

# Redução de Dimensionalidade para Diagnóstico de Falhas em Transformadores de Potência

Thales W. Cabral, Larissa M. de Almeida, Eduardo R. de Lima, José Cândido S. Santos Filho e  
Luís Geraldo P. Meloni

**Resumo**—As subestações de energia enfrentam altos custos operacionais para implantar o monitoramento de falhas em transformadores de potência, o que requer a instalação de equipamentos de monitoramento e o processamento de sinais. Este trabalho investiga duas abordagens para aliviar tais custos no âmbito do diagnóstico de falhas via análise de gases dissolvidos em óleo: (i) técnicas modernas de redução de dimensionalidade ou (ii) redução força-bruta usando menos sensores de gás. Ambas as abordagens incorporam arquiteturas de aprendizado de máquina e são exploradas aqui para uma dimensão mínima (unitária) de saída. Além de menos custosa, uma solução força-bruta com uso exclusivo do gás hidrogênio mostra-se superior para a maioria das métricas de desempenho.

**Palavras-Chave**—Aprendizado de máquina, análise de gás dissolvido, diagnóstico de falhas em transformadores de potência, processamento de sinais.

**Abstract**—Power substations face high operating costs for implementing fault monitoring in power transformers, which requires monitoring equipment installation and signal processing. This work investigates two approaches to alleviate such costs in the context of gas-in-oil-based fault diagnosis: (i) modern dimensionality reduction techniques or (ii) brute-force reduction using fewer gas sensors. Both approaches incorporate machine learning architectures and are explored here for a minimum (unity) output dimension. In addition to being less costly, a brute-force solution with the exclusive use of hydrogen gas proves superior for most performance metrics.

**Keywords**—Machine learning, dissolved gas analysis, power-transformer fault diagnosis, signal processing.

## I. INTRODUÇÃO

Interrupções no fornecimento de energia, como instabilidade e blecautes, causadas por mal funcionamento de equipamentos em subestações elétricas, exigem grande atenção das empresas de energia. Essas falhas podem inutilizar equipamentos essenciais no sistema elétrico, como os transformadores de potencial e de corrente. Por esse motivo, há interesse das empresas de energia por estratégias de monitoramento e diagnóstico dedicados aos transformadores, com o objetivo de prevenir ou minimizar os efeitos causados pelas falhas.

Thales W. Cabral, Departamento de Comunicações, Universidade Estadual de Campinas, Campinas-SP, e-mail: thaleswulcab@gmail.com; Larissa M. de Almeida, Transmissora Aliança de Energia Elétrica S.A. (TAESA), e-mail: larissa.almeida@taesa.com.br; Eduardo R. de Lima, Instituto de Pesquisas Eldorado, e-mail: eduardo.lima@eldorado.org.br; José Cândido S. Santos Filho, Departamento de Comunicações, Universidade Estadual de Campinas, Campinas-SP, e-mail: jcssf@unicamp.br; e Luís Geraldo P. Meloni, Departamento de Comunicações, Universidade Estadual de Campinas, Campinas-SP, e-mail: meloni@unicamp.br. Os autores agradecem ao Instituto de Pesquisas Eldorado e à Transmissora Aliança de Energia Elétrica S.A. pelo financiamento deste trabalho por meio do Programa de P,D&I ANEEL no escopo do projeto PD-07130-0062/2020 intitulado “Análise preditiva de falhas por inteligência artificial”.

Ao longo dos anos, a comunidade científica desenvolveu diversas técnicas para realizar o monitoramento e diagnóstico de falhas em transformadores, como dados de vibração [1], imagens térmicas [2], emissões acústicas [3] e a análise de gases dissolvidos em óleo (DGA, do inglês *Dissolved Gas Analysis*). A DGA baseia-se na concentração de certos gases — os chamados gases-chave — para diagnosticar o estado do transformador: etileno ( $C_2H_4$ ), metano ( $CH_4$ ), etano ( $C_2H_6$ ), acetileno ( $C_2H_2$ ), hidrogênio ( $H_2$ ), monóxido de carbono (CO) e dióxido de carbono ( $CO_2$ ). O  $C_2H_4$  é produzido pela decomposição do óleo em alta temperatura [4]. Por sua vez, o óleo decomposto em baixa temperatura produz  $CH_4$  e  $C_2H_6$ . O  $H_2$  é produzido devido ao efeito corona, e  $C_2H_2$ , devido ao arco elétrico. Por fim, CO e  $CO_2$  são produzidos pela decomposição do papel de isolamento.

