

## Detecção de Falhas de Alta Impedância Baseada em Boosting de Árvores de Decisão

Bernardo A. V. de Souza\* Gustavo S. R. Araújo\* Jonathan N. Gois\*  
Thiago A. Brasil\*

\* Centro Federal de Educação Tecnológica Celso Suckow da Fonseca,  
RJ, (e-mails: [bernardo.souza@aluno.cefet-rj.br](mailto:bernardo.souza@aluno.cefet-rj.br),  
[gustavo.sr.araujo@gmail.com](mailto:gustavo.sr.araujo@gmail.com), [jonathan.gois@cefet-rj.br](mailto:jonathan.gois@cefet-rj.br) e  
[thiago.brasil@cefet-rj.br](mailto:thiago.brasil@cefet-rj.br)).

---

**Abstract:** The occurrence of high impedance faults is an event that is difficult to detect in distribution lines and may cause various hazards to the population. With the conductor falling to the ground, the fault current does not sensitize the protection systems and can cause several accidents. Frequently, this type of fault is confused with other line events, further delaying the detection. The present work aims to insert simulated data of high impedance faults and several other events common to the network, to detect them using a boosting of decision trees, the XGBoost classifier. The obtained results demonstrate the accuracy of the 97.67% classification model.

**Resumo:** O acontecimento de uma falta de alta impedância é um evento de difícil detecção em linhas de distribuição, podendo acarretar em diversos perigos à população. Com a queda do condutor ao solo, a corrente de falta não sensibiliza os sistemas de proteção, podendo causar diversos acidentes. Muitas vezes, esse tipo de falta é confundido com outros eventos da rede, atrapalhando ainda mais a detecção. O presente trabalho tem por objetivo inserir dados simulados de eventos de falta de alta impedância e diversos outros eventos comuns à rede, com o intuito de detectá-los através do uso de *boosting* de árvores de decisão, através do classificador XGBoost. Os resultados obtidos demonstram a acurácia do modelo de classificação de 97,67%.

**Keywords:** High impedance fault; Classification; Decision trees; Boosting; Machine Learning;

**Palavras-chaves:** Falta de Alta Impedância; Classificação; Árvores de decisão; Boosting; Aprendizado de Máquinas;

---

### 1. INTRODUÇÃO

Faltas de alta impedância (FAI) dificilmente costumam sensibilizar as proteções convencionais presentes nas redes de distribuição, tais como fusíveis e relés de sobrecorrente. Em decorrência do rompimento do condutor de uma rede aérea de distribuição, pessoas, animais e a infraestrutura pública próximas a este ramal correm perigo iminente, devido a possibilidade de contato direto ou indireto com o condutor rompido e ainda energizado (Banner and Don Russell, 1997; Moreto, 2005). Devido a baixa magnitude de corrente estabelecida, as FAI tendem a ser confundidas com operações rotineiras de carga, tornando sua correta identificação praticamente impossível (Cabral et al., 1992).

Uma metodologia para se aprimorar a sensibilidade do sistema de proteção convencional seria a de utilização de faltas de alta impedância ensaiadas ao longo do sistema de distribuição (Elkalashy et al., 2007). Entretanto, estes testes vêm sendo evitados, visto que um risco adicional acaba sendo imposto, desnecessariamente, à rede elétrica. Desta forma, uma modelagem matemática que represente com fidelidade os mais diversos tipos de FAI mostra-se primordial no estudo da classificação e, posterior, detecção destas faltas, sem a necessidade do uso de faltas ensaiadas (Brasil et al., 2021; Araujo et al., 2021).

Recentemente, um modelo aprimorado de arco foi proposto por (Wei et al., 2020), tendo sido, inclusive, utilizado em outros trabalhos. Esta implementação inclui um fator randômico no comportamento randômico do arco elétrico, o qual possibilita reproduzir com maior fidelidade os períodos de instabilidade transitória das FAI, onde fatores tais como as condições do solo no qual é realizado o contato com o condutor rompido, sua temperatura e umidade não são adequadamente controlados.

O cerne da dificuldade na detecção de FAI reside na característica inerente das mesmas, a baixa magnitude da corrente de falta, incapaz de sensibilizar as proteções contra sobrecorrente naquele trecho. Desta forma, uma maneira mais confiável e robusta de se realizar a sua detecção é a partir da análise espectral do conteúdo harmônico das grandezas elétricas monitoradas pelo sistema de supervisão, tais como tensões e correntes de fase e/ou corrente residual. Consequentemente, busca-se assim indícios da ocorrência de uma FAI já que as mesmas tendem a possuir características espectrais específicas (Brasil et al., 2021).

Desta forma, é comum o uso de técnicas espectrais que buscam a localização tempo-frequencial dos dados adquiridos. Usualmente, a STFT (do inglês, *Short-Time Fourier Transform*) e as transformadas *Wavelet* são utilizadas

durante o processo de extração de características dos sinais mensurados.

