

Aplicação de Aprendizagem de Máquina para Identificar Motores com Fuga à Terra em Sistemas de Neutro Aterrado com Resistor de Alto Valor

Lucas de Oliveira Soares* Luiz Alberto Pinto*
Marco Antonio de Souza Leite Cuadros*

* Programa de Pós-Graduação em Engenharia de Controle e Automação – Propecaut, Instituto Federal do Espírito Santo, Serra, Brasil (e-mails: lucasosoares@hotmail.com, pintoluizalberto@gmail.com, madsouzacuadros@gmail.com)

Abstract: This paper presents an approach to identify a motor with earth leakage, in an electrical system that has a grounded neutral with a high ohmic resistor, through the application of Machine Learning techniques. The proposed method consists of using residual currents (zero-sequence currents), of the motors in normal operating condition and in the fault condition due to earth leakage. For the construction of the classification models, the following types of descriptors extracted from the zero-sequence current signals were used, (i) descriptors represented by the statistical parameters in the time domain, (ii) descriptors represented by the statistical parameters in the time-frequency domain with the application of the wavelet transform, and (iii) descriptors represented by the Shannon entropy also extracted in the wavelet time-frequency domain. The models were built using the Random Forest and Support Vector Machine, Discriminant Analysis and k Nearest Neighbour algorithms. The best result (98.99% accuracy and f1-score of 97.49%) demonstrates the good capacity of the model used to detect a motor with earth leakage in a system with a grounded neutral with a high value resistor.

Resumo: Esse artigo apresenta uma abordagem para identificação de motor com fuga à terra, em um sistema elétrico que possui o neutro aterrado com resistor de alto valor ôhmico, através da aplicação de técnicas de aprendizado de máquina. O método proposto consiste na utilização das correntes residuais (correntes de sequência zero), dos motores na condição de operação normal e na condição de falha por fuga à terra. Para a construção dos modelos de classificação, foram utilizados os seguintes tipos de descritores extraídos dos sinais de corrente de sequência zero, (i) descritores representados pelos parâmetros estatísticos no domínio do tempo, (ii) descritores representados pelos parâmetros estatísticos no domínio tempo-frequência com a aplicação da transformada *wavelet*, e (iii) descritores representados pela entropia de *Shannon* também extraída no domínio tempo-frequência *wavelet*. Os modelos foram construídos com a aplicação dos algoritmos *Random Forest* e *Support Vector Machine*, *Discriminant Analysis* e *k Nearest Neighbour*. O melhor resultado (98,99% de acurácia e *f1-score* de 97,49%) demonstra boa a capacidade do modelo empregado para detecção de motor com fuga à terra em sistema com neutro aterrado com resistor de alto valor.

Keywords: earth leakage failure; machine learning; zero sequence current; wavelet transform; signal classification.

Palavras-chaves: falha por fuga à terra; aprendizado de máquina; corrente de sequência zero; transformada *wavelet*; classificação de sinais.

1. INTRODUÇÃO

A confiabilidade dos processos produtivos está ligada a técnicas que sejam precisas e rápidas em diagnosticar defeitos em fase incipiente, a fim de garantir a continuidade dos serviços com o mínimo de falhas.

Um dos grandes problemas relacionados a equipamentos elétricos é a ocorrência de curtos circuitos nas instalações industriais sendo que neste contexto, 98% das falhas em sistemas elétricos de potência estão associadas a falhas fase-terra (monofásicas) (Bowen, 2011). Outros tipos de

falhas (bifásicas e trifásicas), em grande parte, são evoluções de falhas monofásicas, ou oriundas de outros modos de falha. Para reduzir o impacto de falhas fase-terra em sistemas industriais de baixa tensão, é utilizado aterramento do neutro através de um resistor de alta impedância, a fim de limitar as correntes entre fase-terra para valores de baixa intensidade. Uma das vantagens da implementação dessa solução é que o mesmo poderá continuar em operação até que haja intervenção para tratativa do problema por parte da operação. Embora a implementação de sistemas de alimentação trifásicos com neutro aterrado por resistor de alta impedância seja vantajosa do ponto de vista fun-

cional, a identificação e o isolamento de falhas à terra é um desafio para a manutenção. As correntes que circulam entre fase e terra podem ser facilmente confundidas com as correntes residuais naturais do sistema. Além disso, a grande quantidade de cargas no mesmo alimentador, pode dificultar a identificação da corrente de fuga. Dessa forma, torna-se necessária a medição pontual da corrente em cada carga, aumentando a probabilidade de erro humano na identificação da carga com fuga à terra (Hussey and Drzymala, 2019).

Este trabalho apresenta uma proposta para a utilização de técnicas de aprendizado de máquina, aplicadas à identificação de motores com falha à terra em instalação com a configuração de aterramento do neutro por resistor de alta impedância. Para a realização dos experimentos foram utilizadas as medições de corrente de sequência zero de 76 motores de um centro de controle de motores de uma indústria de celulose. O conjunto de dados utilizado nos experimentos, é constituído de medições de correntes dos equipamentos na condição de operação normal e com fuga à terra. O método proposto tem o objetivo de identificar de forma individual o motor que apresenta defeito de fuga à terra. Para a construção dos modelos foram extraídos descritores estatísticos dos sinais no domínio tempo e no domínio tempo/escala, obtidos com transformada *wavelet*. Os classificadores foram implementados com a utilização dos algoritmos *Discriminant Analysis (DA)*, *k Nearest Neighbour (k-NN)*, *Support Vector Machine (SVM)* e *Random Forest (RF)*.

Na sequência este trabalho está constituído pelas seguintes seções: a Seção 2 contém a revisão do estado da arte sobre diagnóstico de falhas; a Seção 3 apresenta uma descrição dos sistemas de aterramento utilizando resistor de alto valor ôhmico; A Seção 4 faz uma revisão das técnicas e ferramentas aplicadas ao longo do trabalho; A Seção 5 apresenta a metodologia que orientou a realização do trabalho com uma detalhada descrição do conjunto de dados; A Seção 6 apresenta e discute os resultados obtidos e a Seção 7 traz as conclusões.