Na literatura, duas vertentes são as mais empregadas para gerar diagnósticos via DGA: métodos de interpretação [5] e inteligência artificial (IA) [6]. Os métodos de interpretação são mais antigos, e alguns desses, como em [7], não incluem o indicativo de estado normal de operação do transformador, ao contrário dos métodos baseados em IA. Com o avanço das arquiteturas de aprendizado de máquina, estratégias mais eficazes são constantemente empregadas utilizando modelos como *k-Nearest Neighbor* (k-NN) [8], *Decision Tree* (DT) [9], *Support Vector Machine* (SVM) [10] e Redes Neurais Artificiais (RNAs) [9]. Tal vertente tem alcançado resultados surpreendentes e, atualmente, representa o estado da arte. É importante ressaltar que estas técnicas dependem dos dados de DGA para serem treinadas, tais dados podem ser dinâmicos, isto é, excursionam ao longo do tempo, ou dados estáticos, neste caso, amostras pontuais.

Apesar dos avanços tecnológicos provenientes do uso de IA, ainda há restrições que afetam os sistemas de monitoramento. Devido às características físicas do transformador, como um número limitado de válvulas de monitoramento de óleo, nem sempre é possível conectar uma grande variedade de sensores para monitorar todos os gases mencionados. Além disso, as concessionárias enfrentam altos custos operacionais para implantar e manter o monitoramento, que envolvem a instalação de equipamentos de monitoramento e o processamento de sinais e dados. Por esses motivos, estratégias que diminuam o processamento de dados envolvidos no diagnóstico de falhas ou que reduzam a quantidade de sensores empregados no monitoramento são de grande interesse.

Para atender a demanda acima, o presente trabalho propõe e compara duas abordagens para o diagnóstico de falhas via DGA. A primeira abordagem busca simplificar o processamento através de técnicas de redução de dimensionalidade,

como a clássica análise de componentes principais (PCA, do inglês *Principal Component Analysis*) [11], a moderna técnica *Locally Linear Embedding* (LLE) [12] e a também moderna técnica *t-distributed Stochastic Neighbor Embedding* (*t-SNE*) [13]. Investiga-se aqui uma redução máxima de dimensionalidade, ou seja, com saída unidimensional. Nesse cenário, todos os (sete) gases-chave são empregados como entrada para as técnicas de redução mencionadas, cuja saída unidimensional alimenta então os modelos de IA para o diagnóstico das falhas. A segunda abordagem busca reduzir ao máximo, ou seja, a um, o número de sensores utilizados no monitoramento. Com base em [14], um dos gases mais informativos em relação às falhas em transformadores é o hidrogênio, que por esse motivo é empregado nessa abordagem. Ou seja, o hidrogênio sozinho é usado para alimentar, após algum pré-processamento (transformação logarítmica e normalização), os modelos de IA. Note que o hidrogênio já é um atributo de dimensão unitária.

Ao efetuar pioneiramente a comparação entre as duas abordagens supracitadas, considerando uma dimensão mínima (unitária) de saída, o presente estudo busca responder à seguinte pergunta central: o uso exclusivo de hidrogênio é capaz de rivalizar com o uso de técnicas avançadas de redução de dimensionalidade a partir de todos os gases-chave? De fato, como atestado na Seção V, o uso exclusivo de hidrogênio, além de mais simples, supera as técnicas de redução de dimensionalidade para a maioria das métricas de desempenho.

O restante deste trabalho está organizado como segue. A Seção II apresenta a proposta de diagnóstico de falhas com as abordagens investigadas, contemplando as técnicas de redução de dimensionalidade, o uso exclusivo de hidrogênio e diferentes modelos de aprendizado de máquina. A Seção III elenca as bases de dados e as métricas de desempenho. A Seção IV detalha os resultados alcançados. Finalmente, na Seção V, as principais conclusões são apresentadas.