Além disto, como o sinal de FAI pode possuir espectro harmônico semelhante aos obtidos em decorrência de eventos transitórios normais na rede elétrica (tais como inserção de banco de capacitor e partida de motores, por exemplo), deve-se estabelecer um sistema de classificação que seja capaz de distingui-los ao longo do tempo, evitando desligamentos desnecessários da rede.

No tocante aos classificadores baseados em aprendizado supervisionado, novas soluções têm surgido dia após dia com o intuito de suporte na correta detecção de FAI. Fanucchi et al. (2013), desenvolveu um classificador baseado em Redes Neurais Artificiais (RNA) para suporte à detecção destas faltas. Para isso, foi adotado um perceptron de multicamada e uma rede de função de base radial (RBF) imbuído de classificar uma corrente de fase, tendo como dados de entrada suas componentes espectrais.

Ledesma et al. (2021), por sua vez, apresentou um método aproximado de localização de FAI utilizando uma base algorítmica de RNA. Além disso, outros métodos que diferem dos supracitados foram abordados. Os sistemas imunológicos artificiais foram utilizados por Lima and Minussi (2014) a partir de um algoritmo de seleção negativa. Já Wang et al. (2021) propôs uma solução para a detecção baseada em ressonância estocástica.

Embora os métodos supracitados apresentem resultados de simulação promissores, duas questões ainda permanecem não respondidas. Primeiramente, praticamente nenhum estudo de detecção de FAI a partir do uso de classificadores com aprendizado supervisionado levou em consideração a detecção de faltas em condições de maior intermitência, randomicidade e alta variabilidade estatística. Deste modo, valores de acurácia de classificação irrealisticamente altos foram obtidos por diversos autores. Diante dos diversos classificadores de aprendizado supervisionado apresentados na literatura, o segundo questionamento trata de qual destas soluções é capaz de melhor se adequar a correta classificação de FAI que compreendem uma maior variabilidade que as utilizadas em outros trabalhos.

Diante destas lacunas presentes na literatura, principalmente no que tange a qualidade e tamanho do banco de dados utilizado durante o processo de treinamento do classificador, o intuito desse trabalho é o de promover uma melhor solução de classificação de FAI a partir do uso de um modelo de arco aprimorado proposto por (Wei et al., 2020). Para aumentar a capacidade de generalização dos algoritmos na classificação de faltas de alta impedância, 3600 eventos transientes foram realizados ao longo de várias simulações em um sistema de distribuição de 20 kV, onde aproximadamente metade deles consistiu em FAI com os mais diversos comportamentos, como proposto por Araujo et al. (2021).

O presente artigo propõe, portanto, uma comparação entre três classificadores de aprendizado supervisionados de capacidade semelhantes no suporte à detecção de faltas de alta impedância. Foram testados a floresta aleatória (do inglês *Random Forest*), o XGBoost (abreviado do inglês *extreme gradient boosting*, i.e. boosting em gradiente extremo) e, para referência, a máquina de vetores de

suporte (do inglês *support vector machine* ou *SVM*). Para isso, utilizou-se o banco de dados produzido por Araujo et al. (2021) submetido ao processo de janelamento. Em cada janela, foi extraída a sua transformada de Fourier (3 ciclos de 60Hz e 32 amostras na transformada). Deste modo, os vetores de características foram montados para serem submetidos aos classificadores e posteriormente ao processo de validação cruzada.

## 2. MODELAGEM DA FALTA DE ALTA IMPEDÂNCIA

Para realizar a modelagem das FAI, se faz necessário entender suas principais características físicas. De acordo com (Ghaderi et al., 2017), a principal característica física da falta de alta impedância é a ocorrência de arco elétrico. A ignição do arco acontece quando a diferença de potencial entre o condutor caído e o solo excede um dado limiar. Quando essa diferença de potencial está abaixo desse limiar, o arco é extinto (Brasil et al., 2021). Grande parte dos modelos de arco presente na literatura atual consideram apenas ignições de arco estáveis, desconsiderando tanto a intermitência quanto possíveis aleatoriedades nos parâmetros intrínsecos à falta. A fim de obter uma robustez maior na descrição das FAI, Wei et al. (2020) atribuiu um coeficiente de aleatoriedade  $RC$  ao modelo convencional de arco de Mayr. Essa relação é dada por:

$$\frac{1}{g_{arc}} \left( \frac{\partial g_{arc}}{\partial t} \right) = \frac{1}{\tau_0} \left( \frac{\mu_f i_f}{P_{loss0} RC} - 1 \right), \quad (1)$$

onde  $g_{arc}$  é a condutância equivalente do arco ( $1/\Omega$ ),  $u_f$  e  $i_f$  são a tensão ( $kV$ ) e a corrente ( $kA$ ) do arco,  $\tau_0$  a constante de tempo ( $s$ ) e  $P_{loss0}$  é a potência média dissipada ( $MW$ ) na resistência de contato desconsiderando-se os seus efeitos transitórios. Seja  $RC_{sel}$ ,  $RC_{ign}$  e  $RC_{ext}$  variáveis aleatórias, o coeficiente  $RC$  pode ser definido como (Wei et al., 2020):

