


# Avaliação de técnicas de inteligência artificial na classificação de descargas parciais

## Evaluation of artificial intelligence techniques in the classification of partial discharges

Allan D. C. Silva<sup>1</sup>, Itaiara F. Carvalho<sup>2</sup>, Luiz A. M. M. Nobrega<sup>3</sup>,  
George V. R. Xavier<sup>4</sup>, Edson G. da Costa<sup>5</sup>

Silva, A.D.C; Carvalho, I.F; Nobrega, L.A.M.M;  
Xavier, G.V.R; da Costa, E.G. Avaliação de técnicas de inteligência artificial na classificação de descargas parciais. *Tecnología en Marcha*. Vol. 34, especial. ALTAE. Diciembre 2021. Pág 232-244.

 <https://doi.org/10.18845/tm.v34i7.6047>

- 1 Universidade Federal de Campina Grande, Campina Grande, Brasil. Correo electrónico: [allan.costa@ee.ufcg.edu.br](mailto:allan.costa@ee.ufcg.edu.br)  
 <https://orcid.org/0000-0003-2881-632X>
- 2 Universidade Federal de Campina Grande, Campina Grande, Brasil. Correo electrónico: [itaiara.carvalho@ee.ufcg.edu.br](mailto:itaiara.carvalho@ee.ufcg.edu.br)  
 <https://orcid.org/0000-0002-3662-1955>
- 3 Universidade Federal de Campina Grande, Campina Grande, Brasil. Correo electrónico: [luiz.nobrega@dee.ufcg.edu.br](mailto:luiz.nobrega@dee.ufcg.edu.br)  
 <https://orcid.org/0000-0003-2634-8526>
- 4 Instituto de Telecomunicações, Universidade de Aveiro, Aveiro, Portugal. Correo electrónico: [george.xavier@av.it.pt](mailto:george.xavier@av.it.pt)  
 <https://orcid.org/0000-0002-1056-4948>
- 5 Universidade Federal de Campina Grande, Campina Grande, Brasil. Correo electrónico: [edson@dee.ufcg.edu.br](mailto:edson@dee.ufcg.edu.br)  
 <https://orcid.org/0000-0003-2998-2039>



## Palavras - chave

Aprendizado de máquina; classificação; data mining; descargas parciais; extração de atributos.

## Resumo

A detecção de sinais de Descargas Parciais (DP) na faixa da Ultra Alta Frequência (UHF) permite identificar e classificar, de forma minimamente invasiva, defeitos em equipamentos de alta tensão, bem como estimar o grau de urgência da realização de manutenções preventivas. Neste artigo, técnicas de *machine learning* foram utilizadas para realizar o reconhecimento automático dos padrões obtidos a partir de envoltórias de sinais UHF de DP. Para tanto, foi elaborado um arranjo experimental para emular diferentes fontes de DP: uma cuba de óleo com eletrodos ponta-plano, a barra de um hidrogerador e um transformador de potencial. A partir dos sinais obtidos nesse arranjo, geraram-se envoltórias, a partir das quais foi realizada a extração de uma série de atributos no domínio do tempo, tais como: curtose, amplitude máxima e tempo de subida. Em seguida, realizou-se a seleção dos atributos por meio de uma associação de algoritmos, dentre eles o *k-means*, de forma a reduzir a dimensionalidade dos dados para aumentar a eficiência do algoritmo classificador. Por fim, fez-se a classificação dos sinais de DP a partir de uma rede neural artificial, *decision tree* e *random forest*. Os resultados mostraram que os atributos extraídos das envoltórias foram efetivos na classificação dos sinais de DP, com valores de acurácia média superiores a 95% quando foi utilizado o banco de dados otimizado.

## Keywords

Machine learning; classification; data mining; partial discharge; feature extraction.

## Abstract

The detection of Partial Discharge (PD) signals in the Ultra High Frequency (UHF) range performs identify and classify, in a minimally invasive way, of defects in high voltage equipment, as well as estimating the degree of urgency in carrying out preventive maintenance. In this paper, machine learning techniques were used to perform automatic recognition of patterns obtained from PD UHF signal envelopes. Therefore, an experimental arrangement was designed to emulate different PD sources: an oil vat with flat-tip electrodes, a hydro generator bar, and a potential transformer. From the signals launched in this arrangement, envelopes were generated, from which a series of attributes in the time domain were extracted, such as kurtosis, maximum amplitude, and rise time. Then, the selection of attributes was carried out through an association of algorithms, including k-means, to reduce the dimensionality of the data to increase the efficiency of the classifier algorithm. Finally, a classification of PD signals was performed using an artificial neural network, decision tree, and random forest. The results induced that the attributes extracted from the envelopes were effective in classifying PD signs, with mean accuracy values greater than 95% when the optimized database was used.

## Introdução

O fenômeno das descargas parciais (DP) pode ser definido como descargas de pequena duração que curto circuitam parcialmente a isolação entre dois condutores. Essas descargas geralmente estão associadas ao estresse no meio isolante causado pela presença de um campo elétrico intenso [1].

