

Teoria da Informação no Desenvolvimento de Redes Neurais para Diagnóstico de Falhas em Transformadores Defasadores

Felipe Munaro Lima*, Andre Abel Augusto**, Vitor Hugo Ferreira**

*Programa de Pós-Graduação em Engenharia Elétrica e
Telecomunicações, Universidade Federal Fluminense, Niterói – RJ.

**Departamento de Engenharia Elétrica, Universidade Federal
Fluminense, Niterói – RJ,
(e-mail: aaaugusto@id.uff.br)

Abstract: Machine Learning techniques for fault location and diagnosis in phase shift power transformers have been proposed in the literature. However, most of them adopt loss functions that consider a few moments of the error distribution, not adequately extracting the information stored in data available for diagnosis. Information theoretic criteria can overcome this limitation, resulting in better training and inference, and consequently, enhancing the fault analysis. This work presents an information theoretic learning approach to fault analysis in indirect symmetrical phase-shift transformers. The methodology consists of training an artificial neural network for fault detection using information theoretic cost functions. A comparative study with the traditional mean square error cost function will also be present.

Resumo: Técnicas de Aprendizado de Máquina para localização e diagnóstico de falhas em transformadores defasadores vêm sendo propostas na literatura. No entanto, em sua maioria, adotam funções de perda que consideram poucos momentos da distribuição dos erros, não extraindo adequadamente as informações armazenadas nos dados disponíveis para diagnóstico. Critérios teóricos da informação podem superar essa limitação, resultando em melhor treinamento e inferência e, consequentemente, aprimorando a análise de falhas. Este trabalho apresenta uma abordagem de aprendizagem baseada em teoria da informação para análise de falhas em transformadores defasadores indiretos e simétricos. A metodologia consiste em treinar uma rede neural artificial para detecção de falhas usando funções de custo da teoria da informação, sendo feito o estudo comparativo com a função custo erro quadrático médio tradicional.

Keywords: Fault Diagnosis, Phase Shift Power Transformers, Machine Learning, Information Theory, Artificial Neural Network.

Palavras-chave: Diagnóstico de Falhas, Transformadores de Potência Defasadores, Aprendizado de Máquina, Teoria da Informação, Redes Neurais Artificiais.

1. INTRODUÇÃO

Transformadores defasadores (TDs) foram introduzidos pela primeira vez em 1930 (Bhasker, 2015). Desde então é um componente de sistemas de potência. Os TDs introduzem uma diferença de fase entre as tensões dos terminais emissor e receptor do dispositivo, controlando a potência ou integrando redes independentes.

Existem diferentes tipos de TDs, cada um com características construtivas específicas (Verboomen, 2005). O Transformador Defasador Indireto e Simétrico (*indirect symmetrical phase shift transformer* — ISPST) é amplamente utilizado devido à sua construção simples (Bera, 2018). Considerando a importância e o custo de um ISPST, informações sobre o tipo de falha interna nesse equipamento são necessárias para a identificação da unidade defeituosa, avaliação da quantidade de trabalho de reparo necessário, e estudo das falhas.

Pesquisadores têm proposto diversas técnicas para proteção e classificação de falhas em transformadores de potência. Embora haja literatura suficiente que sustente o uso de técnicas de aprendizado de máquina na classificação de falhas em transformadores, poucos trabalhos endereçam o problema de classificação de falhas em ISPST.

O processo de diagnóstico de falta em transformadores pode ser dividido em três etapas: detecção, responsável por separar os dados coletados em dados pré e pós falta; classificação, que identifica nos dados o tipo e eventualmente a causa do defeito; e finalmente a localização, que determina a partir dos dados disponíveis o componente onde a falta ocorreu. Embora os dispositivos eletrônicos inteligentes voltados à proteção das redes e de TDs permitam o diagnóstico de defeitos nos locais em que estão instalados, sistemas de diagnóstico de faltas são importantes em caso de falhas desses dispositivos, identificação errônea do defeito, e no

suporte para a tomada de decisão por parte de operadores (Ferreira et. al, 2016).

De acordo com Ferreira *et al.* (Ferreira *et al.*, 2016) os métodos de diagnóstico de faltas podem ser agrupados em duas abordagens: clássica e inteligente. Abordagens clássicas empregam conceitos da teoria de circuitos tradicional para construir algoritmos de detecção, classificação e localização de falhas. Os métodos inteligentes utilizam técnicas de inteligência computacional, aprendizado de máquinas, sistemas inteligentes, modelagem baseada em dados (Purarjomandlangrudi, 2014), lógica *fuzzy*, redes neurais artificiais (Souza, 2001) e sistemas multiagentes (Davidson, 2006) para realizar a análise das falhas.

