



UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

Técnica semiautomática de rotulagem e sistema para geração artificial de PRPDs aplicados ao treinamento de classificador de descargas parciais em hidrogeradores

Frederico Henrique do Rosario Lopes

UFPA / ITEC / PPGEE
CAMPUS UNIVERSITÁRIO DO GUAMÁ
BELÉM-PARÁ-BRASIL

2024



UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

Frederico Henrique do Rosario Lopes

Técnica semiautomática de rotulagem e sistema para geração artificial de PRPDs aplicados ao treinamento de classificador de descargas parciais em hidrogeradores

Dissertação de mestrado submetida à Banca Examinadora para a obtenção do Grau de Mestre em Engenharia Elétrica.

Orientador: Prof. Dr. Rodrigo Melo e Silva de Oliveira
Coorientador: Prof. Dr. Ronaldo de Freitas Zampolo

UFPA / ITEC / PPGEE
CAMPUS UNIVERSITÁRIO DO GUAMÁ
BELÉM-PARÁ-BRASIL
2024

**Dados Internacionais de Catalogação na Publicação (CIP) de acordo com ISBD
Sistema de Bibliotecas da Universidade Federal do Pará
Gerada automaticamente pelo módulo Ficat, mediante os dados fornecidos pelo(a) autor(a)**

L864t Lopes, Fredericco.
Técnica semiautomática de rotulagem e sistema para geração artificial de PRPDs aplicados ao treinamento de classificador de descargas parciais em hidrogeradores / Fredericco Lopes. — 2023.
78 f. : il. color.

Orientador(a): Prof. Dr. Rodrigo de Oliveira
Coorientador(a): Prof. Dr. Ronaldo Zampolo
Dissertação (Mestrado) - Universidade Federal do Pará,
Instituto de Tecnologia, Programa de Pós-Graduação em
Engenharia Elétrica, Belém, 2023.

1. PRPD. 2. PCA. 3. PCA. 4. kmeans. 5. label propagation.
I. Título.

CDD 006.3

Frederico Henrique do Rosario Lopes

Técnica semiautomática de rotulagem e sistema para geração artificial de PRPDs aplicados ao treinamento de classificador de descargas parciais em hidrogeradores

Dissertação de mestrado submetida à Banca Examinadora para a obtenção do Grau de Mestre em Engenharia Elétrica.

Data da Defesa: 29 de fevereiro de 2022

Conceito: _____

Banca Examinadora

Prof. Dr. Rodrigo Melo e Silva de Oliveira
Orientador – PPGEE/ITEC/UFPA

Prof. Dr. Ronaldo de Freitas Zampolo
Coorientador – FCT/ITEC/UFPA

Prof. Dr. Victor Dmitriev
Avaliador interno – PPGEE/ITEC/UFPA

Prof. Dr. Licinius Dimitri Sá de Alcantara
Avaliador externo – ICIBE/UFRA

Dr. Fernando de Souza Brasil
Avaliador externo – Eletrobras/Eletronorte

Prof. Dr. Diego Lisboa Cardoso
Coordenador do PPGEE/ITEC/UFPA

*Esta monografia é dedicada aos meus pais
que são as minhas principais inspirações na vida.*

AGRADECIMENTOS

A Deus, pela minha vida e por me ajudar a ultrapassar todos os obstáculos na minha vida.

Aos meus pais que me apoiaram continuamente ao longo do desenvolvimento desse trabalho e do curso todo.

Aos professores que tive ao longo do curso pelos ensinamentos que contribuíram no meu desenvolvimento.

Aos meu amigos e colegas, que compartilharam vários momentos bons durante todo período do curso.

À Eletronorte por disponibilizar a infraestrutura e os dados experimentais que tornaram este trabalho possível. Aos professores Ronaldo Zampolo e Rodrigo Oliveira, que me guiaram e orientaram durante todo o desenvolvimento deste trabalho.

Ao professor Victor Dmitriev por sua liderança, ideias e sugestões ao longo do projeto de P&D com a CPFL, o qual só foi possível graças a ele.

À CPFL e ao CNPq por me agraciarem com uma bolsa de estudos: CPFL até o mês de setembro de 2022 e CNPq a partir de outubro de 2022.

A educação é a arma mais poderosa que você pode usar para mudar o mundo. - Nelson Mandela

RESUMO

Hidrogeradores são ativos cruciais tanto para empresas geradoras de energia quanto para a população que depende desse abastecimento. No entanto, essas máquinas estão suscetíveis a vários tipos de defeitos que podem resultar em interrupções inesperadas, se nada for feito a respeito. A análise de descargas parciais é uma abordagem já consolidada para avaliar a condição de equipamentos de alta tensão, sendo essencial a detecção automática de diferentes tipos de defeitos, uma vez que diferentes níveis de risco à operação variam de acordo com o tipo de descarga. Redes neurais profundas têm sido propostas visando à classificação de descargas parciais usando diagramas PRPD (*phase-resolved partial discharge*). Contudo, a obtenção de conjuntos de dados rotulados com grande número de exemplos é um problema que impacta diretamente no desempenho de classificadores treinados de maneira supervisionada. Neste contexto, nesta dissertação propõe-se uma técnica semiautomática para a rotulagem de PRPDs, baseada em estratégias de redução de dimensionalidade e agrupamento de dados, bem como investiga-se o uso de GAN (*generative adversarial network*) na ampliação artificial do conjunto de treinamento. O conjunto de dados usado no trabalho é composto por PRPDs reais obtidos por meio de procedimentos de monitoramento online de descargas parciais em hidrogeradores. O desempenho dos sistemas propostos é contrastado com resultados recentes representativos do estado da arte na área. Os resultados demonstram que a aplicação da técnica proposta para rotulagem semiautomática tem potencial para reduzir consideravelmente a carga de trabalho e o tempo associados à classificação manual. Além disso, o uso de PRPDs artificiais gerados pela GAN resultou notável melhoria no desempenho do classificador que alcançou 94,72% de acurácia média, em comparação com 89,44% obtido com a melhor técnica concorrente. Foram observados ganhos semelhantes também nas acurácias por classe.

Palavras-chave: PRPD, PCA, PaCMAP, kmeans, label propagation, GAN.

ABSTRACT

Hydrogenerators are crucial assets for both power generating companies and the population that relies on this supply. However, these machines are susceptible to various types of defects that can result in unexpected interruptions if nothing is done about it. Partial discharge analysis is an already established approach for assessing the condition of high-voltage equipment, with the automatic detection of different types of defects being essential, since different levels of risk to operation vary according to the type of discharge. Deep neural networks have been proposed aiming at the classification of partial discharges using phase-resolved partial discharge (PRPD) diagrams. However, obtaining labeled datasets with a large number of examples is a problem that directly impacts the performance of supervised trained classifiers. In this context, this dissertation proposes a semi-automatic technique for PRPD labeling, based on dimensionality reduction strategies and data clustering, as well as investigates the use of Generative Adversarial Networks (GANs) in artificially expanding the training set. The dataset used in the work consists of real PRPDs obtained through online monitoring procedures of partial discharges in hydrogenerators. The performance of the proposed systems is contrasted with recent representative results from the state-of-the-art in the field. The results demonstrate that the application of the proposed semi-automatic labeling technique has the potential to considerably reduce the workload and time associated with manual classification. Furthermore, the use of artificial PRPDs generated by GANs resulted in a notable improvement in the classifier's performance, reaching an average accuracy of 94.72%, compared to 89.44% obtained with the best competing technique. Similar gains were also observed in class accuracies.

Keywords: PRPD, PCA, PaCMAP, kmeans, label propagation, GAN.