Ainda que seja de pequena intensidade, a atividade contínua das DP pode causar a deterioração progressiva do sistema de isolamento dos equipamentos de alta tensão, os quais têm uma relação forte com a confiabilidade do sistema elétrico de potência, uma vez que são responsáveis pelas etapas de geração, transmissão e distribuição de energia elétrica.

Tradicionalmente, o processo de manutenção desses equipamentos ocorre de maneira programada, baseando-se no tempo de operação. Entretanto, esse processo tem sido substituído por técnicas de monitoramento preditivo, dentre as quais, destaca-se a medição da irradiação na faixa da ultra alta frequência (UHF, do inglês: *ultra high frequency*), que constitui um método não invasivo e que permite a detecção, classificação e a localização das DP utilizando sensores UHF [2].

Além da boa imunidade aos ruídos externos presentes em subestações, tais como descargas corona e ruídos de linha [3], proporcionada por esse método, o Grupo de Trabalho A2.27 do Cigré [4] recomendou que os transformadores de potência fossem fabricados com pelo menos quatro janelas dielétricas, a fim de viabilizar o monitoramento via o método UHF sem a necessidade de adaptação dos equipamentos para essa finalidade, o que dá maior respaldo ao uso desse método.

Ressalta-se aqui que a medição de DP associada às técnicas de processamento digital de sinais fornece informações da condição operativa dos equipamentos, sendo possível identificar a natureza e o grau de importância do defeito [2, 5]. Isso tem motivado a avaliação de uma série de técnicas de classificação de sinais de DP, sendo aquelas baseadas em *machine learning* as mais promissoras. Tais técnicas performam a partir da identificação de uma série de parâmetros, os quais são comparados com a assinatura característica de cada defeito [6-7].

Neste artigo são apresentados os resultados da aplicação de técnicas de *machine learning* associadas às estratégias de seleção na classificação de envoltórias de sinais de descargas parciais, obtidos a partir de um arranjo experimental que emulou diferentes fontes de DP: uma cuba de óleo com eletrodos ponta-plano, a barra de um hidrogenador e um transformador de potencial.

O trabalho foi organizado da seguinte forma: na seção II são apresentados os principais conteúdos que dão embasamento ao estudo, na seção III é descrita a metodologia, na seção IV os resultados são discutidos e, por fim, na seção V as conclusões são apresentadas.

## Fundamentação Teórica

### Envoltórias

O detector de envoltória é um demodulador de amplitude ideal, que permite reduzir a frequência de amostragem exigida para detectar a envoltória de um sinal [8]. Tal ferramenta se torna útil, em particular, na detecção de fontes de DP, que exigem equipamentos de medição com altas taxas de amostragem.

A geração de envoltórias via o processamento dos sinais permite suavizar os pulsos de DP, preservando características importantes, como energia, amplitude e duração. Dentre as formas de gerar envoltórias, apresenta-se aqui o algoritmo utilizado para processar o sinal e gerar a envoltória a partir da suavização de densidade por Kernel (KDE, do inglês: *Kernel Density Estimation*) [9]:

1. Remove-se o nível DC do sinal a partir do cálculo do valor médio ( $V_{med}$ ):

$$V_{med} = \frac{1}{N} \sum_{n=0}^{N-1} V_i \quad (1)$$

$$V_k = V_i - V_{med} \quad (2)$$

2. Eleva-se as amostras ao quadrado:

$$V_{sq} = V_k^2 \quad (3)$$

3. Realiza-se a aplicação do KDE para a extração das envoltórias de DP:

$$E_{ni} = \frac{\sum_{j=1}^n K \frac{t_i - t_j}{W} V S_j}{\sum_{j=1}^n K \frac{t_i - t_j}{W}}, \quad (4)$$

em que:

- $E_{ni}$ : é o componente suavizado da envoltória na i-ésima amostra;
- $V S_j$ : é a tensão instantânea ao quadrado na j-ésima amostra;
- $t_i$  e  $t_j$ : são amostras de tempo no i-ésimo e j-ésimo ponto;
- $W$ : representa o parâmetro de suavização para a estimativa de densidade de Kernel  $f(x)$ ;
- $K$ : é a função de suavização Kernel Guassiana.

### Parâmetros

Conforme [7], a classificação dos sinais de DP pode ser feita a partir de parâmetros básicos, deduzidos e estatísticos. Neste artigo, utilizaram-se os parâmetros estatísticos, bem como, atributos associados às envoltórias dos sinais de DP [10-11]:

1. Valor de pico ( $V_p$ ):

$$V_{p(x)} = \max(x); \quad (5)$$

2. Valor RMS ( $V_{rms}$ ):

$$V_{RMS(x)} = \sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2}; \quad (6)$$

3. Amplitude Média ( $V_{med}$ ):

$$V_{med(x)} = \frac{1}{N} \sum_{i=1}^N |x_i|; \quad (7)$$

4. Amplitude de raiz quadrada ( $V_{r2}$ ):

$$V_{r2(x)} = \left( \frac{1}{N} \sum_{i=1}^N \sqrt{|x_i|} \right)^2 ; \quad (8)$$

