

Análise de Incertezas em Amostras de Gases Usadas no Diagnóstico de Falhas em Transformadores de Potência

M.A. Ikeshoji *, A.N. Souza *, P.C. Junior *
D.S. Gastaldello ** J.P. Papa *** C.G. Gonzales ****

* Faculdade de Engenharia, Univ. Estadual Paulista - UNESP, SP,
(e-mail: ikeshoji@uol.com.br, {andre.souza, costa.jr}@unesp.br).

** Universidade Sagrado Coração - USC, SP, (e-mail:
dandangastaldello@gmail.com)

*** Departamento de Computação, Univ. Estadual Paulista - UNESP,
SP, (e-mail: papa@fc.unesp.br)

**** Centro de Manutenções Especiais - ISA CTEEP, SP (e-mail:
cgonzale@isactEEP.com.br)

Abstract: Large power transformers are essential assets of the electrical system. Unexpected departure from transformers due to unobserved failures causes significant financial, social, and environmental losses. Analysis of dissolved gases in the oil and the use of analytical tools for diagnosing incipient transformer failures, widely employed by utilities, may present low assertiveness and/or misdiagnosis. On the other hand, the improvement of computational intelligence techniques, coupled with new measurement technologies, has increased confidence in the diagnostics produced. In this context, to eliminate or reduce errors in the fault classification process, statistical analysis is required to get additional information, assess and treat the presence of outliers in the available dissolved gas data set. Therefore, in order to reduce the deleterious effects of untimely and unpredictable interruptions of transformers, this paper proposes a methodology for eliminating outliers based on limit values applied through identification and processing techniques of machine learning.

Resumo: Grandes transformadores de potência são importantes ativos do sistema elétrico. A saída inesperada de operação desses equipamentos por ocorrência de falhas não observadas, causa grandes prejuízos financeiros, sociais e ambientais. A análise dos gases dissolvidos no óleo e o uso de ferramentas analíticas para o diagnóstico de falhas incipientes nos transformadores, podem apresentar baixa assertividade e/ou ausência de diagnóstico. Por outro lado, o aprimoramento de técnicas de inteligência computacional, aliada a novas tecnologias de medição tem permitido aumentar a confiança nos diagnósticos produzidos. Neste contexto, para eliminar ou reduzir erros no processo de classificação de falhas, empregou-se estatísticas adicionais para obter informações, avaliar e tratar a presença de *outliers* no conjunto disponível de dados dos gases dissolvidos. Portanto, com o objetivo de reduzir os efeitos deletérios de interrupções intempestivas e imprevisíveis dos transformadores, este artigo propõe metodologias de eliminação de *outliers* com base em valores limites determinados através de técnicas de identificação e processamento com algoritmos de aprendizado de máquinas.

Keywords: Dissolved Gas Analysis, Outliers, Machine Learning, Power Transformers.

Palavras-chave: Análise de Gás Dissolvido, *Outliers*, Aprendizado de Máquinas, Transformadores de Potência.

1. INTRODUÇÃO

Considerados como um dos elementos principais do sistema elétrico de potência, os grandes transformadores garantem o fornecimento de energia de modo estável e confiável aos consumidores (Zhao et al., 2014). Os impactos negativos relacionados a falha repentina desses equipamentos provocam, além da interrupção do fornecimento de energia, problemas operacionais com a estabilidade do sistema elétrico (saída de grandes blocos consumidores), comprometimento da operação econômica da concessio-

nária (prejuízos, multas, litígios), aumento dos problemas pela ausência de sobressalente (necessita de meses para fabricação), e possíveis danos ambientais devido ao grande volume de óleo que possuem (Sun et al., 2012). Portanto, detectar a falha em sua fase inicial é fator essencial para se evitar essas consequências negativas.

Dentre os métodos de análises de falhas incipientes mais empregados nos diagnósticos de transformadores, a análise dos gases dissolvidos (DGA - *Dissolved Gas Analysis*) no óleo isolante é uma das ferramentas de auxílio a

manutenção de uso mais comum (Liepniece et al., 2017; Kaur et al., 2019), além de ser possível fazê-la com o equipamento em plena operação. Embora a DGA seja uma técnica já estabelecida (Tenbohlen et al., 2016; Duval, 2012) e normatizada (IEEE C57.104, 2008; IEC 60599, 2007), não é uma ferramenta adequada para prevenção de falhas instantâneas (Sun et al., 2012).

As falhas que ocorrem com os transformadores podem ser decorrentes de interferências externas operacionais (curto circuito, sobrecargas), ambientais (descargas atmosféricas), construtivas (componente com defeito, montagem errada, erro de projeto), ou do próprio término de sua vida útil operacional ocasionada pelas ações naturais decorrentes dos fenômenos térmicos, hidrolíticos e oxidativos inerentes ao funcionamento normal, que acabam degradando o sistema isolante do equipamento (Bechara, 2010; Sokolov, 2005). Dentre esses eventos ou fenômenos, alguns geram internamente no transformador estresses térmicos e elétricos sobre o óleo e papel isolante, produzindo então determinados tipos de gases que acabam dissolvidos no óleo e tornam-se passíveis de serem tipificados e quantificados via análise cromatográfica gasosa.

De posse dessas informações sobre os gases e com uso dos métodos de interpretações e classificações de falhas, torna-se possível diagnosticar de modo incipiente a saúde do transformador, entretanto, os métodos analíticos clássicos de interpretações dos gases (por exemplo, Rogers, Doernenburg) podem apresentar, além de baixa assertividade e casos de ausência de diagnóstico (Zhao et al., 2014), divergências na classificação das falhas nos transformadores quando comparados entre si (Sun et al., 2012) e, provavelmente, seja esta uma das razões de não existir ainda método unânime de diagnóstico entre os mesmos (Ikeshoji et al., 2018).

Em contrapartida, isto tem contribuído com o surgimento de vários outros métodos que visam compensar essas deficiências, como é o caso dos algoritmos de aprendizado de máquinas empregados como classificadores de padrões de falhas. Funcionalmente, alguns desses algoritmos inteligentes requerem inicialmente uma etapa de treinamento de onde se extraem as informações das propriedades e fenômenos que envolvem os dados de um conjunto de amostras de gases rotulados (com falhas e normais) e uma segunda etapa onde são feitas as classificações ou avaliações de novas amostras com base nas informações obtidas à priori na fase de treinamento (Ikeshoji et al., 2018). Essa condição de extração e uso das informações obtidas a partir de um conjunto prévio de dados de gases para a classificação de outros conjuntos, pressupõe a necessidade de avaliar a presença de possíveis valores contaminados com erros (*outliers*) nesses dados e tratá-los, pois os mesmos conduzem a erros no processo de classificações das falhas em transformadores e aos impactos negativos citados anteriormente.