$$\begin{cases} RC = RC_{ign}, & \text{if } RC_{sel} = 0, \\ RC = RC_{ext}, & \text{if } RC_{sel} = 1, \end{cases} \quad (2)$$

onde  $RC_{sel} \in \{0, 1\}$  com probabilidades complementares  $q_0$  e  $q_1$  a cada  $xT$  segundos, onde  $x \in [0, 25; 0, 75]$  e  $T = 0,016$  ( $60 Hz$ ). Além disso,  $RC_{ign}$  e  $RC_{ext}$  são os coeficientes que implementam variações limitadas na potência constante dissipada anteriormente, relacionadas com a ignição e a extinção do arco, respectivamente. Tais valores podem ser obtidos (Wei et al., 2020):

$$\begin{cases} RC_{ign} \in \left( k_1, \frac{k_2 P_{lossb}}{P_{loss0}} \right), \\ RC_{ext} \in \left( \frac{k_3 P_{lossb}}{P_{loss0}}, \frac{k_4 P_{lossb}}{P_{loss0}} \right), \end{cases} \quad (3)$$

onde  $P_{lossb}$  denota a potência dissipada de barreira que representa o coeficiente de potência dissipada a partir do qual o arco se extingue e cujo valor precisa ser determinado empiricamente por meio de simulação.

A aleatoriedade, bem como a intermitência, é definida a partir da escolha dos parâmetros  $k_1$ ,  $k_2$ ,  $k_3$  e  $k_4$ . O modelo de arco de Mayr modificado está representado na Figura 1.  $R_{arco}$  denota a resistência do arco (inverso da condutância) e é responsável pelo comportamento dinâmico da falta.

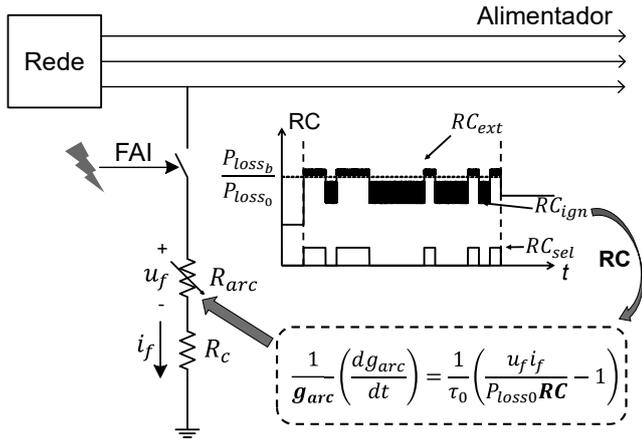


Figura 1. Modelo aprimorado da FAI.

Já  $R_C$  representa a resistência de contato do condutor com a superfície em que ele se projeta, sendo o responsável por efeitos tais como o *buildup* e *shoulder* (Nam et al., 2001). O intervalo a partir do qual ocorre a estabilidade do arco, bem como sua duração, presença/ausência de intermitência/aleatoriedade e a amplitude geral da falta são variáveis inerentes ao modelo de arco melhorado, o qual pode ser arbitrado ou disposto randomicamente para cada FAI emulada, de modo a simular o maior número de comportamentos possíveis na corrente de falta.

### 3. ESTUDO DE CASO

#### 3.1 Sistema de Distribuição de 20kV

Como caso base para verificação da estratégia de detecção de FAI proposta neste trabalho, utilizou-se um sistema de distribuição de 20kV simulado em ambiente SimPowerSystems (MATLAB) cuja topologia de rede é baseada no apresentado por (Elkalashy et al., 2007), conforme apresentado na Figura 2. A rede em questão levou em consideração uma configuração típica de um sistema de distribuição, com o lado de baixa tensão conectado via transformador abaixador. Por outro lado, a rede aérea de média tensão não apresenta ligação para o neutro. A componente residual da corrente é dada como a soma fasorial das correntes de um dado sistema. Para o modelo em questão, pode-se considerar a corrente residual como:

$$i_r = i_a + i_b + i_c \quad (4)$$

#### 3.2 Base de Dados

A construção da base de dados com o intuito de posterior treinamento dos classificadores foi realizada a partir de simulações, com todos os eventos ocorrendo no alimentador F1 (média tensão). Um total de 9 possibilidades foram estipuladas como acontecimentos transitórios que não configuram FAI no referido alimentador: inserção de cargas monofásicas, bifásicas e trifásicas, chaveamento de banco de capacitores monofásicos e trifásicos, energização de transformadores trifásicos sob carga leve (a fim de simular a corrente de magnetização deste), partida de motores monofásicos e trifásicos e variação do carregamento ao final do trecho do alimentador (no lado de baixa tensão). Já para as simulações de FAI, considerou-se também a

