	Base de dados	36
4.4	Transformação do sinal	38
4.5	Técnicas de denoising	40
4.6	Segmentação do sinal	41
4.7	Técnicas de regressão	42
4.8	Métricas de avaliação	43
5	RESULTADOS E DISCUSSÃO	44
5.1	Análise estatística da base de dados	45
5.2	Denoising do sinal	45
5.2.1	Análise de Componentes Principais (PCA)	45
5.2.2	Transformada Discreta de Wavelet (DWT)	47

5.2.3 Filtro Média Móvel.....	49
5.2.4 Comparação entre os processos de denoising	50
5.3 Análise dos algoritmos após o processo de denoising com o PCA	51
5.3.1 <i>K-Nearest Neighbors</i> (KNN)	51
5.3.2 <i>Multilayer Perceptron</i> (MLP)	51
5.3.3 <i>Decision Tree</i> (DT)	52
5.3.4 <i>Random Forest</i> (RF)	53
5.3.5 <i>Extreme Gradient Boosting</i> (XGB).....	53
5.3.6 Tabela comparativa de ER (%) e R^2	54
5.3.7 Desempenho computacional dos regressores	57
5.4 Comparação entre as técnicas de denoising aplicadas	58
5.5 Comparação entre o KNN e o DT	59
6 CONCLUSÃO	60
REFERÊNCIAS.....	64
APÊNDICE.....	72

1 INTRODUÇÃO

1.1 Justificativa da dissertação

A demanda por energia elétrica está crescendo cada vez mais devido à expansão da população e a necessidade de fornecimento contínuo de energia para residências, setores comerciais e industriais (Ahmed et al., 2022). Desta forma, é importante que o Sistema Elétrico de Potência (SEP) tenha um bom desempenho, através do fornecimento contínuo de eletricidade (Moradzadeh et al., 2022) e, para isso, destaca-se o monitoramento das Linhas de Transmissão (LTs), que são essenciais para transmitir energia elétrica da geração até o local de consumo (Lu et al., 2022; Barman, 2022). Nesse sentido, o monitoramento das LTs se faz necessário, devido às faltas resultantes da exposição ao vento, chuva, tempestades, calor e outras condições ambientais (Zerahny et al., 2021). Assim, as faltas devem ser resolvidas o mais rápido possível, a fim de reduzir os danos à rede e evitar a interrupção do serviço, instabilidade, e perdas financeiras impostas a clientes e empresas (Azeroual et al., 2021; Dashti et al., 2021).

1.2 Motivação da dissertação

O processo de detecção e localização de faltas desempenha um papel crucial na proteção das LTs contra ocorrências adversas. Dispositivos de proteção, como relés, monitoram o sistema, detectando alterações nos níveis de tensão e corrente e enviam um sinal de desligamento para o disjuntor isolar a seção com falta, garantindo a integridade do sistema (Bhatnagar; Yadav, 2020). Em (Kumari et al., 2023) foi proposto um método utilizando Rede Neural Artificial (RNA), que compara os valores de pré e pós-falta para detectar se ocorreu uma falta, com rapidez e eficiência. O trabalho apresentado em (Chavan; Kale; Deore, 2022) se mostrou robusto e confiável ao combinar RNA com a Transformada Wavelet (Wavelet Transform - WT) para detectar faltas.

Outras funções importantes de um sistema de proteção são a classificação e a localização da falta. A classificação pode ser monofásica, bifásica ou trifásica, e identifica qual(is) a(s) fase(s) defeituosa(s) (Mukherjee; Kundu, 2021). Trabalhos recentes (Fonseca, 2021; Fonseca, 2022) aprimoraram a classificação com precisão, empregando métodos diversos e destacando vantagens e desvantagens em relação a outras abordagens literárias. A

localização da falta determina o ponto onde a mesma ocorreu, em termos de quilômetros ao longo da LT, sendo este o foco desta Dissertação de Mestrado. A correta localização é desejável para que as concessionárias de energia elétrica possam identificar rapidamente o local onde ocorreu a falta, o que é fundamental para o restabelecimento do SEP o mais rápido possível (Jiang; Wu, 2005). Portanto, novos métodos para localização de faltas em LTs estão sendo continuamente desenvolvidos. A precisão desses métodos dependerá de vários fatores, como sistema de aquisição de dados, sinais medidos e características da falta, como resistência da falta, condição pré-falta e instante de tempo em que ocorrem (Alencar; Santos, 2023; Aftab et al., 2020).

Os métodos de localização de faltas podem ser divididos em duas categorias: métodos convencionais e aqueles baseados em Inteligência Artificial (IA) (Iliyaefar; Hadaeghi, 2023). Os métodos convencionais incluem o método da teoria das ondas viajantes (Traveling Wave - TW) (Joorabian; Monadi, 2023) e o método baseado na impedância da LT (Ortega; Tavares, 2021; Garcia et al., 2021), enquanto os baseados em IA incluem RNA, Lógica Fuzzy, Máquina de Vetor de Suporte (Support Vector Machine - SVM), Algoritmo Genético (AG) e abordagem de correspondência (Latha; Bhimasingu, 2021). Na primeira categoria, os algoritmos que utilizam TW foram reportados em (Naidu; Pradhan, 2023; Ngwenyama; Ngoma, 2021; Cervantes et al., 2019) para localização de faltas. Embora esse método tenha se mostrado eficaz, existem algumas limitações em LTs complexas, pois múltiplos caminhos de propagação de ondas podem prejudicar o desempenho da localização de faltas, além de exigir uma alta taxa de amostragem e, conseqüentemente, alto custo computacional e hardware especial. O método baseado na impedância da LT tem limitações na localização de faltas para LT de alta impedância. Em (Tavoosi et al., 2022) foi proposto o método da impedância da LT combinado com RNA. Diversas ferramentas como Árvore de Decisão (Decision Tree – DT), Rede Neural Convolucional (Convolutional Neural Network - CNN) e Floresta Aleatória (Random Forest) utilizam métodos baseados em dados, conhecidos como Data Driven ou Aprendizado de Máquina (Machine Learning – ML), para localizar faltas através da extração de informações de um grande conjunto de dados. O principal desafio desses métodos é a disponibilidade de dados de treinamento de alta qualidade, que podem cobrir diferentes cenários de faltas, incluindo vários tipos de faltas, ângulos de incidência, resistências de falta, localizações de falta e condições de operações do sistema (Liu et al., 2023).

Os métodos que empregaram Lógica Fuzzy foram baseados em medições de tensão e corrente ao longo da LT, porém, possuem limitações para sistemas complexos, necessitando de combinação com outros métodos para obtenção de melhores resultados (Mustari et al., 2019; Tabari; Sadeh, 2022). Os métodos baseados na Transformada Wavelet (Wavelet Transform - WT) visam extrair as componentes de alta frequência para obter o tempo de chegada da frente de onda em um ou ambos os terminais da LT, assim, pode permitir a estimativa da distância da falta. Existem também métodos baseados no sinal de tensão ou corrente, que utilizam a WT (Adly et al., 2020; Parsi et al., 2020; Shafiullah, 2022) para prever a localização combinada com RNA (Adewole; Tzoneva; Behardien, 2016; Ravesh et al., 2022). Em (Priyadarshini; Gangolu, 2020), o método Transformada Wavelet Discreta (Discrete Wavelet Transform - DWT) foi proposto para extrair as características do sinal de falta, realizando a decomposição em diferentes níveis. Essas técnicas baseadas em WT dependem da análise da forma de onda em diferentes intervalos de tempo e frequência, que podem ser influenciados por distorções ou interferências no sinal.

O SVM é um método de aprendizado de máquina altamente eficaz que apresenta resultados satisfatórios no contexto de classificação de faltas em linhas de transmissão (LT). Sua principal vantagem é a sua capacidade de generalização, ou seja, a habilidade de criar um modelo que funcione bem em dados novos, não vistos durante o treinamento. No entanto, uma desvantagem desse método é que a excelente generalização observada pode ser alcançada apenas quando o modelo é treinado com um número limitado de amostras de treinamento. Com conjunto de amostras muito grande, o SVM pode sofrer de sobrecarga computacional e ter seu desempenho prejudicado, uma vez que a complexidade do modelo pode aumentar significativamente, dificultando a adaptação a novas variabilidades nos dados (Abbasi, 2022).

A Análise de Componente Independente (Independent Component Analysis - ICA) é usada para a separação de componentes independentes de um sinal híbrido que é popular na separação cega de fontes (Blind Source Separation - BSS) (Yuan; Shao; Cai, 2021). O trabalho apresentado em (Almeida et al., 2017) utiliza o algoritmo ICA com WT e SVM para localização de faltas, porém requer um tempo maior de detecção de faltas quando comparado, por exemplo, com RNA (Ogar; Hussain; Gamage, 2023). Em (Psaras; Tzelepis, 2023), foi aplicado um método baseado em AG, usando uma função de média ponderada para obter uma estimativa final mais precisa da distância da falta. Embora os métodos de localização de faltas em LTs anteriormente mencionados tenham apresentado alto desempenho, ainda existem

desafios a serem enfrentados no que diz respeito à influência do ruído e também à complexidade computacional.

A fim de melhorar a precisão da localização de faltas com baixa complexidade computacional e considerando um cenário mais realístico, com a presença de ruído, este trabalho propõe um método baseado no algoritmo Second Order Blind Identification (SOBI). O SOBI é um dos algoritmos da técnica BSS, normalmente aplicado a um conjunto de dados multivariados, e é utilizado para separar sinais estatisticamente independentes utilizando apenas estatísticas de segunda ordem (Second-Order Statistics – SOS), identificando os componentes originais de uma mistura linear. É uma técnica de processamento estatístico de sinais que tem sido amplamente utilizada em diversos estudos envolvendo sinais elétricos do SEP, resultando em alta precisão (Oliveira et al., 2021).

Os sinais trifásicos são segmentados e processados pelo SOBI, que os decompõe em três componentes independentes. As duas primeiras componentes carregam a informação redundante e, portanto, apenas a terceira componente independente é apresentada aos algoritmos de regressão para estimarem a localização da falta. O SOBI apresenta diversas vantagens como: implementação simples, poderosa redução de dimensionalidade, o que o torna aplicável mesmo para dados complexos, robustez a ruídos presentes no sinal e capacidade de extrair fontes gaussianas (Pan, 2020).

As RNAs exigem um extenso processo de treinamento para aprender a generalizar suas saídas, mas podem apresentar resultados precisos na regressão de dados (Braga; Ferreira; Ludemir, 2007). O K-Nearest Neighbors (KNN) é um classificador/regressor baseado em aprendizado supervisionado, onde K é o número de classes vizinhas mais próximas de um teste (Naik; Koley, 2019) e possui a vantagem de não requerer treinamento (zero-shot learning). Com base nos estudos descritos, este trabalho propõe uma abordagem investigativa para o uso do algoritmo SOBI combinado com algoritmos de regressão em sinais de tensão trifásicos, como o KNN, Multilayer Perceptron (MLP – tipo específico de RNA), MLP otimizada por AG, Decision Tree, Random Forest e Extreme Gradient Boosting (XGBoost). O trabalho também visa aplicar a remoção de ruído (denoising) presente nos sinais, através das técnicas de Análise de Componentes Principais (Principal Component Analysis - PCA), Transformada Discreta de Wavelet (Discret Wavelet Transform - DWT) e Filtro Média Móvel. Desta forma, os estimadores irão operar com sinais mais limpos e com menores

dimensões, o que reduz a complexidade computacional, além de obter maior precisão para dados com a presença de ruído.

1.3 Objetivos da dissertação

Os objetivos desta dissertação podem ser subdivididos em objetivos gerais e objetivos específicos.

1.3.1 Objetivos gerais

Propor um método de baixa complexidade computacional na fase operacional utilizando Análise de Componentes Independentes (ICA) por meio do algoritmo Second-Order Blind Identification (SOBI) e aprendizado de máquina para estimar a localização de faltas em Linhas de Transmissão. Essa abordagem busca localizar faltas utilizando algoritmos de regressão, que serão treinados com os dados pré-processados pelo SOBI, visando a melhoria da eficiência e a precisão na identificação das faltas.

1.3.2 Objetivos específicos

- a) implementar o método SOBI para a identificação das componentes independentes;
- b) aprimorar as técnicas de pré-processamento de sinais, incluindo subamostragem e normalização, visando a redução da complexidade computacional sem comprometer a precisão dos modelos;
- c) explorar e comparar estratégias para remoção de ruído, avaliando o impacto de diferentes filtros, como PCA, Wavelet e Média Móvel, na qualidade dos sinais processados.
- d) selecionar regressores apropriados para o problema como o K-Nearest Neighbors (KNN), Multilayer Perceptron (MLP), MLP otimizado por Algoritmo Genético, Decision Tree, Random Forest e Extreme Gradient Boosting (XGBoost);
- e) avaliar o desempenho dos métodos, considerando métricas de erro e eficiência computacional.

2 REFERENCIAL TEÓRICO

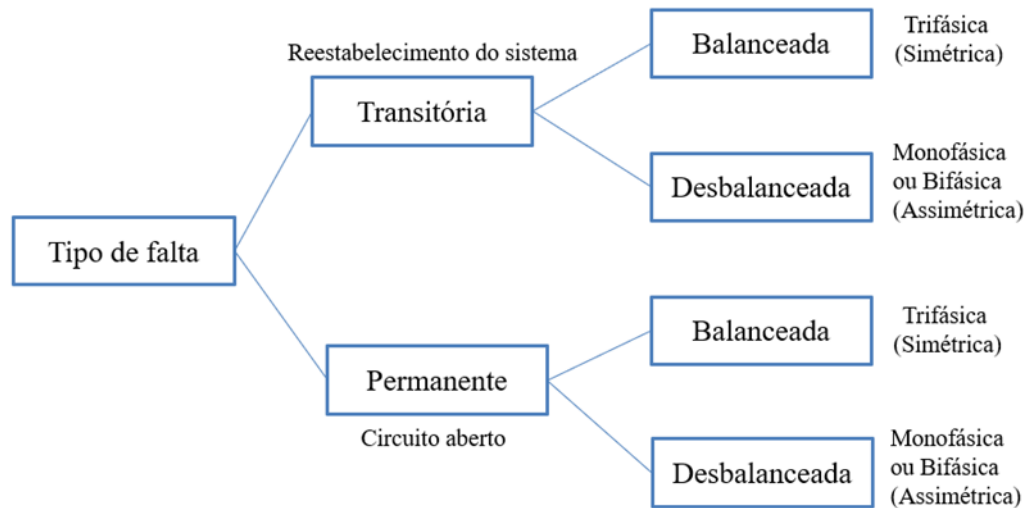
Este capítulo aborda os principais conceitos necessários para a compreensão do trabalho proposto. Aqui, serão apresentadas as definições essenciais do SEP, com foco especial nas LTs, e as faltas que podem ocorrer neste contexto. Adicionalmente, serão detalhados os algoritmos que serão empregados para a localização de faltas.

2.1 Faltas no Sistema Elétrico de Potência (SEP)

O SEP, composto pelas etapas de geração, transmissão e distribuição de energia elétrica, é fundamental para o fornecimento contínuo de energia. No entanto, faltas eventuais nesse sistema podem gerar transitórios perigosos, causar danos a equipamentos e até mesmo causar a interrupção de energia, resultando na redução da confiabilidade do sistema e na insatisfação dos clientes (Dashti et al., 2021). As LTs desempenham um papel importante no SEP de transferir a energia gerada para os centros consumidores. Nelas, muitas vezes ocorrem faltas por estarem localizadas em espaços abertos e expostos a descargas atmosféricas, ventos, tempestades, crescimento de vegetação, queda de árvores, contato com animais, etc., e este problema deve ser resolvido o mais rápido possível (Mampilly, 2020; Gopalarao et al., 2022).

A falta pode ser do tipo falta trifásica, falta bifásica, falta bifásica-terra e falta monofásica, sendo classificadas em faltas: balanceadas ou desbalanceadas. A falta balanceada, também denominada de simétrica, ocorre simultaneamente nas três fases. Já a falta desbalanceada ou assimétrica, pode ser monofásica, bifásica ou bifásica-terra. A falta monofásica é um curto-circuito entre uma fase e o terra devido ao contato físico causado por um raio ou tempestade. A falta bifásica é um curto-circuito entre duas fases, quando as linhas entram em contato físico. A falta bifásica-terra é quando duas fases entram em contato com o solo e entre si, geralmente causado pela tempestade (Ungrad; Winkler; Wisziewski, 2020). As faltas podem ser caracterizadas como permanentes ou transitórias. A falta permanente é assim denominada quando, depois de uma sequência de manobras de religamento realizadas pela concessionária, não ocorre a reenergização da linha, mantendo-a em circuito aberto. A falta é denominada transitória quando depois de uma série de manobras de religamento, é restabelecida a operação da linha (Yadav et al., 2021). A Figura 1 ilustra os tipos de falta no SEP e a Tabela 1 exhibe os exemplos típicos que podem provocar faltas transitórias ou permanentes (Barbosa, 2018).

Figura 1- Denominações de faltas.



Fonte: Barbosa (2018), adaptado.

Tabela 1- Exemplos de causas que provocam faltas transitórias ou permanentes.