A literatura sobre sistemas de diagnóstico de falhas em transformadores é bastante extensa. Behkam et al. usou uma Rede Neural de Regressão Generalizada (GRNN) para detecção de faltas em enrolamentos de transformadores de potência (Benkam, 2022). Da mesma forma, Zimnickas et al. implementou uma proteção para motores de indução utilizando uma Rede neural Convolutiva (CNN) para identificação de defeitos mecânicos e elétricos na máquina (Zimnickas, 2020).

Klomjit et al. realizou a comparação entre Redes Neurais Probabilísticas (PNNs), Redes de Retro propagação (BPNNs) e Máquinas de Vetor de Suporte (SVMs) para detecção de defeitos (Klomjit, 2020). Skowron aplicou uma Rede Neural Auto Organizável (SONN) na identificação de falhas em motores de indução (Skowron, 2020).

O trabalho de Patel *et al.* (Patel, 2018) sugere que a Máquina de Vetor de Relevância desempenha melhor que PNNs e SVMs para diagnóstico de falhas em transformadores. Métodos da teoria de sistemas moderna mostram o uso sistemático da matemática e modelagem de sinais, métodos de identificação e estimação, e de inteligência computacional (Isermann, 2005). Bhasker (Bhasker, 2015) propôs o uso de Transformada Wavelet e Redes Neurais Artificiais para classificação de faltas em PSTs.

Apesar de vasta a literatura sobre o tema, poucos dos trabalhos tratam da investigação de falhas em um componente de complexa manutenção como ISPSTs. Por conter duas unidades trifásicas de transformadores, este dispositivo oferece um desafio maior para a diagnose de falhas.

O emprego de critérios teóricos de informação para aprendizagem, apesar de pouco difundido, podem resultar em melhor treinamento e inferência, por caracterizarem melhor a distribuição de probabilidade dos erros (Príncipe, 2010). Este trabalho apresenta a investigação de critérios estatísticos baseados na teoria da informação para treinamento de redes neurais artificiais para detecção e classificação de falhas em ISPSTs. Os resultados apontam que esses critérios podem apresentar desempenho superior ao dos métodos tradicionais.

2. FUNDAMENTAÇÃO TEÓRICA

2.1 Defeitos em Transformadores ISPSTs

Diferentes causas podem levar a defeitos em ISPSTs em condições operacionais. Correntes de *inrush*, por exemplo, permanecem como sendo um problema, mesmo com os temporizadores de atraso adicionados durante o comissionamento. As correntes de *inrush* surgem após a energização ou acionamento do transformador, ocasionando um pico de corrente, podendo atingir um valor dez a vinte vezes superior do que em condições normais de operação (Rico, 2001). Os temporizadores de atraso permitiram que um único ISPST fosse energizado, porém, quando outro transformador em paralelo é colocado na rede, o primeiro muitas vezes sofre um desarme devido à corrente de *inrush* simpático no relé de neutro (Bladow, 1991). Esse tipo de defeito é denominado energização solidária.

Outra causa conhecida é a saturação ou sobre-excitação, caracterizada por um crescimento atípico do fluxo magnético no núcleo do transformador, quando há uma sobretensão na rede ao qual este está conectado. A consequência deste fenômeno é a elevação da corrente de magnetização, podendo o transformador ser danificado caso não haja um sistema de proteção adequado. Os ISPSTs estão sujeitos a outros defeitos externos, causados por descargas atmosféricas, transitórios de manobra, ou outros danos que não podem ser evitados e precisam de proteção. Esses defeitos, dependendo de sua natureza, apresentam características específicas. Diferentes propostas foram abordadas na literatura visando realizar a discriminação destes casos de defeitos externos (Etumi, 2016).

2.2 Redes Neurais Artificiais

Compreende-se uma rede neural artificial (RNA) por um sistema de processamento massivamente paralelo, constituída de unidades de processamentos simples, capaz de adquirir conhecimento através de experiência, armazená-lo e torná-lo disponível para uso (Haykin, 1999). Diferentes padrões de conexões entre os neurônios originam arquiteturas distintas de RNAs, como a *feedforward*, recorrente, convolutiva, dentre outras.

Denomina-se algoritmo de aprendizagem o procedimento computacional empregado para o aprendizado a partir dos dados. Neste processo, os pesos são modificados em função dos dados apresentados na rede, sendo esse ajuste orientado por uma função de perda ou custo (Haykin, 2009). A função de perda usualmente adotada em problemas de classificação e regressão via RNAs é a de erro quadrático médio (*Mean Square Error* — MSE), dada por:

$$L_2(w) = \frac{1}{N} \sum_{n=1}^N \|y_n - f_w(x_n)\|_2^2 \quad (1)$$

onde y_n é a n -ésima saída desejada, $f_w(x_n)$ é saída produzida pela RNA, w o vetor de peso das sinapses, $\|\cdot\|_2$ a norma 2 (Euclidiana), e N o número de amostras.