5. Área sob a envoltória ( $A_r$ ): área sob a curva da envoltória;

6. Fator de crista ( $F_c$ ):

$$F_{c(x)} = \frac{V_p(x)}{V_{RMS(x)}} ; \quad (9)$$

7. Fator de liberação ( $F_l$ ):

$$F_{l(x)} = \frac{V_p(x)}{V_{r2(x)}} ; \quad (10)$$

8. Fator de impulso ( $F_i$ ):

$$F_{i(x)} = \frac{V_p(x)}{V_{med}} ; \quad (11)$$

9. Fator de forma ( $F_f$ ):

$$F_{f(x)} = \frac{V_{RMS(x)}}{V_{med}} ; \quad (12)$$

10. Skewness ( $S_k$ ):

$$S_{k(x)} = \frac{E(x - \mu)^3}{\sigma^3} ; \quad (13)$$

11. Curtose ( $K_u$ ):

$$K_{u(x)} = \frac{E(x - \mu)^4}{4} ; \quad (14)$$

12. Duração ( $T_{mp}$ ): tempo duração do sinal de DP;

13. Tempo de subida ( $T_s$ ): tempo de duração até o valor de amplitude máxima do sinal.

Em que:

- $x$  é o sinal;
- $i$  é a amostra;
- $N$  é a quantidade de amostras;
- $\mu$  é a média.

## Seleção de Atributos

A seleção de atributos é uma etapa primordial dentro da área do *machine learning*, haja vista a quantidade de dados que é processada pelos diversos algoritmos. Ela envolve o conhecimento da base de dados, detectando: qualidade, padronizações e tipos de variáveis, a fim de reduzir a dimensionalidade do banco de dados. Para tanto, algumas técnicas podem ser utilizadas, dentre elas a análise por variância e correlação.

### Variância

A variância mede a dispersão dos atributos em relação à média. Ela é calculada pela equação (15):

$$Var(x) = \frac{\sum(x - \mu)^2}{N} \quad (15)$$

em que, uma alta variância indica que os dados estão assumindo diferentes posições. Em contrapartida, uma variância baixa significa que os dados estão concentrados.

### Correlação

A correlação é dada pela equação (16) [12]:

$$Cor(x, y) = \frac{Cov(x, y)}{S_x S_y} \quad (16)$$

em que, o coeficiente de correlação varia de -1 a +1 e, quanto maior o seu valor absoluto, mais forte a correlação entre as variáveis. Coeficientes de correção maiores que 0 são um indicativo de que as duas variáveis são positivamente correlacionadas, ou seja, o valor da variável x aumenta conforme o valor da variável y aumenta. Em contrapartida, se o coeficiente é negativo, as duas variáveis se correlacionam negativamente.

## Técnicas de *Machine Learning*

Os algoritmos de *machine learning* utilizam a experiência para desempenhar atividades de previsão e classificação. Esses algoritmos estão separados em: supervisionados, como, por exemplo, as redes neurais artificiais (RNA), o *decision tree* e o *random forest*; e não – supervisionados, como o *k-means* [13]. Em síntese, a diferença está na atribuição de uma identificação ao conjunto de dados, o que é necessário para realizar aplicações de classificação.

O algoritmo da árvore de decisão ganhou espaço como método de classificação, dada a sua simplicidade e baixo custo computacional. Tal técnica realiza a divisão recursiva de um problema de maior complexidade em problemas menores: cada nó da árvore realiza um teste para um determinado atributo e cada galho proveniente desse nó será um valor possível [14]. Além disso, esse método é a base para o algoritmo do *Random Forest*, que se trata de uma combinação de múltiplas árvores, em que cada árvore toma a sua decisão e o resultado é obtido pela votação de todas as árvores [15].

As RNA são arquiteturas computacionais baseadas em estruturas biológicas de neurônios, programadas para: se adaptar, aprender, generalizar, agrupar ou organizar dados. Dentro dessas estruturas, o classificador mais elementar é *perceptron*, o qual é formado por um único neurônio, capaz de resolver problemas com dados linearmente separáveis [7].

O *k-means* é um algoritmo de aprendizagem não supervisionada usado para separar dados em um número *k* de *clusters*. Esse algoritmo pode ser descrito em quatro etapas [13]: inicialização dos centros; atribuição dos pontos médios mais próximos; mudança dos pontos médios para o centro do cluster; avaliação da convergência.

## Metodologia

### Arranjo experimental

Para gerar o banco de dados, definiu-se um arranjo experimental para realizar a geração e a detecção de DP em três diferentes fontes: barra de um hidrogerador, uma cuba de óleo com eletrodos na configuração ponta-plano e um transformador de potencial (TP), conforme mostrado nas Figuras 1 – 2.