Transitórias	Permanentes
Contato momentâneo com galho de árvore, pássaros ou outros animais.	Descarga atmosférica incidindo diretamente ou indiretamente na linha.
Descargas atmosféricas.	Danos ocasionados por vandalismo.
Impacto entre condutores.	Danos mecânicos provocados pelas intempéris do meio ambiente e vida útil do equipamento.

Fonte: Barbosa (2018).

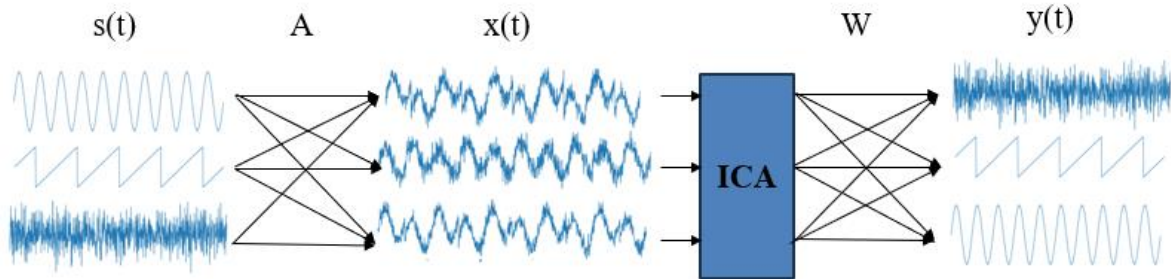
O trabalho apresentado em (Barbosa, 2018) afirma que o tipo de falta mais frequente é a monofásica (fase-terra), cerca de 85% de ocorrência. Outro trabalho apresentado em (Belagoune et al., 2021) também destaca a maior probabilidade de ocorrência das faltas monofásicas, 80%. Vale ressaltar que uma falta permanente pode resultar em danos irreversíveis às LTs e para mitigar o problema deve-se localizar a falta e isolar a linha, para intervir com a manutenção corretiva.

2.2 Análise de Componentes Independentes (ICA)

O ICA é uma ferramenta estatística que faz parte do método BSS, ou seja, o sinal procurado é estimado sem conhecimento prévio do sistema. Sua função é separar

componentes independentes de um sinal considerando que as fontes são estatisticamente independentes entre si (Kumari et al., 2023).

Figura 2- Separação de fontes.



Fonte: Kumari et al. (2023), adaptado.

O recurso em questão é aplicado quando é necessário extrair um vetor de sinais s denominados fontes a partir de um vetor x , composto pelos sinais de mistura, que são formados por uma transformação linear dos sinais das fontes. Para tanto, os métodos do ICA visam encontrar uma matriz de separação W , ao explorar características estatísticas das misturas (observações), conforme a equação (1):

$$\mathbf{s} = \mathbf{W}\mathbf{x} \quad (1)$$

A estimativa das fontes utilizando ICA apenas é possível e confiável se as seguintes condições forem atendidas (Moraes, 2022):

- a) os sinais originais devem ser estatisticamente independentes;
- b) os sinais devem apresentar distribuições de probabilidade não-gaussiana. Se as fontes forem gaussianas, pode haver ambiguidade na separação;
- c) o número de sinais misturados deve ser maior ou igual ao número de fontes.
- d) as misturas das fontes devem ser combinações lineares das fontes independentes;

A estimação das fontes utilizando o método ICA pode gerar duas indeterminações que estão associadas à falta de conhecimento das fontes originais que geraram os sinais e da matriz de mistura. Essas indeterminações estão relacionadas com: as amplitudes, pois não é possível determinar as amplitudes exatas dos sinais originais; e com as ordens dos sinais

estimados, visto que o ICA também não é capaz de identificar a ordem dos sinais independentes (Koker, 2020).

Existem vários algoritmos que implementam o ICA, em que se utilizam diferentes medidas de independência estatística e apresenta diferentes maneiras de maximizá-la a partir da matriz de mistura (observações) (Ferreira et al., 2015; Nagata et al., 2018), tais como o FastICA (Jun et al., 2020), Infomax (Karpiel et al., 2020), JADE (Joint Approximate Diagonalization of Eigenmatrices) (Abdalkareem et al., 2022), AMUSE (Adaptive Mixture Separation) (Alencar; Santos, 2023) e SOBI (Second Order Blind Identification) (Wang; Zhang, 2017). Porém, para aplicação deste trabalho foi utilizado apenas o SOBI. Portanto, apenas este algoritmo será descrito na próxima seção.

2.3 Second Order Blind Identification (SOBI)

O SOBI é um método robusto de BSS por diagonalização conjunta de matrizes de correlação que tem a função de extrair os sinais de origem de uma mistura sob a hipótese de coeficientes de mistura e sinais de origem desconhecidos (Wang; Zhang, 2017). As estatísticas de segunda ordem são adotadas a partir dos dados recebidos e o SOBI tem a capacidade de estimar os sinais de origem mesmo com poucos pontos de dados de mistura.

Suponha N misturas observadas $\mathbf{x}(t) = [x_1(t), x_2(t), \dots, x_N(t)]$, cada um dos $x_i(t)$ pode ser considerado como uma mistura linear instantânea de N componentes desconhecidos ou fontes $s_i(t)$, através de uma matriz de mistura A desconhecida.

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) \quad (2)$$

Os sinais das fontes $\mathbf{s}(t)$ são considerados independentes. Estimar A e $\mathbf{s}(t)$ é conhecido como BSS, em que o objetivo é determinar uma matriz que se aproxima de A^{-1} , conhecida como matriz de desmistura W , $M \times N$, com M sinais de saída. Em que (t) é a série temporal contínua dos componentes ou fontes recuperadas.

$$\mathbf{y}(t) = \mathbf{W}\mathbf{x}(t) = \mathbf{W}\mathbf{A}\mathbf{s}(t) = \mathbf{s}(t) \quad (3)$$

O algoritmo SOBI encontra a solução através de duas etapas. Na primeira etapa, os sinais observados $\mathbf{x}(t)$ são centrados na média e pré-esferados; o vetor transformado é

denotado como $\tilde{\mathbf{x}}(t)$. Uma matriz \mathbf{B} é escolhida de modo que a matriz de correlação de $\tilde{\mathbf{x}}(t)$, $\mathbf{E}(\tilde{\mathbf{x}}(t)\tilde{\mathbf{x}}(t)^T)$, é uma matriz identidade. $\mathbf{E}(\cdot)$ é a operação de média determinística. Isso é realizado movendo-se para a base de Análise de Componentes Principais (PCA).

$$\mathbf{B} = \mathbf{D}^{-\frac{1}{2}}\mathbf{U}^T, \quad (4)$$

em que \mathbf{D} é uma matriz diagonal coletando os autovalores da matriz de correlação $\mathbf{E}(\tilde{\mathbf{x}}(t)\tilde{\mathbf{x}}(t)^T)$ e \mathbf{U} é uma matriz onde as colunas são os autovetores correspondentes, ou seja, os componentes PCA de \mathbf{x} . Na segunda etapa, constrói-se um conjunto de matrizes que, na base separada correta, devem ser diagonais. É definido um conjunto de valores de atraso de tempo para calcular \mathbf{R}_{τ_k} , que são matrizes de correlação simétrica entre o sinal $\tilde{\mathbf{x}}(t)$ e uma versão deslocada temporalmente de si mesmo. Em seguida, é escolhida uma matriz de rotação \mathbf{V} que diagonalize junto com \mathbf{R}_{τ_k} , minimizando a soma dos quadrados das entradas fora da diagonal dos produtos da matriz, descrito como:

$$\sum_{k=1}^p \sum_{i \neq j} (\mathbf{V}^T \mathbf{R}_{\tau_k} \mathbf{V})^2 \quad (5)$$

É possível definir um parâmetro de limite para o seno do ângulo de rotação \mathbf{V} . Caso o seno do ângulo de rotação \mathbf{V} seja menor que um limite pré-definido, o processo iterativo termina. A estimativa final da matriz de separação utilizada para extrair os componentes desejados é:

$$\mathbf{W} = \mathbf{V}\mathbf{B} \quad (6)$$

O SOBI explora a coerência temporal dos sinais da fonte para decompor a mistura de fontes. Ele encontra \mathbf{W} por meio de um processo iterativo que minimiza as correlações cruzadas de soma quadrada entre um componente no tempo t e outro componente no tempo $t+s$, em um conjunto de atrasos de tempo (Xiao; Hu, 2008; Liu et al., 2013).

2.4 K-Nearest Neighbors (KNN)

O algoritmo KNN é um dos algoritmos de Machine Learning (ML) mais simples, onde a classificação do objeto ocorre pela maioria dos votos de seus vizinhos, sendo atribuído à classe com os k -vizinhos mais próximos. k é um inteiro positivo, geralmente pequeno. Para $k = 1$, o objeto é simplesmente atribuído à classe de seu vizinho mais próximo. Em problemas

de classificação binária (duas classes), é útil escolher k como um número ímpar, evitando dificuldades com votos empatados. Este método também é aplicado para regressão, atribuindo o valor da propriedade ao objeto como a média dos valores de seus k vizinhos mais próximos. É útil pesar as contribuições dos vizinhos para que os vizinhos mais próximos contribuam mais para a média do que os mais distantes (Sowah et al., 2018).

Quando ocorre uma falta na LT, o KNN é usado para encontrar o método mais preciso e fornecer a distância da falta, utilizando diferentes técnicas baseadas na análise da frequência de energia do sinal. Se a interferência do ruído nos dados for alta, a precisão da classificação é menor (Cui et al., 2016). Portanto, tem-se a proposta de utilizar o SOBI para tratamento dos dados.

2.5 Redes Neurais Artificiais (RNA)

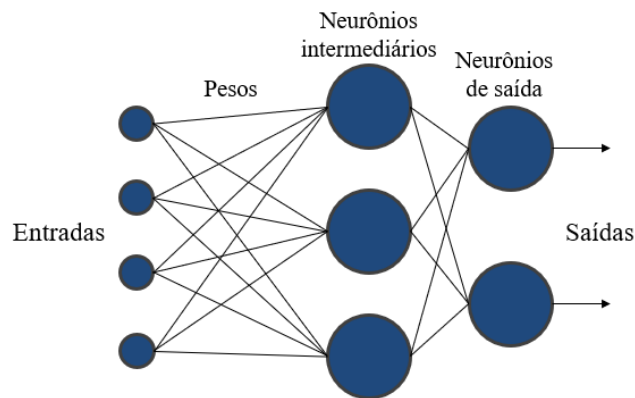
As Redes Neurais Artificiais (RNA) são sistemas de computação adaptativos inspirados nas características de processamento de informação dos neurônios reais (biológicos) e também nas características de suas interconexões, ou seja, imitam o processamento de informação realizado por um neurônio biológico, e quando utilizados em conjunto e interagindo uns com os outros, compõem a chamada rede neural.

As unidades são dispostas em uma ou mais camadas interligadas por um grande número de conexões que, na maioria das vezes, estão associadas aos pesos, que armazenam o conhecimento e ponderam a entrada recebida por cada neurônio. Um dos modelos mais comuns de RNA é o Perceptron Multicamadas (MLP), uma rede do tipo feedforward composta por pelo menos uma camada oculta e treinada por backpropagation. A solução de problemas baseados em RNAs passam por um processo de aprendizagem, onde um conjunto de exemplos é apresentado para a rede, que irá extrair automaticamente as características necessárias para representar a informação fornecida, para posteriormente dar respostas coerentes para dados desconhecidos.

Em termos de representação de conhecimento, as RNAs são o que há de mais avançado. Comporta-se de maneira semelhante sobre como o ser humano adquire e representa seu conhecimento. A partir de estudos neurológicos, procurou-se formalizar um padrão matemático para as reações elétricas dos neurônios humanos para, assim, representar computacionalmente estes comportamentos. A Configuração da rede depende do grau de

precisão do problema a ser resolvido. O conjunto de vários neurônios artificiais denomina-se RNA. Basicamente, serão fornecidos dados de entrada para a rede e implementadas propagações de forma que a alteração dos pesos ao longo dela forneça uma saída compatível com a resposta conhecida àquele estímulo pré-determinado. Logo, os pesos são balanceados até obter um padrão de respostas aos estímulos, compatível com o conhecido (Hernandez, 2005).

Figura 3- RNA típica.



Fonte: Hernandez (2005), adaptado.

2.6 Decision Tree (DT)

As Árvores de Decisão (Decision Tree - DT) são modelos amplamente utilizados em tarefas de classificação e regressão, devido à sua estrutura hierárquica e interpretabilidade intuitiva. Esses modelos seguem a estratégia dividir para conquistar, onde o espaço de entrada é recursivamente particionado em sub-regiões mais homogêneas. Esse processo ocorre com base em uma função de impureza, como a entropia ou o índice de Gini para classificação, ou o erro quadrático médio (MSE) no caso da regressão.

Cada nó interno da árvore representa uma condição de decisão baseada em um atributo específico, enquanto os nós folha armazenam a classe prevista (para classificação) ou um valor numérico (para regressão). Esse modelo tem como vantagem a sua facilidade de interpretação, pois permite entender o fluxo lógico das decisões através de regras do tipo "se-então", tornando-se uma alternativa útil em problemas que exigem transparência e explicabilidade (Tomic; Kusljevic; Vujicic, 2007).

Além de sua interpretabilidade, as árvores de decisão podem ser utilizadas para explicações individuais de previsões, similar a técnicas como LIME e Shapley Values, que auxiliam na análise da influência de cada variável na predição final. No entanto, apesar de sua simplicidade e eficiência, as árvores de decisão podem apresentar tendência ao overfitting, especialmente quando muito profundas. Para mitigar esse problema, técnicas como poda da árvore (pruning) e a definição de hiperparâmetros adequados são frequentemente aplicadas (Costa, 2023).

2.7 Random Forest (RF)

O RF é um algoritmo de aprendizado de máquina baseado na combinação de múltiplas árvores de decisão, o que lhe confere robustez, alta precisão e capacidade de generalização. Sua construção ocorre através do bootstrap aggregating (bagging), onde cada árvore é treinada com um subconjunto aleatório dos dados e um conjunto aleatório de atributos. Esse processo permite que o modelo capture relações complexas entre as variáveis, minimizando overfitting e tornando-o eficaz para problemas de classificação e regressão. Na classificação, a decisão final é obtida por votação majoritária, enquanto na regressão, calcula-se a média das predições individuais.

Uma das principais vantagens do Random Forest é sua resiliência a dados ruidosos e desbalanceados, pois a diversidade entre as árvores melhora a estabilidade do modelo. Além disso, ele fornece medidas de importância das variáveis, ajudando na interpretação dos resultados e na seleção de atributos relevantes. Sua aplicação se estende a diversas áreas, como segurança cibernética, medicina, modelagem financeira e ciências ambientais, onde é essencial um modelo confiável e capaz de lidar com grandes volumes de dados de forma precisa (Santos et al., 2022).

2.8 Extreme Gradient Boosting (XGBoost)

O XGBoost (Extreme Gradient Boosting) é um algoritmo baseado em árvores de decisão que aprimora significativamente o desempenho em tarefas de regressão e classificação. Seu diferencial está na adição de um termo de regularização à função de perda, além da utilização da expansão de Taylor de segunda ordem para ajustar os modelos de forma mais eficiente. Graças a essas características, o XGBoost apresenta alta capacidade de

generalização, evitando overfitting e garantindo um ajuste robusto dos dados. Sua eficiência computacional, habilidade em lidar com valores ausentes e a possibilidade de análise da importância dos atributos fazem com que seja amplamente utilizado em diversas aplicações, como previsão de séries temporais, análise de dados financeiros e diagnósticos médicos. Além disso, o XGBoost pode ser combinado com modelos lineares, como regressão linear, ou não lineares, como redes neurais do tipo MLP, potencializando ainda mais a precisão e o desempenho da modelagem (Wu; Chang, 2024).

Apesar de suas vantagens, o desempenho do XGBoost depende diretamente da qualidade e quantidade dos dados disponíveis. O modelo exige um volume significativo de informações para obter bons resultados, o que pode ser um desafio em cenários com conjuntos de dados limitados. Outra limitação está na complexidade de ajuste de hiperparâmetros, pois diferentes combinações podem influenciar drasticamente os resultados. Além disso, devido à sua estrutura baseada em árvores, não é possível expressar o modelo final como uma equação matemática explícita, o que dificulta a interpretabilidade em algumas aplicações. Ainda assim, o XGBoost representa um avanço significativo em relação aos métodos tradicionais de aprendizado de máquina, oferecendo uma abordagem escalável e eficiente para problemas que exigem alto desempenho computacional e precisão preditiva (Wang et al., 2024).

3 ESTUDOS RELACIONADOS

Diversos estudos têm abordado a localização de faltas, buscando aprimorar as técnicas de detecção e localização. O trabalho apresentado em (Alencar; Santos, 2023) propôs um método de localização de faltas baseado em Single Channel ICA (SCICA), uma variante do ICA que encontra os componentes independentes a partir de apenas uma fonte de misturas. O método não necessita de banco de dados prévio, o que independe das condições da carga. O trabalho foi comparado com a Transformada Discreta de Wavelet (Discrete Wavelet Transform - DWT) e apresentou bom desempenho. No entanto, foi considerado apenas um nível de Signal-to-Noise Ratio (SNR) de 70 dB. Além disso, a técnica SCICA requer uma etapa de agrupamento de dados após a decomposição da matriz de misturas e também requer uma matriz de misturas formada por versões deslocadas do sinal monitorado de alta dimensão. A alta dimensão da matriz de misturas e a etapa de agrupamento elevam a complexidade computacional do método, impedindo sua implementação em tempo real. Explorar outras

variações de SNR auxilia a compreender melhor a sensibilidade do método. Os resultados serão apresentados na Tabela 2, comparados com outros métodos presentes na literatura.