O treinamento das redes Perceptron Multicamadas (MLP) é usualmente feito na forma de um problema de otimização, cujo objetivo é a minimização da função de perda total adotada (Bishop, 2006). A maioria das propostas de redes

neurais para diagnóstico de faltas utilizam como função de perdas o erro quadrático médio. Tal critério, por considerar somente o primeiro e o segundo momento da distribuição de probabilidade do erro do modelo, tem acesso limitado à informação presente nessa. Funções de perda probabilísticas tais como a Entropia, Entropia Cruzada (EC), Correntropia e Informação Mútua (Príncipe, 2010) conseguem contornar esse problema, o que pode melhorar o desempenho dos algoritmos ML.

2.3 Teoria da Informação

A teoria da informação foi concebida por Shannon (Shannon, 1949) para resolver o problema de transmissão ótima de mensagens em canais ruidosos. Para tal, a teoria investiga formas de caracterização e codificação da mensagem transmitida e limites para transmissão de mensagem sem erros (Príncipe, 2010). Embora tenha surgido na área de comunicação, a teoria é derivada da aplicação de conceitos probabilísticos. Tanto a mensagem quanto o ruído de comunicação podem ser entendidos como variáveis aleatórias com certas distribuições de probabilidades. Assim sendo, do ponto de vista matemático, a teoria da informação estuda a caracterização e quantificação da informação (ou da incerteza) contida nas distribuições de probabilidade das variáveis aleatórias de interesse (Rubinstein, 2008). Diferentes medidas de informação e de incerteza podem ser utilizadas para esse propósito. Neste contexto, a entropia reflete o grau de incerteza em relação aos valores de uma variável aleatória, mesmo após a observação de seus possíveis resultados. Note que quanto mais frequente a ocorrência de um certo valor da variável aleatória, menor é a informação obtida sobre a variável quando esse valor é observado. Este conceito pode ser traduzido na definição de quantidade de informação proposta por Shannon:

$$I_k = \log_2 \frac{1}{p_k} = -\log_2 p_k \quad (2)$$

onde I_k é a quantidade de informação trazida pela observação do valor $X = x_k$ e p_k a probabilidade de ocorrência desse valor. A partir de (2), a entropia de Shannon é definida como o valor esperado da quantidade de informação tomado em relação à variável aleatória discreta X :

$$H(X) = -E_X[\log_2 p(X)] = -\sum_k p_k \log_2 p_k \quad (3)$$

onde, $p(X)$ é a distribuição de probabilidade de X .

A entropia mede a quantidade de informação presente em uma variável aleatória. Contudo, quando mais de uma variável aleatória está presente, diferentes medidas de informação podem ser construídas para diversos propósitos. Sejam X e Y duas variáveis aleatórias. A entropia conjunta $H(X, Y)$ mede a incerteza associada às variáveis X e Y conjuntamente. Já a entropia condicionada $H(X|Y)$ quantifica a incerteza de X dada a ocorrência da variável Y . Por fim, a informação mútua $I(X, Y)$ quantifica a informação compartilhada pelas variáveis X e Y , sendo dada por:

$$I(X, Y) = \sum_i \sum_k p_{XY}(x_k, y_i) \log_2 \left(\frac{p_{XY}(x_k, y_i)}{p_X(x_k)p_Y(y_i)} \right) \quad (4)$$

onde p_{XY} é a distribuição de probabilidade conjunta de X e Y , p_X e p_Y as marginais de X e Y , respectivamente.

Por fim, define-se a entropia cruzada como:

$$H(p, q) = -E_p[\log_2 q] = -\sum_k p_k \log_2 q_k \quad (5)$$

onde $p_k = p(x_k)$, $q_k = q(x_k)$.

As distribuições p e q em (5) são duas distribuições de probabilidade sobre o conjunto de valores possíveis para a variável aleatória X . Se p é a distribuição de X e q a assumida, então a entropia cruzada $H(p, q)$ pode ser entendida como a incerteza de X quando se presume ser q a sua distribuição verdadeira. Note que esta interpretação é útil para problemas de classificação e regressão, em que p é a distribuição real dos dados e q aquela segundo o modelo.