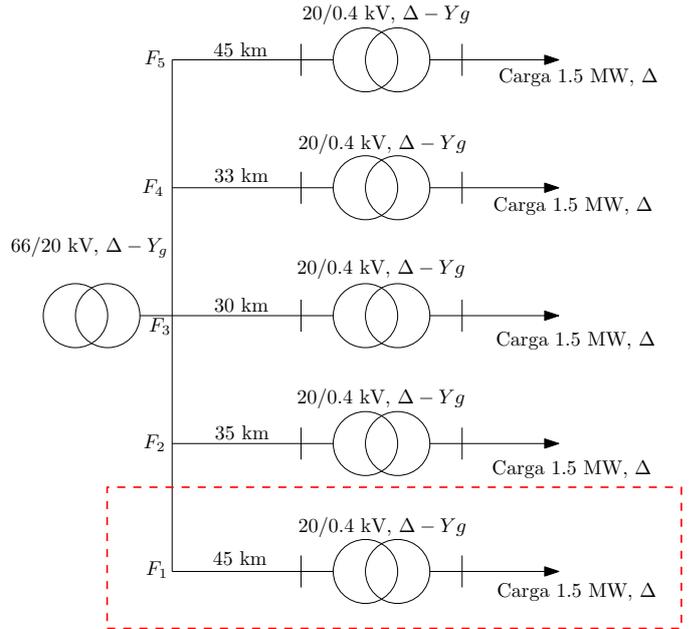


Figura 2. Topologia da rede do sistema teste de 20 kV.

Tabela 1. Eventos simulados e seus respectivos intervalos.

Evento	Intervalo
<i>Carga monofásica</i>	$P_l = [10, 100]$ kW $Q_l = [-5, 5]$ kVAr
<i>Capacitor monofásico</i>	$Q_l = [-200, -20]$ kVAr
<i>Carga bifásica</i>	$P_l = [10, 100]$ kW $Q_l = [-5, 5]$ kVAr
<i>Carga trifásica</i>	$P_l = [10, 100]$ kW $Q_l = [-5, 5]$ kVAr
<i>Energização de transformador trifásico (carga leve)</i>	$P_l = [10, 50]$ kW $Q_l = [-5, 5]$ kVAr
<i>Motor monofásico</i>	$T = [3.1, 19.5]$ N.m
<i>Motor trifásico</i>	$T = [2.1, 280]$ N.m
<i>Variação da carga na BT</i>	$P_l = [10, 50]$ kW
<i>Falta de alta impedância</i>	Ver Tabela 2.

possibilidade de existência de certa magnitude de ruído gaussiano, cujo intervalo utilizado foi de 15 à 5 dB (uniformemente provável).

O resumo de tais eventos, bem como seus respectivos intervalos, é apresentado na Tabela 1. A ocorrência de cada um destes eventos tem como consequência o desbalanço transitório das correntes no sistema modelado. Os eventos foram limitados a ocorrerem aleatoriamente, de um a um, e apenas no alimentador F1 (Figura 2), em qualquer ponto do mesmo (de 0 a 45 km). O tempo definido para cada simulação foi de 5s, com uma média de ocorrência de 9 eventos (estipulados de forma randômica) por simulação e um total de 3600 eventos (400 simulações foram realizadas). Cada simulação realizou a extração de características da corrente residual medida no princípio do alimentador F1, a partir da STFT, dividindo-as também, de forma binária, entre eventos de falta ou não falta, associando cada uma das classificações realizadas a uma janela da STFT.

#### 3.3 Classificação

Após a geração dos sinais de 5 segundos através dos modelos descritos anteriormente, foram utilizados modelos supervisionados para a classificação. Modelos de classifica-