2. REVISÃO BIBLIOGRÁFICA

Todos os processos produtivos buscam a confiabilidade de seus ativos e neste contexto, o desenvolvimento de técnicas para detecção de falhas em sistemas elétricos industriais utilizando análise de sinais de variáveis elétricas, tem sido objeto de pesquisa ao longo das últimas décadas. Instalações elétricas trifásicas com neutro aterrado são muito estudadas e aplicadas devido a característica de tais sistemas poderem operar indefinidamente sob a condição de falha à terra, desde que essa envolva apenas uma das fases. Contudo, a falha deve ser identificada e isolada de forma a evitar a sua propagação pelos outros componentes do sistema, e a sua evolução para condições mais severas, o que, inevitavelmente, provocaria a interrupção do processo produtivo.

Os autores em Lubich (1997) propõem uma solução para a detecção do ponto de ocorrência da fuga, que consiste na injeção de um sinal de corrente pulsante com valor entre 3 A a 4 A no ponto de aterramento entre o neutro do transformador e a terra. A corrente injetada, superior a corrente capacitiva do circuito, circulará, para à terra,

pelo caminho da fase que está com baixa isolamento. Para a identificação do circuito com falha à terra, a corrente de todas as cargas conectadas ao barramento deverão ser medidas, e a carga com falha à terra será identificada quando a corrente medida for igual ao valor da corrente injetada.

Os autores em Baldwin et al. (2003) apresentam um algoritmo para detecção de falha à terra em sistemas aterrados por resistor de alto valor em instalações de baixa tensão. O sistema, que se baseia na detecção da direção da falta em função das variações dos fasores de tensão e corrente do circuito, é constituído por quatro estágios principais: detecção de falha, filtragem da polarização, estimativa de fasor e cálculo do ângulo de fase com estimativa de direção. A medição dos sinais é realizada com o uso de um transformador de corrente residual e um transformador de potencial com conexão secundária "delta aberto". Quando ocorre uma fuga à terra, são geradas tensões e correntes de sequência zero que são detectadas pelo conjunto. Em seguida, é aplicada a transformada rápida de Fourier para encontrar os fasores dos sinais de tensão e corrente, e a partir desses, determina-se o ângulo da impedância, que é a diferença angular entre os fasores de tensão e corrente. Essas informações alimentam um sistema microprocessado embarcado que faz a estimação da direção da falha. De acordo com os autores, a acurácia do sistema embarcado é em média 60% para correntes de falha à terra na faixa de 10 mA e 20 mA, e de 100% para correntes de fuga superiores a 20 mA.

O trabalho publicado por Cai et al. (2014) propõe uma técnica para detecção de fuga à terra em redes de distribuição elétrica, baseado nas correntes de sequência zero. O procedimento consiste na medição das correntes de fase, e no cálculo das correntes de sequência zero com a utilização de computadores. De acordo com o algoritmo proposto, se os valores das correntes de sequência zero forem muito baixos, não há falhas de fuga à terra no sistema, caso contrário, existe fuga à terra em algum ponto do sistema em questão. Para a identificação do nó da rede onde ocorreu a falha utiliza-se o ângulo do componente de corrente de sequência zero. Os autores mencionam que para sistemas onde a impedância de falha é inferior a 1 k Ω o método proposto é altamente eficaz, com acurácia em torno de 100%.

Em Topolanek et al. (2015) é apresentada uma técnica para detecção de fuga à terra baseada em medição de tensão. Esta aplicação é desenvolvida especialmente para sistemas de distribuição compensados, equipados com conexão automática de curta duração de um resistor auxiliar para aumentar o componente ativo da corrente de falha à terra. Com isso, esse procedimento é normalmente utilizado em redes de distribuição europeias para melhorar a sensibilidade da proteção contra fuga à terra direcional. O método consiste na avaliação das quedas de tensão causadas pela conexão de um resistor auxiliar na condição de fuga à terra, registradas por um sistema de medição instalado no lado de baixa tensão dos transformadores abaixadores de 22 kV para 0.4 kV. Neste estudo são utilizadas estações de medição de tensão secundária em cada ramal e após a ocorrência da fuga de fuga à terra, os diagramas de tensão em todas as linhas são analisados, e as possíveis áreas do sistema em que ocorreu a falha são

delimitadas pela análise das quedas de tensão. De acordo com os autores, o método se mostrou eficiente para falhas sólidas ou de baixa impedância, não sendo garantido seu funcionamento no caso de falhas de alta impedância. Os autores consideram falhas de alta impedância, aquelas com resistências superiores a 800 k Ω .

O trabalho de Valdes et al. (2006) propõe uma técnica para detecção de fuga à terra em sistemas aterrados por alta resistência. Na proposição dos autores, um sinal externo da ordem de 4 a 8 A, 600 Hz é injetado no circuito do resistor de aterramento. Sensores de corrente são instalados nos ramais, e as informações sobre o comportamento das correntes são coletadas por um sistema de monitoramento. Quando ocorre uma falha à terra, o sistema de monitoramento identifica o caminho da corrente injetada no circuito e indica com relativa precisão o ponto de ocorrência. De acordo com os autores, comparativamente às demais técnicas existentes, o método proposto é um dos mais economicamente viáveis.

Não foram encontrados na literatura trabalhos sobre a identificação de cargas com fuga à terra em sistemas aterrados com resistor de alto valor ôhmico em baixa tensão, sem a utilização de sinal externo. O presente trabalho se distingue dos demais disponíveis na literatura, por se basear na aquisição e análise direta dos sinais medidos nos equipamentos para localização de pontos com fuga à terra, sem a necessidade de realizar alterações significativas no sistema elétrico. A abordagem proposta contorna os problemas oriundos de modificações de projetos originais e minimiza os custos de implantação dos sistemas de localização dos pontos com falha à terra.

3. SISTEMAS DE ATERRAMENTO POR RESISTOR DE ALTO VALOR

Sistemas aterrados por resistor de alto valor (RAV) ou "HGR" (*High Grounding Resistance*), possuem, entre o neutro e o terra, um elemento que limite as correntes de falha à terra entre 1A e 10A (Blackburn and Domin, 2015). Valores de corrente de fuga à terra nessa faixa reduzem a probabilidade de danos nas instalações elétricas. Uma grande vantagem deste sistema é a capacidade de operar de forma ininterrupta em condição permanente de falha fase-terra, por este motivo é muito comum em sistemas de baixa tensão industrial.