Além das medidas de incerteza, indicadores de dissimilaridade entre distribuições de probabilidade são poderosas ferramentas para caracterização e medição de informação. Um desses indicadores é a Divergência de Kullback—Leibler (DKL) ou entropia relativa. Essa divergência apresenta uma série de propriedades atrativas, além de guardar íntima relação com a entropia, com a informação mútua, e com a estimação de máxima verossimilhança. Sejam p e q duas distribuições de probabilidade sobre um conjunto de valores possíveis para a variável aleatória X . A divergência de Kullback-Leibler é definida pela expressão:

$$D_{KL}(p||q) = \sum_k p_k \log_2 \frac{p_k}{q_k} \quad (6)$$

A divergência de Kullback-Leibler pode ser reescrita na forma (7), onde $H(p)$ é a entropia de X segundo p e $H(p, q)$ a entropia cruzada entre p e q .

$$D_{KL}(p||q) = H(p, q) - H(p) \quad (7)$$

Com base em (7) a DKL pode ser compreendida como o excesso de incerteza devido a assumir-se q como distribuição verdadeira de X quando p é a sua distribuição verdadeira. As medidas de informação apresentadas permitem o desenvolvimento de critérios alternativos para treinamento de redes neurais, que consideram a quantidade de informação presentes nos dados. Seja y_n a saída desejada para a entrada x_n e $\hat{y}_n = f_w(x_n)$ a correspondente saída gerada pelo modelo. Portanto, $e_n = y_n - \hat{y}_n$ representa o erro do modelo. A função de custo da entropia do erro é dada por:

$$L_H(w) = -E_p[\log_2 p(e)] \quad (8)$$

onde $p(e)$ é a distribuição de probabilidade dos erros.

A função de custo de entropia cruzada visa minimizar a entropia cruzada entre as distribuições de Y (verdadeira) e de \hat{Y} (distribuição construída pelo modelo), isto é:

$$L_{CE}(w) = -E_p[\log_2 q_w(y)] \quad (9)$$

Com base nesse critério, o melhor ajuste dos pesos é aquele para o qual a adoção da distribuição $q_w(y)$ como distribuição dos dados ao invés de $p(y)$ incorre em aumento mínimo de incerteza possível. Note que um dos casos em que isto ocorre é quando $p = q_w(y)$. O critério da minimização da entropia do erro ou da entropia cruzada estabelece que o melhor ajuste de w é aquele que minimiza a respectiva entropia. Propõe-se também a comparação das distribuições $p(y)$ e $q_w(y)$ por meio da divergência de Kullback-Leibler. Assim sendo, a função de perdas é dada por:

$$L_{KL}(w) = -D_{KL}[p(y)||q_w(y)] \quad (10)$$

Note que nos problemas de classificação e regressão a distribuição dos dados observados mantém-se inalterada. Assim sendo, a entropia $H(Y)$ é constante e a minimização da DKL corresponde à minimização da entropia cruzada. Por essa razão costuma-se utilizar os termos divergência de Kullback-Leibler e entropia cruzada como correlatos (Rubinstein, 2008). Um ponto a se destacar é que o cômputo das medidas de informação requer o conhecimento das distribuições das saídas observadas e previstas pelo modelo. Técnicas de estimação não paramétrica de distribuições de probabilidade (Príncipe, 2010), métodos de Monte Carlo (Rubinstein, 2008) ou variantes da entropia de Shannon, como a entropia de Reyni, podem ser adotadas para estimá-las diretamente dos dados (Príncipe, 2010).

3. METODOLOGIA

3.1 Definição dos Atributos

A metodologia desenvolvida para comparação de desempenho entre os modelos treinados com o critério estatístico tradicional (MSE) e aqueles utilizando as funções de perda desenvolvidas a partir da teoria da informação (EC e DKL) é resumida na Figura 6. Em síntese, as oscilografias da corrente de cada fase do ISPST são pré-processadas (remoção de outliers, registros duplicados ou vazios) e as características extraídas realizando-se a seleção dos atributos que constituirão as entradas do modelo. Para cada oscilografia s coletada, obteve-se os seguintes atributos:

- Valor máximo: $\max(s)$;
- Valor mínimo: $\min(s)$;
- Mediana: $\text{med}(s)$;
- Média: $\mu_s = E[s]$;
- Desvio padrão: $\sigma_s = E[(s - \mu)^2]^{1/2}$;
- Assimetria: $E[((s - \mu) / \sigma_s)^3]$;
- Curtose: $E[((s - \mu) / \sigma_s)^4]$.