Ma et al. (2012) destacam dois aspectos relevantes na importância da preparação ou limpeza dos dados, o primeiro é que os dados reais são incompletos, inconsistentes e contém erros ou *outliers*, e o segundo é que dados de qualidade produzem conseqüentemente padrões de alta qualidade. No entanto, não é uma tarefa simples limpar os dados dos gases visto que há muitas incertezas inseridas e

que podem envolver: desde o processo de formação desses gases no óleo até sua análise final em laboratório (IEC 60599, 2007; Sun et al., 2012); procedimentos operacionais e ocorrências de anormalidades externas que influenciam a razão gás-óleo como é o caso de curto circuito externo, regeneração do óleo, concentrações de gases residuais após correção de falhas, e outros (Khalyasmaa et al., 2016). Além disso, há outras questões complexas a serem abordadas no tratamento dos *outliers*, como por exemplo: Como identificá-los? Qual o rigor da limpeza a ser feita sem que comprometa a perda de informações? Deve ocorrer sobre todo o conjunto de dados ou apenas para os casos considerados de maior criticidade ou risco para operação do sistema elétrico?

Neste contexto, este artigo analisa os efeitos da remoção de *outliers* sobre os dados dos casos considerados críticos no diagnóstico de falhas em transformadores. Para isto, foram empregados os algoritmos classificadores de padrões: redes neurais artificiais (ANN – *Artificial Neural Network*), máquina de aprendizado extremo (ELM – *Extreme Learning Machine*), K vizinho mais próximos (KNN – *K Nearest Neighborhood*) e floresta de caminhos ótimos (OPF – *Optimum Path Forest*), na avaliação dos métodos de remoção de *outliers*. Como ferramenta de auxílio a análise de presença de *outliers* nos conjuntos de dados de falhas dos transformadores, utilizou-se a matriz de erros ou confusão (Stehman, 1997) uma vez que esta permite analisar com clareza como todas as instâncias envolvidas foram classificadas, mesmo na presença de desbalançamento quantitativo entre os padrões existentes. Em complemento, como critério de seleção e definição dos valores de remoção dos *outliers* foram estabelecidos limites dos gases baseados: nas normas IEC 60599 (2007) e IEEE C57.104 (2008), na distribuição normal de probabilidades sobre os dados brutos de gases dissolvidos no óleo, e uso do algoritmo de agrupamento *Kmeans Cluster*.

O artigo está organizado conforme segue: a seção 2 aborda breve taxonomia dos *outliers* e seu uso na DGA; a seção 3 descreve os algoritmos de aprendizado de máquinas usados neste artigo; a seção 4 descreve os métodos usados na remoção dos *outliers* e os resultados experimentais obtidos, e a seção 5 a conclusão.

2. OUTLIER: FUNDAMENTAÇÃO TEÓRICA

Historicamente, o problema da qualidade e limpeza dos dados é visto como uma etapa sensível à interpretações errôneas e, no entanto, é preciso saber lidar com os erros de várias fontes e seus efeitos nos resultados dos estudos (Van Den Broeck et al., 2005). Em mineração de dados, o estudo de detecção de *outlier* possui um imenso campo de atuação, como por exemplo (Gupta et al., 2014): análise de dados em alta dimensão, incertezas nos dados, rede de dados, dados de séries temporais, entre outros.

Aguinis et al. (2013) realizaram uma ampla revisão bibliográfica de diversos trabalhos científicos sobre *outliers*, e como resultado foram apresentadas 14 definições para esse termo, 39 técnicas de identificações e 20 modos diferentes de manipulações. Baseado nesse estudo, a definição utilizada para *outlier* neste artigo refere-se a amostras de dados que estão distantes de outras por motivo de incertezas, mais precisamente referentes à valores discrepantes por

não fazerem parte da população de interesse seja por erros de manipulação, procedimento de amostragem, ou erros de preparação de dados. Quanto as técnicas empregadas na detecção de *outliers* foram aplicadas a conduta de análise com e sem *outliers*, o agrupamento K (*K-cluster*) e análise da distribuição normal. A manipulação dos dados baseou-se na análise de fenômeno único de interesse (remoção de *outliers* nas amostras de falhas) e ajustes de valores adequados dentro de um intervalo confiável de valores de remoção.

Em se tratando de detecção de *outliers* em DGA para melhorar a assertividade no reconhecimento de falhas, seja com uso de algoritmos analíticos ou de aprendizado de máquinas, diversos trabalhos tem sido desenvolvidos. Em Qin et al. (2019) foi proposto um método de monitoramento contínuo (reduz a introdução de incertezas) empregando a transformação *Wavelet* dos sinais para identificação de ocorrências de anomalias, associado ao algoritmo Fator Ponderado de *Outliers* Locais. O sinal resultante da soma dos gases avaliado, permitiu a identificação de ocorrências de anomalias durante o período de aquisição dos dados, e assim observar os pontos extremos dos sinais (característica comum aos *outliers*). Em Khalyasmaa et al. (2016) é descrito que os atuais sistemas de informações analíticos responsáveis pelo armazenamento, processamento, e análise de dados dos gases dissolvidos no óleo são importantes ferramentas de auxílio a gestão, entretanto, mais do que simplesmente aplicar os métodos analíticos de classificação de falhas (Duval, Rogers, e outros), há necessidade de treinamento para interpretar essas informações que alimentam o sistema. Um exemplo disso, é a interpretação dos dados dos gases, bem como o reconhecimento e remoção de *outliers* existentes visto que induz ao reconhecimento incorreto das falhas. Ma et al. (2012) propuseram a eliminação de *outliers* nas amostras de gases dos transformadores empregando os algoritmos de agrupamento FCM – *Fuzzy C-means* e KFCM – *Kernel Fuzzy C-means* junto com o classificador FSVM – *Fuzzy Support Vector Machine*, ambos FCM e KFCM atribuem diferentes pesos para as amostras em função de suas distâncias ao centro dos grupos que pertence, e posteriormente submete-se ao FSVM para realizar a classificação. Já Subroto et al. (2017) empregaram o método (estatístico) Thompson Tau para eliminar *outliers* existentes nos gases dissolvidos e posteriormente combinaram a técnica FCM com os métodos clássicos para criar também um sistema de pesos e assim melhorar a assertividade da classificação de falhas, entretanto, observa-se que método Thompson Tau apresenta profundo corte nos dados dependendo da dispersão entre os grupos (*clusters*).