Quando ocorrem falhas monofásicas, há um deslocamento do neutro, conforme ilustrado na Fig. 1. O potencial da fase em curto-circuito é igual ao potencial do terra deslocado ao ponto de neutro do sistema elétrico. Nessa configuração, as correntes de falha fase-terra são de baixa intensidade em função da limitação imposta pelo RAV.

Contudo, caso ocorra falha de outra fase para a terra, haverá um curto-circuito bifásico de alta intensidade pela conexão de duas fases do sistema ao mesmo ponto. Como consequência, poderá haver a ocorrência de arcos elétricos e a destruição de componentes (Costa, 2014). Essa condição poderá levar a parada do sistema como um todo em função da atuação das proteções elétricas e pela impossibilidade da limitação das correntes de curto-circuito pelo RAV.

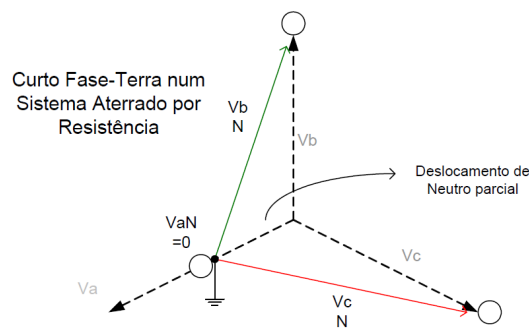


Figura 1. Deslocamento parcial do neutro em sistema aterrado por resistência.

Uma limitação desse sistema é a dificuldade para a identificação de ramais com fuga à terra, em função da pequena intensidade da corrente de falha, que pode ser confundida com correntes provocadas pelo desbalanceamento de fases, por transitórios, distúrbios elétricos e correntes capacitivas que possam ocorrer no sistema.

4. REFERENCIAL TEÓRICO

Essa seção apresenta os algoritmos, *Support Vector Machine - SVM*, *Discriminant Analysis - DA*, *k-Nearest Neighbour - k-NN* e *Random Forest - RF*, que foram aplicados para a obtenção dos modelos de classificação. Além disso, é feita uma descrição das ferramentas utilizadas para a extração de descritores dos sinais de corrente.

4.1 Discriminant Analysis - DA

Este algoritmo é capaz de separar objetos, por um conjunto de variáveis independentes, em duas ou mais categorias mutuamente exclusivas e exaustivas (Morrison, 1969). Segundo Dibennardo and Taylor (1982) a *discriminant analysis* baseia-se em dois princípios gerais: (i) análise - para delinear as dimensões ao longo das quais as populações são diferenciadas ao máximo; (ii) classificação - para atribuir indivíduos a grupos com base em semelhanças compartilhadas. Na análise, ao indicar quais variáveis são altamente ponderadas, os coeficientes da função discriminante destacam as dimensões ao longo das quais as populações mais diferem. A classificação, é baseada em comparação do perfil de um indivíduo com o perfil médios de dois ou mais grupos, em um dos quais ele deve ser atribuído. Essas comparações são computadas para cada grupo, atribuindo o indivíduo para o grupo cuja função possui pontuações mais altas.

4.2 Support Vector Machine - SVM

Support Vector Machine é um algoritmo de classificação supervisionada, linear e binária, que procura o hiperplano de separação entre as classes, de maneira a maximizar a distância entre as amostras mais próximas (Cortes and Vapnik, 1995). A aplicação do *SVM* à problemas não lineares pode ser possível com a utilização de funções *kernel*, que projetam o espaço amostral original sobre espaços de maior dimensão onde os dados sejam linearmente separáveis. Embora seja, em sua forma convencional, um algoritmo binário, que suporta apenas problemas com duas

classes, a utilização de métodos multiclasse torna o *SVM* aplicável a problemas com mais de duas classes.

4.3 *k*-Nearest Neighbour - *k*-NN

Classificadores implementados com base no algoritmo *k*-NN classificam amostras desconhecidas com base na distância entre essas e as amostras atribuídas às classes envolvidas no problema (Duda et al., 2001). Por ser de fácil implementação, esses classificadores são amplamente utilizados em problemas de classificação. Para sua implementação devem ser definidos, o número de vizinhos (*k*), da métrica para o cálculo da distância entre as amostras e a regra de decisão para atribuição de uma classe a uma amostra desconhecida. Uma regra de decisão comumente utilizada é atribuir à amostra desconhecida a classe da maioria das *k* amostras mais próximas. O algoritmo que também é conhecido como *k*-NN, possui uma variável "k" que representa a quantidade de elementos próximos (ou vizinhos) a um elemento e com base nisso, classifica os vetores em um espaço amostral com *k*-vizinhos mais próximos. Normalmente, os valores atribuídos a *k* são de três ou um (Krishnaiah and Kanal, 1982), e pode ser definido com apoio de um conjunto de validação.

4.4 *Random Forest* - *RF*

A *RF* é um algoritmo de aprendizado de máquina supervisionado, que pode ser aplicado à problemas de classificação e regressão. Uma *RF* é constituída por um conjunto de Árvores de Decisão, cujo algoritmo de treinamento mais comumente utilizado é o bagging. Como utiliza combinação das árvores de decisão, a Floresta Aleatória, de forma geral, é mais estável e apresenta melhor desempenho quando comparada a outros algoritmos de *Machine Learning*. A premissa sobre a qual se baseia a *RF* é que, um vetor aleatório v_k para a *k*-ésima árvore, independente dos vetores aleatórios passados $v_1, v_2, v_3, \dots, v_{k-1}$, mas com a mesma distribuição será criado. Com estes vetores aleatórios gerados, em um processo de votação envolvendo as *k* árvores, uma classe será atribuída à amostra desconhecida (Breiman, 2001).