Repare que os atributos gerados constituem estatísticas das oscilografias. Considerando-se que podem ser compreendidas como séries temporais, outros atributos podem ser construídos a partir dos dados. A Transformada Discreta de Fourier (*Discrete Fourier Transform – DFT*) converte um sinal no

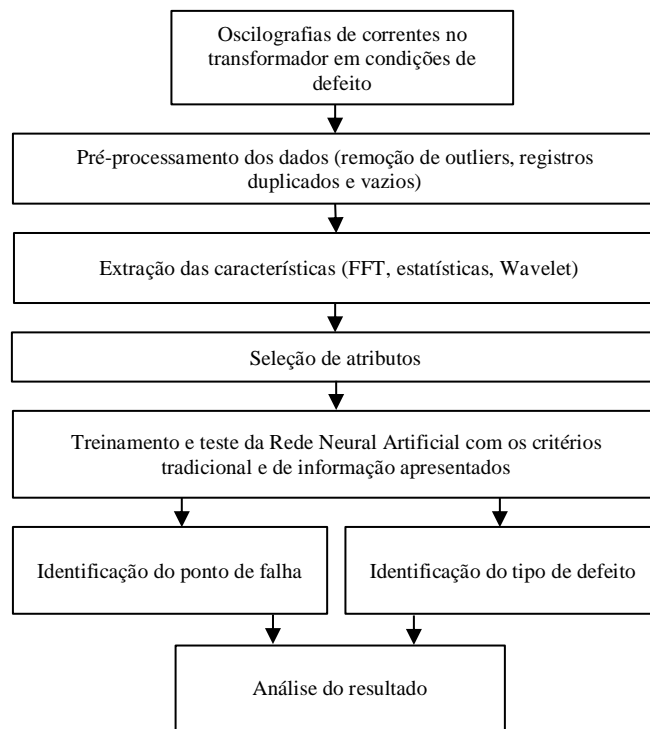


Fig 1 – Fluxograma da metodologia proposta.

domínio do tempo para o domínio da frequência, possibilitando a identificação de novos padrões nos dados. O seu uso para identificação de casos de faltas múltiplas em sistemas de distribuição foi investigado por Devadasu *et al.* (Devadasu, 2016). Na metodologia proposta, as partes real e imaginária das M_h primeiras componentes harmônicas, são adotadas como atributos do vetor de entrada do modelo.

Uma alternativa à DFT é a Transformada *Wavelet* Discreta (*Discrete Wavelet Transform – DWT*), que decompõe o sinal original em um conjunto de *wavelets*, sendo capaz de capturar não só informação de frequência (como a DFT) como de localização no tempo. Essa abordagem justifica-se porque as faltas em ISPSTs produzem sinais transientes, não estacionários (Bhasker, 2015). Além dos atributos anteriores, adotou-se como atributos os coeficientes de aproximação máximos das DWTs de Daubechies de ordens 1 a M_b , isto é, Db1 a Db M_b . Os parâmetros M_h e M_b podem ser definidos durante a etapa de construção do modelo, a depender do nível de ruído e características dos sinais de entrada. Neste trabalho adotou-se $M_h = 40$ (o que corresponde a uma componente fundamental de 60 Hz e componentes harmônicas de até 2400 Hz) e $M_b = 10$.

3.2 Definição e Treinamento do Modelo

O modelo utilizado é a RNA com arquitetura *feedforward* Perceptron de Múltiplas Camadas (*Multilayer Perceptron –*

MLP) implementada com a interface de aplicação Keras. São construídas duas redes: uma para classificação de falta que pode ser superexcitação, magnetização *inrush*, energização solidária (Bronzeado, 1995) e faltas externas com saturação de transformador de corrente e outra para sua localização (unidade série ou de excitação). A entrada do modelo é um vetor cujos elementos são os atributos discutidos na seção 3.1. Para a classificação de falta, adotou-se como saída um vetor binário de 4 posições em que cada elemento corresponde à ocorrência (1) ou não-ocorrência (0) de uma determinada falta. O mesmo procedimento foi adotado para a localização, tendo o vetor de saída associado duas posições. Os modelos propostos são treinados considerando as funções de perda apresentadas na seção II. Para treinamento foi utilizado o otimizador ADAM (Kingma, 2014).

Para evitar que o sobre ajuste do modelo aos dados de treinamento e que o estimador tenha tendência, regularizou-se o mesmo a cada etapa de treinamento removendo-se aleatoriamente cada neurônio e suas conexões de entrada e saída, até que haja melhora do desempenho, se prevenindo desta forma a coadaptação. Esta técnica oferece grandes melhorias em muitas tarefas de classificação em que foi testada (Srivastava, 2014).

3.3 Inferência e Análise dos Resultados

Para avaliar o desempenho dos critérios de informação, será utilizado o número de épocas realizadas pelo algoritmo ADAM como medida da eficiência do treinamento. Para verificar a qualidade do modelo, adotou-se a Acurácia, a Precisão, a Sensibilidade (*recall*) e o F1-Score.