LISTA DE ILUSTRAÇÕES

Figura 1 – Exemplo de gráfico PRPD com eixos de amplitude, fase e barra lateral, indicando cores associadas à contagem de descargas. Fonte: o autor	10
Figura 2 – Exemplo do padrão de PRPD para descarga de cavidade. Fonte: adaptado de (IEC-60270, 2000)	12
Figura 3 – Seção transversal de uma barra estatórica, com delaminação visível. Fonte: adaptado de (IEC-60270, 2000)	12
Figura 4 – Exemplo do padrão de PRPD para delaminação interna. Fonte: adaptado de (IEC-60270, 2000)	13
Figura 5 – Exemplo de arborização gerada por agulha imersa em óleo. Fonte: (CHEN et al., 2015)	13
Figura 6 – Corte transversal em um <i>ranhura</i> : a) bobina de múltiplas espiras; e b) barra Roebel. Fonte: (CRUZ et al., 2022)	14
Figura 7 – Exemplo do padrão de PRPD para ranhura. Fonte: adaptado de (IEC-60270, 2000)	14
Figura 8 – Exemplo de padrão PRPD para descarga do tipo corona. Fonte: (IEC-60270, 2000)	15
Figura 9 – Exemplo de padrão PRPD para descarga do tipo <i>gap</i> . Fonte: (IEC-60270, 2000)	16
Figura 10 – Exemplo de padrão PRPD para <i>surface tracking</i> . Fonte: (IEC-60270, 2000) .	17
Figura 11 – Comparação entre: (A) perceptron de Roseblatt e (B) perceptron moderno. Fonte: o autor	26
Figura 12 – Exemplo de ANN do tipo <i>feedforward</i> . Fonte: o autor	27
Figura 13 – Exemplo de <i>underfitting</i> . Fonte: o autor	28
Figura 14 – Exemplo de boa generalização. Fonte: o autor	29
Figura 15 – Exemplo de <i>overfitting</i> . Fonte: o autor	29
Figura 16 – Esquema simplificado do gradiente descendente. Fonte: o autor	31
Figura 17 – Lenet5. Fonte: (LECUN et al., 1998)	33
Figura 18 – Exemplo de <i>maxpooling</i> com filtro 2×2 em uma matriz 4×4 . Fonte: o autor	34
Figura 19 – Exemplo de camada completamente conectada. Fonte (IN, 2022)	34
Figura 20 – Exemplo do modelo de GAN. Fonte: (GOODFELLOW; BENGIO; COURVILLE, 2016)	35

Figura 21 – Arquitetura do modelo DCGAN. Fonte: (RADFORD; METZ; CHINTALA, 2015)	36
Figura 22 – Número de diagramas PRPD de cada máquina do conjunto de dados	37
Figura 23 – Estratégia de propagação de rótulos para conjunto de dados de PRPD.	38
Figura 24 – Diagrama em blocos, detalhando a estratégia de agrupamento de dados PPRD.	38
Figura 25 – Efeitos do pré-processamento nos PRPDs.	40
Figura 26 – Diagrama detalhado da etapa do agrupamento de PRPDs, incluindo as operações de escalonamento de características (<i>features</i>).	41
Figura 27 – Exemplo de PRPD obtido do conjunto de dados. Fonte: o autor	43
Figura 28 – Arquitetura da rede base. Fonte: (MANTACH et al., 2021)	46
Figura 29 – Gráfico de acurácia para vários valores de k . Fonte: o autor	48
Figura 30 – <i>Silhouette score</i> para cada configuração. $\bar{\alpha}_w$: média ponderada (barras em cinza); σ_α : desvio padrão (curva em preto).	50
Figura 31 – Métrica de Calinski-Harabasz para cada configuração. $\bar{\beta}_w$: média ponderada (barras em cinza); σ_β : desvio padrão (curva em preto).	50
Figura 32 – Métrica de Davies-Bouldin para cada configuração. $\bar{\gamma}_w$: média ponderada (barras em cinza); σ_γ : desvio padrão (curva em preto).	51
Figura 33 – Exemplo de PRPDs artificiais da classe gap gerados com GAN. Fonte: o autor	53
Figura 34 – Exemplo de PRPDs artificiais da classe superficial gerados com GAN. Fonte: o autor	54
Figura 35 – Exemplo de PRPDs artificiais da classe cavidade gerados com GAN. Fonte: o autor	54
Figura 36 – Exemplo de PRPDs artificiais da classe ranhura gerados com GAN. Fonte: o autor	55

LISTA DE TABELAS

Tabela 1	– Número de exemplos para cada classe no conjunto de dados original	44
Tabela 2	– Parâmetros de treinamento parcial da rede base.	45
Tabela 3	– Parâmetros de treinamento parcial da rede base	47
Tabela 4	– As 10 melhores configurações para o <i>silhouette score</i> (α). $\bar{\alpha}_w$: média ponderada de α ; σ_α : desvio padrão de α ; d_{\max} : número máximo de dimensões no espaço latente; e n_{\max} : número máximo de <i>clusters</i> resultantes	50
Tabela 5	– As 10 melhores configurações para a métrica de Calinski-Harabasz (β). $\bar{\beta}_w$: média ponderada de β ; σ_β : desvio padrão de β ; d_{\max} : número máximo de dimensões no espaço latente; e n_{\max} : número máximo de <i>clusters</i> resultantes	51
Tabela 7	– Melhor configuração de acordo com cada métrica. SIL: <i>silhouette score</i> ; CHA: métrica de Calinski-Harabasz; DBO: métrica de Davies-Bouldin; EA: escalonamento de amplitude; RD: redução de dimensionalidade; EC-RD: escalonamento de características antes da redução de dimensionalidade; EC-CL: escalonamento de características antes do clustering	52
Tabela 6	– As 10 melhores configurações para a métrica de Davies-Bouldin (γ). $\bar{\gamma}_w$: média ponderada de γ ; σ_γ : desvio padrão de γ ; d_{\max} : número máximo de dimensões no espaço latente; e n_{\max} : número máximo de <i>clusters</i> resultantes	52
Tabela 8	– Configurações mais comuns. EA: Escalonamento de Amplitude; RD: Redução de Dimensionalidade; EC-RD: Escalonamento de Características antes da Redução de Dimensionalidade; EC-CL: Escalonamento de Características antes do Agrupamento	53
Tabela 9	– Acurácia dos testes com dados não binarizados: média de acurácia de 85,16% ($\pm 3,21\%$)	55
Tabela 10	– Matriz de Confusão do Teste com a Rede Base (Dados não binarizados)	55
Tabela 11	– Acurácia por classe feito com a rede após treinamento com dados não binarizados	56
Tabela 12	– Acurácia dos testes feitos com <i>data augmentation</i> Não-treinamento da rede base: média de acurácia de 91,22% ($\pm 1,66\%$)	56
Tabela 13	– Matriz de Confusão do Teste com a Rede Base (<i>Data augmentation com GAN</i>)	56
Tabela 14	– Acurácia por classe da rede base	57
Tabela 15	– Relação geral das acurácias de todos os testes	57

LISTA DE ABREVIATURAS E SIGLAS

ADALINE	<i>Adaptative linear neuron</i>
ANN	<i>Artificial neural network</i>
CNN	<i>Convolutional neural network</i>
EA	<i>Escalonamento de amplitude</i>
EC	<i>Escalonamento de característica</i>
EC-CL	<i>Escalonamento de característica antes do clustering</i>
EC-RD	<i>Escalonamento de característica antes da redução de dimensionalidade</i>
GAN	<i>Generative adversarial network</i>
HFCT	<i>high frequency current transducer</i>
KPCA	<i>Kernel principal component analysis</i>
MADALINE	<i>Multiple adaptative linear neuron</i>
MAE	<i>Mean absolute error</i>
MSE	<i>Mean squared error</i>
PaCMAP	<i>Pairwise Controlled Manifold Approximation</i>
PCA	<i>Principal component analysis</i>
PD	<i>Partial discharge</i>
PRPD	<i>Phase-resolved partial discharge</i>
RD	<i>Redução de dimensionalidade</i>
ReLU	<i>Rectified linear unit</i>
UHF	<i>Ultra High Frequency</i>

SUMÁRIO

1	INTRODUÇÃO	1
1.1	Contexto	1
1.2	Justificativa	2
1.3	Objetivos	3
1.3.1	Objetivo geral	3
1.3.2	Objetivos específicos	3
1.4	Metodologia	3
1.5	Publicações originárias da presente dissertação	4
1.6	Estrutura do trabalho	5
2	DESCARGAS PARCIAIS	6
2.1	Introdução	6
2.1.1	Definição	6
2.1.2	Condições de ocorrência	6
2.1.3	Importância da detecção de descargas parciais	7
2.2	Técnicas de monitoramento	8
2.2.1	Detecção acústica	8
2.2.2	Detecção com transformadores de corrente de alta frequência (HFCT)	8
2.2.3	Detecção em frequência ultra-alta (Ultra High Frequency-UHF)	9
2.2.4	Diagramas do tipo <i>phase-resolved partial discharge</i> (PRPD)	9
2.2.5	Detecção de ozônio	10
2.2.6	Detecção elétrica	10
2.3	Tipos de descargas	11
2.3.1	Descargas internas	11
2.3.1.1	Descargas de cavidades	11
2.3.1.2	Delaminação interna	12
2.3.2	Arborização elétrica	13
2.3.3	Descargas de ranhura	13
2.3.4	Descargas na cabeça de bobina	15
2.3.4.1	Corona	15
2.3.4.2	<i>Gap</i>	16
2.3.4.3	Descargas superficiais	16