4.5 *Transformada Wavelet*

A *Transformada Wavelet* - *TW* é uma ferramenta matemática que permite uma abordagem de análise de sinais no domínio tempo/escala. A forma como a *TW* localiza as informações do sinal no domínio da frequência e do tempo, simultaneamente, torna-a uma solução adequada para a análise de sinais não-estacionários. Ao contrário da janela de largura fixa utilizada pela *Short Time Fourier Transform* - *STFT*, a *TW*, na forma de suas diversas famílias, por exemplo, Haar, Daubechies, Symlets, Morlets, Coiflets, etc., é escalável, o que significa que a transformação *wavelet* é adaptável a uma ampla gama de resoluções baseadas em frequência e tempo Nandi and Ahmed (2019). Para a extração de descritores dos sinais, nesse trabalho foi utilizada a versão discreta da *Transformada Wavelet*.

4.6 *Descritores Utilizados para Obtenção dos Modelos*

Para a construção dos modelos de classificação, dois tipos de descritores foram extraídos dos sinais de corrente com a utilização da *Transformada Wavelet*.

Entropia de Shannon - A entropia é uma medida da aleatoriedade de um sinal. De acordo com a teoria da informação, ela está ligada a quantidade de informação disponível no sinal Aydın et al. (2009).

A entropia de *Shannon* é o número de dígitos binários necessários para codificar uma mensagem, que depende dos símbolos e de sua frequência Shannon (1948). A equação de *Shannon* para o cálculo da entropia é definida por:

$$H(x) = - \sum_{i=0}^{N-1} p_i(x) \log_2 p_x \quad (1)$$

onde, p_x representa a distribuição de probabilidade da variável x e $p_i(x)$ é a probabilidade do *i*-ésimo resultado para a variável x . A utilização desse descritor para sinais decompostos com a transformada *wavelet* é muito eficiente para a identificação de características únicas dos sinais. Para a representação dos sinais de corrente no domínio tempo-escala, foi calculado o valor da entropia de *Shannon*. Este indicador reflete a distribuição de energia contida num sinal evidenciando características específicas dos mesmos.

Descritores estatísticos - Sinais constituídos por uma quantidade muito grande de variáveis, podem ser representados por um número reduzido de descritores que evidenciem características intrínsecas dos mesmos. Uma forma de redução e representação de sinais é a utilização de parâmetros estatísticos que estejam associados as principais fontes de variabilidade do processo. De forma geral, a escolha dos parâmetros estatísticos para representação dos sinais para fins de construção de modelos, deve considerar que os modelos devem apresentar bom desempenho para falhas incipientes, sem, contudo, ter seu desempenho degradado para falhas consolidadas (Lei, 2016). Os parâmetros estatísticos utilizados foram o valor médio, curtose, desvio padrão, assimetria, raiz média quadrática e fator de crista.

4.7 *Descrição das Instalações Elétricas onde foi obtido o conjunto de dados*

Os dados utilizados nesse estudo foram obtidos em uma planta de celulose onde é adotado o aterramento do neutro por RAV. Todas as cargas em baixa tensão são alimentadas por transformadores de 2 MVA de potência, na configuração triângulo-estrela, sendo o primário alimentado por 13.800V e o secundário com uma tensão de 460V.

Os barramentos de baixa tensão (460V) saem do transformador e são conectados diretamente ao Centro de Distribuição de Cargas (CDC), onde o neutro (centro da estrela secundária) é ligado a dois resistores de 90 Ω associados em série, totalizando 180 Ω interligando o neutro a terra (RAV). Nesse sistema há um circuito que supervisiona a tensão sobre o conjunto resistivo de aterramento indicando quando há circulação de corrente de fuga à terra em alguma das cargas alimentadas por essa instalação. Tal indicação é realizada por um *led* que acende nesse mesmo painel.

Os resistores que conectam o centro da estrela do transformador ao terra, limitam a corrente de fuga à terra em no máximo 1,47 A conforme apresentado na Equação 2.

$$I_{Neutro} = \frac{V_{Fase-Neutro}}{R_{Neutro}} \rightarrow \frac{265V}{180\Omega} = 1,47A \quad (2)$$

onde I_{Neutro} é a corrente máxima que circulará no neutro, $V_{Fase-Neutro}$ é a tensão de fase do circuito (tensão de linha 460 V dividido por $\sqrt{3}$) e R_{Neutro} é o valor total da resistência que interliga o neutro do sistema à terra. Essa corrente irá variar, dependendo da impedância da falha. Por exemplo, caso o contato da fase com a terra tenha uma conexão ruim, com impedância de 120 Ω , haverá uma corrente de fuga à terra da ordem de 885 mA (Equação 3).

$$I_{Neutro} = \frac{265V}{180\Omega + 120\Omega} = 0,885A = 885mA. \quad (3)$$

Essa configuração de falha, é uma das condições mais difíceis para a localização da fuga à terra com a realização da medição direta de corrente utilizando amperímetro nos cabos de saída dos motores, pois, para valores nessa ordem de grandeza, as correntes de fuga podem ser facilmente confundidas com as correntes capacitivas intrínsecas da instalação, conforme previamente explicado na Seção 3.

O CDC alimenta dois Centros de Controle de Motores (CCM), que acionam 148 motores, e caso uma dessas cargas apresente fuga à terra, o sistema do CDC indica a ocorrência da falha elétrica. Contudo, como as gavetas citadas anteriormente não possuem sistema de detecção de fuga à terra individual, não é possível identificar a(s) carga(s) em que a falha ocorreu. Tal configuração de instalação é comum em sistemas antigos, e muitas vezes a instalação de dispositivos para identificação de fuga à terra individuais é economicamente inviável, pois os custos de implantação são elevados.

4.8 Aquisição do Sinais de Fuga à Terra

A primeira etapa para a implementação dos sistemas para identificação dos motores com fuga à terra, foi a coleta dos sinais de correntes de sequência zero. As medições das corrente de sequência zero foram realizadas inserindo uma garra de medição de corrente nos cabos que saem dos alimentadores do CCM, envolvendo todas os condutores, simultaneamente.