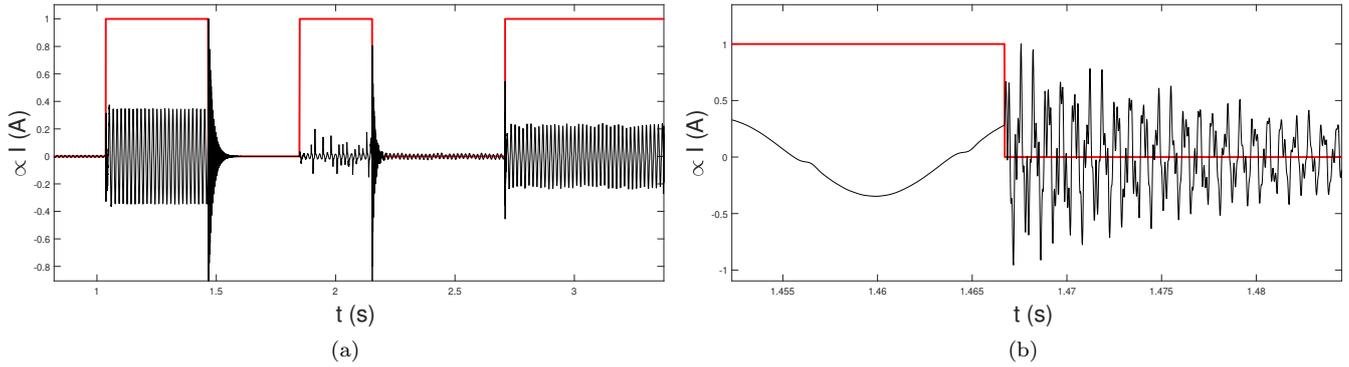


Figura 3. Amostra de um sinal presente no conjunto de dados. A linha vermelha marca onde há ocorrência de falta de alta impedância. Em 3a, temos uma amostra completa e em 3b, o detalhe do fim da ocorrência de um evento.

Tabela 2. Características das FAI simuladas.

FAI	$P_{loss0}$	$P_{lossb}$	$k_1$	$k_2$	$k_3$	$k_4$	$R_c(k\Omega)$
T1(Ws)	7.5-12.5	$P_{loss0}$	0.75	1.25	-	-	0.5-1.0
T1(DA)	1.0-3.0	$P_{loss0}$	0.85	1.00	-	-	0.5-1.0
T1(WC)	2.0-4.0	$P_{loss0}$	0.35	1.00	-	-	0.17-0.33
T2A	10	50	0.01	0.05	1.00	1.05	0.17-0.55
T2B	1	40	0.5	0.75	1	6	0.25-0.55
T3	1	4	0.85	1.00	0.65	1.25	2.5-10.0

ção com aprendizado supervisionado são aqueles em que parte dos dados rotulados são disponíveis no treinamento a fim de se obter a aprendizagem necessária na classificação. Como a corrente residual medida é uma série temporal, cada amostra desse sinal (amostrado originalmente a uma taxa de 50KHz), é marcado com os rótulos 0 (não-falta) e 1 (FAI). Como a FAI tem características espectrais relevantes Brasil et al. (2021), uma janela de tamanho  $L$  foi definida a fim de que seja definido o rótulo por janela. Após vários testes desenvolvidos anteriormente Araujo et al. (2021), a janela que será arbitrada neste trabalho e que apresenta acurácia relevante para a classificação foi a de  $L = 2500$  amostras.

#### 4. METODOLOGIA

A metodologia empregada é explicitada a partir do fluxograma da Figura 4. Para a utilização da base de dados, foram selecionadas 400 sinais distintos do banco que foram decimados por um fator de 10. Logo após, esses sinais foram recortados de tal maneira a obter comprimento 250. Subsequentemente, a STFT de tamanho 32 foi aplicada em cada janela para formar o vetor de características que utilizado na etapa de classificação. Como cada amostra do sinal pode ser rotulado como falta ou não, cada janela foi, portanto, adequadamente rotulada, posteriormente.

Seguindo adiante no fluxograma metodológico deste trabalho, segue a etapa de validação. Foram gerados 400 sinais distintos, utilizando o modelo anteriormente descrito. Para garantir que não haja vieses nos grupo de treino e teste, esta etapa é necessária. Neste trabalho, foi utilizado o algoritmo de validação cruzada  $K$ -Fold. Esse método separa o banco de dados em  $K$  conjuntos onde um conjunto é usado para validação e  $K - 1$  conjuntos para treinamento, de modo que os conjuntos sejam permutados. Dessa forma, é garantido que todos os dados vão participar do treino e do teste, fazendo com que a acurácia total seja a média das acurácias. Separando os conjuntos dessa maneira, é possível prevenir o sobreajuste do algoritmo de

aprendizado aos dados. Para este trabalho, foi arbitrado  $K = 4$ .

Após a divisão dos sinais em conjuntos, o conjunto de treino foi submetido a um dos classificadores (XGBoost, SVM ou *Random Forest*) e avaliado sua acurácia no conjunto de teste. A escolha dos métodos é dado ao fato do *Random Forest* ser difundido na literatura (Samantaray, 2012) e o XGBoost é proposto como uma alternativa para este método (Chen and Guestrin, 2016). Além disso, a máquina de vetores de suporte será utilizada como referência para a comparação de resultados (Abdoos et al., 2016; Brasil et al., 2021; Araujo et al., 2021). Depois de realizado os  $K$  testes, é gerado a acurácia do algoritmo em questão e repetido o processo com os demais classificadores.