Em (Naidu; Pradhan, 2021) foi utilizado um método baseado na teoria das ondas viajantes (Traveling Wave – TW) para localizar faltas em LT, sem depender de comunicação externa, ou seja, não requer a troca de informações entre diferentes pontos do sistema para realizar a detecção de faltas. O método utiliza como entradas os tempos de chegada das primeiras três frentes de onda viajante e a velocidade de propagação das ondas na linha. Os resultados dos testes obtidos são comparados com um método baseado em comunicação de dois terminais e são considerados precisos. A desvantagem é que existem algumas limitações em LTs complexas ao utilizar o método TW, pois múltiplos caminhos de propagação de ondas podem prejudicar o desempenho da localização de faltas, além de exigir uma alta taxa de amostragem.

Em (Niazkar et al., 2024) é apresentado um método de localização de faltas em LTs utilizando medições de corrente não sincronizadas. A seção defeituosa da linha é identificada analisando a polaridade da segunda onda viajante que chega a cada terminal. Em seguida, a falta é discriminada como ocorrendo antes ou depois do ponto médio da seção defeituosa, observando a mudança de polaridade entre as duas primeiras ondas viajantes consecutivas. Por fim, a localização da falta é estimada em um terminal específico ao qual a seção defeituosa está conectada. As etapas são realizadas utilizando a DTW das ondas viajantes incidentes, refletidas e refratadas nas extremidades de medição. O desempenho do método é comparado com outros métodos existentes e apresentou bom desempenho, principalmente para faltas próximas dos terminais da LT, atingindo um erro máximo de 1,22% para o pior caso com 1% de ruído no sinal medido. O trabalho destaca a influência do ruído nos sinais de corrente sobre o desempenho de algoritmos de localização de faltas, principalmente os que são baseados em ondas viajantes. Considerou-se ruído branco gaussiano nos sinais de corrente, apresentando resultados satisfatórios para SNR igual ou maior que 50 dB. No entanto, os autores enfatizam que o desempenho do método pode ser afetado em casos de maior ruído, uma limitação comum a algoritmos de localização de faltas que utilizam ondas viajantes. Além disso, o método requer uma alta frequência de amostragem.

O método descrito em (Gaur; Bhalja, 2022) emprega Redes Neurais Artificiais (RNA), ou Artificial Neural Network (ANN), por meio da técnica de retropropagação padrão de alimentação direta. Após o processamento das informações de corrente e tensão das três fases

retiradas do início da linha, esses dados são divididos em cinco grupos de frequência específicos, os quais são utilizados como entrada para a RNA. O uso de RNA exige maior complexidade computacional, o que requer processamento mais elevado devido à etapa de treinamento do algoritmo. Além disso, não foi considerada a influência de ruído no sinal e o método utiliza medições de corrente ao invés de medições de tensão.

O trabalho apresentado em (Coban; Tezcan, 2020) utiliza duas redes neurais profundas. Uma é a MLP (Multilayer Perceptron) profunda e a outra é a Rede Neural Convolutiva (Convolutional Neural Networks - CNN) combinada com redes do tipo long short-term memory - LSTM. Uma CNN mista com LSTM é treinada para prever a distância da falta com base em medições de tensão e corrente obtidas de uma única extremidade da linha. Para preservar a invariância e capturar correlações temporais nos dados de séries temporais, são utilizadas funções convolucionais, camadas de pooling e a estrutura LSTM. O desempenho dessas abordagens foi comparado com um método convencional baseado na impedância, demonstrando maior precisão para o esquema baseado na CNN-LSTM. Apesar de o trabalho ter apresentado menor erro comparado com os demais estudos, para situação com ruído, da Tabela 2, este método também requer medições de corrente e o nível de ruído é relativamente baixo, o que não causa impactos significativos em comparação com perturbações reais.

A metodologia aplicada em (Fan et al., 2019) emprega técnicas para a extração de dados característicos de sinais de corrente e tensão em relação à localização de faltas em linhas de transmissão, com base na Decomposição Modal Empírica (Empirical Mode Decomposition - EMD) e a Decomposição Modal Variacional (Variational Mode Decomposition - VMD) associada ao operador de energia Teager, que auxilia na identificação da energia instantânea da primeira função de modo intrínseco. A VMD tem um custo computacional mais alto em comparação com o EMD, devido à otimização empregada em seu algoritmo. Os resultados comprovaram que as técnicas são capazes de extrair os dados característicos dos sinais de corrente e tensão e estimar a distância da falta com alta precisão. No entanto, a presença de ruído pode comprometer a precisão da EMD, conhecida por sua sensibilidade a flutuações de alta frequência. A ausência de abordagens de pré-processamento ou filtragem pode resultar na interpretação inadequada de componentes indesejados como partes essenciais do sinal, impactando a confiabilidade do método em ambientes práticos onde a presença de ruídos é comum.

O trabalho apresentado em (Oliveira; Moreira; Picanço, 2023) emprega a DWT combinada com rede neural profunda (Deep Neural Network – DNN) em uma LT com compensação em série. Recursos extraídos de medições síncronas de correntes de falta nos três terminais usando DWT são fornecidos para o DNN. Apesar de este ter apresentado o menor erro comparado aos demais trabalhos, o algoritmo utiliza sinais de tensão e corrente. A saturação dos transformadores de corrente, devido a transitórios de falta de alta frequência, impõe erros na medição de corrente, afetando negativamente o desempenho do algoritmo de localização de faltas na prática. Outro aspecto desfavorável neste trabalho é a ausência da influência do ruído. Além disso, o sistema de compensação e a utilização de três terminais aumentam o custo do sistema, o tornando inviável dependendo da aplicação.

A Tabela 2 apresenta os métodos aplicados em alguns estudos, considerando aspectos relevantes dos métodos, bem como alguns parâmetros da LT que são fundamentais para o desempenho da localização das faltas. As variações na tensão da LT podem influenciar diretamente nas condições do sistema elétrico. Se a tensão for muito alta, pode ocorrer saturação nos transformadores e equipamentos de medição, afetando a precisão das leituras. Além disso, a tensão impacta nas características das ondas elétricas, o que é relevante em métodos que se baseiam nessas características para localização de faltas (Mirzaei et al., 2018; Andrey; Oliveira, 2019; Manassero, 2023). Uma maior frequência de amostragem garante uma análise mais detalhada do sinal, porém requer maior complexidade computacional, resultando em alto custo computacional. O erro médio absoluto é altamente influenciado pelo comprimento da LT, pois quanto maior o comprimento, maior a distorção e atenuação dos sinais (Lopes et al., 2022). Os métodos que exigem treinamento dependem dos dados para ter generalização, além de exigir processamento mais elevado (Prisco et al., 2021). O desempenho do método utilizando um terminal é afetado à medida que a resistência da falta aumenta (Basso, 2020; Holanda et al., 2020). No entanto, quanto mais terminais, maior o custo de implantação. Alguns métodos utilizam medições de corrente, as quais são mais suscetíveis a interferências em comparação com as medições de tensão. Os métodos que utilizam tensão e corrente (V e I) necessitam de mais informações, tornando-os mais complexos (Gaur; Bhalja, 2022). Os erros apresentados na Tabela 2 foram discutidos ao longo deste capítulo. Porém, vale ressaltar que em alguns trabalhos apresentados na Tabela 2 não foram consideradas as influências do ruído no sinal, sendo que as variações de ruído auxiliam na avaliação da aplicabilidade prática do método para condições operacionais reais.

Tabela 2- Avaliação comparativa de métodos de localização de faltas em Linhas de Transmissão.

Referência	Alencar (2023)	Naidu (2023)	Prisco (2021)	Gaur (2022)	Coban (2020)	Fan (2019)	Mirzaei (2018)
Método	SCICA	TW	DWT	RNA	CNN-LSTM	EMD	DNN
Sinal de Entrada	V	I	I	V e I	V e I	V e I	V e I
Tensão da LT (kV)	500 e 440	400	-	230	320	765	400
Comprimento da LT (km)	200 e 330	150	100	100	200	200	120
Resistência da Falta (Ω)	100	100	100	125	10	100	150
Frequência de Amostragem (kHz)	100	250-2000	1000	4	10	200	40
Etapas de Treinamento	Não	Não	Não	Sim	Sim	Não	Sim
Número de Terminais	1 e 2	2	3	1	1	2	3
Erro Médio Absoluto sem Ruído (%)	0,24 ^a	0,47 ^c	-	0,18	-	0,235	0,0458
Erro Médio Absoluto com Ruído (%)	0,72 ^b	1,17 ^d	0,25 ^e	-	0,11 ^f	-	-

Condições:

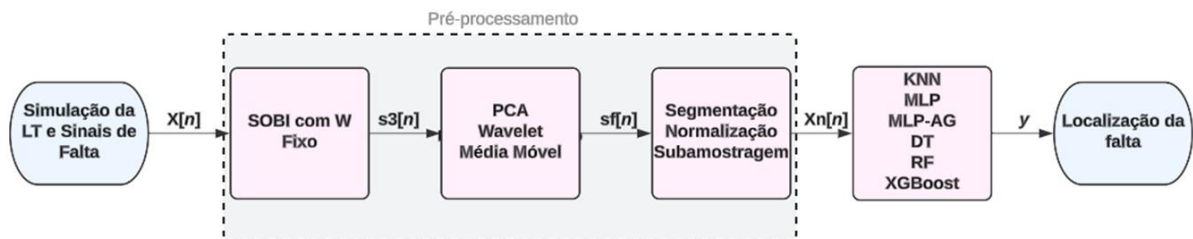
- a) ^amedições de dois terminais e tensão da LT de 440 kV;
- b) ^bmedições de dois terminais, tensão da LT de 500 kV e com SNR de 70 dB;
- c) ^cfrequência de amostragem de 250 kHz;
- d) ^dfrequência de amostragem de 250 kHz e nível de ruído de 2% da magnitude do sinal;
- e) ^eSNR de 60 dB;

f) nível de ruído de 1% da magnitude do sinal.

4 METODOLOGIA

O projeto do método proposto é dividido em seis etapas: simulação dos sinais de faltas, implementação do SOBI, remoção de ruídos, segmentação do sinal, normalização e subamostragem do sinal, projeto do estimador e localização da falta. O diagrama de blocos da metodologia proposta para a localização de faltas é ilustrado na Figura 4.

Figura 4- Fluxograma do Método Proposto.



Fonte: Do autor (2025).

4.1 Ambiente de implementação e simulação dos dados

Neste trabalho, foi utilizada a base de dados simulada no ATP (Alternative Transients Program) por Cardoso et al. (2021), um software amplamente utilizado na análise de transitórios eletromagnéticos em sistemas elétricos. Já as demais implementações foram realizadas utilizando a linguagem de código aberto Python, devido à sua ampla disponibilidade de bibliotecas voltadas para aprendizado de máquina, processamento de sinais e modelagem matemática. Para a implementação dos modelos de aprendizado de máquina, foram utilizados pacotes específicos, conforme listado abaixo:

- a) Pandas e NumPy: Manipulação e estruturação dos dados;
- b) Scikit-Learn: Implementação dos algoritmos de regressão. Também utilizado para normalização dos dados e validação cruzada com K-Fold;
- c) PyWavelets (PyWT): Aplicação da Transformada Wavelet para remoção de ruído nos sinais;

d) SciPy: Implementação de otimizações matemáticas e aplicação de filtros para pré-processamento dos sinais;

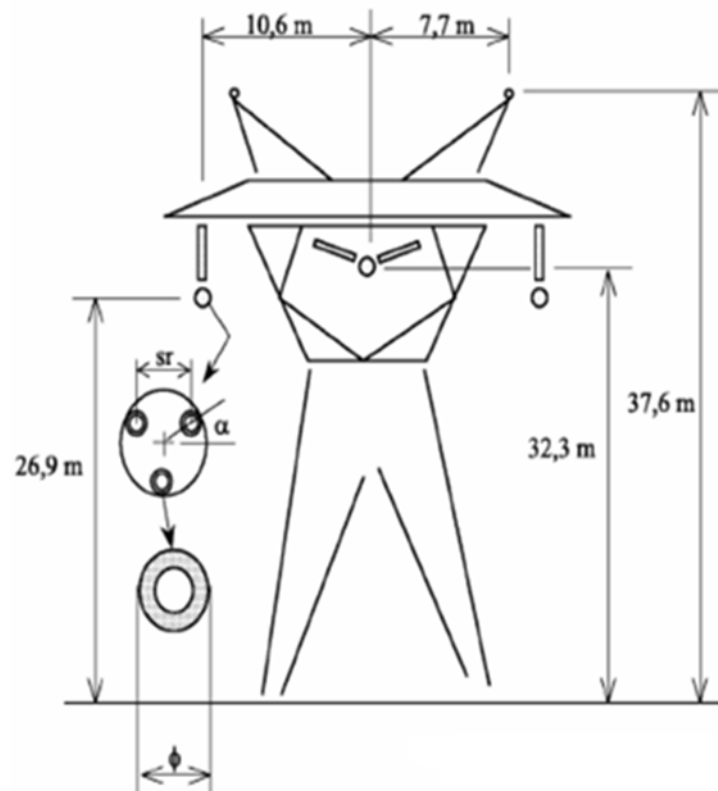
e) Matplotlib e Seaborn: Geração dos gráficos e visualização dos resultados.

O ambiente de execução foi um notebook Intel Core i5 (1 GHz), com 8 GB de RAM, SSD de 512 GB, sistema operacional Windows 11 (64 bits) e placa NVIDIA, garantindo um ambiente controlado para execução dos testes. Além disso, medidas foram tomadas para minimizar interferências externas, como desativação de atualizações automáticas, modo de execução prioritário e ausência de conexão com a internet durante os testes.

4.2 Simulação da Linha de Transmissão

A linha de transmissão simulada é do modelo J. Marti, que é dependente do domínio modal, amplamente utilizado para simular transitórios em LTs (Marti, 1982), com 200 km de extensão, frequência de 60 Hz e tensão de 500 kV, obtendo os dados de um terminal. O modelo da LT utilizada é mostrado na Figura 5. As faltas elétricas são monofásicas do tipo AT, entre um condutor e o terra.

Figura 5- Modelo da Linha de Transmissão de 500 kV.

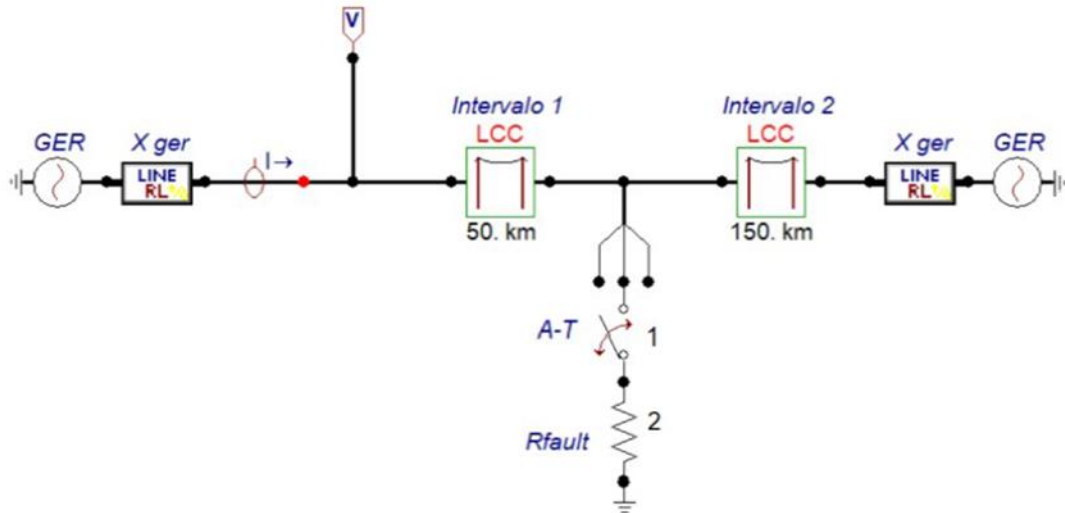


Fonte: Cardoso et al. (2021).

4.3 Base de Dados

Os sinais possuem a relação sinal-ruído (signal-to-noise ratio - SNR) entre 30 dB e 70 dB, variando de 10 em 10 dB, e considerou-se também a situação de sinal sem ruído. A Figura 6 exibe o circuito proposto por Cardoso et al. (2021) para simulação das faltas e simula uma falta em um local de 50 km, uma distância definida no Intervalo 1, do início da linha até a falta, e no Intervalo 2, que é de onde ocorre a falta até o final da linha.

Figura 6- Circuito para simulação das faltas.



Fonte: Cardoso et al. (2021).

Ao utilizar uma frequência de amostragem de 200 kHz e a análise sendo realizada em um período (um ciclo) do sinal elétrico, foram adquiridas 3.334 amostras discretas para cada sinal analisado. O conjunto de dados é composto por 7.920 faltas e foi processado utilizando validação cruzada K-Fold com 10 folds, onde, a cada iteração, 90% dos dados foram utilizados para treinamento e 10% para teste. As características dos dados estão detalhadas na Tabela 3.