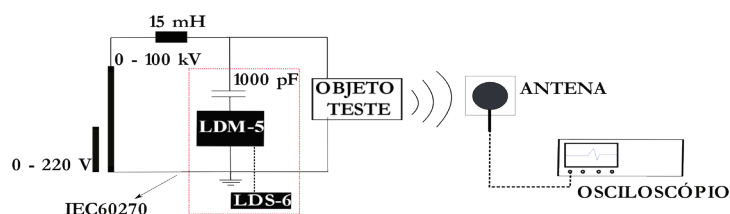


Figura 1. Arranjo experimental.

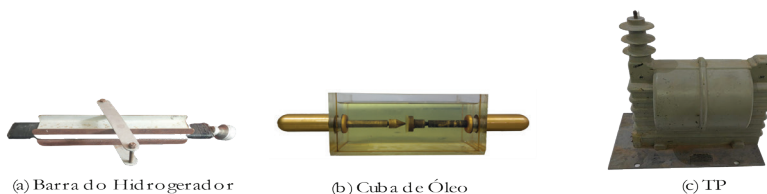


Figura 2. Objetos de teste.

O arranjo, apresentado na figura 1, é formado por um capacitor de acoplamento (1000 pF), uma impedância de medição (LDM-5), um indutor de 15 mH como filtro passa alta e o objeto de teste [3]. A aquisição dos sinais foi realizada a partir de um osciloscópio com taxa de amostragem de 5 GSa/s, largura de banda de 1 GHz e 4 canais. Um desses canais foi conectado à uma antena circular [16] posicionada a 0,65 m do equipamento sob teste, a qual está mostrada na figura 3.

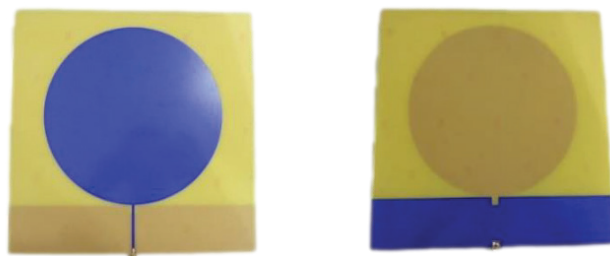
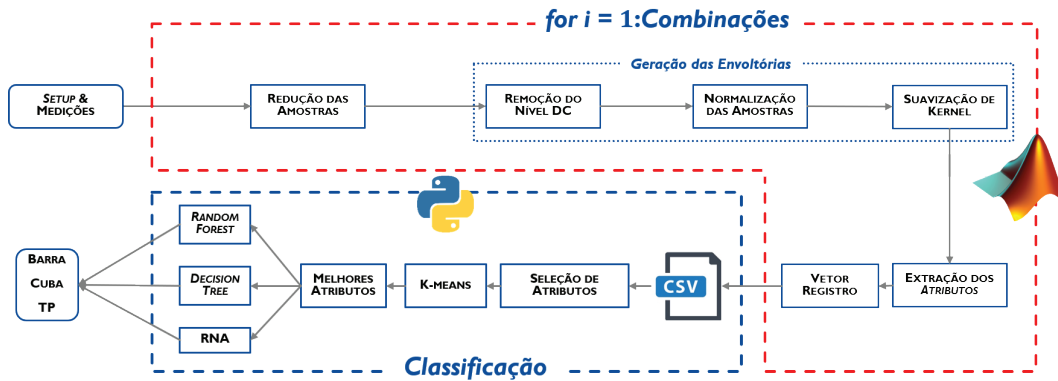


Figura 3. Antena monopolar. Fonte: [16].

Na realização dos ensaios o nível de tensão foi elevado de forma gradativa, por meio de um transformador regulador, até que a atividade de DP fosse gerada, realizando o procedimento de calibração do sistema antes da mudança de cada objeto de teste, haja vista a variação na capacitância total do arranjo.

### Classificação dos defeitos

A classificação das DP é feita conforme o fluxograma mostrado na figura 4.



**Figura 4.** Metodologia utilizada para a classificação de fontes de DP.

Conforme o fluxograma apresentado na figura 4, após a aquisição dos dados, inicia-se o processamento dos sinais. Em primeiro lugar, as amostras dos sinais são reduzidas ao trecho de ocorrência da DP, de forma a diminuir o poder computacional requerido para extrair os atributos. Então, inicia-se o processo de geração das envoltórias, a partir do procedimento descrito na subseção 2.1. Posteriormente, é feita a extração automática de todos os atributos descritos na subseção 2.2. Após essa etapa, parte-se para a etapa de seleção dos atributos via a análise da correlação e da variância, de forma a reduzir a dimensionalidade do banco de dados, preservando apenas 2 atributos.

A eficácia da etapa de seleção dos atributos é verificada a partir da aplicação do método de aprendizagem não supervisionado *k-means*, etapa que precede a classificação e garante a qualidade dos atributos. Então, se dá início à classificação via RNA, *decision tree* e *random forest*.

## Resultados

Apresentam-se nesta seção os principais resultados obtidos com a aplicação das técnicas de classificação utilizadas. Na subseção 4.1 são discutidos os resultados obtidos na etapa de seleção dos atributos; na subseção 4.2 são apresentados os resultados referentes à aplicação do *decision tree* e *random forest* e na subseção 4.3. são apresentados os resultados referentes à RNA.