O CCM é constituído por diversos compartimentos que são responsáveis por realizar o acionamento dos motores denominados “demarradores” ou “gavetas”. Cada gaveta é composta por elementos de força, controle e proteção (contatores, disjuntores, relés térmicos, fusíveis) responsáveis por proteger e ligar/desligar estes equipamentos. Cabos de alimentação são conectados na saída dos demarradores, alimentando os motores em campo.

O kit de medição utilizado neste trabalho foi constituído por um oscilógrafo digital de alto processamento e uma garra de medição de corrente com capacidade de até 50 A.

Ao todo foram realizadas 76 medições úteis, sendo 67 sinais de motores em condições normais e 9 com fuga à terra. Cada medição foi realizada por um período de 10 segundos a uma taxa de amostragem de 100 kHz, totalizando 1000001 pontos amostrais por medição.

Os dados foram coletados entre dezembro de 2019 e novembro de 2020, sendo que as medições nas condições de fuga à terra, foram obtidas no momento em que ocorriam as falhas, de forma natural, ao longo do processo de

funcionamento da planta. O procedimento utilizado para a certificação que, de fato, ocorreu uma falha à terra, consistiu na medição das correntes de todos os motores conectados ao CCM em questão, juntamente com o desligamento voluntário das cargas, quando o equipamento podia ser desligado.

4.9 Segmentação dos Sinais

Cada arquivo com os sinais nas condições normal e falha, foi segmentado em 100 amostras com 10000 pontos. As amostras correspondentes a uma mesma condição (normal ou falha) foram reunidas para constituírem as classes que compõem o conjunto de dados. Dessa forma, as classes normal e falha à terra, possuem, respectivamente, 6700 e 900 amostras, sendo o conjunto completo constituído por 7600 amostras e 10000 variáveis por amostra, como indicado na Tabela 1. Como pode ser notado, existe um desbalanceamento entre o número de amostras das classes “normal” e “falha”.

Tabela 1. Divisão de pontos amostrais por classe

Classe	Nº de Amostras	Nº de Variáveis
Normal	6700	10000
Falha	900	10000

4.10 Extração de Características

Para a realização dos experimentos, os sinais foram decompostos com os filtros *wavelet Daubechies (Db4)*, *Coiflet (Coif1)* e *Symlet (Sym8)*, em todos os níveis de resolução possíveis. Três abordagens para a extração de descritores foram utilizadas. Na primeira, os descritores foram extraídos utilizando os parâmetros estatísticos no domínio do tempo apresentados na Seção 4.6. Na segunda abordagem, os mesmos parâmetros estatísticos foram utilizados para a extração de descritores a partir dos coeficientes resultantes da decomposição dos sinais pelos filtros *wavelet*. Por último, foi calculado o valor da entropia de *Shannon* também descrito na Seção 4.6 a partir dos coeficientes *wavelet*.

4.11 Treinamento do Classificador

Após a segmentação e normalização dos sinais, o conjunto foi particionado em subconjuntos de validação (70%) e teste (30%) das amostras. Os modelos de classificação foram obtidos utilizando os algoritmos de classificação *DA*, *SVM*, *k-NN* e *RF* tanto para os dados no domínio do tempo, quanto no domínio *wavelet*. Para avaliar o desempenho dos modelos na etapa de validação foi utilizado método *k-Fold Cross-Validation* com 10 subconjuntos. Para avaliação do desempenho na etapa de testes foi utilizada a métrica *f1-score* e acurácia.

Para a construção dos modelos foi necessário a configuração de alguns parâmetros dos algoritmos de classificação. As configurações utilizadas em cada caso estão descritas a seguir.

- *k-NN*: os valores de *k* avaliados no algoritmo foram *k = 1, 3 e 5*. A métrica de distância utilizada foi a euclidiana.

- *DA*: Utilizado como tratamento preditor de covariância o discriminante linear.
- *SVM*: Por definição de projeto, foi utilizado o *kernel* do tipo gaussiano.
- *RF*: Para floresta aleatória foi utilizada uma coleção de 100 árvores de classificação para o treinamento do algoritmo.

5. RESULTADOS

Essa seção apresenta o desempenho dos modelos de classificação construídos com os descritores estatísticos no domínio do tempo e no domínio *wavelet*, e com a entropia de Shannon.

5.1 Resultado da Classificação com os Modelos Construídos com Descritores Estatísticos no Domínio do Tempo

Nessa abordagem, foram extraídos seis descritores estatísticos dos sinais que constituíram as amostras do conjunto de dados: valor médio, curtose, desvio padrão, assimetria, valor *RMS* e fator de crista. Na Tabela 2 são apresentadas as dimensões das matrizes, original (coluna 2) e após a extração dos descritores estatísticos (coluna 3), por classe. Pode-se observar, inicialmente, que a estratégia de extrair descritores a partir de parâmetros estatísticos, apresenta a vantagem visível da redução do número de variáveis que serão utilizadas para a construção dos modelos.

Tabela 2. Dimensões das matrizes por classe

Classes	Matrizes de Atributos	
	Original	Estatístico
Normal	6700x10000	6700x6
Falha	900x10000	900x6

A Tabela 3 mostra os resultados dos testes dos modelos de classificação. Todos os classificadores, em todas as configurações de teste obtiveram *f1-score* superior a 80%, sendo que o melhor resultado global foi alcançado pela RF e pelo k-NN com k=1, com 88,96% de *f1* e 96,2% de acurácia em ambos, e os piores foram obtidos com k-NN em k=3 e 5, com valor de 82,11% de *f1*. O menor valor acurácia foi obtida com o classificador DA, com 94,33%.

Tabela 3. Resultado classificação no domínio do tempo com descritores estatísticos

Classificador	Accu.	f1
DA	94,33	84,78
1-NN	96,2	88,96
3-NN	95,45	82,11
5-NN	95,45	82,11
SVM	95,45	82,11
RF	96,2	88,96

Analisando os resultados da classificação obtidos utilizando os descritores estatísticos no domínio do tempo, nota-se que, de forma geral, os modelos construídos apresentam boa capacidade de predição, mesmo considerando o desbalanceamento das classes.