Tabela 3- Propriedades das Faltas.

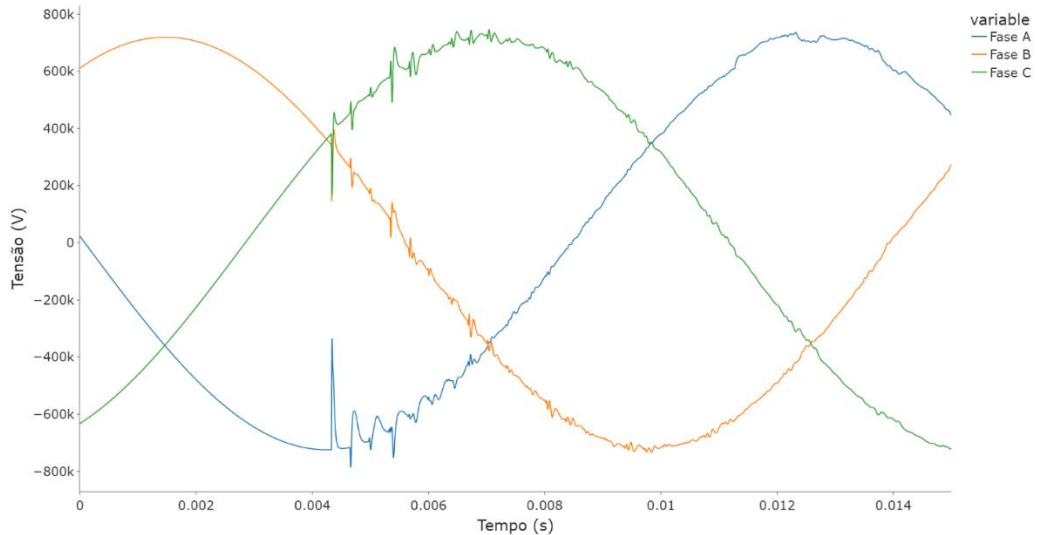
Características	Valores
Localização	de 2 a 198 com passos de 2 (km)
Resistência	20, 70, 120, 150, 180, 200, 240, 280 (Ω)
Ângulo	45, 90, 115, 135, 155, 180, 205, 225, 250, 270 (graus)

Fonte: Cardoso et al. (2021).

Os sinais de tensão apresentados na Figura 7 correspondem ao terminal da linha de transmissão simulada, onde ocorreu uma falta localizada a 48 km, com resistência de 20 Ohms e ângulo de incidência de 90°. A intensidade do transitório varia conforme os

parâmetros da falta, refletindo as características específicas do evento e suas influências na propagação do sinal ao longo da linha.

Figura 7- Falta Elétrica a 48 km, 20 ohms e incidência em 90°.



Fonte: Do autor (2025).

4.4 Transformação do sinal

Utilizando a base de dados apresentada anteriormente, os registros foram processados pelo algoritmo SOBI. As três fases de tensão monitoradas foram organizadas em um vetor \mathbf{x} , formando uma abordagem multicanal:

$$\mathbf{x} = [V_A[n]V_B[n]V_C[n]]^T \quad (7)$$

em que V_A, V_B e V_C são as tensões nas fases A, B e C do sistema trifásico monitorado, $n = 1, \dots, N$, é o índice de amostragem e T sobrescrito indica a operação transposta.

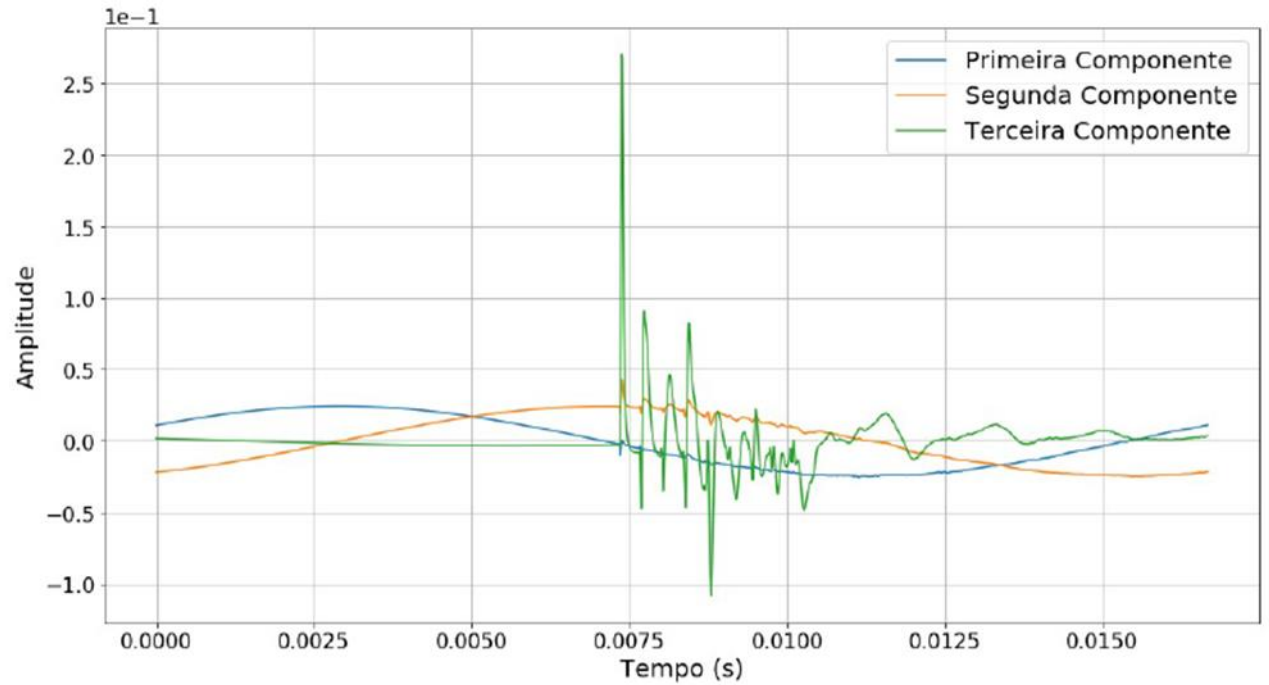
O algoritmo SOBI é aplicado ao vetor \mathbf{x} , considerando uma janela de sinal com $N = 3.334$ amostras discretas. O SOBI retorna neste caso uma matriz de separação \mathbf{W} de dimensão 3×3 , já que o vetor \mathbf{x} possui dimensão 3×3.334 . No método adotado neste estudo, a matriz \mathbf{W} é calculada previamente na fase de treinamento e utilizada de forma fixa na fase operacional. Isso elimina a necessidade de recalculá-la utilizando o algoritmo SOBI a cada nova janela do

sinal, reduzindo significativamente o custo computacional, pois assim a separação dos sinais na fase operacional ocorrerá por meio de uma multiplicação algébrica simples, conforme já apresentado na equação (1).

Por outro lado, a abordagem alternativa com \mathbf{W} variável recalcula a matriz \mathbf{W} para cada janela processada, permitindo que o método se adapte a novas variações nos sinais, mesmo que não tenham sido observadas no conjunto de treinamento. Essa característica pode ser benéfica para cenários altamente dinâmicos, mas acarreta um custo computacional elevado. Sendo assim, dado que o objetivo deste trabalho é propor um método simples computacionalmente e eficiente, e os estudos apresentados em (Cardoso et al., 2021; Pereira et al., 2024) demonstraram que as duas abordagens não apresentaram diferenças significativas em termos de desempenho, neste trabalho foi adotada a versão com \mathbf{W} fixo, garantindo maior eficiência computacional sem comprometer a qualidade da separação dos sinais.

Como o SOBI ordena as componentes de saída da maior para a menor energia, observaram-se, a partir da base de dados de treinamento (vide exemplo da Figura 8), que a terceira componente, de menor energia, é a que melhor discrimina os transitórios causados pelos eventos de falta, carregando as informações de mais alta frequência, enquanto as duas primeiras componentes preservam predominantemente as frequências mais baixas e informações redundantes do sinal, como a componente fundamental.

Figura 8- Três componente do SOBI referentes a uma falta a 48 km.

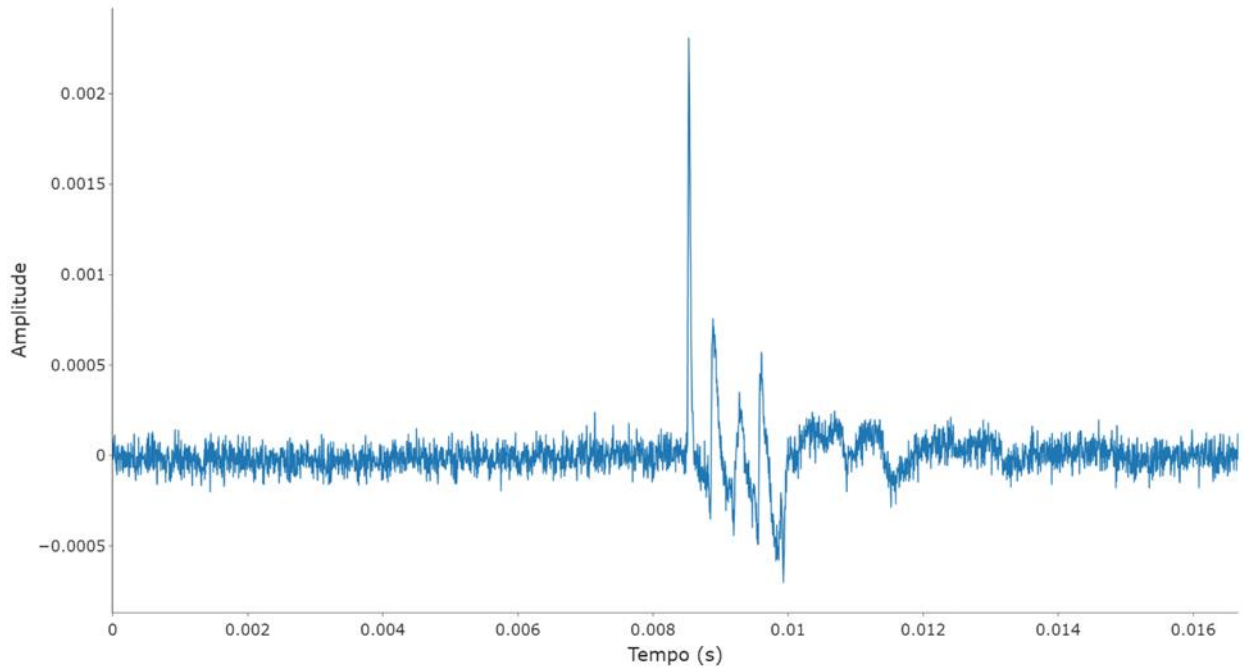


Fonte: (Cardoso et al., 2021; Pereira et al., 2024).

4.5 Técnicas de denoising

Observou-se que a componente de saída do SOBI continha boa parte de ruídos além dos transitórios causados pela falta. Portanto, estratégias para atenuar os ruídos foram empregadas. Após a aplicação do SOBI para identificar as componentes independentes dos sinais e a seleção da terceira componente (que contém as informações mais relevantes), as janelas de sinal foram processadas por meio da construção de séries temporais deslocadas no tempo. A Figura 9 exibe um exemplo de uma série do sinal obtido.

Figura 9- Terceira componente do SOBI referente a uma falta a 48 km, 20 Ohms, incidência em 180° e SNR 50dB.



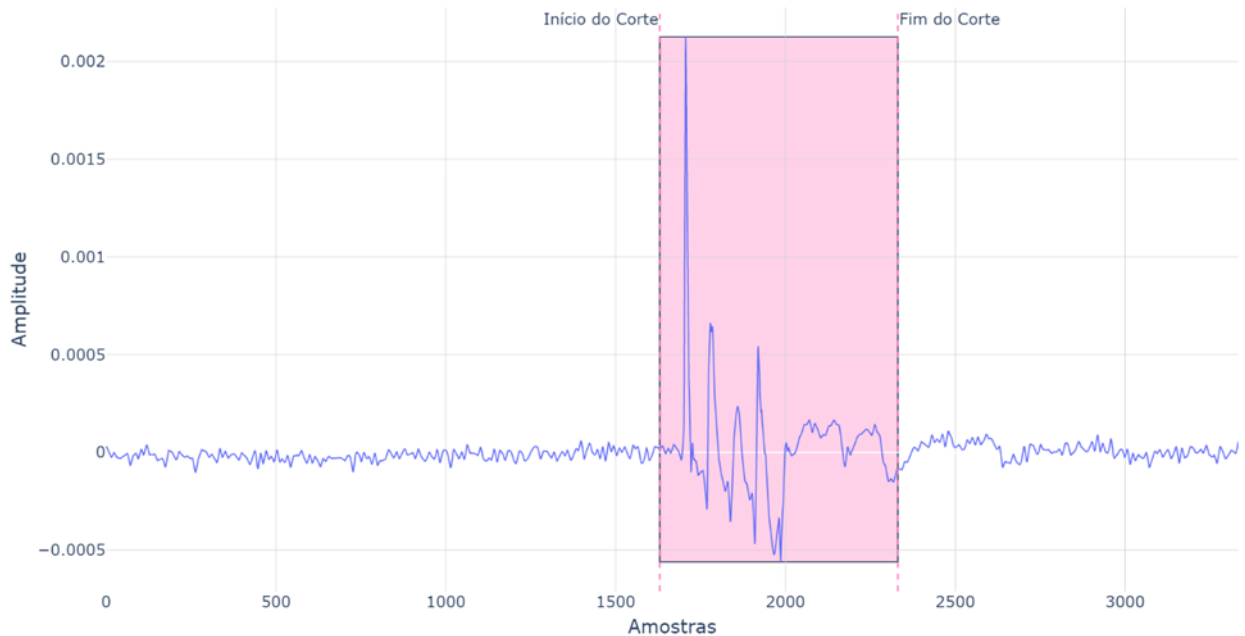
Fonte: Do autor (2025).

Devido à presença de ruído, que foi inserido para simular condições reais, foram aplicadas três abordagens distintas: PCA (Principal Component Analysis), DWT (Discrete Wavelet Transform) e Filtro Média Móvel. Essas técnicas serão detalhadas na seção de resultados, com o objetivo de comparar sua eficiência em remover o ruído e preservar as características essenciais do sinal para uma localização precisa de faltas.

4.6 Segmentação do sinal

Após o processo de denoising, os sinais resultantes foram segmentados para isolar as regiões de interesse, correspondendo às instâncias de falta detectadas. Essa segmentação foi realizada com base no índice do pico máximo da falta e ajustada para capturar tanto o início do transiente quanto a duração completa do evento de falta, visando capturar apenas as informações mais relevantes do transiente. A Figura 10 ilustra esse processo, destacando o sinal original com 3.334 amostras e a região segmentada utilizada na modelagem. A área marcada no gráfico indica o intervalo selecionado para análise.

Figura 10- Região de interesse do sinal.



Fonte: Do autor (2025).

Após a segmentação, cada janela de sinal passou a conter 700 amostras, o que reduziu significativamente a complexidade computacional do processamento e da modelagem dos algoritmos, visto que antes o volume de dados era quase cinco vezes maior. Por fim, o sinal segmentado foi apresentado aos algoritmos de aprendizado de máquina para estimarem a localização da falta na linha de transmissão, otimizando o desempenho dos modelos na predição da distância do ponto de falta.

4.7 Técnicas de regressão

Para realizar a estimativa das faltas, foram testados os seguintes regressores: KNN, MLP, MLP otimizado por AG, Decision Tree, Random Forest e Extreme Gradient Boosting - XGBoost.

Os dados de entrada com as informações relevantes das características das faltas foram normalizados para garantir que todas as variáveis tenham a mesma escala, evitando que atributos com valores numéricos maiores dominem a modelagem dos algoritmos (Santos et al., 2024). Para isso, essa normalização ajusta os dados para que tenham média zero e desvio

padrão igual a 1, o que facilita o treinamento, além de contribuir para uma melhor convergência dos modelos e estabilidade numérica, principalmente quando há grande variação na escala dos atributos. Assim, a normalização aplicada é dada pela equação (10):

$$x_{\text{Normalizado}} = \frac{x - \mu}{\sigma} \quad (10)$$

onde:

x é o valor original da variável de entrada; μ é a média dos valores da variável, e σ é o desvio padrão dos valores da variável.

Após a normalização, os dados foram apresentados aos modelos de regressão. Nesta etapa, diversos parâmetros foram testados e ajustados em cada regressor, incluindo número de vizinhos no KNN, arquitetura de camadas no MLP, número de árvores no Random Forest e hiperparâmetros específicos do XGBoost.

A escolha das melhores configurações foi feita de forma sistemática, utilizando técnicas como validação cruzada, ajustes de parâmetros e aplicação de otimizadores para a busca de melhores parâmetros, visando garantir a robustez dos resultados. Os desempenhos dos modelos foram comparados e as análises serão detalhadas na seção de resultados, destacando as vantagens e limitações de cada abordagem.

4.8 Métricas de Avaliação

Para validação dos resultados, foram utilizadas as métricas da Tabela 4 para avaliar o desempenho dos modelos.

Tabela 4- Métricas de avaliação para regressão.