3. ALGORITMOS INTELIGENTES E ANÁLISES DE FALHAS EM TRANSFORMADORES

O rápido desenvolvimento da tecnologia e da inteligência computacional tem permitido aos pesquisadores desenvolverem novos métodos e técnicas de diagnósticos de falhas em transformadores que visam compensar as causas e efeito das falhas, julgamento dos dados dos gases e medições inadequadas existentes com métodos analíticos clássicos (Cheng e Yu, 2018). No âmbito de desenvolvimento dessas técnicas, existem suas variações que buscam ainda melhorarem os resultados das classificações de padrões através do emprego de recursos de outras técnicas

associadas (Afonso et al., 2017; Amora et al., 2012), e métodos de otimização e associações híbridas (Barbosa e Costa, 2018; Nur et al., 2014; Shawe-taylor e Sun, 2011; Suguna e Thanushkodi, 2010). Nesta seção são abordados brevemente o princípio das técnicas de aprendizado de máquinas empregados nos diagnósticos de falhas em transformadores com base na análise da DGA.

3.1 Redes Neurais Artificiais - ANN

A ANN é um modelo matemático que funciona de modo similar aos neurônios do cérebro, ou seja, um neurônio biológico recebe informações de diversas fontes externas, combina-os e executa uma operação não linear de ativação, e em seguida toma uma decisão (Cheng e Sutariya, 2012). Como classificador de padrões, a ANN faz o mapeamento de um conjunto de dados de interesse e gera um espaço de atributos constituído de regiões definidas por classes distintas separadas por hiperplanos (fase de treino), e as novas amostras são classificadas com base nessas diferentes regiões do espaço de atributos em que ocupam (fase de teste). Sua estrutura basicamente é constituída por camadas de neurônios, pesos, *bias* e funções de ativações interligadas entre si convenientemente, entretanto, a escolha de sua configuração não é uma tarefa simples, pois, depende de diversas variáveis de projeto como por exemplo, topologia, estratégia de aprendizado, dados de interesse, funções de ativações, e outros (Ikeshoji et al., 2018).

Na maioria dos trabalhos que envolvem diagnóstico de falhas em transformadores via DGA e aprendizado de máquinas, as variáveis de análises envolvidas (gases) são identificadas com as classes ou padrões de falhas encontrados durante as inspeções e manutenções dos equipamentos, permitindo assim adotar uma estratégia de aprendizado supervisionado como por exemplo, em Farooque et al. (2016) e Zakaria et al. (2013). Os atributos dessas variáveis são definidos segundo os tipos e quantidades de gases (ou suas possíveis combinações/razões) com seus devidos rótulos. Na etapa de treinamento da ANN, são feitas as apresentações sucessivas de amostras do conjunto de treino que, por meio de processos iterativos, ajustam os pesos da rede para que a maioria dos seus respectivos rótulos de falhas sejam corretamente identificados. Já na fase de teste, são apresentadas novas amostras para a rede treinada e os resultados obtidos são comparados com os rótulos reais das falhas para avaliar sua capacidade de generalização.

3.2 Máquina de Aprendizado Extremo - ELM

Com base na ANN *feedforward* clássica de uma camada oculta e visando reduzir o tempo de treinamento supervisionado desse tipo de rede, Huang et al. (2006) propuseram o algoritmo ELM, que diferentemente do processo iterativo de ajuste dos pesos da ANN, usa o processo analítico tornando-a mais rápida (Ikeshoji et al., 2018). Para melhorar a capacidade de generalização e a assertividade da ELM clássica na classificação de padrões, Huang et al. (2012) modificaram seu algoritmo introduzindo a regularização e o uso *kernel* (alta dimensionalidade). Denominado de ELM *kernel*, em sua parametrização não é necessário configurar o número de neurônios da camada escondida devido generalização (dados de entradas levado a alta

dimensão) da ANN *feedforward*, entretanto, deve-se escolher uma função de ativação infinitamente diferenciável, e ajustar os coeficientes de regularização C e parâmetro *kernel* γ associado à função de ativação empregada. Os valores ótimos para C e γ devem ser escolhidos no intervalo de $[2^{-25}, 2^{-24}, \dots, 2^{24}, 2^{25}]$ (Huang et al., 2012). Trabalhos relacionados à ELM e DGA podem ser vistos em Venkatasami e Latha (2016), e Ikeshoji et al. (2018).

3.3 Floresta de Caminhos Ótimos - OPF

Diferentemente dos algoritmos ANN e ELM que utilizam hiperplanos separadores para classificar padrões, o OPF supervisionado (Papa et al., 2009) é um classificador baseado em grafos que particiona o espaço de atributos com base na relação de adjacência entre os dados de interesse e uma heurística de competição e conquista. Basicamente, a partir de um diagrama de grafo completo, constituído por amostras do conjunto de treinamento conectados entre si através de arcos ponderados pela métrica distância, determina-se nós protótipos (amostras representantes de diferentes classes) responsáveis pelo início do processo de competição e conquista para particionamento do grafo completo em várias árvores de caminhos ótimos. Cada classe é representada por, ao menos, uma árvore e o conjunto de todas essas árvores constituem a floresta de caminhos ótimo. Na fase de teste ou avaliação, novas amostras são inseridas no espaço de atributos e submetidas ao processo de competição e conquista com todas as amostras do espaço e, como resultado, cada nova amostra recebe o rótulo daquela que lhe ofereceu o menor custo (Ikeshoji et al., 2018). O OPF é um algoritmo de fácil utilização, pois, não necessita de parametrização, pode ser aplicado em problemas multiclases e não assume forma ou separabilidade das classes (Papa, 2008; Barbosa e Costa, 2018). Aplicações envolvendo diagnóstico de falhas em transformadores, podem ser encontrados em Souza et al. (2012a) e Ikeshoji et al. (2019).