4. TESTES E RESULTADOS

Para realização do teste de comparação entre as funções de perda MSE, entropia cruzada e DKL, para distinção do desempenho para o problema de classificação e localização de falta, utilizou-se oscilografias das correntes diferenciais de cada fase de um ISPST. Esses dados foram retirados da

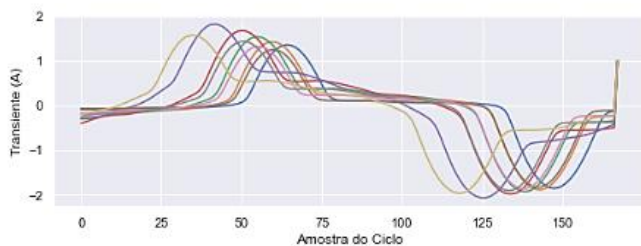
plataforma IEEE *data storage* (Pallav, 2020) e correspondem a simulações do dispositivo no software PDCAD/EMTDC. Cada registro apresenta 167 amostras da oscilografia de um ciclo de corrente, sendo que, para o caso de localização de falta há 46872 registros amostrais, sendo estes exemplos divididos entre defeitos na unidade de série e excitação. E para o caso de identificação do tipo de defeito há 13680 casos de correntes diferenciais, sendo estas divididas em 5 classes de falhas (sobre excitação, corrente de *inrush*, energização solidária e falhas externas).

As figuras 2 e 3 apresentam exemplos de correntes de falta para diversas situações estudadas neste trabalho. A Figura 2, apresenta as oscilografias para sobre excitação (a), *inrush* (b), energização solidária (c) e falhas externas (d). E a Figura 3 apresenta as oscilografias para defeitos no enrolamento de série (a) e de excitação (b). Os gráficos nas figuras apresentam observações de cada tipo de defeito. Cada curva em determinado gráfico corresponde a uma oscilografia observada.

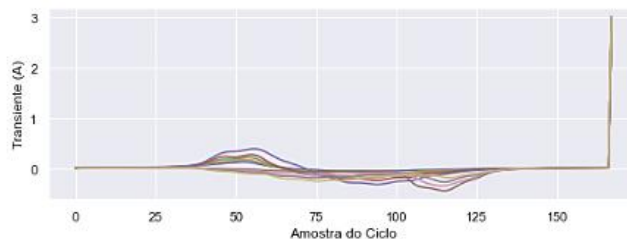
As RNAs construídas para classificação e localização de faltas são densamente conectadas e com quatro camadas: uma de entrada, duas escondidas e a camada de saída. A estrutura é constituída por camadas de entrada e ocultas, contendo 256 unidades de neurônios e função de ativação *ReLU*. A camada de saída da rede classificadora de falhas tem 5 neurônios, e função de ativação do tipo sigmoide logística.

No caso da RNA para localização das falhas utilizou-se 2 neurônios na camada de saída com função de ativação do tipo *softmax* e a arquitetura das redes definida empiricamente, com testes realizando a variação do número de neurônios em cada camada. 20% do conjunto de dados foi separado para o teste dos modelos e os demais foram utilizados para o aprendizado.

Para a reprodução dos resultados de comparação, foi definida a semente do gerador de números aleatórios como constante, permitindo inicialização dos treinamentos em condições semelhantes.



(a)



(b)

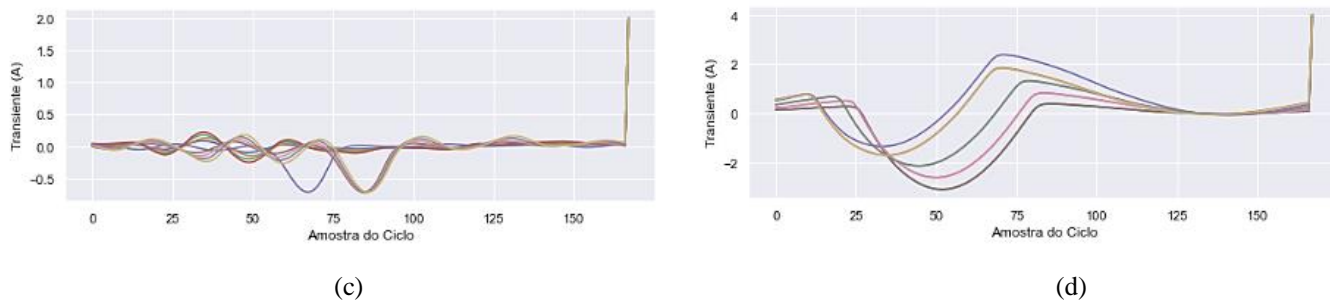


Fig 2 – Amostras das correntes diferenciais transitórias trifásicas de uma fase do circuito (fase a) - identificação do tipo de falha.