A máquina de vetores de suporte, por sua vez, tem como objetivo definir um hiperplano que sirva de fronteira para a separação das amostras em classes. Desse modo é possível prever qual classe pertence a amostra. Seja uma distribuição de dados  $x \in \mathbb{R}^n$  onde cada amostra  $x_i$  tem um rótulo  $y_i \in \{-1, 1\}$  que se trata da componente  $i$  do vetor de rótulos  $y \in \mathbb{R}^n$ . O hiperplano gerado pela SVM de *kernel* linear tem a equação definida por (Steinwart and Christmann, 2008):

$$f(x) = \sum_{i=1}^n w_i \cdot x_i + b = 0. \quad (5)$$

onde  $w_i$  é o componente  $i$  do vetor de pesos  $w \in \mathbb{R}^n$  e  $b$  é o termo de vies. Desse modo, para que o método cumpra seu objetivo, o hiperplano deve satisfazer a Equação 6 (Steinwart and Christmann, 2008):

$$y_i(f(x_i)) = y_i(w \cdot x_i + b) \geq 1, \quad i = 1, 2, \dots, n. \quad (6)$$

A metodologia utilizada para determinar o modelo ótimo do hiperplano que prediz a classe de cada amostra é a minimização da expressão dada na Equação 7 através das variações dos pesos (Steinwart and Christmann, 2008):

$$\min \left( \frac{1}{2} \|w\|^2 + C \sum_{i=1}^m \xi_i \right), \quad i = 1, 2, \dots, n. \quad (7)$$

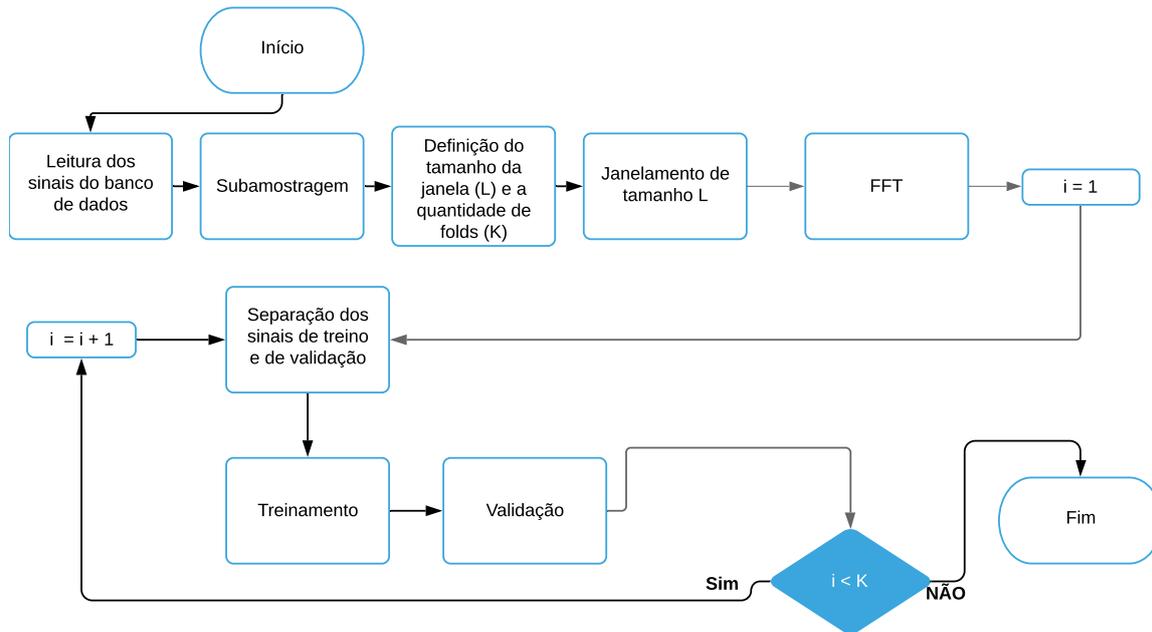


Figura 4. Fluxograma metodológico do algoritmo proposto.

onde  $C$  é a penalidade por erro e  $\xi_i$  é a distância entre a amostra  $x_i$  e o hiperplano. Entretanto, o valor de  $\xi_i$  precisa satisfazer as inequações descritas em 8 (Steinwart and Christmann, 2008):

$$\begin{cases} y_i(w \cdot x_i + b) \geq 1 - \xi_i, \\ \xi_i > 0. \end{cases} \quad (8)$$

O interesse nos algoritmos *Random Forest* e XGBoost, por sua vez, vem do fato de eles serem desenvolvidos a partir do conceito de Árvores de Decisão. As árvores de decisão são seqüências de seleções binárias que subdividem as amostras avaliadas tendo como base as características do modelo. Com esse tipo de abordagem se torna possível a realização de métodos de classificação e de regressão (Bishop, 2006). A Figura 5 exemplifica esse processo. A árvore é iniciada no nó raiz, onde é avaliado a característica  $x_1$ , verificando se  $x_1 > \theta_1$ . Caso a proposição for verdadeira, o algoritmo segue o fluxo da direita até o próximo nó, chamado de *nó folha*. O algoritmo vai seguindo dessa forma, de folha em folha, até chegar na menor subdivisão estabelecida. Essa metodologia tem como principal vantagem a alta velocidade de execução (Ho, 1995).