3.4 K Vizinhos mais Próximos - KNN

O KNN trata-se de um algoritmo também baseado em grafo, não paramétrico e com heurística operacional simples de entender e implementar (Hu et al., 2016). Para classificar o padrão de uma nova amostra, o KNN considera a influência de K elementos de uma determinada classe mais próxima para lhe atribuir o rótulo. Por exemplo, seja $K=1$ vizinho, o algoritmo atribui o rótulo do único elemento mais próximo (menor distância) para a nova amostra. Caso o número de vizinhos selecionado seja maior que um, o classificador considera o maior número de vizinhos de mesma classe mais próximos (sistema de voto majoritário) e então rotula a amostra, tornando-o assim menos suscetível a presença de *outliers* nos dados (Souza et al., 2012b). Para melhorar o desempenho do KNN existem diversas propostas de técnicas, como por exemplo, ponderar a distância (Gou et al., 2012), ponderar os K vizinhos mais próximos, usar conjunto de dados sintéticos ou usar o classificador médio *Cmeans* (Sarma et al., 2011). Aplicações que envolvem DGA, podem ser encontradas em Benmahamed et al. (2018) e Samirmi et al. (2013).

3.5 Kmeans Cluster

Trata-se de um algoritmo que busca encontrar semelhanças entre dados não rotulados de um conjunto de amostras e agrupá-los em K diferentes grupos. O *Kmeans* é um algoritmo que funciona somente com dados numéricos, pois as variáveis envolvidas são avaliadas por medidas escalares (por exemplo, distância) e o agrupamento é feito com base na minimização do custo por meio da mudança média dos K centroides dos grupos envolvidos (Huang, 1998). Basicamente, o algoritmo inicialmente distribui aleatoriamente os K grupos e seus centroides no espaço de atributos dos dados, posteriormente os utiliza como ponto partida para fazer n iterações de correção e otimização das posições desses centroides usando a minimização da soma do erro quadrático com relação a cada amostra (Fränti e Sieranoja, 2019). A interrupção do processo iterativo ocorre quando o número máximo de iterações (n) é alcançado ou quando as posições dos centroides se estabilizam. Em Selim e Ismail (1984) é provado de forma generalizada a convergência do algoritmo *Kmeans*, sendo investigado e proposto meios para se obter a otimalidade local das soluções ou grupos (*clusters*). Também são apresentadas as notações e formulações matemáticas que envolvem os problemas de generalização, bem como são profundamente investigadas. Para definição do valor adequado de K no algoritmo *Kmeans*, Syakur et al. (2018) utilizaram o método *Elbow* para fazer o agrupamento dos dados de acordo com os atributos definidos. Este método usa o percentual da variância dos dados em função do número de grupos K , ou seja, o algoritmo testa diferentes valores de K . A partir do momento em que os valores de K não apresentam um aumento expressivo da variância, tem-se o valor ideal de K .

4. IMPLEMENTAÇÃO, ANÁLISE E RESULTADO

4.1 Base de dados

O conjunto de dados dos gases dos transformadores usados nesse artigo possui 1516 amostras, dos quais 1141 caracterizados como normal, 168 como falha térmica e 207 como falha elétrica. A Fig. 1 mostra uma projeção gráfica dos componentes principais (PCA - *Principal Component Analysis*) PC1, PC2 e PC3 (Richardson, 2009), extraídos desse conjunto onde é possível observar a distribuição e a zona de confusão entre as 3 diferentes classes. A caracterização em apenas 3 classes se deve as variadas fontes de dados (Ikeshoji et al., 2018; Arantes, 2005; Morais, 2004) considerarem em suas análises diferentes quantidades de classes de falhas. Os gases usados nesse artigo são: o hidrogênio (H_2), metano (CH_4), acetileno (C_2H_2), etileno (C_2H_4) e etano (C_2H_6), não sendo considerado os gases monóxido (CO) e dióxido (CO_2) de carbono pela ausência desses dados em algumas das fontes e por serem relacionados a degradação do papel (IEEE C57.104, 2008). O método *holdout* (Kohavi, 1995) foi usado para a formação dos conjuntos de dados de treinamento e validação dos algoritmos inteligentes para avaliar os métodos propostos de eliminação de *outliers*, a proporção definida foi de 70% e 30% respectivamente, e foram feitas 30 simulações para cada algoritmo de aprendizado de máquinas.

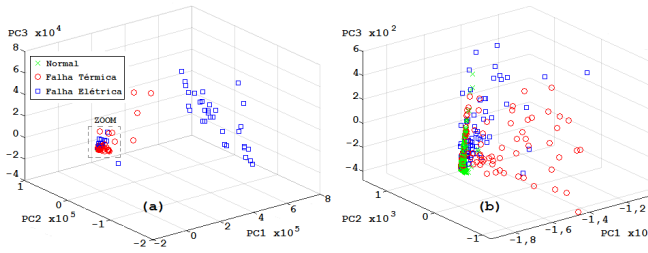


Figura 1. (a) Representação gráfica das 1516 amostras via PCA, (b) Ampliação da zona de confusão ZOOM.

4.2 Implementação dos métodos usados na determinação e remoção dos outliers

Como os dados de DGA, obtidos aleatoriamente, não possuíam os registros dos motivos e observações que provocaram as correspondentes falhas, principalmente para eventos com concentrações atípicas de gases, foram tomados como bases para os métodos de remoção de outliers (Fig. 2) os seguintes fatores: a condição de normalidade das concentrações de gases da IEEE C57.104 (2008); o valor percentual típico dos gases da IEC 60599 (2007) para os quais não se mostra falha detectável; a distribuição normal de probabilidades sobre os dados brutos e dados dos grupos determinado pelo algoritmo *Kmeans*.

A norma IEEE C57.104 (2008) consiste de um guia de interpretação de gases produzidos e dissolvidos no óleo isolante dos transformadores. Seu escopo compreende desde a teoria de formação dos gases e sua relação com as falhas, até sugestões de procedimentos para acompanhamento, análise, diagnóstico, e manutenção. Em seu procedimento de análise denominado de total de gases combustíveis dissolvidos no óleo (TDCG – *Total Dissolved Combustible Gas*) são observadas as taxas de evoluções dos gases (por tipo e valor total), e então relacionado a critérios de tomadas de decisão de acompanhamento e manutenção durante a vida operacional do transformador. Os valores dos gases em $\mu\text{l/l}$ (ou ppm) determinados para a condição de operação normal e sem histórico de gases são: $H_2=100$, $CH_4=120$, $C_2H_2=1$, $C_2H_4=50$ e $C_2H_6=65$.