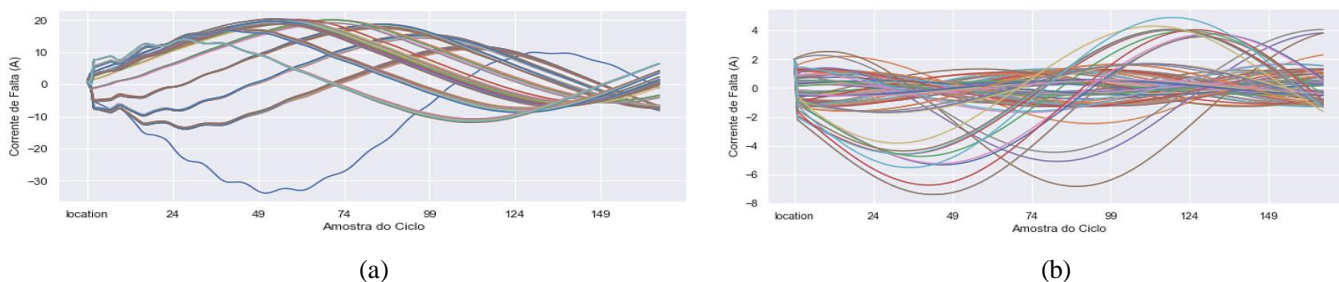


Fig 3 – Amostra da oscilografia das correntes diferenciais transitórias trifásicas de uma fase do circuito (fase a) – localização da falha.

As Tabelas 1 e 2 apresentam os resultados de treinamento para a diagnose e identificação da localização do ponto de falha levando em consideração cada uma das funções de custo. Para a identificação do tipo de falha, o conjunto treinado com a função de perda DKL superou o resultado dos demais classificadores, atingindo um F1-Score de 92% e convergindo 11 épocas antes do que a função de Entropia Cruzada. A diferença de exatidão ao final do treinamento entre as funções de perda MSE e DKL, para este caso, foi de 3,19%. Já para a localização da falha, o modelo que foi treinado com a função de perda de Entropia Cruzada obteve melhor F1-Score, tendo diferença de exatidão ao final do treinamento, entre as funções de perda MSE e Entropia Cruzada, de 2,56% e a convergência na mesma época. Neste caso, a função de perda DKL apresentou sobre ajuste no treinamento.

Por outro lado, o modelos treinados segundo critérios de informação apresentaram melhor resultado nas métricas de performance de teste do modelo. Em relação à convergência do treinamento, não se observou comportamentos consistentes que possibilitem concluir sobre a eficiência do treinamento.

Tabela 1. Comparativo entre os critérios de treinamento para a classificação de faltas.

Função de perda	MSE	EC	DKL
Nº de épocas	16	31	20
Acurácia(%)	94	97	97
Precisão(%)	88	93	94
Sensibilidade(%)	84	91	91
F1-Score(%)	86	92	92

Tabela 2. Comparativo entre os critérios de treinamento para a localização de faltas.

Função de perda	MSE	EC	DKL
Nº de épocas	22	22	6
Acurácia(%)	78	80	71
Precisão(%)	78	80	71
Sensibilidade(%)	95	95	100
F1-Score(%)	85	87	83

5. CONCLUSÕES

Levando-se em consideração a hipótese inicial que motivou o estudo, a aplicação de funções de perda probabilísticas proporciona a geração de um resultado superior para os dados de treinamento nas métricas avaliadas, sendo utilizadas a mesma arquitetura e parâmetros de rede e com a inicialização dos treinamentos em condições semelhantes. Esses resultados são observados tanto para o problema que busca realizar a diagnose do tipo de transiente quanto para a identificação e localização do enrolamento com defeito.

O critério de Divergência de Kullback—Leibler obteve um resultado de desempenho semelhante à Entropia Cruzada em exatidão para a diagnose do tipo de defeito, porém mais rápida convergência. O benefício do uso dos critérios de EC e DKL é obtido tanto em eficiência quanto em desempenho. Esse resultado evidencia que, para determinados conjuntos de dados, a escolha da função de custo, principalmente na classificação de padrões e em problemas como a identificação e localização de falhas, é um parâmetro importante a ser explorado na construção de um modelo classificador.

A aplicação de funções que se baseiam em distribuições de probabilidade dos resultados, apesar de, em muitos casos e aplicações estarem longe de serem exploradas mais a fundo, podem trazer resultados melhores para diversos problemas de aprendizado e modelos, mesmo que utilizadas como parâmetro para testes empíricos. Para trabalhos futuros é sugerida também uma análise estatística mais aprofundada dos resultados.