## II. PROPOSTA DE DIAGNÓSTICO DE FALHAS

Este trabalho propõe duas abordagens para o diagnóstico de falhas. A primeira abordagem, ilustrada na Figura 1(a), emprega as  $n$  amostras das concentrações de todos os gases-chave:  $H_2$  ( $x_{1,1} \cdots x_{1,n} \triangleq \mathbf{x}_1$ ),  $CH_4$  ( $x_{2,1} \cdots x_{2,n} \triangleq \mathbf{x}_2$ ),  $C_2H_4$  ( $x_{3,1} \cdots x_{3,n} \triangleq \mathbf{x}_3$ ),  $C_2H_2$  ( $x_{4,1} \cdots x_{4,n} \triangleq \mathbf{x}_4$ ),  $C_2H_6$  ( $x_{5,1} \cdots x_{5,n} \triangleq \mathbf{x}_5$ ),  $CO$  ( $x_{6,1} \cdots x_{6,n} \triangleq \mathbf{x}_6$ ) e  $CO_2$  ( $x_{7,1} \cdots x_{7,n} \triangleq \mathbf{x}_7$ ). A segunda abordagem, representada na Figura 1(b), emprega apenas as  $n$  amostras das concentrações de  $H_2$  ( $x_{1,1} \cdots x_{1,n} \triangleq \mathbf{x}_1$ ). Ambas as abordagens possuem um estágio de redução de dimensionalidade e um estágio de classificação, que resulta no diagnóstico entre falha (F) ou operação normal (N). No entanto, como cada gás ocupa uma dimensão, a estratégia apresentada na Figura 1(a) emprega técnicas de redução de dimensionalidade para reduzir a dimensão dos dados de 7 para 1, com o objetivo de reduzir o volume de dados e minimizar o esforço computacional e de hardware. Ao final, encontra-se o vetor de dimensão reduzida ( $c_{1,1} \cdots c_{1,n}$ ). Por outro lado, o caminho representado na Figura 1(b) emprega o uso exclusivo de  $H_2$  e, por isso, não necessita do emprego das técnicas mencionadas. Nesse caso, apenas um pré-processamento, via função logaritmo e

normalização, é aplicado aos dados, produzindo o vetor final ( $c_{1,1} \cdots c_{1,n}$ ). O objetivo é demonstrar as vantagens do uso exclusivo de  $H_2$  frente ao emprego das técnicas de redução de dimensionalidade a partir de todos os gases. As seções seguintes detalham os estágios das referidas abordagens: a redução de dimensionalidade, na Seção II-A, composta pelas técnicas de redução de dimensionalidade e pelo uso exclusivo de  $H_2$ , e os modelos adotados de aprendizado de máquina, na Seção II-B, onde apresentamos brevemente as arquiteturas  $k$ -NN, SVM e RNA.

### A. Redução de Dimensionalidade

**Principal Component Analysis (PCA):** A PCA é uma técnica que preserva a informação principal dos dados [11]. Tal técnica pode ser empregada para realizar a extração de características dos dados com a vantagem de reduzir a dimensionalidade dos mesmos. Seja uma matriz de dados  $\mathbf{X} \in \mathbb{R}^{d \times n}$ , em que  $n$  é o número de amostras e  $d$  é a dimensão por amostra. Geralmente, no processo de redução de dimensionalidade, é necessário definir um valor limiar  $\varphi_r$  de acordo com a *Accumulative Contributory Ratio* (ACR) e os respectivos autovalores  $\lambda_j$ , conforme  $\varphi_r = \frac{\sum_{j=1}^r \lambda_j}{\sum_{j=1}^d \lambda_j}$  para  $r < d$  [11], de forma que se obtenham os  $r$  componentes principais mais representativos. No entanto, como neste trabalho estressamos a técnica à menor dimensão possível (unitária), não há necessidade de estabelecer um valor limite para  $\varphi_r$ ; apenas considera-se  $r = 1$ . Mais informações acerca do algoritmo da PCA podem ser encontradas em [11].

**Locally Linear Embedding (LLE):** A técnica de LLE é uma ferramenta de redução de dimensionalidade baseada no aprendizado de *manifold*, em que as características que residem em um espaço de alta dimensionalidade são “desdobradas” de forma a ter uma representação mais próxima da linear em um espaço de dimensão reduzida. Segundo [12], o algoritmo LLE mapeia a matriz de dados  $\mathbf{X} \in \mathbb{R}^{d \times n}$  para uma matriz  $\mathbf{Y} \in \mathbb{R}^{r \times n}$  com dimensão reduzida ( $r < d$ ). No processo de redução de dimensionalidade, o LLE obtém os  $r + 1$  autovetores associados aos  $r + 1$  menores autovalores. O primeiro autovetor cujo autovalor é próximo de zero é excluído. Os  $r$  autovetores restantes produzem  $\mathbf{Y}$ . Aqui também, estressamos a técnica à dimensão unitária, ou seja,  $r = 1$ . Mais informações sobre o algoritmo LLE podem ser encontradas em [12].