A norma IEC 60599 (2007) consiste também de um guia para interpretação dos gases dissolvidos no óleo e diagnóstico de falhas. Essa norma abrange análise de equipamentos, tais como: transformadores, reatores, buchas, entre outros. Seu escopo diferencia-se da IEEE C57.104 (2008), entre outros, em alguns procedimentos de análise, na determinação dos valores mínimos e típicos dos gases, e nos limites das razões dos gases do seu método de interpretação de falhas.

No anexo da IEC 60599 (2007), há uma tabela contendo o valor de referência dos gases para diversos tipos de equipamentos, entretanto, recomenda-se seu uso apenas como informativo para auxílio à manutenção, ou seja, não devem ser tratados como limites. Como valor de referência típico para concentração de gases normais, a norma relaciona um percentual de 90% calculado sobre o conjunto de dados das amostras consideradas normais. A margem de 10% restante destina-se a uma faixa de transição entre falhas onde uma análise mais cuidadosa é necessária. Entretanto, variar o percentual de 90% para

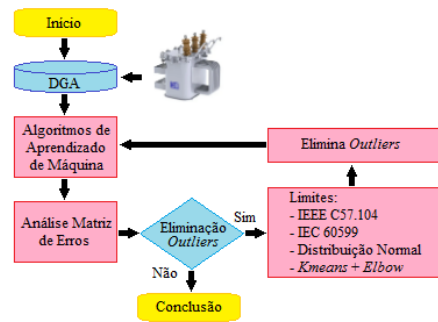


Figura 2. Fluxograma dos métodos de remoções aplicados.

menos envolve riscos de colocar muitos equipamentos sob suspeita de falhas, ou para mais pode ocorrer uma falha sem aviso prévio, portanto, os valores típicos devem ser considerados como diretrizes iniciais quando nenhuma outra informação estiver disponível (IEC 60599, 2007), logo, adotou-se o valor típico de 95% calculados de acordo com Duval (2008).

O algoritmo *Kmeans* foi empregado como fator de seleção dos maiores grupos das amostras analisadas em torno de 95% dos dados, com uma segunda filtragem associada a 99,7% da distribuição normal. Quanto a distribuição normal de probabilidade sobre os dados brutos adotou-se o valor de 95% de possibilidade de ocorrência.

Para avaliação das técnicas e métodos de remoção dos outliers propostos neste artigo, empregou-se o MATLAB (2010) com os dados dos gases normalizados e parametrizações dos algoritmos inteligentes para melhor assertividade conforme apresenta a Tabela 1.

Tabela 1. Parametrizações implementadas.

Algoritmo	Parametrizações
ANN	1 camada escondida com 9 neurônios, Sigmóide, Levenberg Marquardt, Épocas= 10^3 , Erro= 10^{-4} , Taxa Aprendizado= 10^{-3}
ELM	Função <i>kernel Gaussian</i> , Regularização $C=2^9$, Parâmetro <i>kernel</i> $\gamma=2^{-7}$
KNN	$K=3$ vizinhos mais próximos

4.3 Resultados obtidos e análise

Inicialmente foram avaliados a qualidade dos dados brutos (amostras) de DGA obtidos de diferentes fontes através da assertividade dos algoritmos classificadores inteligentes. Os resultados são apresentados na Tabela 2 sob a forma de matriz de erros, junto com as taxas de acertos por classe global (AG) e ponderada (AP). A taxa AG se refere a taxa de acerto dos classificadores que independe do fator desbalanceamento das amostras e classes; já a classe AP, conforme proposto por Papa et al. (2009), considera a influência desse fator.

Comparando-se ambas as taxas AG e AP da Tabela 2, é possível verificar o impacto sobre o percentual de assertividade dos classificadores quando há uma classe dominante sobre as demais, o que pode não refletir a qualidade do resultado de classificação sobre todas as classes como no caso AG. Complementarmente, a análise dessa matriz também permite observar que as taxas de acertos das falhas térmicas e elétricas em todos os algoritmos estão

baixas e com muitas amostras classificadas como normal. Na análise de falhas incipientes em transformadores via DGA, as diferentes classes de falhas determinam o nível da gravidade e urgência no atendimento da correção (Duval, 2012). O erro de classificação de uma amostra normal como falha, a princípio é menos grave que o oposto, pois, mesmo que a primeira situação envolva custos, o valor financeiro é muito menor do que a decorrente da segunda situação quando o transformador sai de operação abruptamente gerando perdas da ordem de milhares ou milhões de dólares (Ikeshoji et al., 2019).

Nesse contexto, uma avaliação nos dados das amostras e remoção dos *outliers* podem contribuir para reduzir os erros de classificações das falhas incipientes, no entanto, há necessidade de identificá-las e avaliar a profundidade da limpeza para que não descaracterize a realidade das informações contidas nos dados das amostras. Considerando-se apenas as incertezas inseridas nas medições dos gases, Jafar et al. (2011) realizaram um trabalho de avaliação com a colaboração de 24 laboratórios de 14 diferentes países e observou uma variação entre 3 a 7,5% de *outliers* como sendo normal. Assim, priorizando melhorar os resultados dos erros de classificações de falhas térmicas e elétricas com a classe normal por sua criticidade no diagnóstico de falhas, tornou-se foco deste artigo avaliar a presença (com base somente nos dados dos gases apresentados) de possíveis *outliers* nos dados somente das falhas.

Tabela 2. Matriz de erros com dados brutos.

	N	T	E	A(%)	N	T	E	A(%)
ANN: AP=89,4%, AG=93,8%					ELM: AP=83,3%, AG=91,2%			
N	337	4	1	98,5	340	1	1	99,4
T	12	35	3	70,0	16	29	5	58,0
E	5	3	54	87,1	15	3	44	71,0
KNN: AP=84,3%, AG=90,1%					OPF: AP=82,5%, AG=86,8%			
N	333	2	7	97,4	321	10	11	93,9
T	13	29	8	58,0	12	29	9	57,5
E	10	4	48	77,4	10	7	45	72,6
N – Normal; T – Térmica; E – Elétrica; A – Taxa de acerto por classe; AP – Acerto Ponderado; AG – Acerto Global.								