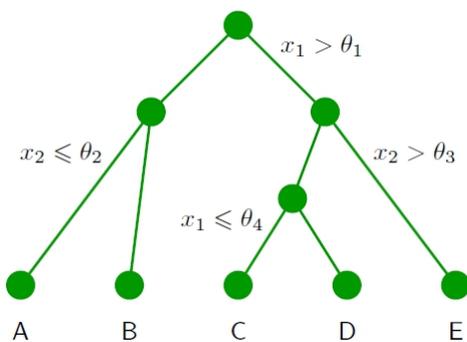


Figura 5. Árvore de decisão genérica (Bishop, 2006).

O algoritmo *Random Forest*, por sua vez, se trata de várias árvores de decisão sendo processadas paralelamente com diferentes critérios de subdivisão, de modo que é feita uma comparação entre as predições para que seja feita a predição final. Esse tipo de processamento em paralelo é chamado de *bagging* (Bishop, 2006).

De forma análoga, o algoritmo XGBoost tem como base o processamento de várias árvores de decisão em série, fazendo o que é chamado de *boosting* (Bishop, 2006). Dentre as diferentes formas de realizar essa metodologia, o XGBoost utiliza uma versão aprimorada do *Gradient Boost* (Chen and Guestrin, 2016). Seja  $L(\hat{y}_i, y_i)$  a função de erro entre a  $i$ -ésima predição  $\hat{y}_i$  e o  $i$ -ésimo valor real  $y_i$ , seja  $g_i = \partial_{\hat{y}^{(t-1)}} L(y_i, \hat{y}^{(t-1)})$  e  $h_i = \partial_{\hat{y}^{(t-1)}}^2 L(y_i, \hat{y}^{(t-1)})$  os gradientes de primeira e segunda ordem respectivamente, a função a ser minimizada é definida por (Chen and Guestrin, 2016):

$$\mathcal{L}^{(t)} = \sum_{i=1}^n [g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)] + \Omega(f_t), \quad (9)$$

onde  $f_t$  é a estrutura de árvores da  $t$ -ésima iteração e  $\Omega(f_t)$  é dado por (Chen and Guestrin, 2016):

$$\Omega(f_t) = \gamma T + \frac{1}{2} \lambda \|w\|^2. \quad (10)$$

A Equação 9 demonstra a forma a partir da qual o método obtém a modelagem ótima. Para isso, o termo  $\Omega(f_t)$  é inserido como uma penalidade para a complexidade do modelo. Para isso, este termo apresenta o número de folhas na árvore ( $T$ ), o peso da folha ( $w$ ) e as constantes  $\gamma$  e  $\lambda$  que ajudam o termo  $\Omega(f_t)$  a evitar o *over-fitting* (Chen and Guestrin, 2016).

## 5. RESULTADOS

De modo comparativo ao método empregado neste artigo, Brasil et al. (2021), utilizando um banco de dados mais simplificado e menor que este presente no artigo, chegou ao resultado de detecção em cerca de 90% de acurácia utilizando SVM e seleção das componentes harmônicas. Araujo et al. (2021), utilizando o mesmo conjunto de dados apresentados neste trabalho, com a janela de 2500 amostras e método de classificação *Weighted KNN* atingiu a acurácia média de 94%.

Na tabela a seguir, estão dispostos os resultados comparativos entre os métodos testados: XGBoost, *Random Forest* e SVM 3.

Tabela 3. Desempenho dos algoritmos.

Algoritmo	XGBoost	Random Forest	SVM
Acurácia	<b>97,67%</b>	96,10%	91,17%
F1-Score	<b>97,87%</b>	84,27%	90,79%
Sensibilidade	<b>98,14%</b>	83,82%	96,99 %
Precisão	97,60%	<b>97,85%</b>	85,35%

Dos resultados observa-se 4 métricas distintas: acurácia, F1-Score, Sensibilidade e Precisão. A acurácia mede a taxa de acertos totais (falta ou não-falta) em relação ao total de dados. A métrica F1-Score consiste no balanceamento no valor da acurácia em relação ao total de dados de falta e dados de não-falta. A sensibilidade representa o poder do algoritmo de acertar na condição de interesse, neste caso, na condição de falta. A precisão, por outro lado, contra-balança o acerto na classe de interesse em relação as demais amostras.

Como visto nos resultados, na maioria das 4 métricas, o algoritmo proposto neste trabalho apresentou resultado superior. A exceção está na métrica precisão, onde o algoritmo *Random Forest* apresenta melhor resultado.

Em vias de comparação com os trabalhos anteriormente publicados, o algoritmo proposto possui desempenho superior as técnicas utilizadas. É importante notar que neste trabalho não foi utilizado nenhum seletor de características (como nos anteriores). Tanto a seleção quanto a importância de cada uma delas foi definida na etapa de treinamento das árvores.