3	APRENDIZADO DE MÁQUINA	18
3.1	Redução de dimensionalidade	18
3.1.1	PCA	18
3.1.2	Kernel PCA	20
3.1.3	PaCMAP	20
3.2	Algoritmo de agrupamento de dados <i>k-means</i>	22
3.3	Redes neurais artificiais	23
3.3.1	Histórico	23
3.3.2	Perceptron	24
3.3.3	Redes neurais do tipo <i>feedforward</i>	27
3.3.4	Função de custo	27
3.3.5	Algoritmo de <i>backpropagation</i>	29
3.3.6	Gradiente descendente	30
3.3.7	Funções de ativação	31
3.3.8	Regularização em redes neurais	32
3.4	Redes neurais convolucionais	32
3.4.1	Camada convolucional	33
3.4.2	Camada de <i>pooling</i>	33
3.4.3	Camada completamente conectada	34
3.5	<i>Generative artificial network</i> (GAN)	35
4	SISTEMA DE ROTULAGEM E USO DE GAN PARA TREINAMENTO DE CLASSIFICADORES DE PD	37
4.1	Técnica semiautomática de rotulagem de PRPDs	37
4.1.1	Pré-processamento	39
4.1.2	Redução de dimensionalidade	39
4.1.3	Clustering	41
4.1.4	Escalonamento padrão de características	41
4.1.5	Avaliação de desempenho do sistema de agrupamento de PRPDs	41
4.2	Treinamento de classificadores usando PRPDs gerados por GAN	43
4.2.1	Geração de PRPDs artificiais	43
4.2.2	Arquitetura do classificador	45
4.2.3	Metodologia de treinamento do classificador	47
5	RESULTADOS	49

5.1	Avaliação do sistema de agrupamento automático de PRPDs	49
5.2	Resultados da geração artificial com rede GAN	53
5.3	Resultados da classificação de PRPDs	53
5.3.1	Classificador treinado com PRPDs não binarizados	53
5.3.2	Classificador treinado mediante <i>data augmentation</i> usando GAN	56
5.3.3	Comparação de resultados	57
6	CONCLUSÕES	58
	REFERÊNCIAS	60

1 INTRODUÇÃO

1.1 Contexto

Hidrogeradores assumem um papel central e de grande valor no sistema de geração de energia elétrica do Brasil, uma vez que aproximadamente 57% da eletricidade consumida no país tem sua origem em usinas hidrelétricas (Empresa de Pesquisa Energética, 2022).

A relevância dos hidrogeradores para a estabilidade e sustentabilidade do sistema energético brasileiro sublinha a necessidade premente de uma monitorização regular desses equipamentos. Essa prática não apenas garante a eficiência operacional dos geradores, mas também se revela crucial na prevenção de falhas que poderiam comprometer a continuidade do fornecimento de energia elétrica.

Nesse contexto, a avaliação do estado dessas máquinas torna-se uma prioridade, e a classificação de descargas parciais surge como uma abordagem valiosa para alcançar esse objetivo. Esta classificação oferece uma maneira eficaz e precisa de diagnosticar potenciais irregularidades nos hidrogeradores, permitindo intervenções proativas para evitar falhas mais severas e impactos significativos no sistema.

Ao integrar a monitorização regular e a classificação de descargas parciais no ciclo de vida operacional dos hidrogeradores, não apenas fortalecemos a confiabilidade do sistema elétrico, mas também contribuímos para a sustentabilidade da matriz energética brasileira. A implementação dessas práticas promove a segurança operacional, e também otimiza os custos associados à manutenção, maximizando, assim, a eficiência global dos hidrogeradores.

O presente estudo visa explorar e desenvolver uma abordagem para a detecção e classificação de descargas parciais em hidrogeradores, mediante a aplicação de uma técnica semiautomática de rotulagem, modelos generativos para aumento do conjunto de dados e aprendizado profundo para classificação de descargas parciais obtidas de um ambiente de funcionamento real, enfatizando a necessidade de práticas avançadas de monitorização para assegurar uma contribuição contínua e confiável desses equipamentos para o fornecimento de energia elétrica renovável.

1.2 Justificativa

O monitoramento constante do estado dos hidrogeradores por meio da análise de descargas parciais desempenha uma função essencial na prevenção de paradas não programadas e na implementação de estratégias de manutenção mais eficientes (CRUZ et al., 2022). A capacidade de identificar com precisão diversos tipos de descargas parciais não apenas aprimora a eficácia na otimização das ações de manutenção, elevando a qualidade do serviço prestado, mas também resulta em uma redução nos custos operacionais. A identificação precisa e eficiente de fatores desgastantes predominantes é crucial para a otimização das práticas de manutenção em máquinas elétricas rotativas. No contexto específico, observa-se que o desgaste elétrico, associado principalmente à intensidade elevada de descargas parciais (PD, do inglês *partial discharge*) na bobina do estator, é um dos principais contribuintes para a degradação desses sistemas complexos (Iris Power, 2016). A atividade constante de descargas parciais é identificada como um fator significativo que intensifica a degradação (KHAN et al., 2020). A compreensão aprofundada desses processos é essencial, pois a degradação pode levar a falhas, resultando em perdas econômicas substanciais, seja por paralisações nos processos ou por danos físicos aos ativos (HASSAN et al., 2020). A habilidade de identificar com precisão vários tipos de descargas parciais não apenas aprimora a eficácia na otimização das práticas de manutenção. Atualmente essa identificação é ainda realizada preferencialmente de forma manual em máquinas rotativas, comparando medições com documentos técnicos (IEEE, 2014). Essa abordagem manual, embora baseada em padrões confiáveis, destaca a necessidade de avanços tecnológicos. A implementação de técnicas avançadas, como o uso de aprendizado profundo (HASSAN et al., 2020), emerge como uma abordagem promissora para automatizar esse processo de identificação, superando as limitações da análise manual. A transição para métodos automatizados não apenas resultaria em uma redução significativa nos custos operacionais, mas também aprimoraria a eficiência do serviço, proporcionando uma resposta mais rápida e precisa diante das complexidades das máquinas elétricas rotativas.

Esse avanço tecnológico representa um marco significativo na área, pois não só otimiza as práticas de manutenção, mas também promove a eficiência operacional, a sustentabilidade e a confiabilidade de máquinas elétricas rotativas em ambientes de alta demanda. A pesquisa atual busca explorar e aprofundar o entendimento sobre a identificação e monitoramento de descargas parciais, visando contribuir para o avanço contínuo nesse campo crucial para o funcionamento eficaz desses sistemas críticos.

1.3 Objetivos

1.3.1 Objetivo geral

O objetivo geral desta dissertação consiste na proposição de técnicas para rotulagem semi-automática de diagramas PRPD (do inglês, *phase-resolved partial discharge*) e para ampliação de conjunto de dados baseada em modelos generativos visando o treinamento de classificador de descargas parciais.

1.3.2 Objetivos específicos

- Investigar estratégia de agrupamento de PRPDs baseada em técnicas de redução de dimensionalidade e algoritmo *kmeans*;
- Implementar modelos generativos para produção de PRPDs artificiais;
- Avaliar a influência da aplicação de técnicas de *data augmentation* no desempenho de classificadores de descargas parciais a partir de diagramas PRPD.

1.4 Metodologia

A metodologia adotada neste estudo visa o desenvolvimento de classificadores multi-classe destinados à identificação de descargas parciais em hidrogeradores. Para alcançar esse objetivo, são explorados diversos modelos de redes neurais de aprendizado profundo, considerando a complexidade e a diversidade dos dados disponíveis.

A abordagem adotada baseia-se em conjuntos de dados pré-existentes, os quais foram originalmente coletados em campo durante a operação real de hidrogeradores em usinas hidrelétricas por meio de equipamentos especializados em monitoramento de descargas parciais. Ressalta-se, assim, que o registro de dados foi feito em condições não controladas e representativas das operações reais nas usinas hidrelétricas. Esse enfoque confere ao estudo uma base sólida e aplicável, considerando a natureza dinâmica e desafiadora dos ambientes em que os hidrogeradores operam.

Na avaliação dos resultados de classificação, são empregadas análises comparativas entre os modelos propostos e os representativos do estado da arte na área, observando matrizes

de confusão, cálculo de acurácias, e outros indicadores típicos que permitem uma apreciação abrangente do desempenho.