Métrica	Definição	Fórmula
Erro Absoluto Médio <i>(Mean Absolute Error - MAE)</i>	Média das diferenças absolutas entre os valores preditos e reais.	$MAE = \frac{1}{n} \sum_{i=1}^n y_{\text{Predito}} - y_{\text{Real}} $
Erro Relativo Percentual (ER %)	Mede o erro absoluto médio relativo à faixa de variação dos valores reais, expresso em percentual.	$ER = \frac{1}{n} \sum_{i=1}^n \frac{ y_{\text{Predito}} - y_{\text{Real}} }{y_{\text{Real}}} \times 100$
Coeficiente de Determinação <i>(R squared - R²)</i>	(Quadrado da correlação) Indica a proporção da variabilidade na variável dependente que é explicada pelo modelo. Quanto mais próximo de 1, melhor o modelo se ajusta aos dados.	$R^2 = 1 - \frac{\frac{1}{n} \sum_{i=1}^n (y_{\text{Predito}} - y_{\text{Real}})^2}{\frac{1}{n} \sum_{i=1}^n (y_{\text{Real}} - \bar{y})^2}$

Fonte: (Naser; Alavi, 2020), adaptado.

Onde:

n é o número de observações no conjunto de dados; \bar{y} é a média dos valores observados; y_{Predito} é o valor predito pelo modelo, e y_{Real} é o valor real da variável dependente.

5 RESULTADOS E DISCUSSÃO

Neste capítulo serão apresentados e discutidos os resultados obtidos na simulação da localização de faltas em linhas de transmissão. Foram avaliados os seguintes métodos de regressão: K-Nearest Neighbors (KNN), Multilayer Perceptron (MLP), MLP otimizado por AG, Decision Tree (DT), Random Forest (RF) e XGBoost (XGB). Além disso, foram exploradas as seguintes técnicas de denoising: Análise de Componentes Principais (PCA), Transformada Discreta de Wavelet (DWT) e Filtro Média Móvel. Ademais, a influência de diferentes níveis de ruído (30 dB a 70 dB e sem ruído) foi avaliada, e as métricas utilizadas

para análise do desempenho foram o Erro Relativo (ER), Erro Absoluto Médio (MAE) e o coeficiente de determinação (R^2).

5.1 Análise estatística da base de dados

Para avaliar as características gerais da base de dados utilizada neste estudo, foram calculadas a média e o desvio padrão das amostras, considerando as 7920 janelas de sinais e suas respectivas 700 amostras, após denoising e segmentação. Essas métricas oferecem uma visão inicial da estabilidade e da consistência dos dados, além de servirem como referência para a análise do desempenho dos algoritmos aplicados.

A média geral obtida foi $0,02 \times 10^{-4}$, o que demonstra que os valores estão centralizados em torno de zero. Esse resultado é altamente desejável em aplicações de aprendizado de máquina, especialmente em métodos que dependem de dados balanceados e centralizados. Valores próximos de zero evitam problemas de escala e ajudam a estabilizar os cálculos durante o treinamento de modelos.

O valor do desvio padrão foi $2,74 \times 10^{-4}$, indicando que a dispersão dos valores em relação à média é relativamente baixa, um resultado positivo considerando a presença do transitório da falta, que normalmente introduz variações significativas. Isso significa que a maioria dos dados está concentrada próxima à média, resultando em uma distribuição mais homogênea. Para modelos de aprendizado de máquina, essa característica é vantajosa, pois reduz o risco de influências excessivas de valores extremos (outliers) e facilita a convergência de algoritmos de otimização, especialmente em métodos sensíveis à variação, como gradiente descendente.

5.2 Denoising do sinal

5.2.1 Análise de Componentes Principais (PCA)

Na primeira abordagem foi utilizado o PCA para selecionar e extrair características do sinal sem perda significativa de informação relevante, otimizando a representação dos dados e melhorando a confiabilidade do modelo. O método baseia-se na suposição de que os sinais de origem são não correlacionados, permitindo que a transformação ortogonal maximize a variabilidade explicada com o menor número de componentes. Isso possibilita a construção de

um modelo mais compacto e eficiente, reduzindo a complexidade computacional e eliminando redundâncias.

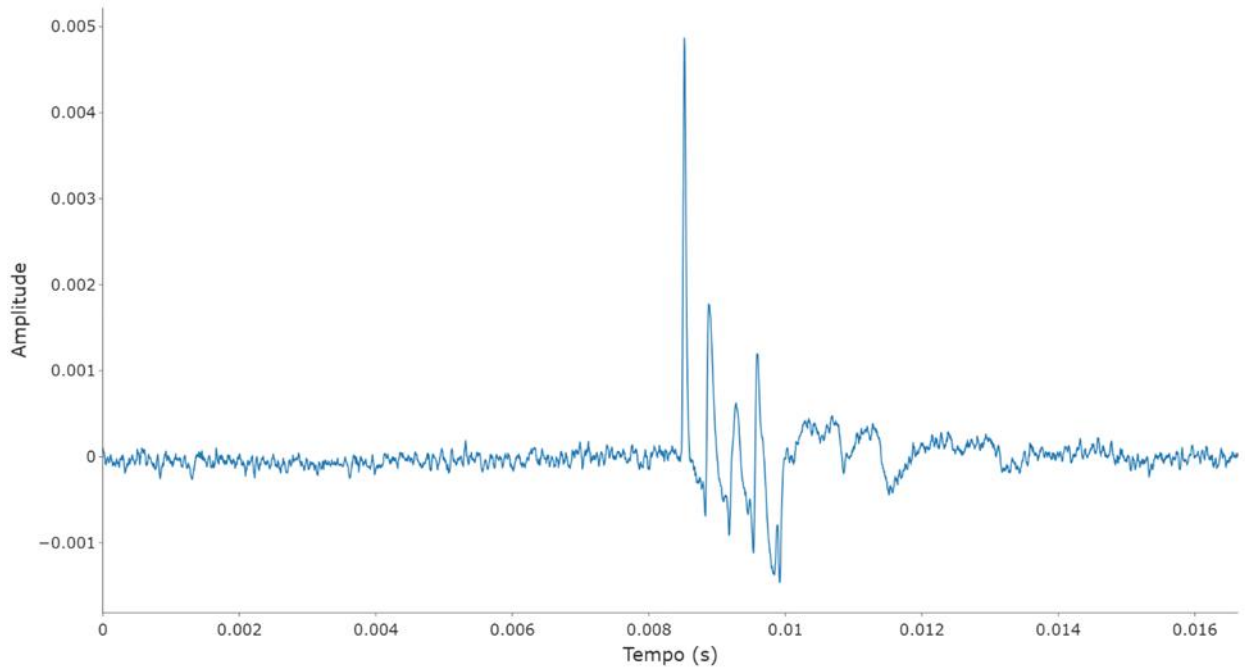
Para isso, o procedimento é dividido em três estágios. Primeiro, um vetor de atraso temporal é gerado a partir do sinal observado, criando versões deslocadas no tempo para capturar informações dinâmicas do sinal. Em seguida, aplica-se o PCA linear sobre esse vetor, extraíndo um conjunto de componentes principais, das quais a primeira é selecionada para reconstrução do sinal filtrado, pois contém a maior parte da variância do sinal original (Soother et al., 2021). No presente estudo, foram utilizadas sete componentes principais, garantindo um equilíbrio entre a redução de ruído e a preservação das características do sinal. Para isso, foi considerada a terceira saída do algoritmo SOBI, representada como $s_3[n]$, que contém as informações de transitório associadas às faltas elétricas. Em seguida, construiu-se uma matriz de versões deslocadas dessa componente, formando um vetor de atraso temporal Z , definido como na equação (10):

$$\mathbf{Z} = \begin{bmatrix} \mathbf{s}_3[n] \\ \mathbf{s}_3[n - 1] \\ \mathbf{s}_3[n - 2] \\ \mathbf{s}_3[n - 3] \\ \mathbf{s}_3[n - 4] \\ \mathbf{s}_3[n - 5] \\ \mathbf{s}_3[n - 6] \end{bmatrix} \quad (10)$$

Essa matriz representa uma estrutura multicanal do sinal, capturando as relações temporais entre as versões deslocadas. O PCA é então aplicado sobre Z , resultando em sete componentes principais, das quais a primeira é selecionada para reconstrução do sinal filtrado, pois contém a maior parte da variância do sinal original.

Além de atuar na remoção de ruído e na redução da dimensionalidade dos dados, o PCA também pode ser interpretado como um filtro FIR estatístico, no qual os coeficientes do filtro são determinados com base na variabilidade dos dados (Garza et al., 2022). Dessa forma, a transformação realizada pelo PCA permite a supressão de ruído sem comprometer as informações essenciais da falta elétrica, ajustando-se automaticamente às características do sinal analisado. O resultado desse processo pode ser observado na Figura 11.

Figura 11- Sinal filtrado pelo PCA referente a uma falta a 48 km, 20 Ohms, incidência em 180° e SNR 50dB.



Fonte: Do autor (2025).

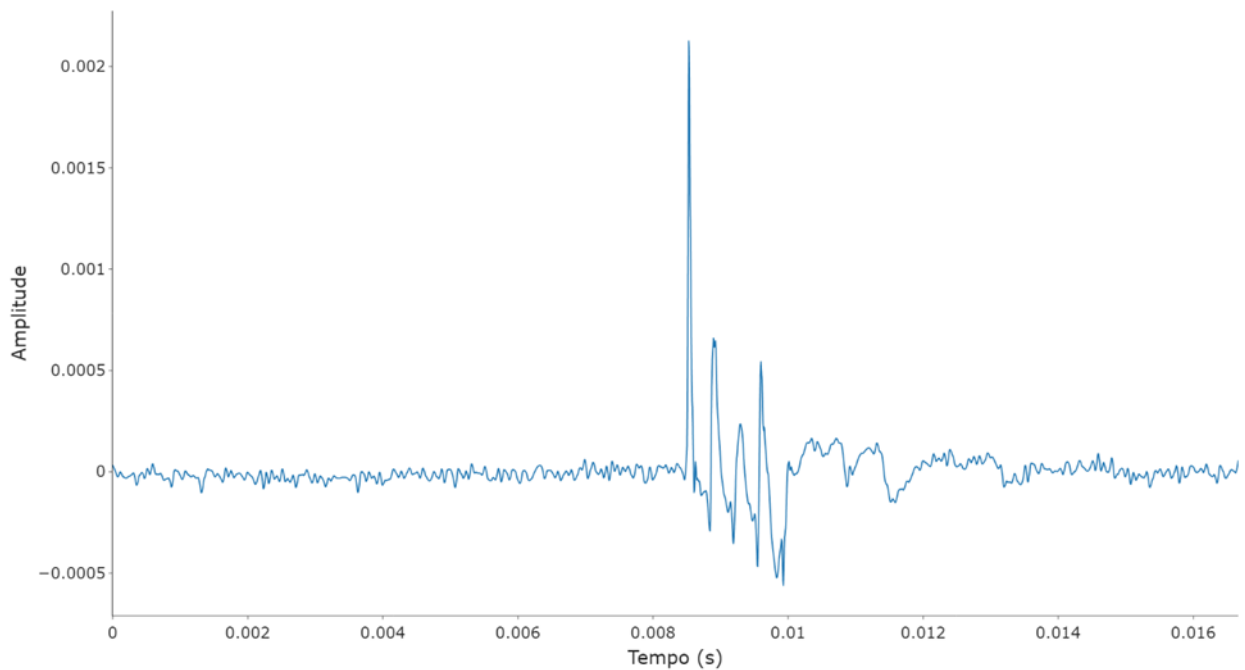
Ao comparar as Figuras 9 e 11, observa-se uma redução significativa do ruído, evidenciada pela maior clareza e suavidade do sinal após o processamento. Essa melhoria na qualidade do sinal demonstra que o PCA é uma abordagem eficaz para a remoção de ruídos, preservando as informações essenciais. Ao eliminar componentes irrelevantes e manter as características principais do sinal, o PCA prepara os dados para a aplicação de algoritmos de aprendizado de máquina, garantindo maior precisão e confiabilidade nas etapas subsequentes, que serão detalhadas mais adiante.

5.2.2 Transformada Discreta de Wavelet (DWT)

Na segunda abordagem, foi utilizada a DWT, que é uma técnica de análise de sinais que permite decompor um sinal em diferentes escalas de frequência e tempo, capturando tanto características globais quanto locais (Basheer, 2024). A DWT foi aplicada com os seguintes parâmetros: a Wavelet escolhida foi a Daubechies 4, conhecida por seu equilíbrio entre suavidade e localização temporal; o número de níveis de decomposição foi definido como 3,

permitindo uma análise detalhada sem comprometer a eficiência computacional; e o método de thresholding utilizado foi o soft thresholding, que remove o ruído de forma suave, preservando as características importantes do sinal. O resultado é apresentado na Figura 12.

Figura 12- Sinal filtrado pela Wavelet referente a uma falta a 48 km, 20 Ohms, incidência em 180° e SNR 50dB.



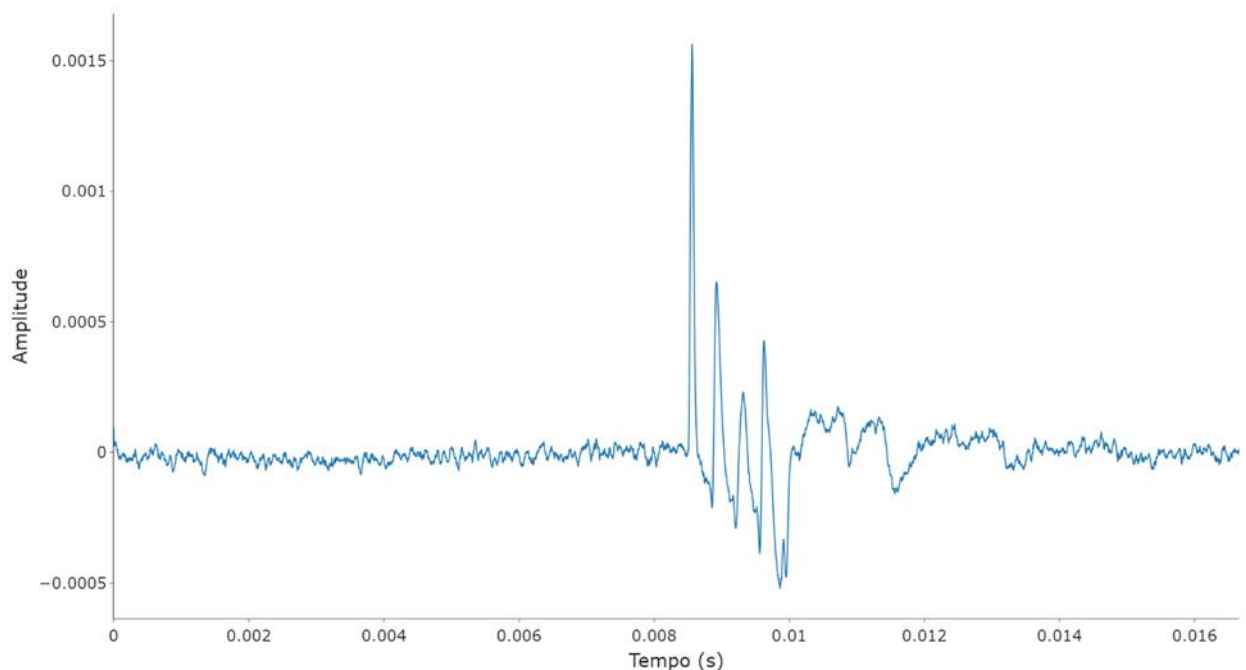
Fonte: Do autor (2025).

Ao comparar as Figuras 9 e 12, observa-se que a aplicação da DWT resultou em uma redução significativa do ruído, também evidenciada pela maior clareza e suavidade do sinal processado. Assim como o PCA, a Wavelet demonstrou ser uma abordagem eficaz para a remoção de ruídos, preservando as informações essenciais do sinal. A análise detalhada dos resultados, incluindo a comparação entre as duas técnicas em termos de eficiência na remoção de ruído e preservação das características do sinal, será apresentada na seção de resultados. Essa avaliação permitirá determinar qual método é mais adequado para a preparação dos dados, garantindo maior confiabilidade e precisão na localização de faltas em linhas de transmissão.

5.2.3 Filtro Média Móvel

O filtro de média móvel é uma técnica amplamente utilizada para suavizar sinais e reduzir o ruído presente em dados temporais. Ele funciona calculando a média de um conjunto de valores dentro de uma janela deslizante, que se move ao longo do sinal (Junior et al., 2020). No caso deste estudo, foi utilizada uma janela de 10 pontos, o que significa que, para cada ponto do sinal, a média foi calculada com base nos 10 pontos anteriores. Esse processo ajuda a eliminar flutuações rápidas e indesejadas, destacando as tendências principais do sinal. O resultado da aplicação do filtro de média móvel foi claramente visível (vide Figura 13), onde o sinal suavizado apresentou uma redução significativa do ruído, mantendo as características essenciais dos dados.

Figura 13- Sinal filtrado pelo Filtro Média Móvel referente a uma falta a 48 km, 20 Ohms, incidência em 180° e SNR 50dB.