## 6. CONCLUSÃO

O presente trabalho avaliou a eficiência de alguns dos principais algoritmos de classificação na detecção de faltas de alta impedância em sistemas de distribuição de energia elétrica. Mostrou-se que a capacidade de classificação dos algoritmos XGBoost e Random Forest foi superior a alcançada em publicações supracitadas, as quais fizeram uso de soluções mais conhecidas, tais como SVM e KNN. Vale ressaltar que a comparação com estes algoritmos foi a mais justa possível, levando em consideração que o banco de dados utilizado foi exatamente o mesmo para todos os casos. A partir do classificador baseado em algoritmo XGBoost, obteve-se uma acurácia média de 97,67%, sendo o melhor desempenho dos autores até o momento.

## REFERÊNCIAS

Abdoos, A.A., Khorshidian Mianaei, P., and Rayatpanah Ghadikolaie, M. (2016). Combined VMD-SVM based feature selection method for classification of power quality events. *Applied Soft Computing Journal*, 38, 637–646. doi:10.1016/j.asoc.2015.10.038.

Araujo, G.d.S.R., Brasil, T.A., Souza, B.A.V., Gois, J.N., and Neto, J.A.M. (2021). Análise e detecção de faltas de alta impedância em sistemas de distribuição com a utilização de classificadores por aprendizado supervisionado. *Anais do XV Simpósio Brasileiro de Automação Inteligente, SBAI. Aceito para publicação.*

Banner, C.L. and Don Russell, B. (1997). Practical high-impedance fault detection on distribution feeders. *IEEE Transactions on Industry Applications*, 33(3), 635–640. doi:10.1109/28.585852.

Bishop, C.M. (2006). *Pattern recognition and machine learning*. springer.

Brasil, T.A., Gois, J.N., and Neto, J.A. (2021). Detection of high impedance faults in primary distribution grid using support vector machines classification. In *2021 14th IEEE International Conference on Industry Applications (INDUSCON)*, 1138–1145. IEEE.

Cabral, S., Senger, E., and Santos, J. (1992). Falta de alta impedância-proposta de solução através de sistemas supervisórios. *Seminário Nacional de Distribuição de Energia Elétrica, Blumenau, SC.*

Chen, T. and Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 785–794.

Elkalashy, N.I., Lehtonen, M., Darwish, H.A., Izzularab, M.A., and Abdel-maksoud, I.T. (2007). Modeling and experimental verification of high impedance arcing fault in medium voltage networks. *IEEE Transactions on Dielectrics and Electrical Insulation*, 14(2), 375–383.

Fanucchi, R.Z., Oleskovicz, M., and Barbosa, D. (2013). Análise da detecção de faltas de alta impedância utilizando redes neurais artificiais com topologias baseadas em perceptron multicamadas e redes rbf. *Simpósio Brasileiro de Automação Inteligente (SBAI), Fortaleza.*

Ghaderi, A., Ginn III, H.L., and Mohammadpour, H.A. (2017). High impedance fault detection: A review. *Electric Power Systems Research*, 143, 376–388.

Ho, T.K. (1995). Random decision forests. In *Proceedings of 3rd international conference on document analysis and recognition*, volume 1, 278–282. IEEE.

Ledesma, J.J.G., de Araujo, L.R., and de Araújo, D.R.R.P. (2021). A method for the approximate location of high impedance faults using neural networks. *IEEE Latin America Transactions*, 19(3), 351–358.

Lima, F. and Minussi, C. (2014). Diagnóstico de faltas de alta impedância em sistemas de distribuição de energia elétrica usando um algoritmo imunológico artificial.

Moreto, M. (2005). Localização de faltas de alta impedância em sistemas de distribuição de energia: uma metodologia baseada em redes neurais artificiais.

Nam, S., Park, J., Kang, Y., and Kim, T. (2001). A modeling method of a high impedance fault in a distribution system using two series time-varying resistances in emtp. In *2001 Power Engineering Society Summer Meeting. Conference Proceedings (Cat. No.01CH37262)*, volume 2, 1175–1180 vol.2. doi:10.1109/PESS.2001.970231.

Samantaray, S. (2012). Ensemble decision trees for high impedance fault detection in power distribution network. *International Journal of Electrical Power & Energy Systems*, 43(1), 1048–1055.

Steinwart, I. and Christmann, A. (2008). *Support vector machines*. Springer Science & Business Media.

Wang, X., Wei, X., Gao, J., Song, G., Kheshti, M., and Guo, L. (2021). High-impedance fault detection method based on stochastic resonance for a distribution network with strong background noise. *IEEE Transactions on Power Delivery*.

Wei, M., Shi, F., Zhang, H., Jin, Z., Terzija, V., Zhou, J., and Bao, H. (2020). High Impedance Arc Fault Detection Based on the Harmonic Randomness and Waveform Distortion in the Distribution System. *IEEE Transactions on Power Delivery*, 35(2), 837–850. doi:10.1109/TPWRD.2019.2929329.