O mesmo conjunto de dados é usado na investigação e implementação de uma estratégia de rotulagem semiautomática de diagramas PRPD, com ênfase na otimização da etapa de agrupamento de dados. A avaliação de desempenho dos algoritmos testados é feita mediante o uso das métricas *silhouette score*, *Calinski-Harabasz* e *Davies-Bouldin*.

A linguagem de programação Python foi escolhida para a implementação dos sistemas propostos por oferecer flexibilidade e eficiência, permitindo a utilização de bibliotecas poderosas de aprendizado de máquina, processamento de dados e visualização. A plataforma Google Colab também foi usada, pois facilita a reprodutibilidade de resultados e a colaboração no trabalho.

1.5 Publicações originárias da presente dissertação

A produção desenvolvida no âmbito deste mestrado é composta por três artigos, sendo o primeiro de tema mais diretamente relacionado a esta dissertação, ao passo que os outros dois abordam tópicos correlatos:

- F. H. R. Lopes, R. F. Zampolo, R. M. S. Oliveira and V. Dmitriev, “Evaluation of transfer learning approaches for partial discharge classification in hydrogenerators,” 2022 Workshop on Communication Networks and Power Systems (WCNPS), Fortaleza, Brasil, 2022, pp. 1-6, doi: 10.1109/WCNPS56355.2022.9969682.
- R. M. S. Oliveira, R. F. Zampolo, L. D. S. Alcantara, G. G. Giroto, F. H. R. Lopes, N. M. Lopes, F. S. Brasil, J. A. S. Nascimento, and V. Dmitriev. 2023. “Analysis of Ozone Production Reaction Rate and Partial Discharge Power in a Dielectric-Barrier Acrylic Chamber with 60 Hz High-Voltage Electrodes: CFD and Experimental Investigations” *Energies* 16, no. 19: 6947. <https://doi.org/10.3390/en16196947>
- R. F. Zampolo, F. H. R. Lopes, R. M. S. Oliveira, V. Dmitriev, F. S. Brasil, M. F. Fernandes. “Análise experimental sobre a viabilidade do uso de sensores eletroquímicos de ozônio em sistemas de monitoramento de descargas parciais em hidrogeradores”, XXVII Seminário Nacional de Produção e Transmissão de Energia Elétrica (SNPTEE), Brasília, Brasil, 2023.

1.6 Estrutura do trabalho

O restante desta dissertação está organizado da seguinte forma:

Capítulo 2 – Descargas parciais: introduz conceitos fundamentais relacionados às descargas parciais, explorando seus mecanismos e fornecendo uma base teórica essencial para compreender o contexto do estudo.

Capítulo 3 – Aprendizado de máquina: aborda conceitos fundamentais para a implementação de classificadores com base em redes neurais artificiais, incluindo redes neurais convolucionais. Além disso, discute técnicas de redução de dimensionalidade, agrupamento não supervisionado de dados e modelos para geração artificial de dados.

Capítulo 4 – Sistema de rotulagem e uso de GAN para treinamento de classificadores de PD: apresenta detalhadamente o sistema de rotulagem semiautomática e o modelo para geração artificial de PRPDs, bem como os procedimentos adotados na avaliação de desempenho dos mesmos.

Capítulo 5 – Resultados: fornece uma análise detalhada dos resultados obtidos. Serão apresentadas interpretações e discussões críticas sobre as descobertas, oferecendo uma visão aprofundada do desempenho dos sistemas propostos.

Capítulo 6 – Conclusões: resume os principais entendimentos derivados da pesquisa, destacando contribuições significativas, limitações identificadas e possíveis direções para futuras pesquisas na área.

Referências: esta seção lista as principais fontes bibliográficas consultadas durante o desenvolvimento da pesquisa, proporcionando a base teórica e metodológica que sustenta o trabalho.

A estrutura delineada busca proporcionar uma abordagem organizada e abrangente, permitindo uma compreensão completa do estudo conduzido. Esta estrutura visa ainda garantir que os principais aspectos do problema sejam considerados, desde a preparação dos dados até a análise detalhada dos resultados, oferecendo uma visão ampla do desenvolvimento e resultados obtidos. Ao seguir essa abordagem, espera-se fornecer uma base sólida para futuras pesquisas e aplicações práticas na detecção e classificação de descargas parciais em hidrogeradores.

2 DESCARGAS PARCIAIS

2.1 Introdução

2.1.1 Definição

A caracterização das descargas parciais pode ser abordada de diferentes perspectivas. Esse fenômeno é descrito como uma descarga elétrica de baixa intensidade que ocorre dentro de um dielétrico sujeito a um campo elétrico. Nessa descrição, destaca-se que o caminho formado pela descarga não atravessa completamente o espaço entre dois condutores (ZNIDARICH, 2008). Essas descargas ocorrem quando a intensidade do campo elétrico excede a rigidez dielétrica do meio, resultando em múltiplas descargas durante um ciclo de voltagem, que é o período completo de variação da tensão elétrica

Uma outra abordagem (IEC-60270, 2000) caracteriza as descargas parciais como tipos de descargas elétricas localizadas que causam curto-circuito parcial na isolação entre condutores, podendo ou não ocorrer adjacente a um condutor. Nessa definição, a ênfase recai sobre o impacto localizado dessas descargas na isolação, podendo afetar a integridade da mesma.

Em geradores hidroelétricos resfriados a ar, as elevadas tensões presentes nas bobinas do estator podem levar à degradação da isolação devido a descargas parciais. Essas descargas ocorrem quando há a quebra localizada da rigidez dielétrica do ar e/ou outros materiais isolantes, podendo ocorrer tanto nos vazios da isolação quanto nas superfícies das bobinas. O termo “descarga parcial” é utilizado porque o fenômeno é limitado apenas à parte da isolação, não ocorrendo sua quebra completa ou violação.

É importante destacar que, no contexto de geradores hidroelétricos resfriados a ar, o calor gerado pelas partículas de ar ionizadas contribui para a degradação térmica e erosão da isolação. Um efeito colateral desse processo é a conversão do oxigênio em ozônio e a formação de óxidos de nitrogênio nos vazios da isolação (NAILEN, 1999).

2.1.2 Condições de ocorrência

Descargas parciais representam eventos elétricos complexos que ocorrem em sistemas dielétricos sob condições específicas. Esta seção explora resumidamente as condições necessárias para desencadear PDs, abrangendo desde a presença de cargas elétricas até os mecanismos de

feedback que sustentam o efeito avalanche.

A primeira condição essencial para a ocorrência de PDs é a presença de cargas elétricas livres em uma região específica. Essas cargas desempenham um papel crucial ao criar um ambiente propício para a ionização, que é o ponto de partida para a formação das PDs. A segunda condição requer um campo elétrico suficientemente forte para iniciar o processo, desencadeando o efeito avalanche. Conforme definido por (DION; MARTOFF; HOSACK, 2010), o efeito avalanche é um processo de ionização de gases, no qual elétrons livres colidem com moléculas de gás, gerando mais elétrons livres e permitindo a condução no gás. Por fim, é necessário algum tipo de mecanismo de *feedback* para manter o efeito de avalanche, garantindo a continuidade do processo de ionização e a propagação das PDs. Além dessas condições fundamentais, a rigidez dielétrica do meio que separa dois pontos com potenciais elétricos diferentes desempenha um papel importante no início, crescimento e finalização das descargas elétricas. Variáveis como o meio dielétrico, tipo de tensão aplicada, condições ambientais, distribuição do campo elétrico, forma de onda e intensidade do campo elétrico influenciam diretamente o comportamento das PDs. Formação de bolhas de gás; e nos gases, da ionização de moléculas nos pontos de maior potencial elétrico.

2.1.3 Importância da detecção de descargas parciais

A detecção de PDs é crucial para avaliar o estado operativo dos dielétricos em equipamentos de alta tensão. A medição destas descargas visa verificar e, se possível, quantificar as atividades de microdescargas no interior dos dielétricos. Essa medição é essencial, pois a presença de PDs indica deterioração do material isolante. Dependendo da intensidade das PDs, a vida útil do material pode ser drasticamente reduzida, justificando a necessidade de intervenções preventivas. A medição de PDs busca ainda determinar a relação entre suas grandezas e a expectativa de duração útil do dielétrico. Em síntese, o estudo das descargas parciais é essencial para garantir a integridade dos sistemas isolantes em equipamentos de alta tensão. A compreensão desses fenômenos e a implementação de técnicas de medição adequadas são fundamentais para prevenir danos, prolongar a vida útil dos materiais dielétricos e assegurar a confiabilidade do sistema elétrico interligado.