### Seleção dos atributos

A extração dos atributos gerou um banco de dados com 13 colunas e 30 linhas, ou seja, 13 atributos para cada sinal de DP captado, sendo 10 amostras referentes a cada um dos objetos sob teste. A partir da aplicação e da análise da variância entre as amostras, tabela 1,



destacaram-se os seguintes atributos: área sob a envoltória, fator de liberação e curtose, os quais apresentaram as maiores diferenças e, por isso, há a possibilidade de caracterizar bem os diferentes tipos de DP.

**Tabela 1.** Aplicação da análise por variância.

Atributo	Variância
Pico (F1)	$8,87 \cdot 10^{-1}$
Valor médio (F2)	$4,32 \cdot 10^{-3}$
Amplitude de raiz quadrada (F3)	$4,67 \cdot 10^{-4}$
Valor RMS (F4)	$2,90 \cdot 10^{-2}$
Área (F5)	$1,39 \cdot 10^5$
Fator de crista (F6)	6,29
Fator de liberação (F7)	$1,39 \cdot 10^2$
Fator de impulso (F8)	$6,49 \cdot 10$
Fator de forma (F9)	$8,42 \cdot 10^{-2}$
Skweness (F10)	3,93
Curtose (F11)	$7,80 \cdot 10^2$
Tempo de duração (F12)	$1,00 \cdot 10^{-13}$
Tempo de subida (F13)	$8,18 \cdot 10^{-14}$

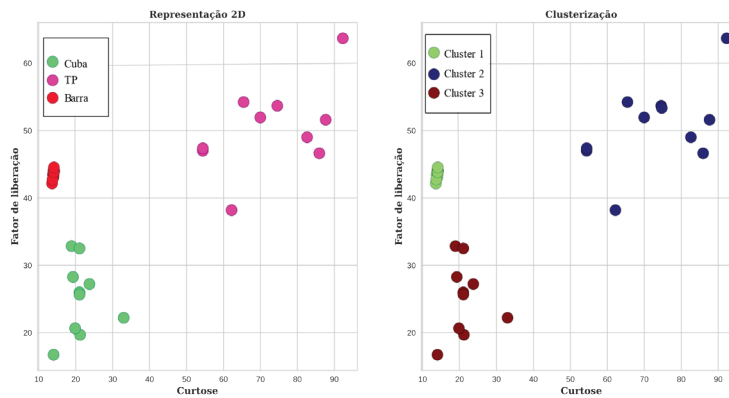
Com relação à análise a partir da técnica da correlação, em primeiro lugar, foi gerada a matriz de correlação apresentada na tabela 2. Então, buscaram-se os atributos com alta correlação com a variável alvo da classificação e baixa correlação com as demais, haja vista que o alto coeficiente de correlação indica semelhança entre as informações [11].

**Tabela 2.** Coeficientes de correlação entre os atributos.

	F <sub>1</sub>	F <sub>2</sub>	F <sub>3</sub>	F <sub>4</sub>	F <sub>5</sub>	F <sub>6</sub>	F <sub>7</sub>	F <sub>8</sub>	F <sub>9</sub>	F <sub>10</sub>	F <sub>11</sub>	F <sub>12</sub>	F <sub>13</sub>	Alvo
F <sub>1</sub>	1,0	1,0	1,0	1,0	1,0	-0,7	0,2	-0,5	0,2	-0,6	-0,6	0,7	-0,9	-0,9
F <sub>2</sub>	1,0	1,0	1,0	1,0	1,0	-0,7	0,2	-0,5	0,2	-0,6	-0,6	0,7	-0,9	-0,9
F <sub>3</sub>	1,0	1,0	1,0	1,0	1,0	-0,7	0,2	-0,5	0,2	-0,6	-0,6	0,7	-0,9	-0,9
F <sub>4</sub>	1,0	1,0	1,0	1,0	1,0	-0,7	0,2	-0,5	0,2	-0,6	-0,6	0,7	-0,9	-0,9
F <sub>5</sub>	1,0	1,0	1,0	1,0	1,0	-0,7	0,2	-0,5	0,2	-0,6	-0,6	0,7	-0,9	-0,9
F <sub>6</sub>	-0,7	-0,7	-0,7	-0,7	-0,7	1,0	0,5	1,0	0,5	1,0	1,0	-0,5	0,7	0,9
F <sub>7</sub>	0,2	0,2	0,2	0,2	0,2	0,5	1,0	0,7	1,0	0,6	0,6	0,2	-0,1	0,2
F <sub>8</sub>	-0,5	-0,5	-0,5	-0,5	-0,5	1,0	0,7	1,0	0,7	1,0	1,0	-0,4	0,6	0,9
F <sub>9</sub>	0,2	0,2	0,2	0,2	0,2	0,5	1,0	0,7	1,0	0,5	0,6	0,2	-0,1	0,2
F <sub>10</sub>	-0,6	-0,6	-0,6	-0,6	-0,6	1,0	0,6	1,0	0,5	1,0	1,0	-0,5	0,6	0,9
F <sub>11</sub>	-0,6	-0,6	-0,6	-0,6	-0,6	1,0	0,6	1,0	0,6	1,0	1,0	-0,4	0,6	0,9
F <sub>12</sub>	0,7	0,7	0,7	0,7	0,7	-0,5	0,2	-0,4	0,2	-0,5	-0,4	1,0	-0,7	-0,7
F <sub>13</sub>	-0,9	-0,9	-0,9	-0,9	-0,9	0,7	-0,1	0,6	-0,1	0,6	0,6	-0,7	1,0	0,8
Alvo	-0,9	-0,9	-0,9	-0,9	-0,9	0,9	0,2	0,9	0,2	0,9	0,9	-0,7	0,8	1,00