Portanto, sobre o conjunto de 1141 amostras normais que constitui o maior número de dados por classe avaliada, determinou-se os valores limites dos gases para corte dos *outliers* existentes nos dados brutos das falhas, e também o número de grupos ideal ($K=12$) do algoritmo *Kmeans* calculado para uma variância menor que 10% pelo método *Elbow* (Fig. 3).

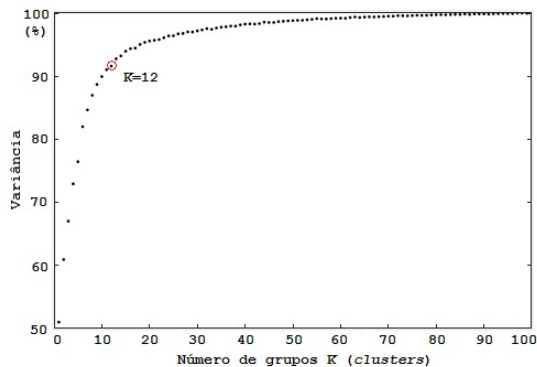


Figura 3. Gráfico *Elbow* para determinação do K ideal.

Tabela 3. Valores limites de corte dos *outliers*.

Técnica	H ₂	CH ₄	C ₂ H ₂	C ₂ H ₄	C ₂ H ₆	FTE	FEE
DNP	94	105	1,4	87,5	188,4	21	8
IEEE	100	120	1	50	65	15	7
IEC	56	118	1	67	187	16	7
<i>kmeans</i>	60,5	128	2,1	86,7	202,4	19	8
Valores em ppm; DNP – Distribuição Normal de Probabilidade; FTE – Falha Térmica Extraída; FEE – Falha Elétrica Extraída							

A Tabela 3 apresenta os limites dos gases encontrados segundo as técnicas e método relacionados nesse artigo, e aplicados na remoção dos *outliers* sobre os conjuntos de falhas térmica e elétrica, e a Tabela 4 mostra o resultado percentual da remoção dos *outliers* sobre 168 amostras com falhas térmicas, 207 com falhas elétricas e o total de ambas.

Tabela 4. Percentual de amostras extraídas.

Técnica	Falha térmica	Falha elétrica	Falha Total
DNP	12,5	3,8	7,7
IEEE	8,9	3,4	5,9
IEC	9,5	3,4	6,1
<i>Kmeans</i>	11,3	3,8	7,2
Valores em (%); DNP – Distribuição Normal de Probabilidade			

A maior quantidade de amostras eliminadas no conjunto de falhas térmicas (Tabela 4) reflete a estimativa de um número maior de candidatos a *outliers*. Analisando-se os valores dos atributos (gases) dessas falhas eliminadas, todos apresentam semelhança com os valores das amostras normais, e como não há informações adicionais à priori sobre cada uma das mesmas, acabam sendo classificadas erroneamente como normais. Embora, individualmente essas falhas possuam elevado percentual de amostras extraídas, globalmente encontram-se dentro dos limites de 3 a 7,5% (Jafar et al., 2011) esperado para as incertezas.

A Tabela 5 apresenta os resultados das classificações de falhas em transformadores após a remoção dos *outliers* nos conjuntos de dados dos gases de falhas. De um modo geral, observa-se que houve melhoras nas taxas de acertos AP e AG, bem como nas taxas individuais de falhas de cada algoritmo quando se compara com os valores simulados da Tabela 1. Dentre as técnicas empregadas para remoção dos *outliers* a DNP se mostrou mais eficiente pelos próprios limites dos gases determinados pelo método, entretanto, as demais técnicas contribuíram para identificar amostras em comum detectadas como *outliers*, aumentando a certeza sobre as mesmas. Avaliando o desempenho das ferramentas inteligentes, nota-se que a ANN foi a melhor classificadora, provando ser menos sensível à presença de *outliers*.

5. CONCLUSÃO

A partir do uso das amostras rotuladas como normal e de técnicas e métodos de remoção de *outliers* propostas nesse artigo, foi possível diminuir o efeito negativo da interferência de ruídos presentes nos dados sobre os erros críticos de classificações em falhas dos transformadores, ou seja, diagnóstico de falhas sendo classificadas como normal. Isto demonstra a importância de se tratar os dados antes de submetê-los à classificação. Neste contexto, a matriz de erros (ou confusão) demonstrou ser uma importante ferramenta de auxílio a análise, pois permite conhecer a eficiência das ferramentas inteligentes na classificação

Tabela 5. Matriz de erros após extrair outliers.

	N	T	E	A(%)	N	T	E	A(%)
DNP								
ANN: AP=92,8%, AG=95,5%				ELM: AP=85,6%, AG=92,6%				
N	337	4	1	98,5	340	1	1	99,4
T	6	35	3	79,6	11	28	5	63,3
E	3	3	54	88,5	12	3	45	75
KNN: AP=87,0%, AG=91,9%				OPF: AP=85,5%, AG=90,1%				
N	334	1	7	97,7	329	4	9	96,2
T	7	29	8	65,9	7	29	8	65,9
E	9	4	47	78,3	9	7	44	73,3
IEEE C57.104 (2008)								
ANN: AP=91,4%, AG=94,8%				ELM: AP=85,2%, AG=92,4%				
N	337	4	1	98,5	340	1	1	99,4
T	8	35	3	76,1	13	29	4	63,0
E	4	3	53	88,3	12	3	45	75
KNN: AP=86,5%, AG=92,0%				OPF: AP=84,6%, AG=89,3%				
N	335	3	4	98,0	327	5	10	95,6
T	10	30	6	65,2	9	29	8	63,0
E	9	4	47	78,3	8	8	44	73,3
IEC 60599 (2007)								
ANN: AP=90,9%, AG=94,4%				ELM: AP=85,6%, AG=92,6%				
N	336	5	1	98,2	340	1	1	99,4
T	8	34	3	75,6	12	29	4	64,4
E	4	4	52	86,7	12	3	45	75,0
KNN: AP=86,8%, AG=91,9%				OPF: AP=84,9%, AG=89,7%				
N	334	2	6	97,7	328	5	9	95,9
T	9	29	7	64,4	9	27	9	60,0
E	8	4	48	80,0	7	7	46	76,7
Kmeans								
ANN: AP=91,8%, AG=95,1%				ELM: AP=85,8%, AG=92,6%				
N	337	4	1	98,5	340	1	1	99,4
T	8	34	3	75,9	11	30	4	66,7
E	3	3	54	90	12	4	44	73,3
KNN: AP=85,9%, AG=91,9%				OPF: AP=84,9%, AG=89,9%				
N	336	2	4	98,2	330	4	8	96,5
T	11	28	6	62,2	7	28	10	62,2
E	9	4	47	78,3	9	7	44	73,3
N – Normal; T – Térmica; E – Elétrica; A – Taxa de acerto por classe; AP – Acerto Ponderado; AG – Acerto Global.								