2.2 Técnicas de monitoramento

A detecção e medição de PDs constituem um aspecto fundamental na análise da integridade de sistemas elétricos. Esses eventos elétricos localizados geram uma variedade de sinais, cada um proveniente de processos associados distintos. Entre esses processos, destacam-se a emissão de luz, ruídos acústicos, ondas eletromagnéticas, liberação de gases e reações químicas. Cada um desses fenômenos representa uma manifestação singular das PDs, proporcionando uma janela única para compreender e identificar esses eventos. Para efetuar a detecção desses sinais, é comum empregar sensores especializados, cada um projetado para capturar um tipo específico de sinal associado a PDs. Essa abordagem multifacetada não apenas permite a identificação precoce de possíveis falhas no sistema elétrico, mas também possibilita uma análise mais abrangente da natureza e origem das descargas parciais.

2.2.1 Detecção acústica

Tanto as partículas quanto as descargas parciais têm a capacidade de gerar ondas acústicas que podem ser detectadas e analisadas. O método de detecção acústica, conforme descrito por (YAACOB et al., 2014), envolve a conversão de vibrações causadas por sinais acústicos em sinais elétricos. Essa abordagem é frequentemente implementada com o uso de sensores piezoelétricos para capturar as vibrações resultantes das descargas parciais. Esse método de detecção acústica segue um procedimento não invasivo, onde a detecção é realizada por detectores internos que integram e predizem a posição da descarga parcial. A utilização de vários sensores maximiza a precisão da detecção. Uma das principais vantagens desse método é a obtenção de resultados em tempo real. Além disso, a posição da descarga parcial pode ser determinada utilizando vários sensores. No entanto, algumas desvantagens associadas a esse método incluem baixa sensibilidade, como discutido abordado em (THAKOR; SHAH, 2020).

2.2.2 Detecção com transformadores de corrente de alta frequência (HFCT)

Sensores do tipo *high frequency current transducer* (HFCT) geralmente são instalados em cada fase dentro da caixa de terminais para capturar a corrente de PDs de máquinas rotatórias (FRANCO; RICHARDS; SELTZER-GRANT, 2019). Alternativamente, alguns métodos envolvem o uso de transformadores de corrente no sistema de aterramento da máquina na conexão ao neutro (MITICHE et al., 2018).

Para realizar a detecção de PDs recomenda-se seguir diretrizes internacionais para resultados consistentes. A norma IEC 60034-27 (IEC-60034-27-2, 2012) é comumente usada como referência para medições de PDs, especificando o uso de sensores de capacitor para coleta de dados.

2.2.3 Detecção em frequência ultra-alta (Ultra High Frequency-UHF)

A detecção de descargas parciais através da análise de ondas eletromagnéticas na faixa de frequência ultra-alta (UHF) representa um método eficaz, com frequências superiores a 300 MHz (YAACOB et al., 2013). Este procedimento é fundamentado na avaliação das ondas eletromagnéticas emitidas no local das descargas, conforme detalhado por (SOOMRO; RAMDON, 2014).

A efetividade desse método reside na capacidade de analisar sinais eletromagnéticos em uma faixa de frequência específica, proporcionando uma detecção sensível e precisa de eventos de descargas parciais. O sensor UHF, desempenha um papel de grande relevância na identificação precoce dessas descargas, contribuindo para a manutenção preventiva e prolongamento da vida útil de equipamentos elétricos. Essa técnica, ao explorar as características distintivas do espectro UHF, oferece uma abordagem valiosa para monitorar a integridade de sistemas elétricos e garantir a confiabilidade operacional ao longo do tempo.

2.2.4 Diagramas do tipo *phase-resolved partial discharge* (PRPD)

Um PRPD é uma representação da atividade de descargas parciais relativas aos 360° do ciclo de corrente alternada. O PRPD é formado pela contagem de pares pico de amplitude - ângulo de fase, obtidos a partir de sinais de descargas parciais registrados ao longo de um certo tempo de medição. Na formação do diagrama PRPD, o pico de amplitude da descarga parcial corresponde a valores na ordenada, a fase de ocorrência é referenciada a valores na abcissa, e informação sobre a quantidade de pulsos é indicada por cores. Verifica-se que diferentes tipos de defeitos do sistema de isolamento estão associados a distintos padrões PRPD. Este tipo de diagrama é bem útil para a avaliação de uma máquina. A Fig. 1 mostra um exemplo de diagrama PRPD.

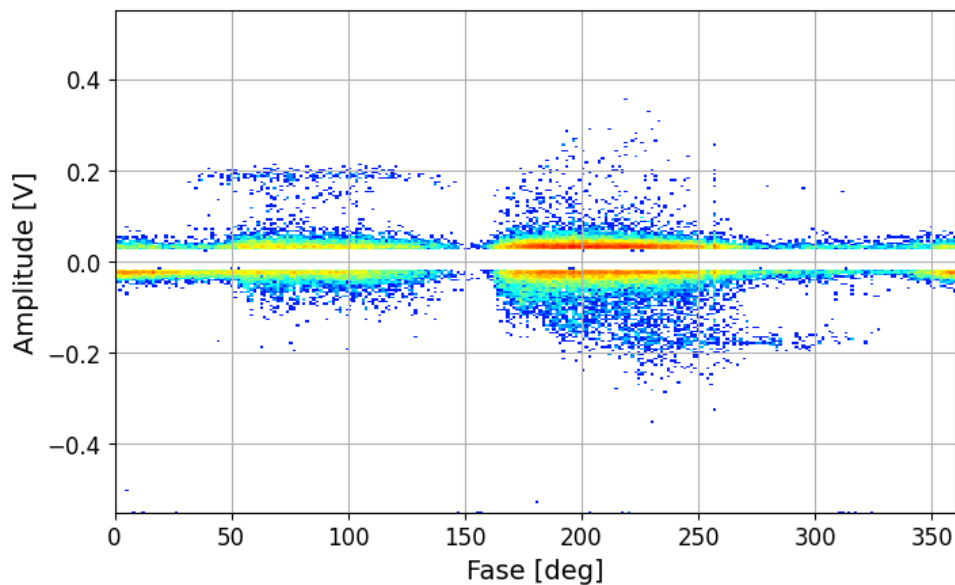


Figura 1 – Exemplo de gráfico PRPD com eixos de amplitude, fase e barra lateral, indicando cores associadas à contagem de descargas. Fonte: o autor

2.2.5 Detecção de ozônio

Estudos recentes têm destacado uma correlação significativa entre a incidência de descargas parciais e a presença de concentrações elevadas de ozônio (JAVED; LI, 2017). Nas máquinas resfriadas a ar, as descargas que ocorrem nas superfícies das bobinas ou barras contribuem para a geração de ozônio. Essa geração de ozônio é resultado da quebra das moléculas de oxigênio presentes no ar pela energia liberada durante a descarga elétrica, formando o gás ozônio (O₃). Sensores empregados neste método comumente utilizam a absorção de raios ultravioleta (UV) para medir concentrações de ozônio. Outros sensores utilizam reações eletroquímicas para realizar a detecção de ozônio. Outros tipos de sensores empregam princípios eletroquímicos para a detecção de ozônio no ambiente.

2.2.6 Detecção elétrica

A detecção elétrica de descargas parciais é comumente realizada por meio de sensores capacitivos acoplados às fases do objeto de teste. A norma (IEC-60270, 2000) estabelece alguns circuitos básicos para garantir o melhor desempenho possível em um circuito acoplador, visando a repetibilidade e confiabilidade das medidas. Esses circuitos se diferenciam pelo arranjo físico, conferindo características específicas a cada um. Este trabalho concentra-se especificamente no circuito com acoplamento capacitivo, devido à sua alta eficiência e maior indicação para a

realização das medidas. O acoplamento dos capacitores forma um divisor de tensão, reduzindo o valor da tensão aplicada no objeto testado. Essa configuração proporciona uma metodologia eficaz para avaliar e medir descargas parciais em sistemas elétricos, permitindo uma análise precisa das características dos pulsos associados a esse fenômeno.

2.3 Tipos de descargas

Este trabalho utiliza como base para classificar os tipos de descargas parciais, a norma IEC 60034-27-2, a qual trata de monitoramento de máquinas elétricas rotativas. A seguir, serão descritas as principais classes tratadas neste estudo.