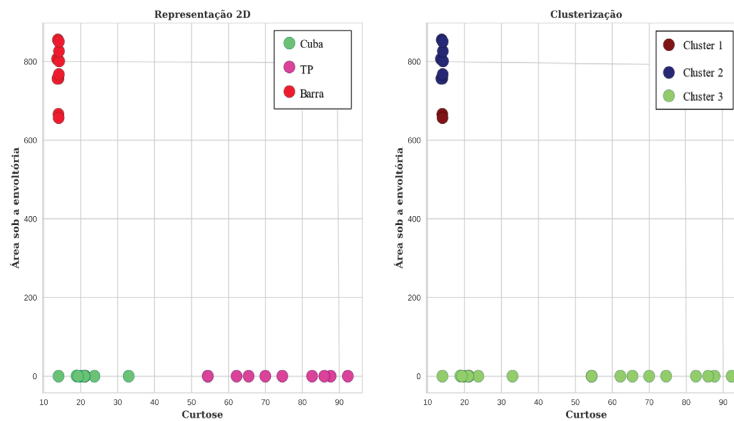
Notou-se, a partir da tabela 2, que os atributos que corresponderam a tais características foram: área sob a envoltória e curtose, reafirmando a análise inicial, via variância.

Para validar a escolha dos atributos selecionados, foi feita a aplicação do método *k-means*. Ressalta-se aqui, que há 3 atributos possíveis e, é necessário selecionar dois que caracterizem bem o banco de dados, logo, trata-se de um problema em que há 3 termos combinados 2 a 2, ou seja, há 3 possibilidades. Nas figuras 5 – 7, há a representação 2D dessas 3 possibilidades, entretanto apenas um dos casos separou os defeitos de forma correta.

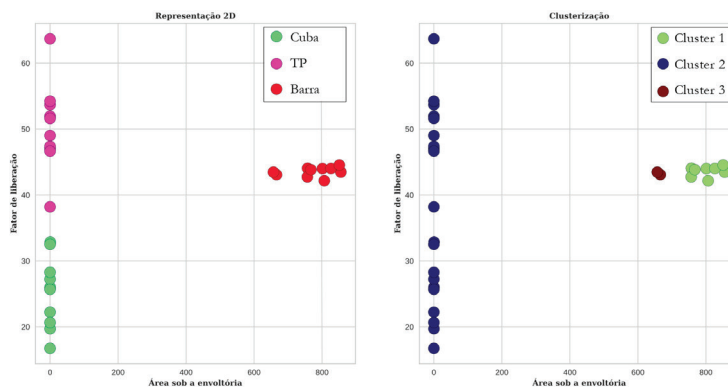


**Figura 5.** Fator de liberação e curtose.

Observou-se, na figura 5, que o método foi capaz de agrupar corretamente todas as amostras, sendo dependente apenas do hiperparâmetro *k*.



**Figura 6.** Curtose e área.



**Figura 7.** Área e fator de liberação.

Já nas figuras 6 e 7, notou-se que o método associou um mesmo *cluster* para diferentes tipos de defeito. Assim, utilizaram-se os atributos curtose e fator de liberação para realizar a classificação.

#### Aplicação dos Métodos *Decision Tree* e *Random Forest*

Para a aplicação do *decision tree*, o banco de dados foi subdividido conforme a seguir: 30% das amostras foram utilizadas para teste e 70% para o treinamento da árvore. Para tanto, utilizaram-se o banco de dados otimizado, ou seja, com os atributos fator de liberação e curtose; o banco de dados íntegro e o banco de dados contendo dois atributos com as menores variâncias. A eficácia do método foi avaliada utilizando a acurácia média de 5 simulações e a matriz de confusão, conforme a tabela 3.