***t-distributed Stochastic Neighbor Embedding* (*t-SNE*):** A ferramenta *t-SNE* é uma técnica de redução de dimensionalidade não-linear comumente empregada para visualização de dados em aprendizado de máquina. Conforme [13], o algoritmo *t-SNE* também mapeia uma matriz de dados  $\mathbf{X} \in \mathbb{R}^{d \times n}$  no espaço de alta dimensionalidade para uma matriz  $\mathbf{Y} \in \mathbb{R}^{r \times n}$  em um espaço de dimensionalidade reduzida ( $r < d$ ). Ao mapear os dados do espaço de alta dimensionalidade para o espaço de dimensão reduzida, o *t-SNE* emprega a minimização da divergência de Kullback-Leibler  $KL(P||Q)$ , onde  $P$  é a medida de similaridade entre os pontos no espaço de alta dimensão e  $Q$  é a medida de similaridade entre os pontos no espaço de dimensão reduzida. Novamente, para estressar a técnica à menor dimensão possível, empregamos

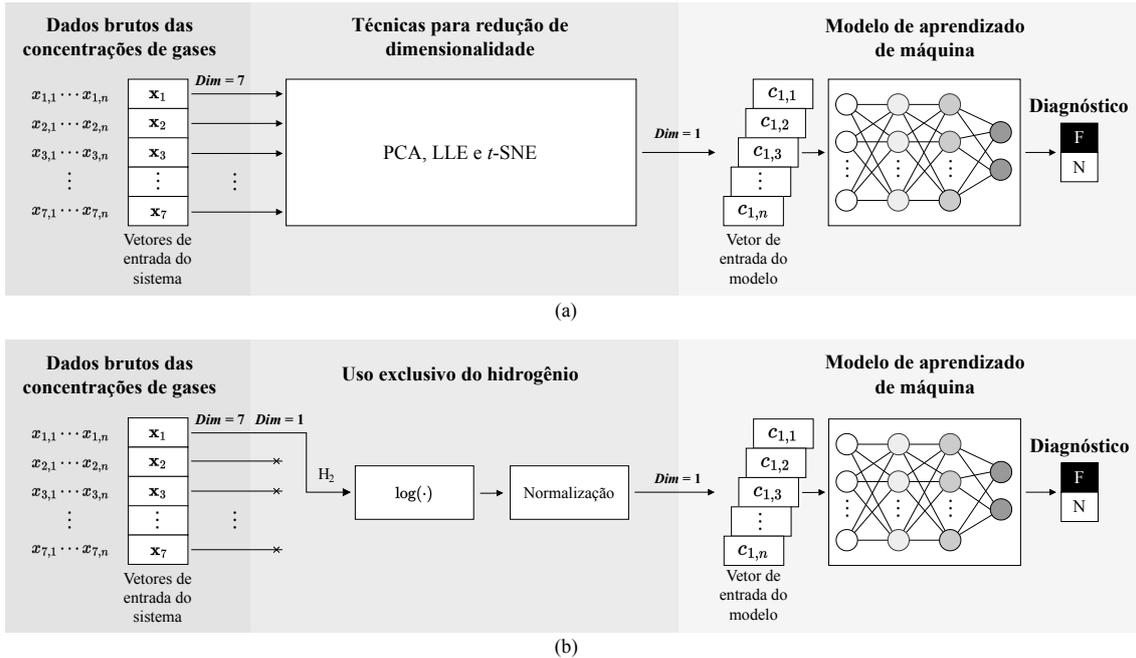


Fig. 1. Diagnóstico de falhas em transformadores via DGA: (a) técnicas de redução de dimensionalidade *versus* (b) uso exclusivo de hidrogênio.

$r = 1$ . Os principais passos do algoritmo estão sumarizados em [13].

**Uso exclusivo do hidrogênio:** De acordo com [8], uma característica altamente informativa acerca das concentrações de gases é a ordem de grandeza. Um modo de informar ao modelo de IA as variações de ordem grandeza é reescalar os dados através de uma transformação logarítmica. Para uma dada concentração de  $H_2$ , digamos  $x$ , emprega-se a função logaritmo, produzindo  $x' \triangleq \log_{10}(x)$ . Buscando-se assegurar operações numericamente estáveis, os dados provenientes da transformação logarítmica são então normalizados, resultando em  $x'' \triangleq (x' - \mathbb{E}[x']) / \sqrt{\mathbb{V}[x']}$ , onde  $\mathbb{E}[\cdot]$  é o operador de esperança amostral, e  $\mathbb{V}[\cdot]$ , de variância amostral.