2.3.1 Descargas internas

As descargas internas surgem devido a falhas no isolamento, que podem ocorrer durante o processo de fabricação ou se desenvolver ao longo do tempo devido a tensões mecânicas, térmicas ou elétricas constantes durante a operação da máquina. Essas tensões resultam na formação de vazios e delaminações dentro do material dielétrico, comprometendo a integridade do isolamento e tornando-o mais suscetível a falhas. A presença dessas falhas cria regiões de menor resistência dielétrica, aumentando a probabilidade de descargas parciais quando expostas a campos elétricos elevados. Assim, as descargas internas não apenas enfraquecem o sistema de isolamento, mas também servem como precursores de eventos mais graves de falha elétrica (STONE; WU, 2009).

2.3.1.1 Descargas de cavidades

São descargas que ocorrem em cavidades no sistemas de isolamento de fita de mica impregnados com resina, comumente utilizada em máquinas rotativas de alta tensão. A presença de mica previne que as descargas desenvolvam uma quebra da rigidez dielétrica do sistema de isolamento. Contanto que essas cavidades tenham um tamanho pequeno e não sofram um alargamento, o equipamento continuará a operar de maneira confiável. A Fig. 2 demonstra o padrão do PRPD desse tipo de descarga.

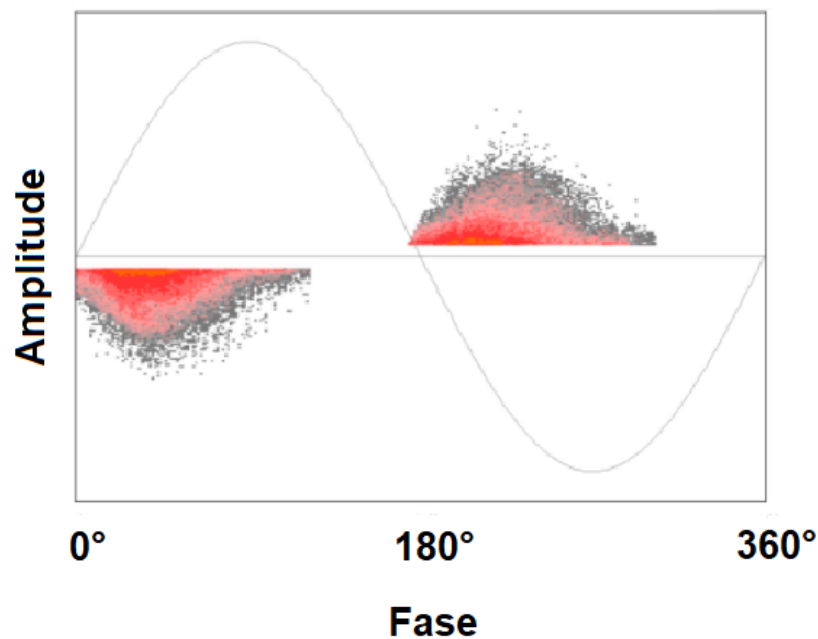


Figura 2 – Exemplo do padrão de PRPD para descarga de cavidade. Fonte: adaptado de (IEC-60270, 2000)

2.3.1.2 Delaminação interna

Dois tipos de delaminação ocorrem: um na fita isolante da parede do solo e o outro próximo aos filamentos internos do condutor. Barras Roebel com isolamento epóxi-mica raramente são afetadas, exceto aquelas com a primeira geração de flocos de poliéster-mica, que são mais propensas a delaminar internamente. A delaminação é mais comum em bobinas de múltiplas voltas. Em um caso testado, foi encontrada extensa delaminação no isolamento da parede do solo e da bobina de um gerador. A Fig. 3 ilustra este tipo de defeito presente em barra estatórica.

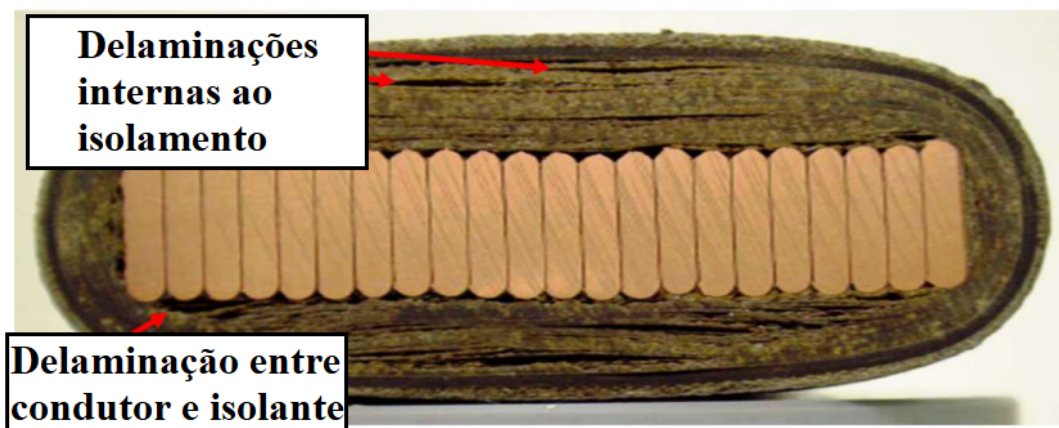


Figura 3 – Seção transversal de uma barra estatórica, com delaminação visível. Fonte: adaptado de (IEC-60270, 2000)

Este tipo de descarga pode gerar envelhecimento acelerado na máquina e precisa ser

considerado na sua avaliação. A Fig. 4 demonstra o padrão do PRPD desse tipo de descarga.

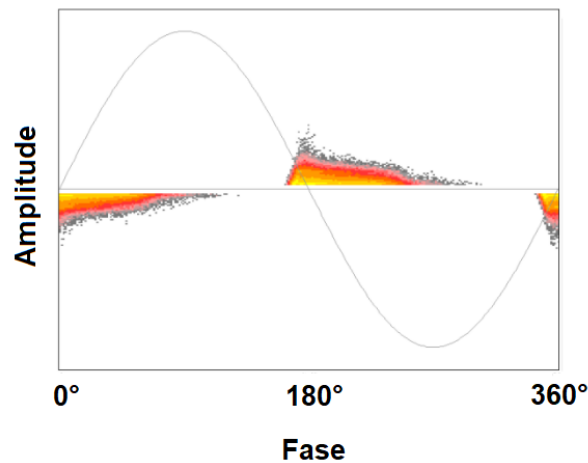


Figura 4 – Exemplo do padrão de PRPD para delaminação interna. Fonte: adaptado de (IEC-60270, 2000)

2.3.2 Arborização elétrica

Trata-se de um processo de envelhecimento no qual canais finos de erosão se propagam através do material isolante, podendo resultar na ruptura elétrica do isolamento principal. Esta ruptura, também conhecida como ruptura dielétrica, acontece quando um material isolante elétrico passa a conduzir eletricidade sob uma tensão suficientemente alta. A arborização elétrica pode ocorrer em cavidades do isolamento preenchidas por gases, ar, impurezas ou pela delaminação do material.

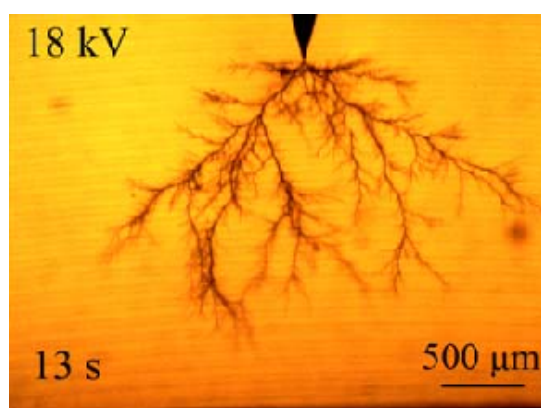


Figura 5 – Exemplo de arborização gerada por agulha imersa em óleo. Fonte: (CHEN et al., 2015)

2.3.3 Descargas de ranhura

Existem três grandes causas para ocorrência de descargas de ranhura, essas são: Barras soltas permitem a vibração da barra na ranhura, o que pode desgastar e destruir o revestimento

semi-condutivo da ranhura (STONE et al., 2008). O revestimento semi-condutivo da ranhura é mal fabricado quando não cumpre sua função devido a uma resistência inicial excessivamente alta ou à aplicação inadequada do revestimento na parede do solo (STONE et al., 2008). Má conexão do revestimento semi-condutivo à terra, ou seja, quando a barra não está devidamente aterrada devido à presença de um filme isolante ou um material de isolamento lateral entre o revestimento semi-condutivo da ranhura e o núcleo do estator (STONE et al., 2008). A Fig. 7 ilustra este tipo de defeito.