5.2 Resultado da Classificação dos Modelos Construídos com a Utilização de Descritores Estatísticos no Domínio Wavelet

Nessa etapa dos experimentos, os sinais de corrente foram decompostos utilizando os filtros *wavelet* das famílias,

Daubechies-4 (Db4), *Symlet-8 (Sym8)* e *Coiflet-1 (Coif1)*. Para os três filtros os sinais foram decompostos até o nono nível de resolução. Dessa forma, o conjunto de dados após a decomposição, ficou constituído pelos coeficientes de aproximação (ca1), e pelos coeficientes de detalhe de todos os níveis de resolução do primeiro até o nono (cd1 a cd9). Após a decomposição, foram extraídos os mesmos seis descritores estatísticos utilizados na etapa anterior; valor médio, curtose, desvio padrão, assimetria, valor *RMS* e fator de crista.

As Tabelas 4 e 5 apresentam os resultados da classificação dos algoritmos *DA*, *SVM*, *RF* e *k-NN*. A coluna 2 (Coef.) indica o nível de resolução, sendo "ca" e "cd" referentes a coeficiente de aproximação e coeficiente de detalhe, respectivamente. O melhor resultado global foi alcançado com o classificador *RF* no coeficiente de detalhe 2 (cd2), com o filtro *Coiflet 1*, com 98,82% de acurácia e 97,04% de *f1*. O pior desempenho foi obtido com o modelo *DA*, o coeficiente de detalhe 7 (cd7), com o filtro *Coiflet 1*, com 89,30% de acurácia e 65,95% de *f1*.

Tabela 4. Resultados de desempenho nos testes dos classificadores DA, SVM e RF com descritores estatísticos obtidos dos coeficientes no domínio tempo-escala

Clas.	Coef.	Db4		Sym8		Coif1	
		Accu.	f1	Accu.	f1	Accu.	f1
DA	ca1	95,00	85,66	95,26	86,84	94,91	85,76
	cd1	98,38	95,87	98,90	97,26	98,73	96,80
	cd2	98,38	95,87	98,60	96,46	98,82	97,03
	cd3	98,38	95,87	98,95	97,38	98,60	96,46
	cd4	98,38	95,87	98,73	96,80	98,55	96,34
	cd5	97,02	92,05	98,07	95,02	97,37	93,01
	cd6	93,99	82,52	94,17	83,10	94,39	83,77
	cd7	89,78	66,63	89,87	66,92	89,30	65,95
	cd8	92,41	77,56	92,85	78,76	92,98	78,93
	cd9	95,48	87,90	95,61	87,67	95,66	87,92
SVM	ca1	96,90	92,04	97,06	92,33	97,11	92,43
	cd1	98,38	95,87	98,90	97,27	98,73	96,80
	cd2	98,38	95,87	98,60	96,46	98,77	96,93
	cd3	98,38	95,87	98,90	97,27	98,60	96,46
	cd4	98,38	95,87	98,60	96,70	98,51	96,24
	cd5	97,94	94,69	98,25	95,57	97,68	94,00
	cd6	97,24	92,73	98,23	95,84	98,25	95,59
	cd7	94,91	85,70	94,96	85,80	94,91	85,53
	cd8	93,33	80,24	94,47	83,69	94,91	85,41
	cd9	96,45	90,34	96,67	91,12	96,67	91,19
RF	ca1	97,28	92,94	96,84	91,92	97,02	92,23
	cd1	98,38	95,87	98,90	97,27	98,68	96,70
	cd2	98,38	95,87	98,60	96,46	98,82	97,04
	cd3	98,38	95,87	98,95	97,38	98,60	96,46
	cd4	98,38	95,87	98,73	96,80	98,55	96,34
	cd5	97,98	94,80	98,07	95,16	97,68	94,00
	cd6	97,46	93,33	98,40	96,12	98,46	96,15
	cd7	94,87	85,94	97,12	84,23	94,61	85,10
	cd8	94,04	83,10	94,43	84,50	94,82	85,62
	cd9	96,32	90,37	96,58	91,15	96,62	91,22

Os modelos construídos com os descritores estatísticos no domínio *wavelet*, apresentaram melhor desempenho que os modelos obtidos com descritores estatísticos no domínio do tempo. Esse resultado demonstra que a utilização da transformada *wavelet* pode extrair informações importantes para a modelagem, que não são reveladas pelos descritores no domínio do tempo.

Tabela 5. Resultados de desempenho nos testes do classificador k-NN (k=1, 3 e 5) com descritores estatísticos obtidos dos coeficientes no domínio tempo-escala

Clas.	Coef.	Db4		Sym8		Coif1	
		Accu.	f1	Accu.	f1	Accu.	f1
1-NN	ca1	94,87	87,32	95,04	88,08	95,44	88,56
	cd1	97,02	92,79	97,85	94,86	97,59	94,22
	cd2	96,40	91,96	97,89	94,83	97,44	93,95
	cd3	96,80	92,28	97,11	93,18	96,67	92,07
	cd4	96,49	91,26	96,75	92,10	96,58	91,73
	cd5	95,88	89,80	95,66	89,75	94,91	87,78
	cd6	95,39	88,51	96,32	91,01	95,75	89,66
	cd7	90,31	75,05	89,91	74,70	90,48	76,25
	cd8	89,96	74,59	90,57	76,31	90,7	77,22
cd9	93,95	85,17	94,47	86,77	94,56	87,02	
3-NN	ca1	96,10	90,05	96,36	90,56	96,49	90,58
	cd1	98,16	95,32	98,64	96,64	98,68	96,70
	cd2	98,20	95,45	98,38	95,90	98,55	96,39
	cd3	98,07	95,11	98,55	96,39	98,20	95,47
	cd4	97,15	92,53	97,76	94,30	97,63	94,00
	cd5	96,71	91,45	97,02	92,50	96,75	91,69
	cd6	95,88	89,29	96,40	91,08	96,84	92,00
	cd7	92,72	78,78	91,93	75,88	92,28	78,06
	cd8	92,76	79,04	93,77	81,85	93,25	80,79
cd9	95,88	89,29	96,01	89,69	96,40	90,79	
5-NN	ca1	96,00	90,76	96,40	90,49	96,49	90,51
	cd1	98,29	95,63	98,82	97,05	98,68	96,69
	cd2	98,29	95,45	98,38	95,88	98,51	96,26
	cd3	98,07	95,08	98,06	96,49	98,07	95,11
	cd4	97,11	92,35	97,89	94,61	97,76	94,32
	cd5	96,80	91,73	96,93	92,19	96,93	92,03
	cd6	95,88	89,22	96,58	91,15	96,84	91,86
	cd7	92,32	77,00	92,46	76,75	92,15	76,98
	cd8	93,11	79,54	93,42	79,63	93,64	81,50
cd9	95,75	88,81	96,17	89,82	96,18	90,00	