### B. Modelos de Aprendizado

**$k$ -Nearest Neighbors ( $k$ -NN):** Para uma instância de teste, digamos  $c_{\text{teste}}$ , o algoritmo busca os  $k$  vizinhos mais próximos de  $c_{\text{teste}}$  no conjunto de treinamento. Em seguida, o algoritmo determina a categoria de  $c_{\text{teste}}$  como sendo a classe que ocorre com mais frequência entre os  $k$  vizinhos. É importante salientar que o número  $k$  de vizinhos é um parâmetro que pode ser inicialmente ajustado conforme a necessidade do problema [15]. No presente estudo, após realizar uma validação cruzada de  $K$ -pastas, com  $K=10$ , obteve-se  $k = 5$ .

**Support Vector Machine (SVM):** A SVM realiza um mapeamento não-linear dos dados de entrada para um espaço de características de alta dimensão, onde um hiperplano ótimo é responsável por separar as classes. Desse modo, o modelo SVM encontra a melhor forma de separar as classes e classifica novas observações com base no lado do hiperplano em que elas se encontram [15]. Nessa aplicação, empregou-se o kernel *Radial Basis Function* (RBF), em que  $\gamma$  é um hiperparâmetro que controla a largura RBF. Nesse caso, após realizar a validação

cruzada de  $K$ -pastas, com  $K=10$ , obteve-se  $\gamma = 0,1124$ .

**Rede Neural Artificial (RNA):** No presente trabalho, empregamos uma estrutura de RNA [15] composta por duas camadas ocultas, ambas com 90 neurônios e funções de ativação ReLu. Para a camada de saída, de único neurônio, empregamos a função de ativação sigmoide. Além disso, a função de perda a ser otimizada é a entropia cruzada binária, com o emprego do otimizador ADAM. Por fim, através da validação cruzada de  $K$ -pastas, com  $K=10$ , obteve-se a quantidade de 3000 épocas adequadas para o treinamento.

## III. BASES DE DADOS E MÉTRICAS DE DESEMPENHO

### A. Bases de Dados

Um número limitado de conjuntos de dados estáticos de DGA, com amostras pontuais, está disponível na literatura. O presente trabalho emprega a fusão de seis conjuntos de dados: dados da concessionária do Egito [16], [17], de Duval de 2001 [5], de Duval de 2002 [18], de Ganyun [19], de Mang-Huy [20] e de Gouda [21]. Conforme a norma padrão IEC [5], as falhas em transformadores são rotuladas em seis tipos: descargas parciais (PD), descarga de baixa energia ( $D_1$ ), descarga de alta energia ( $D_2$ ), falhas térmicas ( $T_1$ ) de baixa temperatura ( $T < 300\text{ C}^\circ$ ), falhas térmicas ( $T_2$ ) de média temperatura ( $300\text{ C}^\circ < T < 700\text{ C}^\circ$ ) e falhas térmicas ( $T_3$ ) de alta temperatura ( $T > 700\text{ C}^\circ$ ). No entanto, algumas dessas bases de dados consideram a união entre classes. Logo, para realizar a fusão de todos os conjuntos de dados mencionados, é necessário mesclar alguns rótulos. Os novos rótulos são estes: PD para descargas parciais (209 amostras), D para descargas de energia (compreende  $D_1$  e  $D_2$ , com 206 amostras), T para falhas térmicas (compreende  $T_1$ ,  $T_2$  e  $T_3$ , com 90 amostras) e N para operação normal (46 amostras). Por fim, com a fusão, obtém-se um total de 551 amostras.

### B. Métricas de Desempenho

O presente trabalho emprega acurácia, F1-Score, precisão e especificidade como métricas de verificação de desempenho [15]. A acurácia mede a fração de instâncias corretamente classificadas:

$$\text{Acurácia} = \frac{TP + TN}{TP + FP + TN + FN}, \quad (1)$$

em que TP, TN, FP e FN indicam as quantidades de verdadeiros positivos, verdadeiros negativos, falsos positivos e falsos negativos, respectivamente. Por outro lado, o F1-Score é a média harmônica entre a precisão e a revocação:

$$\text{F1-Score} = \frac{2 \times TP}{2 \times TP + 1 \times (FN + FP)}. \quad (2)$$