6. REFERÊNCIAS

- Behkam, R. et al. (2022) "Generalized regression neural network application for fault type detection in distribution transformer windings considering statistical indices", COMPEL. The International Journal for Computation and Mathematics in Electrical and Electronic Engineering, 41(1), p. 381–409. doi: 10.1108/compel-06-2021-0199.
- Bera, P. K., Kumar, R. e Isik, C. (2018) "Identification of internal faults in indirect symmetrical phase shift transformers using ensemble learning", em 2018 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT). IEEE.
- Bhasker, S. K. et al. (2015), "Differential protection of indirect symmetrical phase shift transformer and internal faults classification using wavelet and ANN," in TENCON 2015 - 2015 IEEE Region 10 Conference. IEEE.
- Bishop M., C. (2006), Pattern Recognition and Machine Learning. New York, NY: Springer
- Bronzeado, H. and Yacimini, R. (1995), "Phenomenon of sympathetic interaction between transformers caused by *inrush* transients," IEE Proceedings - Science Measurement and Technology, 142(4), pp. 323–329. doi: 10.1049/ip-smt:19951953.
- Davidson, E. M. et al. (2006), "Applying multi-agent system technology in practice: Automated management and analysis of SCADA and digital fault recorder data," IEEE transactions on power systems: a publication of the Power Engineering Society, 21(2), pp. 559–567. doi: 10.1109/tpwrs.2006.873109.
- Devadasu, G. and Sushama, M. (2016), "A novel multiple fault identification with fast fourier transform analysis," in 2016 International Conference on Emerging Trends in Engineering, Technology and Science (ICETETS). IEEE.
- Etumi, A. A. A. e Anayi, F. J. (2016) "The application of correlation technique in detecting internal and external faults in three-phase transformer and saturation of current transformer", IEEE transactions on power delivery, 31(5), p. 2131–2139. doi: 10.1109/tpwrdr.2016.2572608
- Ferreira, V. H. et al. (2016), "A survey on intelligent system application to fault diagnosis in electric power system transmission lines," Electric power systems research, 136, pp. 135–153. doi: 10.1016/j.epsr.2016.02.002.
- Haykin, S. O. (1999), Neural networks: Comprehensive foundation. 2nd ed. Piscataway, NJ: I.E.E.E. Press.
- Haykin, S. O. (2009), Neural networks and learning machines: International edition. 3rd ed. Upper Saddle River, NJ: Pearson.
- Kingma, D. P., & Ba, J. (2014), "Adam: A method for stochastic optimization," in arXiv [cs.LG]. Available at: <http://arxiv.org/abs/1412.6980>.
- Klomjit, J. e Ngaopitakkul, A. (2020) "Comparison of artificial intelligence methods for fault classification of the 115-kV hybrid transmission system", Applied sciences (Basel, Switzerland), 10(11), p. 3967. doi: 10.3390/app10113967.
- Pallav Bera, Can Isik (2020), "Data: Transients in Indirect Symmetrical Phase Shift Transformers," IEEE Dataport, doi: <https://dx.doi.org/10.21227/d8fv-6257>.
- Patel, D. et al. (2018), "Design and development of fault classification algorithm based on relevance vector machine for power transformer," IET electric power applications, 12(4), pp. 557–565. doi: 10.1049/iet-epa.2017.0562.
- Príncipe Jose, C. (2010), Information Theoretic Learning. New York, NY: Springer.
- Purarjomandlangrudi, A., Ghapanchi, A. H. and Esmalifalak, M. (2014), "A data mining approach for fault diagnosis: An application of anomaly detection algorithm," Measurement: journal of the International Measurement Confederation, 55, pp. 343–352. doi: 10.1016/j.measurement.2014.05.029.
- Rico, J. J., Acha, E. e Madrigal, M. (2001) "The study of *inrush* current phenomenon using operational matrices", IEEE transactions on power delivery, 16(2), p. 231–237. doi: 10.1109/61.915488.
- Rubinstein, R. Y. and Kroese, D. P. (2008), Simulation and the Monte Carlo method. 2nd ed. New York: Wiley-Interscience.
- Shannon, C. E. and Weaver, W. (1949), The mathematical theory of communication. Baltimore, MD: University of Illinois Press.
- Skowron, M. et al. (2019) "Application of self-organizing neural networks to electrical fault classification in induction motors", Applied sciences (Basel, Switzerland), 9(4), p. 616. doi: 10.3390/app9040616.
- Srivastava, N. et al, (2014), "Dropout: a simple way to prevent neural networks from overfitting," The Journal

of Machine Learning Research, 15, pp. 1929–1958. doi:
10.5555/2627435.2670313.

Souza, J. C. S. et al. (2001), “Fault location in electrical power systems using intelligent systems techniques,” IEEE transactions on power delivery, 16(1), pp. 59–67. doi: 10.1109/61.905590.

Verboomen, J. et al. (2005), “Phase Shifting Transformers: Principles and Applications,” in 2005 International Conference on Future Power Systems. IEEE.

Zimnickas, T. et al. (2020) “A technique for frequency converter-fed asynchronous motor vibration monitoring and fault classification, applying continuous wavelet transform and convolutional neural networks”, Energies, 13(14), p. 3690. doi: 10.3390/en13143690.