de padrões, bem como avaliar como todas as instâncias envolvidas foram classificadas e também a qualidade dos dados de concentrações de gases. Como continuidade deste trabalho, pretende-se avaliar a eficiência de outras técnicas e métodos de detecção e remoção de outliers, bem como o uso de outros algoritmos classificadores de falhas em transformadores.

REFERÊNCIAS

Afonso, L.C.S., Pereira, D.R., and Papa, J.P. (2017). A Kernel-based Optimum-Path Forest Classifier. In *Iberoamerican Congress on Pattern Recognition*, 1–8.

Aguinis, H., Gottfredson, R., and Joo, H. (2013). Best-Practice Recommendations for Defining, Identifying, and Handling Outliers. *Organizational Research Methods*, 16(2), 270–301.

Amora, M., Almeida, O., Braga, A., Barbosa, F., Lisboa, L., and Pontes, R. (2012). Improved DGA method based on rules extracted from high-dimension input space. *Electronics Letters*, 48(17), 1048–1049.

Arantes, J.G. (2005). *Diagnóstico De Falhas Em Transformadores De Potência Pela Análise De Gases Dissolvidos Em Óleo Isolante Através De Redes Neurais*. Ph.D. thesis, Dissertação (Mestrado em Ciências), UNIFEI.

Barbosa e Costa, G.S. e Costa, S. (2018). A New Genetic Algorithm-based Pruning Approach for Optimum-Path Forest. In *2018 7th Brazilian Conference on Intelligent Systems*, 13–18. BRACIS.

Bechara, R. (2010). *Análise de falhas de transformadores de potência*. Ph.D. thesis, Dissertação (Mestrado em Engenharia Elétrica), Escola Politécnica da USP.

Benmahamed, Y., Kemari, Y., Teguar, M., and Bouba-keur, A. (2018). Diagnosis of Power Transformer Oil Using KNN and Naïve Bayes Classifiers. In *2018 IEEE 2nd International Conference on Dielectrics*, 3, 1–4.

Cheng e Sutariya, F. e Sutariya, V. (2012). Applications of Artificial Neural Network Modeling in Drug Discovery. *Clinical and Experimental Pharmacology*, 2(3), 2–3.

Cheng e Yu, Lefeng e Yu, T. (2018). Dissolved gas analysis principle-based intelligent approaches to fault diagnosis and decision making for large oil-immersed power transformers: A survey. *Energies*, 11(4), 1–69.

Duval, M. (2008). Calculation of DGA limit values and sampling intervals in transformers in service. *IEEE Electrical Insulation Magazine*, 24(5), 7–13.

Duval, M. (2012). New frontiers of DGA interpretation for power transformers and their accessories. URL http://www.seeei.org.il/prdFiles/2922{_}desc2.pdf.

Farooque, M., Wani, S., and Khan, S. (2016). Artificial neural network (ANN) based implementation of Duval pentagon. In *2015 International Conference on Condition Assessment Techniques in Electrical Systems*.

Fränti e Sieranoja, Pasi e Sieranoja, S. (2019). How much can k-means be improved by using better initialization and repeats? *Pattern Recognition*, 93, 95–112.

Gou, J., Du, L., Zhang, Y., and Xiong, T. (2012). A new distance-weighted k-nearest neighbor classifier. *Journal of Information & Computational Science*, 9, 1429–1436.

Gupta, M., Gao, J., Aggarwal, C., and Han, J. (2014). Outlier Detection for Temporal Data: A Survey. *IEEE Transactions on Knowledge and Data Engineering*, 26(9), 2250–2267.

Hu, L.Y., Huang, M.W., Ke, S.W., and Tsai, C.F. (2016). The distance function effect on k-nearest neighbor classification for medical datasets. *SpringerPlus*, 5(1).

Huang, G.B., Zhou, H., Ding, X., and Zhang, R. (2012). Extreme learning machine for regression and multiclass classification. *IEEE transactions on systems, man, and cybernetics. Part B, Cybernetics*, 42(2), 513–29.

Huang, G., Zhu, Q., and Siew, C. (2006). Extreme learning machine: Theory and applications. *Neurocomputing*, 70(1-3), 489–501.

Huang, Z. (1998). Extensions to the k-Means Algorithm for Clustering Large Data Sets with Categorical Values. *Data Mining and Knowledge Discovery* 2, 283–304. *Data Mining and Knowledge Discovery*, 2(3), 283–304.

IEC 60599 (2007). *International Electrotechnical Commission - IEC 60599: Mineral oil-impregnated electrical equipment in service – Guide to the interpretation of dissolved and free gases analysis*. IEC, Geneva.

IEEE C57.104 (2008). *Institute of Electrical and Electronics Engineers - IEEE C57.104: Guide for the Interpretation of Gases Generated in Oil-Immersed Transformers*. February. IEEE, New York, NY.