A precisão é uma métrica importante na avaliação de modelos quando o objetivo é maximizar a proporção de previsões corretas:

$$\text{Precisão} = \frac{TP}{TP + FP}. \quad (3)$$

Por fim, a especificidade mede a capacidade do modelo em identificar corretamente os casos que não possuem a condição ou a característica que está sendo estudada:

$$\text{Especificidade} = \frac{TN}{TN + FP}. \quad (4)$$

## IV. RESULTADOS

Inicialmente, é pertinente ressaltar que as amostras foram divididas em conjuntos de treinamento e teste, distribuídos na proporção de 70% e 30%, respectivamente. Apenas dados destinados ao treinamento foram empregados na validação cruzada, conduzida com 10 grupos distintos (10 pastas). Para o desenvolvimento do presente estudo, adota-se a linguagem Python, a biblioteca scikit-learn e a API Keras. As Tabelas I, II e III apresentam resultados médios, em porcentagem, para as métricas de desempenho acurácia, F1-Score, precisão, e especificidade. A Tabela I reporta os resultados obtidos para o modelo SVM; a Tabela II, para  $k$ -NN; e a Tabela III, para RNA. Por outro lado, a Tabela IV apresenta o tempo médio de treinamento para cada modelo aliado a cada proposta de redução de dimensionalidade. Os resultados médios, tanto para as métricas de desempenho quanto para o tempo de treinamento, foram obtidos através de 50 execuções dos experimentos de classificação. Por fim, em cada uma das tabelas mencionadas, lista-se um esquema de referência, denominado *benchmark*, que emprega transformação logarítmica, normalização e todos os sete gases, ou seja, sem redução de dimensionalidade. É importante ressaltar que esse esquema de referência não é objeto deste trabalho, figurando aqui apenas como base de comparação para as abordagens propostas de dimensionalidade reduzida.

Ao se compararem as duas abordagens propostas — técnicas de redução de dimensionalidade e uso exclusivo de  $H_2$  — para o SVM na Tabela I, nota-se que os maiores valores de acurácia (92,22%), F1-Score (89,23%) e precisão (89,18%) foram obtidos via  $H_2$ . Porém, nessa comparação, a maior especificidade foi obtida via LLE (100%). Essa tendência se

repete em relação ao *benchmark*. Comparada ao *benchmark*, a proposta via  $H_2$  foi a que apresentou a menor perda de desempenho para a maioria das métricas, com uma perda aproximada de 4 pontos percentuais para acurácia e uma perda inferior a 7 pontos percentuais para F1-Score e precisão.

TABELA I  
DESEMPENHO MÉDIO DO MODELO SVM

Técnicas	Acurácia (%)	F1-Score (%)	Precisão (%)	Especificidade (%)
Benchmark	96,23	96,04	96,12	98,67
PCA	88,81	86,41	84,64	96,82
$t$ -SNE	91,66	88,35	85,63	99,34
LLE	91,84	87,94	84,37	100
$H_2$	92,22	89,23	89,18	99,63

A Tabela II aborda a comparação entre as duas abordagens propostas para o modelo  $k$ -NN. Nesse caso, a maior especificidade foi alcançada via PCA (99,07%). No entanto, os maiores valores para acurácia (92,47%), F1-Score (90,73%) e precisão (90,66%) foram novamente obtidos através do uso exclusivo de  $H_2$ . Aqui também, quando comparada ao *benchmark*, a abordagem via  $H_2$  apresentou a menor perda de desempenho para a maioria das métricas: aproximadamente 3 pontos percentuais para acurácia, e inferior a 5 pontos percentuais para F1-Score e precisão.

TABELA II  
DESEMPENHO MÉDIO DO MODELO  $k$ -NN

Técnicas	Acurácia (%)	F1-Score (%)	Precisão (%)	Especificidade (%)
Benchmark	95,75	95,43	95,61	98,76
PCA	90,54	86,94	83,93	99,07
$t$ -SNE	91,36	90,29	90,30	97,28
LLE	91,20	88,83	88,55	98,40
$H_2$	92,47	90,73	90,66	98,61

A Tabela III traz a comparação entre as duas abordagens para o modelo RNA. Novamente, através da PCA obteve-se o maior valor de especificidade (100%). Porém, a abordagem via  $H_2$  vence outra vez para acurácia (92,29%), F1-Score (90,21%) e precisão (91,38%), com menor perda de desempenho em relação ao *benchmark*: cerca de 4 pontos percentuais para acurácia, e inferior a 6 e 5 pontos percentuais para F1-Score e precisão, respectivamente.