**Tabela 3.** Resultados de classificação utilizando o *decision tree*.

		$K_u \times F_l$				Dataset íntegro					$T_{mp} \times T_s$			
		Estimado				Estimado					Estimado			
Real		Cuba	TP	Barra	Real		Cuba	TP	Barra	Real		Cuba	TP	Barra
	Cuba	4	0	0		Cuba	2	0	0		Cuba	1	0	0
	TP	1	2	0		TP <td>0</td> <td>3</td> <td>0</td> <td>TP <td>1</td> <td>2</td> <td>4</td> </td>	0	3	0		TP <td>1</td> <td>2</td> <td>4</td>	1	2	4
	Barra	0	0	2		Barra <td>0</td> <td>0</td> <td>4</td> <td>Barra <td>0</td> <td>1</td> <td>1</td> </td>	0	0	4		Barra <td>0</td> <td>1</td> <td>1</td>	0	1	1
Acurácia média = 0,95					Acurácia média = 1,00					Acurácia média = 0,48				

De forma análoga ao *decision tree*, o *random forest* utilizou a mesma subdivisão do banco de dados e 100 árvores. A aplicação de tal método gerou os resultados apresentados na tabela 4.

**Tabela 4.** Resultados de classificação utilizando a técnica *random forest*.

		$K_u \times F_l$				Dataset íntegro					$T_{mp} \times T_s$			
		Estimado				Estimado					Estimado			
Real		Cuba	TP	Barra	Real		Cuba	TP	Barra	Real		Cuba	TP	Barra
	Cuba	3	0	0		Cuba <td>3</td> <td>0</td> <td>0</td> <td>Cuba <td>3</td> <td>0</td> <td>0</td> </td>	3	0	0		Cuba <td>3</td> <td>0</td> <td>0</td>	3	0	0
	TP	0	3	0		TP <td>0</td> <td>2</td> <td>0</td> <td>TP <td>0</td> <td>0</td> <td>4</td> </td>	0	2	0		TP <td>0</td> <td>0</td> <td>4</td>	0	0	4
	Barra	0	0	3		Barra <td>0</td> <td>0</td> <td>4</td> <td>Barra <td>0</td> <td>0</td> <td>2</td> </td>	0	0	4		Barra <td>0</td> <td>0</td> <td>2</td>	0	0	2
Acurácia média = 0,98					Acurácia média = 1,00					Acurácia média = 0,50				

Nota-se que ambos os métodos *decision tree* e *random forest* apresentaram valores de acurácia próximos, entretanto, o *random forest* é menos suscetível ao *overfitting* e exige maior esforço computacional [15]. Destaca-se aqui, que a redução do banco de dados foi suficiente para classificar os sinais de DP, diminuindo o esforço computacional requerido para o processamento dos dados. Nota-se ainda, que os atributos com baixa variância não resultaram em uma boa taxa de acerto, o que ressalta a importância da etapa de seleção dos atributos.

### Aplicação da RNA

Na aplicação da RNA, o banco de dados foi subdividido conforme a seguir: 30% das amostras para teste e 70% para o treinamento da rede. Além disso, utilizaram-se 1000 épocas e 3 camadas ocultas ( $C_{oc}$ ), as quais foram calculadas conforme a equação (17):

$$C_{oc} = \frac{\text{Atributos} + \text{Classes}}{2} \quad (17)$$

Para haver maior segurança ao resultado das simulações, bem como, diminuir a possibilidade de *overfitting*, foi associado um fator de aleatoriedade à seleção dos dados, assim, a cada simulação o processo foi feito com amostras diferentes nas etapas de teste e treinamento. As acurácias médias, bem como as matrizes de confusão são apresentadas na tabela 5.

**Tabela 5.** Resultados de classificação utilizando a RNA.

$K_u \times F_l$					Dataset íntegro					$T_{mp} \times T_s$				
Estimado					Estimado					Estimado				
Real	Cuba	TP	Barra		Real	Cuba	TP	Barra		Real	Cuba	TP	Barra	
	Cuba	3	0	0		Cuba	2	0	0		Cuba	0	0	3
	TP	1	4	0		TP	0	3	0		TP	4	0	0
	Barra	0	0	2		Barra	0	0	4		Barra	0	0	2
Acurácia média = 1,00					Acurácia média = 1,00					Acurácia média = 0,22				

Nota-se que a RNA apresentou altos valores de acurácia para os dois primeiros casos, entretanto não foi capaz de classificar corretamente os defeitos quando foram inseridos atributos com baixa variância na entrada, o que também ocorreu com os métodos apresentados na subseção 4.2. Ressalta-se, entretanto, que o pequeno número de amostras limitou a quantidade de técnicas aplicadas para excluir a possibilidade de *overfitting*.

### Conclusões

Foram apresentados os resultados da classificação de sinais de DP via uma série de métodos de *machine learning*. Destaca-se aqui, que a extração automática dos atributos acrescentou mais segurança à metodologia, evitando erros na tomada de decisão. No que concerne ao método *k-means*, evidenciou-se a importância do processo de clusterização na etapa de seleção dos atributos, haja vista que o banco de dados otimizado foi suficiente para realizar a classificação correta das amostras e demandou menor esforço computacional. Com relação aos métodos de classificação, o *random forest* se mostrou mais robusto ao *overfitting*, quando comparado ao *decision tree*. Já a RNA apresentou resultados promissores para a sua aplicação em bancos de dados com um número maior de amostras.