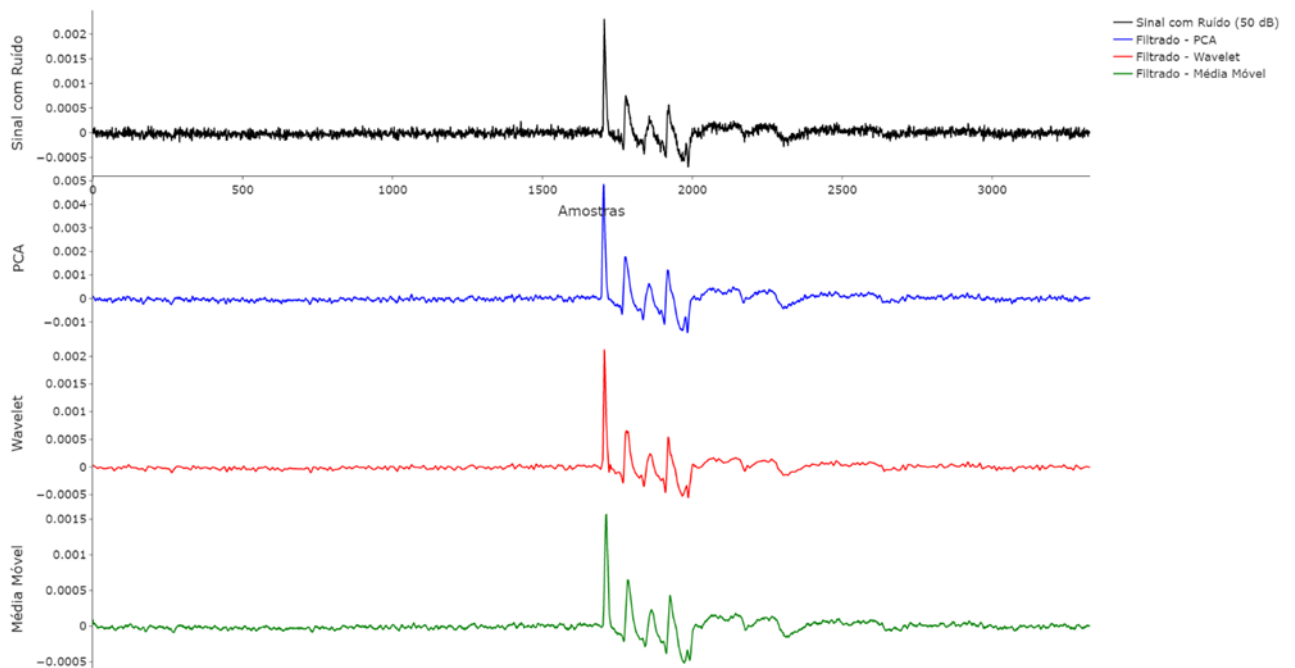


Fonte: Do autor (2025).

5.2.4 Comparação entre os processos de denoising

A fim de permitir uma comparação entre os três métodos apresentados para a remoção de ruído, o gráfico da Figura 14 apresenta os resultados obtidos do desempenho de cada abordagem. Todos os métodos, a princípio, demonstraram eficácia na extração de informações relevantes dos dados, sem remover características importantes ou apresentar distorções significativas. Na seção de resultados, será realizada uma análise quantitativa e qualitativa desses métodos, considerando métricas de desempenho e o impacto na localização das faltas.

Figura 14- Comparação entre os métodos de denoising.



Fonte: Do autor (2025).

5.3 Análise dos algoritmos após o processo de denoising com o PCA

5.3.1 *K*-Nearest Neighbors (KNN)

O regressor KNN foi avaliado utilizando diferentes valores de k e diversas métricas de distância: Euclidiana, Minkowski e Manhattan. Através do Grid Search, os melhores parâmetros identificados foram $k = 10$ (para SNR 30 dB), $k = 6$ (para SNR 40 dB), e $k = 2$ para os demais cenários, com a distância de Minkowski apresentando o melhor desempenho geral. Além disso, foi aplicada a técnica de validação cruzada K-Fold para dividir o conjunto em 10 subconjuntos, a fim de avaliar a capacidade de generalização do modelo.

A subamostragem dos sinais por um fator de 2 foi aplicada para reduzir ainda mais a dimensionalidade dos dados, resultando em 7920 janelas de sinais com 350 amostras em cada, diminuindo significativamente a complexidade computacional sem impacto relevante no desempenho do modelo. Essa abordagem demonstrou que as características mais importantes para a identificação das faltas, ainda assim, foram preservadas. Desta forma, será utilizada essa mesma base de dados para todos os algoritmos.

Os testes revelaram que o KNN obteve o melhor resultado entre todos os algoritmos implementados, atingindo excelente desempenho para sinais com SNR entre 50 dB e 70 dB, e para o sinal sem ruído, com uma degradação moderada no desempenho para 30 dB e 40 dB. Essa degradação não influencia na aplicação prática do modelo, visto que, de acordo com Tomic et. al. (2007), estes cenários são atípicos e a faixa comumente observada em sistemas elétricos reais varia entre 50 dB e 70 dB (Alcântara et al., 2019).

5.3.2 *Multilayer Perceptron* (MLP)

O MLP foi configurado variando-se parâmetros, como por exemplo, o número de elementos de processamentos (neurônios) na camada intermediária, quantidade de camadas intermediárias, número de iterações e taxas de aprendizagem. As configurações do modelo para o melhor desempenho foram: duas camadas intermediárias contendo 50 e 35 neurônios, respectivamente, treinado com o otimizador Adam, taxa de aprendizado de 0,0005 e 1000 iterações. A técnica de validação cruzada K-Fold também foi aplicada. O modelo demonstrou menor precisão em relação ao KNN para todos os cenários.

A necessidade de maior ajuste na arquitetura ou mais iterações pode ter impactado no desempenho do MLP. No entanto, isso exigiria uma maior complexidade computacional, incluindo a necessidade de um maior número de neurônios, camadas e ajustes finos de hiperparâmetros para melhorar a precisão do modelo (Alves, 2024). Esse aumento de complexidade contraria o objetivo deste trabalho, que é propor uma ferramenta simples e eficaz para estimativa de localização de faltas, capaz de operar de forma eficiente mesmo em cenários de ruído moderado a alto, com baixo custo computacional e fácil implementação em ambientes reais.

5.3.3 *Decision Tree (DT)*

Os parâmetros selecionados para o modelo DT incluem o critério de divisão que minimiza o erro quadrático médio, e o número mínimo de amostras para dividir um nó, definido como 2, permitindo que qualquer nó com pelo menos 2 amostras seja dividido. Além disso, o modelo utiliza o valor padrão para o número mínimo de amostras em um nó folha, que é 1, garantindo que cada folha tenha pelo menos uma amostra. Não foi definida uma profundidade máxima da árvore, permitindo que a árvore cresça até que todas as folhas sejam puras ou contenham menos amostras que o especificado. Além disso, nenhuma limitação foi aplicada ao número de características consideradas para divisão, o que significa que todas as características disponíveis foram usadas em cada nó.

Ao utilizar os parâmetros mencionados acima, o DT apresentou menor esforço computacional comparado a todos os algoritmos aqui apresentados, oferecendo simplicidade e flexibilidade. Em relação ao desempenho, o resultado se aproximou do MLP para níveis de ruído elevados (SNR 30 dB e 40 dB), mas mostrou maior eficácia para menores intensidades de ruído (60 dB, 70 dB e sinal sem ruído).

Esse comportamento indica que o DT é uma solução simples e eficaz para cenários de baixa intensidade de ruído. No entanto, sua performance menos robusta em sinais com maior intensidade de ruído limita sua aplicação em cenários mais desafiadores. Mesmo assim, a simplicidade e o baixo custo computacional tornam o modelo uma opção viável para situações específicas. Vale ressaltar que os parâmetros podem ser ajustados para melhorar a generalização e o desempenho do modelo em diferentes cenários.

5.3.4 *Random Forest (RF)*

O modelo de regressão RF foi configurado com 100 árvores, o que permite que múltiplos estimadores sejam combinados para melhorar a robustez e a generalização das previsões. O número de árvores utilizado foi escolhido por sua capacidade de equilibrar precisão e eficiência computacional, enquanto os outros parâmetros seguiram os valores padrão, como o uso do erro quadrático médio para minimizar a perda durante o treinamento. Todas as características disponíveis foram consideradas em cada divisão, maximizando a capacidade do modelo de capturar padrões complexos nos dados. Essa abordagem torna o Random Forest uma solução altamente robusta e eficaz, especialmente para lidar com ruído e outliers, ao mesmo tempo em que oferece previsões mais estáveis ao combinar os resultados de várias árvores.

Quanto aos resultados, o RF apresentou desempenho superior ao DT e ao MLP para quase todas as SNRs. A robustez do modelo em condições de ruído elevado (SNR 30 dB) foi evidente, com um ER de 10,99%, significativamente melhor que o DT (14,63%) e o MLP (14,38%). Para condições com menor intensidade de ruído, como SNR 50 dB, 60 dB, 70 dB e sinal sem ruído, o RF manteve desempenho competitivo. Ajustes adicionais nos parâmetros relacionados à profundidade das árvores ou ao número mínimo de amostras podem ser explorados para refinar ainda mais o desempenho do modelo.

A combinação de múltiplas árvores no RF permite maior capacidade de generalização em relação ao DT, reduzindo o impacto de outliers e ruídos nos dados. Enquanto o DT é mais suscetível ao overfitting, o RF demonstra ser uma solução robusta e confiável, especialmente em cenários mais ruidosos. No entanto, essa maior robustez vem com um custo computacional mais elevado, o que pode limitar sua aplicabilidade em sistemas com restrições de processamento (Ferreira et al., 2021; Fonseca et al., 2022).

5.3.5 *Extreme Gradient Boosting (XGB)*

O modelo XGB foi configurado utilizando o booster do tipo "gbtree", uma taxa de aprendizado ajustada para 0,01 e 350 árvores para otimizar a performance. A profundidade máxima das árvores foi definida como 20, permitindo maior complexidade na representação dos dados. A subamostragem foi configurada para 80%, enquanto 70% das colunas foram

usadas em cada árvore, promovendo generalização. A configuração foi finalizada com o uso de uma semente aleatória fixa para garantir reprodutibilidade.

O XGB apresentou desempenho competitivo com o Random Forest (RF) e superior ao DT e MLP em sinais com SNR de 30 dB, evidenciando sua capacidade de lidar com cenários de ruído mais elevado. Para SNRs de 40 dB, 50 dB e 60 dB, também superou o DT, o RF e o MLP, mostrando-se mais eficiente na identificação de padrões em níveis de ruído moderado. No entanto, para SNR de 70 dB e sinal sem ruído, o XGB teve desempenho inferior ao KNN, DT e RF, superando apenas o MLP. Esse comportamento pode ser atribuído ao fato de o XGB ser mais adequado para situações em que técnicas avançadas de modelagem são necessárias, enquanto algoritmos mais simples, como KNN e DT, se beneficiam da simplicidade e da ausência de interferências em cenários de baixo ruído. Mesmo assim, o XGB demonstrou ser uma solução robusta e confiável, especialmente para ruídos elevados.

5.3.6 Tabela comparativa de ER (%) e R^2

Os resultados obtidos permitem uma análise comparativa detalhada do desempenho dos algoritmos na estimativa da localização de faltas, considerando diferentes níveis de ruído. O KNN apresentou o melhor desempenho geral, especialmente para SNRs de 50 dB ou superiores, enquanto o MLP, mesmo com ajustes de hiperparâmetros, demonstrou menor precisão e exigiu maior complexidade computacional. O DT, embora eficiente e de baixo custo computacional, teve desempenho inferior em cenários com maior ruído. Já o RF mostrou-se mais robusto que o DT e o MLP, mantendo uma boa capacidade de generalização mesmo em condições adversas. O XGB destacou-se em sinais com ruído moderado a alto, superando os demais modelos, com exceção do KNN, em SNRs de 30 dB a 60 dB, mas apresentou uma leve degradação para sinais com menor ruído. A Tabela 5 sintetiza esses resultados, apresentando o Erro Relativo Percentual (ER%) de cada modelo para os diferentes níveis de SNR avaliados, permitindo uma comparação objetiva da precisão de cada abordagem.

Tabela 5- Comparação da Métrica ER (%).

SNR (dB)	KNN	MLP	DT	RF	XGB
30	9,35	14,69	14,63	10,99	10,13
40	2,53	6,55	5,62	4,73	4,39
50	0,49	2,59	2,70	2,29	1,95
60	0,17	1,55	1,12	1,13	1,06
70	0,14	1,25	0,66	0,70	0,72
SR	0,17	1,81	0,46	0,60	0,75

Fonte: Do autor (2025).

O R^2 indica o quanto o modelo representa os dados e mede a qualidade geral do ajuste. Nesse sentido, observando a Tabela 6, percebe-se que o KNN alcançou os maiores valores de R^2 (próximos de 1) em 50 dB ou mais, indicando que o modelo ajusta bem os dados. O MLP superou o DT em todas as SNRs, teve melhor ajuste que RF e XGB em 50 dB, e empatou com XGB em 60 dB. A DT teve R^2 significativamente menor em 30 dB, o que mostra que o modelo explica menos a variância dos dados em cenários ruidosos. O RF e XGB apresentaram valores de R^2 mais estáveis, mostrando robustez no ajuste dos dados mesmo em baixa SNR.

Tabela 6- Comparação da Métrica R^2 .

SNR (dB)	KNN	MLP	DT	RF	XGB
30	0,712	0,464	0,331	0,668	0,717
40	0,928	0,841	0,759	0,883	0,909
50	0,993	0,977	0,898	0,951	0,973

60	0,998	0,991	0,957	0,982	0,991
70	0,998	0,993	0,978	0,991	0,995
SR	0,998	0,988	0,987	0,994	0,996

Fonte: Do autor (2025).

A Tabela 7 exibe uma comparação geral entre os regressores em termos de vantagens e desvantagens, bem como os principais parâmetros que foram selecionados para cada modelo. Essa visão geral permite compreender como cada modelo se comportou e fornece uma base para a escolha da melhor abordagem para a localização de faltas em LTs.

Tabela 7- Vantagens e desvantagens de cada método.

Método	Principais parâmetros	Vantagens	Desvantagens
KNN	K: 2, 10 (30dB) e 4 (40dB)	Melhor desempenho em todas as SNRs.	Lento na inferência.
MLP	2 camadas intermediárias (50,35) Iterações: 1000	Bom desempenho para 50dB, comparado com o DT. Rápida inferência.	Lento no treinamento. Requer ajustes de parâmetros.
DT	Divisão do nó: 2 amostras Folha: 1 amostra	Rápido no treinamento e na inferência. Apresentou bom desempenho para situações acima de 60dB.	Requer ajustes de parâmetros.
RF	100 árvores Divisão do nó: 2 amostras Folha: 1 amostra	Apresentou bom desempenho para situações acima de 60dB. Rápida inferência.	Requer ajustes de parâmetros. Lento no treinamento.
XGB	350 árvores Profundidade: 20 Regularização	Apresentou bom desempenho para situações acima de 60dB. Rápida inferência.	Requer ajustes de parâmetros. Lento no treinamento.

Fonte: Do autor (2025).

Vale ressaltar que foi realizada uma análise detalhada, tanto qualitativa quanto quantitativa, por meio de diversos testes investigativos, incluindo a variação no número de

réplicas do algoritmo SOBI, subamostragem com fatores que variaram de 4 a 12, e ajustes nos parâmetros dos algoritmos utilizados. Esses experimentos foram conduzidos com o objetivo de otimizar o desempenho e a precisão dos métodos propostos. No entanto, apesar dos esforços e da investigação, não foram observadas melhorias expressivas ou ganhos significativos nos resultados. Diante disso, os resultados apresentados neste trabalho refletem as condições e configurações que demonstraram maior estabilidade e confiabilidade, mesmo que dentro de limites computacionais considerados modestos.

5.3.7 Desempenho computacional dos regressores

A comparação do tempo de pré-processamento, treinamento e inferência é essencial para avaliar a viabilidade computacional dos modelos, especialmente em aplicações em tempo real. No entanto, como o pré-processamento, incluindo o denoising com o PCA, foi o mesmo para todos os modelos, essa etapa não foi analisada.

Analisando a Tabela 8, é possível verificar que o KNN não requer tempo de treinamento, mas apresentou inferência mais lenta. Por outro lado, o MLP teve tempo moderado de treinamento e inferência quase instantânea, enquanto o DT foi o mais eficiente em ambas as etapas. Já o RF e o XGB exigiram tempos mais longos de treinamento devido à complexidade de suas múltiplas árvores, mas mantiveram inferência relativamente rápida. Esses resultados reforçam a necessidade de equilibrar precisão e custo computacional, garantindo modelos eficazes e viáveis para implementação prática.

Tabela 8- Desempenho computacional dos modelos.

Tempo (s)	KNN	MLP	DT	RF	XGB
Treinamento	-	27,263	2,467	839,182	474,103
Inferência	16,370	0,002	0,001	0,047	0,105

Fonte: Do autor (2025).

5.4 Comparação entre as técnicas de denoising aplicadas

Como o KNN apresentou o melhor desempenho entre os modelos testados, optou-se por utilizá-lo nos testes com os filtros de denoising, juntamente com a árvore de decisão (DT), que também demonstrou baixa complexidade computacional. Portanto, embora refinamentos nos modelos possam superar o KNN, o objetivo deste estudo é propor uma abordagem simples e eficiente. Assim, os testes com a Wavelet e o filtro de Média Móvel foram realizados apenas com o KNN e o DT, permitindo uma comparação direta do impacto dos diferentes processos de denoising neste estudo.