TABELA III  
DESEMPENHO MÉDIO DO MODELO RNA

Técnicas	Acurácia (%)	F1-Score (%)	Precisão (%)	Especificidade (%)
Benchmark	96,29	96,19	96,29	98,44
PCA	91,73	87,79	84,19	100
$t$ -SNE	90,95	88,14	87,07	98,71
LLE	91,67	88,66	88,11	99,42
$H_2$	92,29	90,21	91,38	99,05

A Tabela IV mostra que os modelos RNA e  $k$ -NN são, por larga margem, aqueles com maiores e menores tempos médios de treinamento, respectivamente. Além disso, mostra também que os tempos de treinamento dos modelos RNA e  $k$ -NN são pouco dependentes da técnica de redução de dimensionalidade, diferentemente do modelo SVM.

TABELA IV  
TEMPO MÉDIO DE TREINAMENTO (EM SEGUNDOS)

Modelos	Benchmark	PCA	t-SNE	LLE	H <sub>2</sub>
k-NN	0,0002	0,0003	0,0002	0,0002	0,0002
SVM	0,0045	0,0149	0,0164	0,0036	0,0091
RNA	16,2773	15,9035	16,2921	15,7366	15,8289

A. Modelos Vencedores

A Tabela V apresenta, para cada métrica de desempenho, os dois melhores modelos de aprendizado e as abordagens associadas de redução de dimensionalidade (excetuando-se o esquema *benchmark*).

Do ponto de vista do modelo, nota-se que o *k*-NN obteve o maior valor para as métricas acurácia e F1-Score, ou seja, figurou como primeiro colocado duas vezes. A RNA também figurou em primeiro lugar, para a métrica precisão. Além disso, a RNA apresentou os segundos maiores valores de acurácia, F1-Score e especificidade, ou seja, figurou em segundo lugar três vezes. O *k*-NN figurou como segundo colocado uma vez, apresentando o segundo maior valor de precisão.

Do ponto de vista da abordagem de redução, nota-se que o uso exclusivo de H<sub>2</sub> apresentou os maiores valores de acurácia, F1-Score e precisão, figurando como primeiro colocado três vezes. Além disso, apresentou os segundos maiores valores para essas métricas, figurando também três vezes como segundo colocado. Esse feito transforma a proposta de uso exclusivo de H<sub>2</sub> em recomendação imediata e robusta de procedimento dentro do estágio de redução de dimensionalidade para diagnóstico de falhas em transformadores.

TABELA V  
COMPARATIVO ENTRE MODELOS E ABORDAGENS

Métrica	Primeiro colocado	Abordagem empregada no primeiro colocado	Segundo colocado	Abordagem empregada no segundo colocado
Acurácia	k-NN	H <sub>2</sub>	RNA	H <sub>2</sub>
F1-Score	k-NN	H <sub>2</sub>	RNA	H <sub>2</sub>
Precisão	RNA	H <sub>2</sub>	k-NN	H <sub>2</sub>
Especificidade	SVM	LLE	RNA	PCA

V. CONCLUSÃO

Neste trabalho, obtivemos uma nova recomendação para o diagnóstico de falhas em transformadores de potência via análise de gases dissolvidos em óleo, combinando redução de dimensionalidade e aprendizado de máquina. Apresentamos duas abordagens diferentes e descobrimos que o uso exclusivo de H<sub>2</sub> alcança valores superiores para a maioria das métricas de desempenho, quando comparado ao uso dos sete gases-chave em conjunto com técnicas modernas de redução de dimensionalidade. Além disso, a abordagem via H<sub>2</sub> contribui para um baixo tempo de treinamento das arquiteturas e para o desempenho geral dos modelos, apresentando pequenas perdas de desempenho quando comparada ao *benchmark* com dimensão plena. Por fim, observa-se que o uso exclusivo de H<sub>2</sub> reduz significativamente o gasto com sensores e a complexidade no processamento de sinais envolvidos nas subestações de energia para o diagnóstico de falhas.