Tabela 6. Resultados de desempenho nos testes dos classificadores DA, SVM e RF com descritores de entropia de Shannon obtidos dos coeficientes no domínio tempo-escala

Clas.	Coef.	Db4		Sym8		Coif1	
		Accu.	f1	Accu.	f1	Accu.	f1
DA	ca1	93,29	79,72	93,60	80,94	93,55	80,52
	cd1	98,77	96,92	98,68	96,69	98,68	96,69
	cd2	98,82	97,04	98,95	97,39	98,60	96,47
	cd3	98,86	97,15	98,46	96,10	98,55	96,34
	cd4	98,60	96,46	98,64	96,58	98,51	96,22
	cd5	97,11	92,32	97,02	92,05	97,02	92,05
	cd6	95,00	85,24	95,48	86,97	95,75	87,92
	cd7	88,82	52,90	88,60	50,86	88,99	53,64
	cd8	89,39	56,83	89,91	61,35	89,39	56,83
cd9	93,73	80,27	92,94	76,83	93,46	79,16	
SVM	ca1	93,42	80,66	93,68	81,28	93,82	81,71
	cd1	98,77	96,92	98,68	96,69	98,68	96,69
	cd2	98,86	97,15	98,99	97,49	98,64	96,57
	cd3	98,86	97,15	98,46	96,10	98,55	96,34
	cd4	98,60	96,46	98,68	96,69	98,51	96,22
	cd5	97,24	92,62	97,15	92,36	97,11	92,26
	cd6	97,72	94,06	97,54	93,55	97,72	94,06
	cd7	93,20	79,46	93,20	78,90	93,82	81,47
	cd8	91,89	72,17	91,62	71,84	92,24	74,16
cd9	95,13	86,40	94,78	85,36	95,22	86,80	
RF	ca1	88,82	73,92	89,74	76,03	90,09	75,88
	cd1	97,89	94,91	97,32	93,62	97,98	95,12
	cd2	97,76	94,65	97,85	94,89	97,54	94,10
	cd3	97,63	94,35	97,37	93,60	97,19	93,30
	cd4	97,59	94,14	97,15	93,25	97,41	93,71
	cd5	94,39	86,68	95,31	88,41	94,21	86,18
	cd6	95,75	89,89	96,40	91,13	96,18	90,85
	cd7	89,30	74,12	89,78	73,69	89,30	74,45
	cd8	87,76	69,77	86,80	67,71	87,54	70,27
cd9	92,19	80,81	91,62	79,78	91,14	78,92	

6. CONCLUSÃO

5.3 Resultado da Classificação com os Modelos Construídos com Descritores Obtidos pela entropia de Shannon no domínio Wavelet

Na terceira etapa dos experimentos, o valor da entropia foi calculada para cada amostra do conjunto de dados constituída pelos coeficientes *wavelet* após a decomposição com os filtros *Db4*, *Sym8* e *Coif1*.

As Tabelas 6 e 7 apresentam os resultados dos classificadores implementados com os descritores obtidos pela entropia de *Shannon*. O melhor resultado global foi alcançado com o classificador *SVM* no coeficiente de detalhe 2 (cd2), com o filtro *Symlet 8*, com 98,99% de acurácia e 97,49% de f1. O pior desempenho foi obtido com o modelo *DA*, o coeficiente de detalhe 7 (cd7), com o filtro *Symlet 8*, com 88,60% de acurácia e 50,86% de f1.

É possível observar o mesmo comportamento dos testes da Seção 5.2 onde os coeficientes de detalhe dos primeiros níveis de resolução apresentam maiores valores de *f1-score* e vão reduzindo até o oitavo nível. Os piores resultados são observados nos níveis de resolução cd7 e cd8 em todos os modelos e filtros.

Os resultados dos modelos construídos com o descritor obtido pela entropia de *Shannon*, são comparáveis aos resultados dos modelos construídos com os descritores estatísticos no domínio *wavelet*.

Este trabalho apresenta uma proposta para identificação de falhas em motores com fuga à terra um uma instalação com aterramento do neutro por resistor de alto valor, aplicando modelos de aprendizado de máquina. O conjunto de dados utilizado é constituído de sinais de corrente de sequência zero de motores em condição de operação normal e com falha à terra. Para a classificação, foram construídos modelos com descritores extraídos no domínio do tempo e no domínio *wavelet*, utilizando diferentes filtros, em diferentes níveis de decomposição. Os resultados obtidos mostram que os classificadores construídos com os descritores do domínio *wavelet* apresentaram desempenho significativamente superior em relação aos modelos construídos no domínio do tempo. O melhor desempenho foi obtido com o algoritmo *SVM* em conjunto com o filtro *Sym8* quando utilizado os descritores extraídos pela entropia de *Shannon* (97,49% de *f1-score* e 98,99% de acurácia). Esses resultados podem ser justificados pela boa performance da transformada *wavelet* na análise de sinais não-estacionários, como é o caso dos sinais de corrente de sequência zero. Pode ser observado ainda que os melhores resultados foram obtidos com os modelos construídos nos primeiros níveis de decomposição e isso se deve ao fato de os primeiros níveis de detalhe estarem associados aos maiores valores de frequência dos sinais, onde geralmente estão as informações mais relevantes nas correntes de curto-circuito. De uma forma geral, todos os testes em sua grande maioria

Tabela 7. Resultados de desempenho nos testes do classificador k-NN (k=1, 3 e 5) com descritores de entropia de Shannon obtidos dos coeficientes no domínio tempo-escala