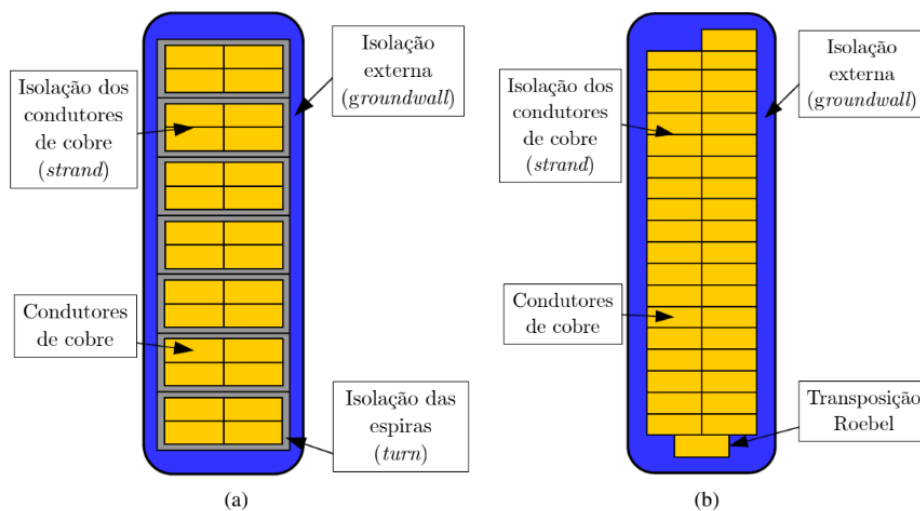


Figura 6 – Corte transversal em um ranhura: a) bobina de múltiplas espiras; e b) barra Roebel. Fonte: (CRUZ et al., 2022)

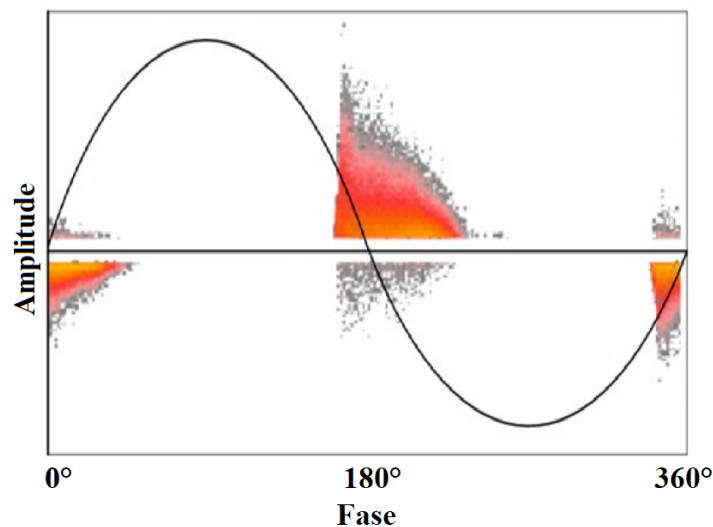


Figura 7 – Exemplo do padrão de PRPD para ranhura. Fonte: adaptado de (IEC-60270, 2000)

Este tipo de descarga apresenta maior intensidade quando o equipamento já passa por problema mecânico sério, podendo aumentar ainda mais os danos já presentes e a probabilidade

de falha. É de extrema importância que esse tipo de descarga seja detectada nos seus estágios iniciais para que sejam tomadas as medidas apropriadas.

2.3.4 Descargas na cabeça de bobina

2.3.4.1 Corona

As descargas superficiais, conhecidas como "descargas de efeito corona", podem fragilizar a isolação das barras do enrolamento estatórico dos geradores, levando à falha do sistema. Existem duas manifestações principais desse fenômeno: o efeito corona anular e o efeito corona linear. O efeito corona anular ocorre na região de interfase entre os recobrimentos semicondutores de ranhura e o de alívio nas cabeças das bobinas. Isso acontece devido à falha na continuidade dos materiais, causada pelo contato galvânico entre partículas de carvão de silício, grafite e negro de fumo. Esse problema resulta em aquecimento localizado excessivo, levando à formação de eflúvios elétricos na forma de efeito corona. Por outro lado, o efeito corona linear pode surgir em regiões não planejadas no projeto, geralmente devido a falta de distanciamento elétrico adequado entre as barras. Essas manifestações lineares podem ocorrer em locais inesperados e são desafiadoras de serem controladas. Uma característica presente nos PRPDs é a assimetria a favor de descargas positivas que ocorrem durante o ciclo negativo da voltagem, combinada com um formato arredondado das nuvens do PRPD como presente na Fig. 8.

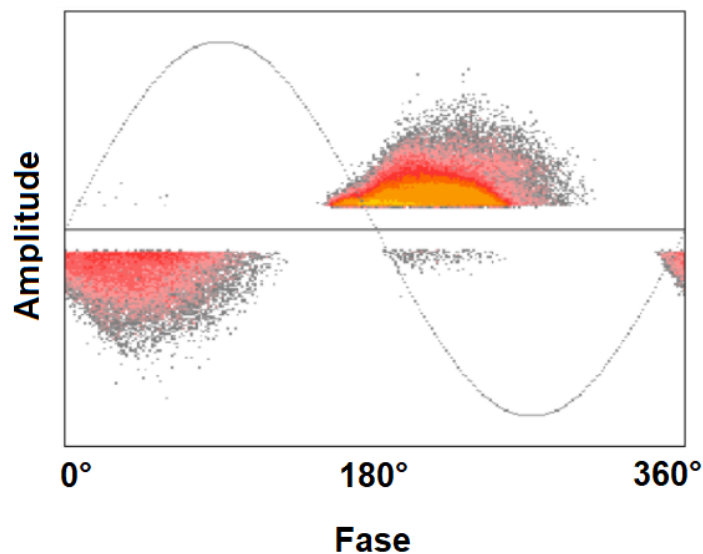


Figura 8 – Exemplo de padrão PRPD para descarga do tipo corona. Fonte: (IEC-60270, 2000)

2.3.4.2 Gap

Essa atividade ocorre na do enrolamento entre duas barras e é causada quando o espaço de ar entre as barras é muito pequeno para suportar o estresse elétrico. As descargas entre barras podem ocorrer entre barras de fases diferentes, entre uma barra de alta voltagem e uma barra de neutro da mesma fase, ou até mesmo entre barras do mesmo plano. Essas descargas deixam um pó branco na superfície das barras, visível durante a inspeção visual.

Produz um padrão PRPD característico, com espaçamento de lacuna muito maior do que as descargas internas que ocorrem em vazios minúsculos (HUDON; BÉLEC, 2005). Esse padrão é caracterizado por um sinal de amplitude quase constante na representação PRPD. Na maioria das vezes, as descargas são registradas durante ambos os semiciclos da voltagem aplicada. A magnitude da descarga está intimamente relacionada ao campo local e às dimensões da lacuna de ar. A Fig. 9 ilustra o padrão de uma descarga do tipo *gap*.

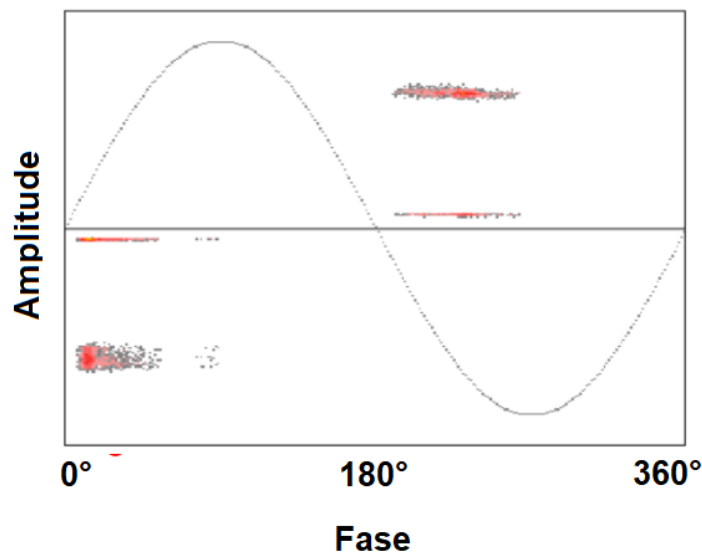


Figura 9 – Exemplo de padrão PRPD para descarga do tipo *gap*. Fonte: (IEC-60270, 2000)

2.3.4.3 Descargas superficiais

Essas descargas ocorrem quando a intensidade do campo elétrico excede a resistência dielétrica do ar ao redor. Geralmente, elas se manifestam na superfície do isolamento elétrico próximo às pontas dos enrolamentos. Essas descargas são frequentemente desencadeadas por efeitos térmicos ou contaminação no revestimento protetor das extremidades dos enrolamentos. Isso contribui para a deterioração do isolamento e pode eventualmente levar à sua falha se não for tratado. Este tipo de descarga é caracterizada pela ocorrência de nuvens verticais. A Fig. 10

ilustra o padrão de uma descarga de superfície (IEC-60270, 2000).