REFERÊNCIAS

- [1] K. Hong, H. Huang, Y. Fu, and J. Zhou, "A vibration measurement system for health monitoring of power transformers," *Measurement*, vol. 93, pp. 135–147, 2016.
- [2] I. Ullah, F. Yang, R. Khan, L. Liu, H. Yang, B. Gao, and K. Sun, "Predictive maintenance of power substation equipment by infrared thermography using a machine-learning approach," *Energies*, vol. 10, no. 12, 2017.
- [3] T. Boczar, S. Borucki, A. Cichon, and D. Zmarzly, "Application possibilities of artificial neural networks for recognizing partial discharges measured by the acoustic emission method," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 16, no. 1, pp. 214–223, 2009.
- [4] V. Tra, B.-P. Duong, and J.-M. Kim, "Improving diagnostic performance of a power transformer using an adaptive partial sampling method for imbalanced data," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 26, no. 4, pp. 1325–1333, 2019.
- [5] M. Duval and A. DePabla, "Interpretation of gas-in-oil analysis using new iec publication 60599 and iec tc 10 databases," *IEEE Electrical Insulation Magazine*, vol. 17, no. 2, pp. 31–41, 2001.
- [6] Y. Zhang, H. C. Chen, Y. Du, M. Chen, J. Liang, J. Li, X. Fan, and X. Yao, "Power transformer fault diagnosis considering data imbalance and data set fusion," *High Voltage*, vol. 6, no. 3, pp. 543–554, 2021.
- [7] G. Irungu, A. Akumu, and J. Munda, "A new fault diagnostic technique in oil-filled electrical equipment; the dual of duval triangle," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 23, no. 6, pp. 3405–3410, 2016.
- [8] P. Mirowski and Y. LeCun, "Statistical machine learning and dissolved gas analysis: a review," *IEEE Transactions on Power Delivery*, vol. 27, no. 4, pp. 1791–1799, 2012.
- [9] M. E. A. Senoussaoui, M. Brahama, and I. Fofana, "Combining and comparing various machine-learning algorithms to improve dissolved gas analysis interpretation," *IET Generation, Transmission & Distribution*, vol. 12, no. 15, pp. 3673–3679, 2018.
- [10] X. Zhang, G. Zhang, P. Paul, J. Zhang, T. Wu, S. Fan, and X. Xiong, "Dissolved gas analysis for transformer fault based on learning spiking neural p system with belief adaboost," *International Journal of Unconventional Computing*, vol. 16, 2021.
- [11] D. A. M. Lemes, T. W. Cabral, G. Fraidenaich, L. G. P. Meloni, E. R. De Lima, and F. B. Neto, "Load disaggregation based on time window for hems application," *IEEE Access*, vol. 9, pp. 70746–70757, 2021.
- [12] B.-Y. Sun, X.-M. Zhang, J. Li, and X.-M. Mao, "Feature fusion using locally linear embedding for classification," *IEEE transactions on neural networks*, vol. 21, no. 1, pp. 163–168, 2009.
- [13] L. Gao, D. Gu, L. Zhuang, J. Ren, D. Yang, and B. Zhang, "Combining t-distributed stochastic neighbor embedding with convolutional neural networks for hyperspectral image classification," *IEEE geoscience and remote sensing letters*, vol. 17, no. 8, pp. 1368–1372, 2019.
- [14] J. Golarz, "Understanding dissolved gas analysis (dga) techniques and interpretations," in *2016 IEEE/PES Transmission and Distribution Conference and Exposition (T&D)*, pp. 1–5, IEEE, 2016.
- [15] A. Géron, *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow*. "O'Reilly Media, Inc.", 2022.
- [16] S. I. Ibrahim, S. S. Ghoneim, and I. B. Taha, "Dgalab: an extensible software implementation for dga," *IET Generation, Transmission & Distribution*, vol. 12, no. 18, pp. 4117–4124, 2018.
- [17] S. S. Ghoneim and I. B. Taha, "A new approach of dga interpretation technique for transformer fault diagnosis," *International Journal of Electrical Power & Energy Systems*, vol. 81, pp. 265–274, 2016.
- [18] M. Duval, "A review of faults detectable by gas-in-oil analysis in transformers," *IEEE electrical Insulation magazine*, vol. 18, no. 3, pp. 8–17, 2002.
- [19] M.-H. Wang, "A novel extension method for transformer fault diagnosis," *IEEE transactions on power delivery*, vol. 18, no. 1, pp. 164–169, 2003.
- [20] L. Ganyun, C. Haozhong, Z. Haibao, and D. Lixin, "Fault diagnosis of power transformer based on multi-layer svm classifier," *Electric Power Systems Research*, vol. 74, no. 1, pp. 1–7, 2005.
- [21] O. E. Gouda, S. Salem, and S. H. El-Hoshy, "Power transformer incipient faults diagnosis based on dissolved gas analysis," *TELKOMNIKA Indonesian J. Electr. Eng*, vol. 17, no. 1, pp. 10–16, 2016.