Clas.	Coef.	Db4		Sym8		Coif1	
		Accu.	f1	Accu.	f1	Accu.	f1
1NN	ca1	88,82	73,92	89,69	75,97	90,09	75,88
	cd1	97,85	94,81	97,32	93,62	97,98	95,12
	cd2	97,76	94,65	97,85	94,89	97,54	94,10
	cd3	97,63	94,35	97,37	93,60	97,19	93,30
	cd4	97,59	94,14	97,15	93,25	97,41	93,71
	cd5	94,39	86,68	95,31	88,41	94,21	86,18
	cd6	95,75	89,89	96,40	91,13	96,18	90,85
	cd7	89,30	74,12	89,74	73,63	89,30	74,45
	cd8	87,76	69,77	86,80	67,71	87,54	70,27
3NN	ca1	92,76	80,32	93,25	82,00	92,81	80,40
	cd1	97,85	94,81	97,32	93,62	97,98	95,12
	cd2	97,76	94,65	97,85	94,89	97,54	94,10
	cd3	97,63	94,35	97,37	93,60	97,19	93,30
	cd4	97,59	94,14	97,15	93,25	97,41	93,71
	cd5	94,39	86,68	95,31	88,41	94,21	86,18
	cd6	95,75	89,89	96,40	91,13	96,18	90,85
	cd7	89,30	74,12	89,74	73,63	89,30	74,45
	cd8	87,76	69,77	86,80	67,71	87,54	70,27
5NN	ca1	93,51	81,68	94,25	84,01	93,77	82,63
	cd1	97,85	94,81	97,32	93,62	97,98	95,12
	cd2	97,76	94,65	97,85	94,89	97,54	94,10
	cd3	97,63	94,35	97,37	93,60	97,19	93,30
	cd4	97,59	94,14	97,15	93,25	97,41	93,71
	cd5	94,39	86,68	95,31	88,41	94,21	86,18
	cd6	95,75	89,89	96,40	91,13	96,18	90,85
	cd7	89,30	74,12	89,74	73,63	89,30	74,45
	cd8	87,76	69,77	86,80	67,71	87,54	70,27
cd9	92,19	80,81	91,54	79,63	91,10	78,91	

apresentaram valores das métricas aplicadas superiores a 90%, o que demonstra a boa capacidade de separação dos dados por meio do conjunto de técnicas utilizado.

Os resultados obtidos mostram que, em sistemas elétricos com aterramento do neutro com resistor de alto valor ôhmico, é possível realizar a detecção de falhas à terra utilizando sinais de corrente de sequência zero e técnicas de aprendizado de máquina com sinais de corrente que podem ser facilmente obtidos sem interferir no processo produtivo.

REFERÊNCIAS

Aydın, S., Saraoğlu, H.M., and Kara, S. (2009). Log energy entropy-based eeg classification with multilayer neural networks in seizure. *Annals of biomedical engineering*, 37(12), 2626.

Baldwin, T., Renovich, F., and Saunders, L.F. (2003). Directional ground-fault indicator for high-resistance grounded systems. *IEEE Transactions on Industry Applications*, 39(2), 325–332.

Blackburn, J.L. and Domin, T.J. (2015). *Protective relaying: principles and applications*. CRC press.

Bowen, J. (2011). *Grounding for Industrial Systems*. Institute of Electrical and Electronics Engineers (IEEE).

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. doi:10.1023/A:1010933404324. URL <https://doi.org/10.1023/A:1010933404324>.

Cai, X.W., Jin, T., Jia, H.B., and Qi, Y.L. (2014). The design of location system for single-phase grounded fault in distribution network based on zero-sequence current. In *Advanced Materials Research*, volume 986, 1677–1681. Trans Tech Publ.

Cortes, C. and Vapnik, V. (1995). Support-vector networks. In *Machine Learning*, 273–297.

Costa, P.F. (2014). Capítulo iii-avanços na especificação e aplicação dos resistores de aterramento do neutro dos sistemas elétricos industriais em baixa tensão. *Revista O Setor Elétrico, Setembro 2014*.

Dibennardo, R. and Taylor, J.V. (1982). Classification and misclassification in sexing the black femur by discriminant function analysis. *American journal of physical anthropology*, 58(2), 145–151.

Duda, R.O., Hart, P.E., and Stork, D.G. (2001). *Pattern Classification*. Wiley, New York, 2 edition.

Hussey, M.R. and Drzymala, S. (2019). Ground fault protection applications in low voltage motor control systems for process industries. In *2019 IEEE-IAS/PCA Cement Industry Conference (IAS/PCA)*, 1–8. IEEE.

Krishnaiah, P.R. and Kanal, L.N. (1982). *Classification, Pattern Recognition, and Reduction of Dimension*, volume 1 of *Handbook of Statistics*. North-Holland, first edition.

Lei, Y. (2016). *Intelligent fault diagnosis and remaining useful life prediction of rotating machinery*. Butterworth-Heinemann.

Lubich, D. (1997). High resistance grounding and fault finding on three phase three wire (delta) power systems. In *1997 IEEE Annual Textile, Fiber and Film Industry Technical Conference*, 10–5. IEEE.

Morrison, D.G. (1969). On the interpretation of discriminant analysis. *Journal of marketing research*, 6(2), 156–163.

Nandi, A. and Ahmed, H. (2019). *Condition Monitoring with Vibration Signals: Compressive Sampling and Learning Algorithms for Rotating Machine*, chapter 5, 79–114. John Wiley and Sons, Ltd. doi: <https://doi.org/10.1002/9781119544678.ch5>. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119544678.ch5>.

Shannon, C.E. (1948). A mathematical theory of communication. *The Bell system technical journal*, 27(3), 379–423.

Topolanek, D., Lehtonen, M., Adzman, M.R., and Toman, P. (2015). Earth fault location based on evaluation of voltage sag at secondary side of medium voltage/low voltage transformers. *IET Generation, Transmission & Distribution*, 9(14), 2069–2077.

Valdes, M., Papallo, T., and Premerlani, B. (2006). Ground fault location in low-voltage high-resistance grounded systems via the single-processor concept for circuit protection. In *Conference Record of 2006 Annual Pulp and Paper Industry Technical Conference*, 1–8. IEEE.