Para isso, a Tabela 9 apresenta a comparação do ER% entre os três métodos de remoção de ruído aplicados. Observa-se que o KNN manteve melhor desempenho em relação ao DT, independentemente do método de denoising. Para o KNN, o Média Móvel demonstrou maior robustez em cenários com alto nível de ruído, apresentando o menor erro para 30 dB e 40 dB, enquanto a Wavelet e/ou o PCA obtiveram melhor desempenho em níveis de ruído moderado e baixo. No caso do DT, o comportamento foi mais variável. O Média Móvel apresentou menor erro apenas em 30 dB, enquanto nos demais cenários seu desempenho foi inferior ou intermediário em relação ao PCA e à Wavelet.

O fato de o Média Móvel ter apresentado melhor desempenho para o maior nível de intensidade de ruído (30dB) em ambos os algoritmos, pode ser explicado pela natureza do filtro, que suaviza o sinal ao calcular a média das amostras vizinhas, como já mencionado na seção 4.4.3. Nesse sentido, em cenários com alto nível de ruído, essa suavização reduz a variação indesejada, tornando-o mais eficaz. No entanto, à medida que o nível de ruído diminui, o Média Móvel pode comprometer detalhes importantes do sinal, reduzindo sua capacidade de preservar características essenciais, o que prejudica o desempenho dos regressores. Para mais, de maneira geral, os resultados demonstram que a escolha do método de remoção de ruído impacta diretamente a precisão dos algoritmos de localização de falta.

Tabela 9- Comparação do ER (%) entre as técnicas de denoising.

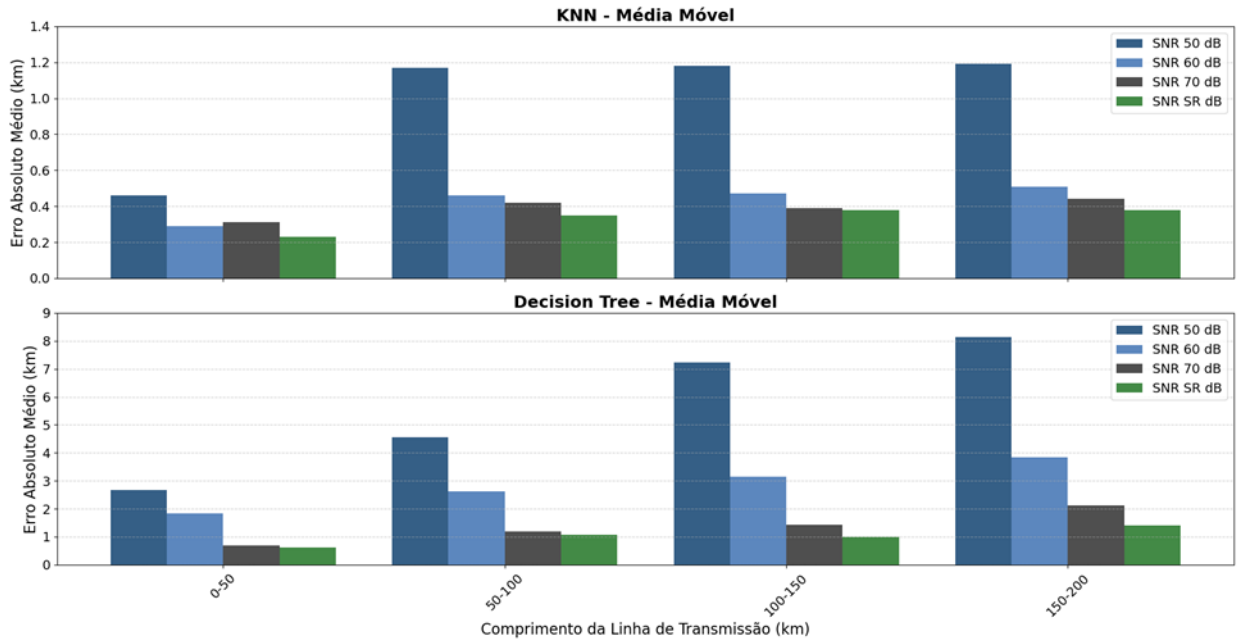
SNR (dB)	KNN			DT		
	PCA	Wavelet	Média Móvel	PCA	Wavelet	Média Móvel
30	9,35	9,27	8,57	14,63	15,07	14,62
40	2,53	2,48	2,28	5,62	6,60	6,85
50	0,49	0,45	0,48	2,70	3,43	2,81
60	0,17	0,19	0,25	1,12	1,41	1,41
70	0,14	0,17	0,23	0,66	0,78	0,68
SR	0,17	0,15	0,16	0,46	0,41	0,50

Fonte: Do autor (2025).

5.5 Comparação entre o KNN e o DT

O gráfico da Figura 15 apresenta uma comparação entre os métodos KNN e DT ao longo do comprimento da LT, considerando valores de SNR entre 50 dB e 70 dB, além do sinal sem ruído. Consolidando as análises anteriores, observa-se que o KNN apresentou melhor desempenho comparado ao DT para todas as SNRs. Além disso, é possível notar que, na maioria dos casos, quanto maior a influência do ruído, maior o MAE em quilômetros (km). Outra observação importante é que à medida que o comprimento da LT aumenta, o erro também tende a crescer. Esse comportamento pode ser atribuído à influência de fatores como a atenuação do sinal, interferências eletromagnéticas e a dispersão causada pela propagação em longas distâncias, que são amplificadas na presença de ruído.

Figura 15- Comparação ao longo do comprimento da LT.



Fonte: Do autor (2025).

6 CONCLUSÃO

O KNN destacou-se como o algoritmo com o melhor desempenho, obtendo os menores valores de ER em todos os níveis de SNR, especialmente em condições de baixo ruído (50 dB ou mais), demonstrando alta precisão na localização de faltas independente da técnica de denoising aplicada. O RF e o XGB apresentaram desempenhos intermediários, com o XGB exibindo leve vantagem em condições de maior intensidade de ruído, o que pode ser atribuído à sua maior capacidade de generalização e redução da variância.

De acordo com Tomic (2007), sinais reais medidos em sistemas elétricos geralmente apresentam relações sinal-ruído (SNR) entre 50 e 70 dB, sendo essa a faixa mais relevante para análise prática. Nesse contexto, os algoritmos testados – KNN, DT, RF e XGBoost – demonstraram resultados satisfatórios, com ERs significativamente baixos. O menor valor de ER foi alcançado pelo KNN, apresentando apenas 0,14% em 70dB, o que equivale a uma imprecisão de 0,28 km em uma linha de transmissão de 200 km, como foi simulada. Por outro lado, o maior valor de ER nessa faixa de SNR foi registrado pelo DT, com 2,7%, correspondendo a 5,4 km de desvio na mesma linha. Esses resultados evidenciam a eficácia

dos métodos em condições de ruído moderado a baixo, reforçando sua aplicabilidade em sistemas reais.

Embora o MLP tenha demonstrado potencial, seu desempenho ficou aquém dos algoritmos em todos os cenários analisados. A arquitetura de redes neurais pode oferecer vantagens em conjuntos de dados mais complexos, mas isso exige ajustes refinados e elevado esforço computacional, dificultando, inclusive, a simulação de um modelo de deep learning neste trabalho. Além disso, ao tentar implementar um otimizador por algoritmo genético ao MLP, este não atingiu a convergência devido, também, ao elevado esforço computacional. Nesse sentido, limitações relacionadas à memória restringiram os testes desses algoritmos em sua máxima eficiência, o que talvez tenha comprometido o potencial desempenho do MLP, bem como de sua otimização. Assim, os resultados ressaltam que, para aplicações práticas e cenários que demandam soluções rápidas e computacionalmente eficientes, algoritmos como KNN, RF e XGB são opções robustas e adequadas.

A comparação entre os métodos de remoção de ruído se fez importante, visto que a escolha do filtro impacta diretamente no desempenho dos modelos de aprendizado de máquina. Esta análise indicou que o Média Móvel demonstrou maior robustez para cenários de alto ruído, apresentando o menor erro em 30 dB, enquanto o PCA e/ou a Wavelet foram mais eficazes à medida que o nível de ruído diminuía, preservando melhor as características do sinal. No geral, a Wavelet obteve um equilíbrio entre remoção de ruído e preservação de informações relevantes, sendo uma alternativa viável para diferentes cenários em que não há alta interferência de ruído. Em termos de complexidade, o Média Móvel é simples de implementar e tem baixa complexidade computacional. O PCA exige mais processamento devido à decomposição estatística dos dados. Já a Wavelet é a mais complexa, pois requer transformações multiescala e ajustes finos para otimizar a filtragem sem comprometer os transientes do sinal. Assim, essas diferenças evidenciam a importância da escolha adequada do método de denoising, considerando o nível de ruído do sistema e o algoritmo de regressão utilizado, a fim de garantir um compromisso entre precisão e eficiência computacional.

Outra observação relevante foi a relação direta entre o comprimento da linha de transmissão e o aumento do erro nos algoritmos. Quanto maior a distância, maior o erro de localização, devido à atenuação do sinal e às interferências acumuladas ao longo da propagação. Isso quer dizer que faltas mais próximas do terminal tendem a ser localizadas com maior precisão, já que o sinal sofre menos degradação em distâncias curtas.

Para trabalhos futuros, refinamentos nos processos de denoising podem ser explorados, visto que as três abordagens testadas demonstraram potencial na remoção de ruído sem comprometer informações relevantes para o aprendizado de máquina. Além disso, a convergência do MLP otimizado por AG pode ser alcançada por meio do aumento do escopo da otimização e da exploração de técnicas alternativas, como o Tree-based Pipeline Optimization Tool (TPOT), além da incorporação de métodos de regularização para melhorar a estabilidade e o desempenho do modelo. A otimização de hiperparâmetros nos modelos testados, utilizando abordagens automatizadas como Bayesian Optimization, também pode contribuir para um melhor ajuste dos algoritmos. Ademais, a aplicação de técnicas de deep learning, como redes neurais convolucionais (CNNs) ou redes recorrentes (RNNs), pode ser investigada, dada sua capacidade de extrair padrões temporais complexos dos sinais elétricos. O uso dessas técnicas pode melhorar a generalização dos modelos em diferentes condições operacionais, permitindo maior precisão na estimativa da localização de faltas. Por fim, a implementação eficiente dos métodos em hardware especializado, como GPUs, pode viabilizar a aplicação em tempo real, garantindo maior eficiência computacional sem comprometer a precisão dos resultados.

REFERÊNCIAS

- ABBASI, A. R. **Fault detection and diagnosis in power transformers: a comprehensive review and classification of publications and methods.** *Electric Power Systems Research*, v. 209, p. 107990, 2022. DOI: <https://doi.org/10.1016/j.epsr.2022.107990>.
- ADEOLE, A. C.; R. TZONEVA; S. BEHARDIEN. **Distribution network fault section identification and fault location using wavelet entropy and neural networks.** *Applied Soft Computing*, v. 46, p. 296–306, 2016. DOI: <https://doi.org/10.1016/j.asoc.2016.05.013>.
- ADLY, A. R. et al. **A novel protection scheme for multi-terminal transmission lines based on wavelet transform.** *Electric Power Systems Research*, v. 183, p. 106286, 2020. DOI: <https://doi.org/10.1016/j.epsr.2020.106286>.
- AFTAB, M. A. et al. **Dynamic protection of power systems with high penetration of renewables: A review of the traveling wave based fault location techniques.** *International Journal of Electrical Power & Energy Systems*, v. 114, p. 105410, 2020. DOI: <https://doi.org/10.1016/j.ijepes.2019.105410>.
- AHMED, S. I. et al. **Deep neural network based fault classification and location detection in power transmission line.** In: 2022 12th International Conference on Electrical and Computer Engineering (ICECE), 2022, p. 252–255. DOI: [10.1109/ICECE57408.2022.10088794](https://doi.org/10.1109/ICECE57408.2022.10088794).
- ALCÂNTARA, I. F. P.; B. H. G. BARBOSA; D. D. FERREIRA. **Detecção de Distúrbios em Qualidade de Energia usando modelos NARX Neurais.** Simpósio Brasileiro de Automação Inteligente (SBAI), 2019.
- ALENCAR, G. T. et al. **A new robust approach for fault location in transmission lines using single channel independent component analysis.** *Electric Power Systems Research*, v. 220, p. 109281, 2023. DOI: <https://doi.org/10.1016/j.epsr.2023.109281>.
- ALENCAR, G. T.; R. C. DOS SANTOS; A. NEVES. **Signal Processing Technologies Applied in Smart Grids.** In: *Smart Grids—Renewable Energy, Power Electronics, Signal Processing and Communication Systems Applications*. Cham: Springer, 2023. p. 299–323.
- ALMEIDA, A. O. et al. **ICA feature extraction for the location and classification of faults in high-voltage transmission lines.** *Electric Power Systems Research*, v. 148, p. 254–263, 2017. DOI: <https://doi.org/10.1016/j.epsr.2017.03.030>.
- ALVES, L. da S. **Otimização de parâmetros de redes neurais do tipo feedforward com algoritmos meta-heurísticos.** 2024.
- ANDREY, M.; S. E. M. DE OLIVEIRA. **Compensação da Saturação de Transformadores de Corrente Aplicados à Proteção de Distância Utilizando Redes Neurais Artificiais.** 2019.

- AZEROUAL, M. et al. **Fault location and detection techniques in power distribution systems with distributed generation: Kenitra city (Morocco) as a case study.** *Electric Power Systems Research*, v. 209, p. 108026, 2022. DOI: <https://doi.org/10.1016/j.epsr.2022.108026>.
- BARBOSA, G. V. D. **Técnicas de Localização de Falhas em Linhas de Transmissão Empregando o Método dos Mínimos Quadrados e o Método de Takagi: Análise Comparativa.** 2018.
- BARMAN, P. K. et al. **A statistical feature based transmission fault identification and classification.** In: 2022 IEEE 2nd International Conference on Sustainable Energy and Future Electric Transportation (SeFeT), 2022, p. 1–6. DOI: 10.1109/SeFeT55524.2022.9908816.
- BASHEER, B. G. et al. **Fault classification and localization in microgrids: Leveraging discrete wavelet transform and multi-machine learning techniques considering single point measurements.** *Electric Power Systems Research*, v. 231, p. 110362, 2024. DOI: <https://doi.org/10.1016/j.epsr.2024.110362>.
- BELAGOUNE, S. et al. **Deep learning through LSTM classification and regression for transmission line fault detection, diagnosis and location in large-scale multi-machine power systems.** *Measurement*, v. 177, p. 109330, 2021. DOI: <https://doi.org/10.1016/j.measurement.2021.109330>.
- BHATNAGAR, M.; A. YADAV. **Fault Detection and Classification in Transmission Line Using Fuzzy Inference System.** In: 2020 5th IEEE International Conference on Recent Advances and Innovations in Engineering (ICRAIE), 2020, p. 1–6. DOI: 10.1109/ICRAIE51050.2020.9358386.
- BRAGA, A.; A. C. P. DE LEON FERREIRA; T. B. LUDERMIR. **Redes neurais artificiais: teoria e aplicações.** Rio de Janeiro: LTC, 2007.
- CARDOSO, L. A. et al. **Método de Localização de Falhas em Sistemas de Potência baseado em Análise de Componentes Independentes.** In: Anais da XIV Conferência Brasileira sobre Qualidade da Energia Elétrica, 2021.
- CERVI, M. et al. **A traveling wave based fault location method using unsynchronized current measurements.** In: 2019 IEEE Power & Energy Society General Meeting (PESGM), 2019, p. 1–1. DOI: 10.1109/PESGM40551.2019.8974132.
- CHAVAN, J. N. et al. **Transmission Line Fault Detection Using Wavelet & Artificial Neural Network.** In: 2022 Interdisciplinary Research in Technology and Management (IRTM), 2022, p. 1–5. DOI: 10.1109/IRTM54583.2022.9791621.
- COBAN, M. S.; S. S. TEZCAN. **Artificial neural network based fault location on 230 kV transmission line using voltage and current signals.** In: 2020 4th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), 2020, p. 1–4. DOI: 10.1109/ISMSIT50672.2020.9255337.

COSTA, S. F. **Detecção de Fraude em Seguros de Saúde: Aplicação de Árvores de Decisão, K-Médias, Redes Sociais**. 2023. Dissertação (Mestrado em Análise e Engenharia de Big Data) – Universidade NOVA de Lisboa, 2023.

CUI, L. et al. **The study of comprehensive transmission line fault location method for smart substations**. In: *2016 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC)*, 2016, p. 825–829. DOI: 10.1109/APPEEC.2016.7779611.

DASHTI, R. et al. **A survey of fault prediction and location methods in electrical energy distribution networks**. *Measurement*, v. 184, p. 109947, 2021. DOI: <https://doi.org/10.1016/j.measurement.2021.109947>.

FAN, R. et al. **Transmission line fault location using deep learning techniques**. In: 2019 North American Power Symposium (NAPS), 2019, p. 1–5. DOI: 10.1109/NAPS46351.2019.9000224.

FAYAZI, M. et al. **A single-ended traveling wave based fault location method using DWT in hybrid parallel HVAC/HVDC overhead transmission lines on the same tower**. *Electric Power Systems Research*, v. 220, p. 109302, 2023. DOI: <https://doi.org/10.1016/j.epsr.2023.109302>.