## Agradecimentos

Os autores agradecem ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo financiamento nesta pesquisa.

## Referências

- [1] IEC 60270. (2000). High Voltage Test Techniques: Partial Discharge Measurement
- [2] Nobrega, L., Costa, E., Serres, A., Xavier, G., & Aquino, M. (2019). UHF Partial Discharge Location in Power Transformers via Solution of the Maxwell Equations in a Computational Environment. *Sensors*, 19(15), 3435. <https://doi.org/10.3390/s19153435>
- [3] Silva, A. D. C., Nobrega, L. A. M. M., da Costa, E. G., Xavier, G. V. R., da Cruz, A. R. D., & Gomes, M. P. A. (2020). Evaluation of Partial Discharge Sources Location Algorithms on Signals with Different Signal to Noise Ratio. In *2020 IEEE International Conference on High Voltage Engineering and Application (ICHVE)*. IEEE. <https://doi.org/10.1109/ichve49031.2020.9279940>
- [4] Working Group A2.27 (2008). Recommendations for condition monitoring and condition assessment facilities for transformers. *Electra*, n. 237, p. 48–57.
- [5] Nattrass, D. A. (1988). Partial discharge measurement and interpretation. *IEEE Electrical Insulation Magazine*, 4(3), 10–23. <https://doi.org/10.1109/57.830>
- [6] Kreuger, F. H., Gulski, E., & Krivda, A. (1993). Classification of partial discharges. *IEEE Transactions on Electrical Insulation*, 28(6), 917–931. <https://doi.org/10.1109/14.249365>
- [7] Macedo, E. C. T., Villanueva, J. M., da Costa, E. G., Freire, R. C. S., Araujo, D. B., de Souza Neto, J. M. R., & Glover, I. A. (2012). Assessment of dielectric degradation by measurement, processing and classification of Partial Discharges. In *2012 IEEE International Power Modulator and High Voltage Conference (IPMHVC)*. IEEE. <https://doi.org/10.1109/ipmhvc.2012.6518812>
- [8] de Souza Neto, J. M. R., Rocha Neto, J. S. d., Macedo, E. C. T., Glover, I. A., & Judd, M. D. (2014). An envelope detector as a trading cost technique for radiometric partial discharge detection. In *2014 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*. IEEE. <https://doi.org/10.1109/i2mtc.2014.6861013>
- [9] Pinpart, T., & Judd, M. D. (2010). Differentiating between partial discharge sources using envelope comparison of ultra-high-frequency signals. *IET Science, Measurement & Technology*, 4(5), 256–267. <https://doi.org/10.1049/iet-smt.2009.0064>
- [10] Han, L., Yan, J., Fan, S., Xu, M., Liu, Z., Geng, Y., & Guan, C. (2019). Feature Extraction of UHF PD Signals Based on Diode Envelope Detection and Linear Discriminant Analysis. In *2019 5th International Conference on Electric Power Equipment - Switching Technology (ICEPE-ST)*. IEEE. <https://doi.org/10.1109/icepe-st.2019.8928690>
- [11] Liu, W., Liu, S., & Hu, X. (2009). Feature extraction and pattern recognition of signals radiated from partial discharge. In *2009 5th Asia-Pacific Conference on Environmental Electromagnetics (CEEM 2009)*. IEEE. <https://doi.org/10.1109/ceem.2009.5304189>
- [12] Wu, W.-J., & Xu, Y. (2010). Correlation analysis of visual verbs' subcategorization based on Pearson's correlation coefficient. In *2010 International Conference on Machine Learning and Cybernetics (ICMLC)*. IEEE. <https://doi.org/10.1109/icmlc.2010.5580507>
- [13] Xiaosheng Peng, Chengke Zhou, Hepburn, D. M., Judd, M. D., & Siew, W. H. (2013). Application of K-Means method to pattern recognition in on-line cable partial discharge monitoring. *IEEE Transactions on Dielectrics and Electrical Insulation*, 20(3), 754–761. <https://doi.org/10.1109/tdei.2013.6518945>
- [14] Yazdandoust, A. R., Haghjoo, F., & Shahrtash, S. M. (2008). Insulation status assessment in high voltage cables based on decision tree algorithm. In *Energy Conference (EPEC)*. IEEE. <https://doi.org/10.1109/epc.2008.4763312>
- [15] Peng, X., Yang, G., Zheng, S., Xiong, L., & Bai, J. (2016). Optimal feature selection for partial discharge recognition of cable systems based on the random forest method. In *2016 China International Conference on Electricity Distribution (CICED)*. IEEE. <https://doi.org/10.1109/ciced.2016.7576360>
- [16] Xavier, G. V. R., da Costa, E. G., Serres, A. J. R., Nobrega, L. A. M. M., Oliveira, A. C., & Sousa, H. F. S. (2019). Design and Application of a Circular Printed Monopole Antenna in Partial Discharge Detection. *IEEE Sensors Journal*, 19(10), 3718–3725. <https://doi.org/10.1109/jsen.2019.2896580>