Ikeshoji, M., Souza, A., Junior, P., Gastaldello, D., Gonzales, C., and Papa, J. (2019). Qualidade no Fornecimento de Energia: Contribuições de uma Ferramenta

- Inteligente para Gestão de Falhas Incipientes em Transformadores. In *XIII CBQEE*.
- Ikeshoji, M., Souza, A., Junior, P., Gastaldello, D., Silva, M., and Maginador, J. (2018). Desenvolvimento de uma Ferramenta de Diagnóstico de Falhas em Transformadores Utilizando à Abordagem de Aprendizagem de Máquina Extrema. In *2018 INDUSCON*.
- Jafar, M., Visser, R., and Starink, R. (2011). Results of Proficiency Test Dissolved Gas Analysis. URL <http://www.iisnl.com/pdf/iis10LO3DGA.pdf>.
- Kaur, A., Brar, Y., and Leena, G. (2019). Fault detection in power transformers using random neural networks. *International Journal of Electrical and Computer Engineering (IJECE)*, 9(1), 78.
- Khalymasaa, A., Bliznyuk, D., Verkhozin, A., and Ovchinnikov, K. (2016). The problems of dissolved in oil gases analysis results' interpretation in information analytical systems. In *Proceedings of the 2016 International Conference and Exposition on Electrical and Power Engineering*, 743–747.
- Kohavi, R. (1995). A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. In *International Joint Conference on Artificial Intelligence*, volume 5, 7.
- Liepniece, R., Vitolina, S., and Marks, J. (2017). Study of approaches to incipient fault detection in power transformer by using dissolved gas analysis. *Energetika*, 63(2), 66–74.
- Ma, H., Ekanayake, C., and Saha, T. (2012). Power transformer fault diagnosis under measurement originated uncertainties. *IEEE Transactions on Dielectrics and Electrical Insulation*, 19(6), 1982–1990.
- MATLAB (2010). *version (R2010b)*. The MathWorks Inc., Natick, Massachusetts.
- Morais, D.R. (2004). *Ferramenta inteligente para detecção de falhas incipientes em transformadores baseada na análise de gases dissolvidos no óleo isolante*. Ph.D. thesis, Dissertação (Mestrado em Engenharia Elétrica), UFSC, Florianópolis.
- Nur, A.S., Haizan, N.H.M., and Osman, A. (2014). Artificial Neural Network Weight Optimization : A Review. *TELKOMNIKA Indonesian Journal of Electrical Engineering*, 12(March 2017), 6897 – 6902.
- Papa, J.P. (2008). *Classificação supervisionada de padrões utilizando floresta de caminhos ótimos*. Tese de doutorado, Tese (Doutorado em Ciência da Computação), Universidade Estadual de Campinas, UNICAMP.
- Papa, J., Falcão, A., and Suzuki, C. (2009). Supervised pattern classification based on optimum-path forest. *International Journal of Imaging Systems and Technology*, 19(2), 120–131.
- Qin, J., Yang, Y., Du, H., and Hong, Z. (2019). Outlier detection for on-line monitoring data of transformer based on wavelet transform and weighted LOF. In *2019 International Conference on New Energy and Future Energy System*, volume 354, 11.
- Richardson, M. (2009). Principal Component Analysis. URL <http://www.dsc.ufcg.edu.br/~hmg/disciplinas/posgraduacao/rncopin-2014.3/material/SignalProcPCA.pdf>.
- Samirmi, F.D., Tang, W., and Wu, H. (2013). Feature selection in power transformer fault diagnosis based on dissolved gas analysis. In *2013 4th IEEE/PES Innovative Smart Grid Technologies Europe*, 1–5.
- Sarma, T., Viswanath, P., Reddy, D., and Raghava, S. (2011). An improvement to k-nearest neighbor classifier. *2011 IEEE Recent Advances in Intelligent Computational Systems*, 227–231.
- Selim e Ismail, Shokri Z. e Ismail, M.A. (1984). K-Means-Type Algorithms: A Generalized Convergence Theorem and Characterization of Local Optimality. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(1), 81–87.
- Shawe-taylor e Sun, John e Sun, S. (2011). A review of optimization methodologies in support vector machines. *Neurocomputing*, 74(17), 3609–3618.
- Sokolov, V. (2005). Understanding failure modes of transformers. In *Euro TechCon 2005*, 43–66.
- Souza, A., Ramos, C., Gastaldello, D., Nakamura, R., and Papa, J. (2012a). Fast fault diagnosis in power transformers using optimum-path forest. In *INES 2012 - IEEE 16th International Conference on Intelligent Engineering Systems, Proceedings*, 209–212.
- Souza, R., Lotufo, R., and Rittner, L. (2012b). A comparison between optimum-path forest and κ -nearest neighbors classifiers. In *Brazilian Symposium of Computer Graphic and Image Processing*, 260–267.
- Stehman, S. (1997). Selecting and Interpreting Measures of Thematic Classification Accuracy. *Remote Sensing of Environment*, 62(1), 77–89.
- Subroto, C., Suwarno, Trianto, and Zhang, G. (2017). Artificial intelligence for DGA interpretation methods using weighting factor. In *1st International Conference on Electrical Materials and Power Equipment*, 85–88.
- Suguna e Thanushkodi, N e Thanushkodi, K. (2010). An Improved k-Nearest Neighbor Classification Using Genetic Algorithm. *IJCS*, 7(4), 7–10.
- Sun, H.C., Huang, Y.C., and Huang, C.M. (2012). A Review of Dissolved Gas Analysis in Power Transformers. In *2nd International Conference on Advances in Energy Engineering*, 2011, 1220–1225.
- Syakur, M., Khotimah, B., Rochman, E., and Satoto, B. (2018). Integration K-Means Clustering Method and Elbow Method for Identification of the Best Customer Profile Cluster. *IOP Conference Series: Materials Science and Engineering*, 336(1).
- Tenbohlen, S., Coenen, S., Djamali, M., Müller, A., Samimi, M., and Siegel, M. (2016). Diagnostic measurements for power transformers. *Energies*, 9(5), 1–25.
- Van Den Broeck, J., Cunningham, S., Eeckels, R., and Herbst, K. (2005). Data cleaning: Detecting, diagnosing, and editing data abnormalities. *PLoS Medicine*, 2(10).
- Venkatasami e Latha, A. e Latha, P. (2016). Application of Extreme Learning Machine in Fault Classification of Power Transformer. *Circuits and Systems*, 7(10).
- Zakaria, F., Johari, D., and Musirin, I. (2013). Artificial neural network (ANN) application in dissolved gas analysis (DGA) methods for the detection of incipient faults in oil-filled power transformer. In *Proceedings - 2012 IEEE International Conference on Control System, Computing and Engineering*, 328–332.
- Zhao, A., Tang, X.J., Zhang, Z.H., and Liu, J.H. (2014). The DGA interpretation method using relative content of characteristic gases and gas-ratio combinations for fault diagnosis of oil-immersed power transformers. In *ISEIM*, 124–127.