FELLIPE, M. S. et al. **Localização de Falhas em Linhas de Transmissão pelo Método das Componentes Fundamentais**. *Simpósio Brasileiro de Sistemas Elétricos (SBSE)*, v. 1, n. 1, 2020.

FERREIRA, D. D. et al. **Method based on independent component analysis for harmonic extraction from power system signals**. *Electric Power Systems Research*, v. 119, p. 19–24, 2015. DOI: <https://doi.org/10.1016/j.epsr.2014.09.001>.

FERREIRA, L. O. et al. **Detecção de falhas mecânicas através de diferentes técnicas de machine learning**. *Observatório de La Economía Latinoamericana*, v. 22, n. 9, p. e6780, 2024.

FILHO, F. M. G. **Aprendizado de máquina aplicado à predição de viabilidade técnica para conexão de cargas em rede de distribuição de média tensão**. 2022.

FONSECA, G. A. **Classificação de faltas em linhas de transmissão utilizando métodos de aprendizado de máquina**. 2021. Dissertação (Mestrado em Engenharia de Sistemas e Automação) – Universidade Federal de Lavras, 2022.

FONSECA, G. A. et al. **Fault classification in transmission lines using random forest and notch filter**. *Journal of Control, Automation and Electrical Systems*, 2021. DOI: <https://doi.org/10.1007/s40313-021-00844-4>.

GARCIA, J. D. et al. **High impedance fault modeling and location for transmission line**. *Electric Power Systems Research*, v. 196, p. 107202, 2021. DOI: <https://doi.org/10.1016/j.epsr.2021.107202>.

GARZA-ABDALA, J. A. et al. **Interfaz cerebro-computadora para la codificación de clave morse mediante señales cerebrales.** *Pädi Boletín Científico de Ciencias Básicas e Ingenierías del ICBI*, v. 10, n. Especial4, p. 196-203, 2022.

GAUR, V. K. et al. **New ground fault location method for three-terminal transmission line using unsynchronized current measurements.** *International Journal of Electrical Power & Energy Systems*, v. 135, p. 107513, 2022. DOI: <https://doi.org/10.1016/j.ijepes.2021.107513>.

GOPALA RAO, G. T. S. et al. **Fault detection and localization for overhead 1-kV distribution lines with direct and indirect measurements.** In: 2022 IEEE 2nd Mysore Sub Section International Conference (MysuruCon), 2022, p. 1–6. DOI: 10.1109/MysuruCon55714.2022.9972678.

GURURAJAPATHY, S. et al. **Fault location and detection techniques in power distribution systems with distributed generation: A review.** *Renewable and Sustainable Energy Reviews*, v. 74, p. 949–958, 2017. DOI: <https://doi.org/10.1016/j.rser.2017.03.021>.

HERNANDEZ, E. D. M. **Inteligência computacional e redes neurais em engenharia elétrica.** PSI-2222. *Práticas de Eletricidade e Eletrônica II*. EPUSP, 2005.

HOLANDA, J. L. D. et al. **Análise de sensibilidade de fatores que influenciam o desempenho de métodos de localização de faltas em linhas de transmissão.** 2020.

ILIAYEAFAR, M. M.; A. HADAEGHI. **Extreme learning machine-based fault location approach for terminal-hybrid LCC-VSC-HVDC transmission lines.** *Electric Power Systems Research*, v. 221, p. 109487, 2023. DOI: <https://doi.org/10.1016/j.epsr.2023.109487>.

JUNIOR, A. K. T.; C. GRANDE. **Algoritmo Adaptativo para Cancelamento de Ruídos em Sinais Bioelétricos Baseado em Rastreadores de Fase e Média Móvel.** 2020. Dissertação (Mestrado em Engenharia Elétrica) – Universidade Federal de Mato Grosso do Sul, 2020.

KARPIEL, I.; Z. DRZAZGA. **Study of Sensimotor Cortex in Multiple Sclerosis (MS) Using Independent Component Analysis (ICA) in GIFT Toolbox with Infomax Algorithm.** In: *Innovations in Biomedical Engineering*. Springer, 2021. p. 337–350.

KOKER, C. **Validation of independent components using a hypothesis testing approach.** 2020. Dissertação – Stellenbosch University, 2020.

KUMARI, S. et al. **Fault Detection in Transmission Line Using ANN.** In: 2023 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS), 2023, p. 1–5. DOI: 10.1109/SCEECS57921.2023.10063045.

LATHA, A. H.; R. BHIMASINGU. **A new transmission line fault location identification using fault current DC transients.** In: 2021 IEEE 18th India Council International Conference (INDICON), 2021, p. 1–6. DOI: 10.1109/INDICON52576.2021.9691611.

- LIU, J. et al. **High-precision identification of power quality disturbances under strong noise environment based on FastICA and random forest**. IEEE Transactions on Industrial Informatics, v. 17, n. 1, p. 377–387, 2020.
- LIU, Y. et al. **Model-based transmission line fault location methods: A review**. International Journal of Electrical Power & Energy Systems, v. 153, p. 109321, 2023. DOI: <https://doi.org/10.1016/j.ijepes.2023.109321>.
- LOPES, F. V. et al. **Análise de Ondas Viajantes em Linhas de Transmissão de Interconexão de Fontes Interfaceadas por Conversores: Um Estudo Sobre Aplicações de Proteção e Localização de Faltas**. 2022.
- LU, Y. et al. **An improved noniterative parameter-free fault location method on untransposed transmission lines using multi-section models**. IEEE Transactions on Power Delivery, v. 37, n. 3, p. 1356–1369, 2022. DOI: 10.1109/TPWRD.2021.3085746.
- MAMPILLY, B. J.; S. V. S. **Transmission lines fault detection using empirical mode decomposition in a grid-connected power system**. In: 2020 International Conference on Power Electronics and Renewable Energy Applications (PEREA), 2020, p. 1–6. DOI: 10.1109/PEREA51218.2020.9339814.
- MANASSERO JUNIOR, G. **Proteção de linhas de transmissão utilizando a técnica de alinhamento temporal dinâmico**. 2023. Dissertação – Universidade de São Paulo, 2023.
- MARTI, J. R. **Accurate modelling of frequency-dependent transmission lines in electromagnetic transient simulations**. IEEE Transactions on Power Apparatus and Systems, v. PAS-101, n. 1, p. 147–157, 1982. DOI: 10.1109/TPAS.1982.317332.
- MEDEIROS, S. A. **Política pública de acesso aberto à produção científica: o caso do Repositório Institucional da Universidade Federal de Lavras**. 2013. 258 p. Dissertação (Mestrado em Administração Pública) – Universidade Federal de Lavras, 2013.
- MIRZAEI, M. et al. **Fault location on a series-compensated three-terminal transmission line using deep neural networks**. IET Science, Measurement & Technology, v. 12, p. 746–754, 2018. DOI: <https://doi.org/10.1049/iet-smt.2018.0036>.
- MORAES, M. A. **Uma proposta para a atribuição de responsabilidades sobre desequilíbrios de tensão através da análise de componentes independentes**. 2022. Tese (Doutorado em Engenharia Elétrica) – Universidade Federal de Uberlândia, 2022. DOI: <http://doi.org/10.14393/ufu.te.2022.5040>.
- MORADZADEH, A. et al. **Hybrid CNN-LSTM approaches for identification of type and locations of transmission line faults**. International Journal of Electrical Power & Energy Systems, v. 129, p. 107563, 2021. DOI: <https://doi.org/10.1016/j.ijepes.2021.107563>.
- MUKHERJEE, A. et al. **Transmission Line Faults in Power System and the Different Algorithms for Identification, Classification and Localization: A Brief Review of Methods**. Journal of The Institution of Engineers (India): Series B, v. 102, p. 855–877, 2021. DOI: <https://doi.org/10.1007/s40031-020-00530-0>.

MUSTARIM, M. R. et al. **Fault location estimation on transmission lines using neuro-fuzzy system.** *Procedia Computer Science*, v. 163, p. 591–602, 2019. DOI: <https://doi.org/10.1016/j.procs.2019.12.141>.

NAIDU, O. D.; A. K. PRADHAN. **Precise traveling wave-based transmission line fault location method using single-ended data.** *IEEE Transactions on Industrial Informatics*, v. 17, n. 8, p. 5197–5207, 2021. DOI: [10.1109/TII.2020.3703027](https://doi.org/10.1109/TII.2020.3703027).

NAIK, S.; E. KOLEY. **Fault detection and classification scheme using KNN for AC/HVDC transmission lines.** In: 2019 International Conference on Communication and Electronics Systems (ICCES), 2019, p. 1131–1135. DOI: [10.1109/ICCES45898.2019.9002230](https://doi.org/10.1109/ICCES45898.2019.9002230).

NASER, M. Z.; A. ALAVI. **Insights into performance fitness and error metrics for machine learning.** arXiv preprint arXiv:2006.00887, 2020.

NGWENYAMA, M. et al. **Traveling wave fault location detection technique for high voltage transmission lines.** In: 2021 2nd International Conference for Emerging Technology (INCET), 2021, p. 1–7. DOI: [10.1109/INCET51464.2021.9456334](https://doi.org/10.1109/INCET51464.2021.9456334).

NIAZKAR, M. et al. **Applications of XGBoost in water resources engineering: A systematic literature review.** *Environmental Modelling & Software*, v. 174, p. 105971, 2024. DOI: <https://doi.org/10.1016/j.envsoft.2024.105971>.

OGAR, V. N. S. et al. **The use of artificial neural network for low latency of fault detection and localisation in transmission line.** *Heliyon*, v. 9, n. 2, p. e13376, 2023. DOI: <https://doi.org/10.1016/j.heliyon.2023.e13376>.

OLIVEIRA, D. R. de M. A. A. et al. **Second order blind identification algorithm with exact model order estimation for harmonic and interharmonic decomposition with reduced complexity.** *International Journal of Electrical Power & Energy Systems*, v. 125, p. 106415, 2021. DOI: <https://doi.org/10.1016/j.ijepes.2020.106415>.

OLIVEIRA, F.; F. MOREIRA; A. PICANÇO. **Accuracy analysis using the EMD and VMD for two-terminal transmission line fault location based on traveling wave theory.** *Electric Power Systems Research*, v. 224, p. 109667, 2023. DOI: <https://doi.org/10.1016/j.epsr.2023.109667>.

ORTEGA, J. S.; M. C. TAVARES. **Fault impedance analysis and non-conventional distance protection settings for half-wavelength transmission line applications.** *Electric Power Systems Research*, v. 198, p. 107361, 2021. DOI: <https://doi.org/10.1016/j.epsr.2021.107361>.

PAN, M. et al. **A review of second-order blind identification methods.** *WIREs Computational Stats*, v. 14, 2022. DOI: [10.1002/wics.1550](https://doi.org/10.1002/wics.1550).

PARIKH, U. B. et al. **Fault classification technique for series compensated transmission line using support vector machine.** *International Journal of Electrical Power & Energy Systems*, v. 32, n. 6, p. 629–636, 2010. DOI: <https://doi.org/10.1016/j.ijepes.2009.09.008>.

PARSI, M. et al. **Wavelet based fault location on power transmission lines using real-world travelling wave data**. *Electric Power Systems Research*, v. 186, p. 106261, 2020. DOI: <https://doi.org/10.1016/j.epsr.2020.106261>.

PEREIRA, T. K. et al. **Localização de Falhas em Linhas de Transmissão Baseado em Processamento Estatístico de Sinais e Aprendizado de Máquina**. 2024.

PRISCO, R. K. et al. **Localização de faltas em linhas de transmissão baseada no espectro de ondas viajantes**. 2021.

PRIYADARSHINI, E. S. G. **Local end data based fault detection technique in transmission line using DWT**. In: 2020 IEEE Students Conference on Engineering Systems (SCES), 2020, p. 1–6. DOI: 10.1109/SCES50439.2020.9236709.

PSARAS, V. et al. **HVDC grid fault location method using genetic algorithm on reconstructed frequency-domain voltage profiles**. *International Journal of Electrical Power & Energy Systems*, v. 144, p. 108429, 2023. DOI: <https://doi.org/10.1016/j.ijepes.2022.108429>.

RAVESH, N. et al. **A hybrid artificial neural network and wavelet packet transform approach for fault location in hybrid transmission lines**. *Electric Power Systems Research*, v. 204, p. 107721, 2022. DOI: <https://doi.org/10.1016/j.epsr.2021.107721>.

ROSENBERG, M. B. **Comunicação não violenta: técnicas para aprimorar relacionamentos pessoais e profissionais**. 5. ed. São Paulo: Ágora, 2021.

SANTOS, S. S. et al. **Indução de árvores de decisão oblíquas como explicadores de predições por modelos de aprendizado de máquina**. 2022. Dissertação (Mestrado em Engenharia Elétrica) – Universidade Federal de Minas Gerais, 2022.

SHAFIULLAH, M. et al. **Wavelet-based extreme learning machine for distribution grid fault location**. *IET Generation, Transmission & Distribution*, v. 11, n. 17, p. 4256–4263, 2017. DOI: <https://doi.org/10.1049/iet-gtd.2017.0656>.

SOOTHER, D. K. et al. **The Importance of Feature Processing in Deep-Learning-Based Condition Monitoring of Motors**. *Mathematical Problems in Engineering*, v. 2021, p. 1–23, 2021. DOI: <https://doi.org/10.1155/2021/9927151>.

SOWAH, R. A. et al. **Design of power distribution network fault data collector for fault detection, location and classification using machine learning**. In: 2018 IEEE 7th International Conference on Adaptive Science & Technology (ICAST), 2018, p. 1–8. DOI: 10.1109/ICASTECH.2018.8506774.

TABARI, M.; J. SADEH. **Fault location in series-compensated transmission lines using adaptive network-based fuzzy inference system**. *Electric Power Systems Research*, v. 208, p. 107800, 2022. DOI: <https://doi.org/10.1016/j.epsr.2022.107800>.

- TAVOOSI, J. et al. **A hybrid approach for fault location in power distributed networks: Impedance-based and machine learning technique.** *Electric Power Systems Research*, v. 210, p. 108073, 2022. DOI: <https://doi.org/10.1016/j.epsr.2022.108073>.
- TOMIC, J. et al. **A New Power System Digital Harmonic Analyser.** *IEEE Transactions on Power Delivery*, v. 22, n. 2, p. 772–780, 2007.
- UNGRAD, H. et al. **Protection techniques in electrical energy systems.** CRC Press, 2020.
- WANG, J. et al. **Operational modal analysis for slow linear time-varying structures based on moving window second order blind identification.** *Signal Processing*, v. 133, p. 169–186, 2017. DOI: <https://doi.org/10.1016/j.sigpro.2016.10.016>.
- WANG, Q. et al. **XGBoost algorithm assisted multi-component quantitative analysis with Raman spectroscopy.** *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, v. 323, p. 124917, 2024. DOI: <https://doi.org/10.1016/j.saa.2024.124917>.
- WU, Y.-C.; Y.-L. CHANG. **Ransomware detection on Linux using machine learning with random forest algorithm.** *IEEE TechRxiv*, 2024. DOI: <http://dx.doi.org/10.36227/techrxiv.171778770.06550236/v1>.
- XIAO, D.; J. HU. **Application of second-order blind identification in motor imagery EEG classification.** In: 2008 Fourth International Conference on Natural Computation, v. 2, 2008, p. 310–313. DOI: 10.1109/ICNC.2008.601.
- YUAN, J. et al. **Energy efficiency state identification of milling processing based on EEMD-PCA-ICA.** *Measurement*, v. 174, p. 109014, 2021. DOI: <https://doi.org/10.1016/j.measurement.2021.109014>.
- ZERAHNY, C. K. Y. et al. **Fault classification and location in three-phase transmission lines using wavelet-based machine learning.** In: 2020 8th International Conference on Intelligent and Advanced Systems (ICIAS), 2021, p. 1–5. DOI: 10.1109/ICIAS49414.2021.9642641.

APÊNDICE

O desenvolvimento do projeto desta pesquisa até o estágio atual resultou no seguinte trabalho, publicado em congresso nacional:

Localização de Faltas em Linhas de Transmissão Baseado em Processamento Estatístico de Sinais e Aprendizado de Máquina. Apresentado no XXV Congresso Brasileiro de Automática, 2024.

Resumo: Este trabalho apresenta um método de localização de faltas em Linhas de Transmissão baseado em Análise de Componentes Independentes (ICA) com ênfase no algoritmo Identificação Cega de Segunda Ordem (SOBI). O banco de dados é composto por 7920 sinais de tensão trifásicos com faltas monofásicas simuladas via ATP, a uma taxa de amostragem de 200 kHz, com diversas configurações de resistência, ângulo de incidência, localização e também com inserção de ruído, em uma linha de transmissão com comprimento de 200 km. Os sinais foram segmentados e processados pelo SOBI para serem apresentados ao algoritmo de regressão K-Nearest Neighbors (KNN) para estimar a localização da falta. Duas abordagens foram propostas: a primeira utiliza-se a matriz de separação obtida pelo SOBI fixo e a segunda aplica o SOBI a cada janela de sinal processada, gerando uma matriz de separação adaptativa. Os resultados apresentaram um erro médio relativo inferior a 1 km para algumas situações com a presença de ruído.