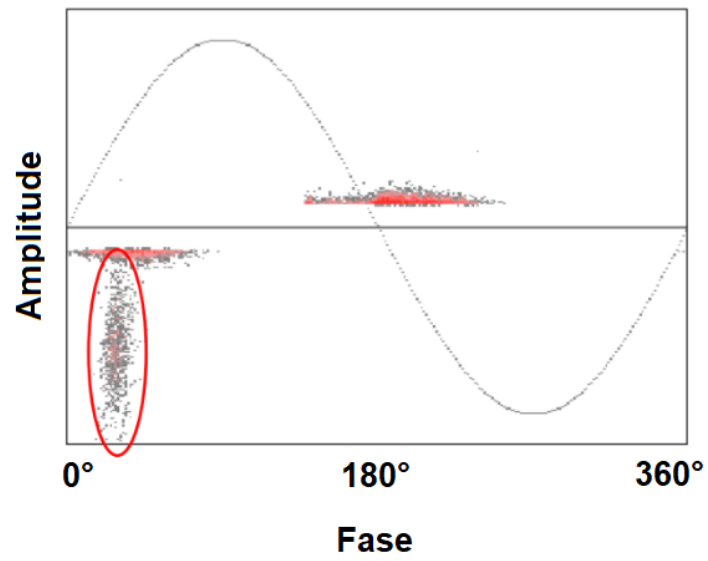


Figura 10 – Exemplo de padrão PRPD para *surface tracking*. Fonte: (IEC-60270, 2000)

3 APRENDIZADO DE MÁQUINA

Neste capítulo, serão introduzidos conceitos gerais sobre aprendizado de máquina, abordando conhecimentos sobre técnicas para redução de dimensionalidade, agrupamento de dados, redes neurais artificiais e modelos generativos

3.1 Redução de dimensionalidade

A análise de dados multidimensionais complexos apresenta desafios relacionados à implementação e execução computacional eficiente. Essa problemática ressalta a dificuldade inerente à manipulação de informações em dimensões elevadas, exigindo abordagens específicas para garantir uma análise eficaz. A visualização de dados, por sua vez, figura como uma ferramenta essencial na análise de conjuntos de dados, proporcionando a capacidade de identificar padrões significativos. Contudo, essa tarefa se torna notavelmente mais desafiadora quando se lida com dados de alta dimensão. O aumento na complexidade, inerente a um maior número de características, impõe limitações à capacidade humana de interpretar e compreender eficazmente a informação, tornando a visualização uma tarefa complexa e intrincada. Nesse contexto, algoritmos de redução de dimensionalidade desempenham um papel fundamental. Métodos como o PCA (JOLLIFFE; CADIMA, 2016) e o PaCMAP (WANG et al., 2021) são utilizados para reduzir as dimensões dos dados, preservando ao máximo as informações relevantes.

3.1.1 PCA

A *análise de componentes principais* (PCA, do inglês *principal component analysis*) é uma técnica estatística utilizada para redução de dimensionalidade em conjuntos de dados multivariados. Seu objetivo é encontrar componentes principais ordenados em termos de sua importância. Esses componentes são então usados para projetar os dados em um novo espaço de dimensão reduzida (JOLLIFFE; CADIMA, 2016).

A história do PCA remonta ao início do século XX, quando foi inventada por Karl Pearson em 1901 (PEARSON, 1901), como uma analogia ao teorema dos eixos principais na mecânica. Posteriormente, na década de 1930, Hotelling lhe conferiu o nome pelo qual é conhecida atualmente (HOTELLING, 1936).

Ao longo do tempo, o PCA ganhou diversas designações em diferentes campos de

aplicação, sendo chamado de transformada discreta de Karhunen–Loève (KLT), transformada de Hotelling, decomposição ortogonal adequada (POD, do inglês *proper orthogonal decomposition*), decomposição de valores singulares (SVD, do inglês *single value decomposition*), decomposição de autovalores (EVD, do inglês *eigen value decomposition*), o que reflete a versatilidade e aplicabilidade do PCA em diversas disciplinas científicas e áreas de engenharia.

A transformação PCA é fundamentada no cálculo dos autovetores e autovalores da matriz de covariância dos dados originais. Para entendermos melhor, consideremos a representação matricial dos dados originais X (ANDRADE; PINTO, 2003). Onde as linhas representam as observações e as colunas representam as variáveis.

A matriz de covariância, denotada por C_X , é calculada a partir da matriz de dados. Seja X esta matriz, X_i uma observação (vetor linha da matriz de dados), \bar{X} o vetor de média dos dados, a matriz de covariância é dada por (3.1):

$$C_X = \frac{1}{M} \sum_{i=1}^M (X_i - \bar{X})^T (X_i - \bar{X}), \quad (3.1)$$

onde M é o número de observações.

O próximo passo é calcular os autovalores λ_i e autovetores correspondentes v_i da matriz de covariância. A equação característica é dada por $C_X v_i = \lambda_i v_i$. Os autovetores representam as direções dos eixos principais, e os autovalores são usados para selecionar os componentes mais importantes nos dados. Componentes com autovalores maiores contribuem mais para a variabilidade dos dados ao longo dessas direções. Os autovetores são ordenados de acordo com a magnitude dos autovalores associados, e os k primeiros autovetores (os mais significativos) são escolhidos para formar a matriz de projeção W . Essa matriz é composta pelos autovetores como se segue:

$$W = [v_1, v_2, \dots, v_k], \quad (3.2)$$

onde k é o número de componentes principais selecionados.

Finalmente, a transformação dos dados originais X para o novo espaço de características é realizada multiplicando a matriz de dados X pela matriz de projeção W

$$T = XW, \quad (3.3)$$

onde a matriz T contém as projeções nos k componentes principais.

3.1.2 Kernel PCA

A *análise de componentes principais com kernel* (KPCA, do inglês *kernel principal component analysis*) é uma extensão da PCA que permite lidar com conjuntos de dados não linearmente separáveis, projetando-os em um espaço de maior dimensionalidade antes de aplicar a PCA padrão. Essa técnica é especialmente útil quando os dados têm uma estrutura complexa e não podem ser adequadamente representados por uma transformação linear.

O KPCA funciona através da aplicação de uma função (*kernel*), que é usada para calcular o produto escalar dos vetores no espaço de características, sem a necessidade explícita de calcular as coordenadas no espaço de características. Isso permite capturar relações não lineares entre os dados. Existem vários tipos de *kernel*, dentre os principais se pode citar:

- *Kernel* linear:

$$K(X_i, X_j) = X_i^T X_j, \quad (3.4)$$

onde X_i e X_j são os vetores coluna de entrada.

- *Kernel* de função de base radial (RBF, inglês *radial basis function*):

$$K(X_i, X_j) = \exp\left(-\frac{\|X_i - X_j\|^2}{2\sigma^2}\right), \quad (3.5)$$

onde X_i e X_j são os vetores de entrada, σ é um parâmetro que controla a largura do *kernel*.

Quanto menor o valor de σ , menor a contribuição de pontos distantes.

- *Kernel* cosseno:

$$K(X_i, X_j) = \frac{X_i^T X_j}{\|X_i\| \|X_j\|}, \quad (3.6)$$

onde X_i e X_j são os vetores de entrada,

3.1.3 PaCMAP

O PaCMAP (do inglês, *pairwise controlled manifold approximation projection*) é outra técnica para redução de dimensionalidade (WANG et al., 2021). Este método utiliza na sua formulação três tipos de pares de dados: pares de vizinhos, pares intermediários e pares distantes.

O primeiro tipo consiste nos nNB vizinhos mais próximos de cada observação no espaço de alta dimensão. O processo inicia com o cálculo das distâncias entre pares de observações

$$d_{ij}^2 = \frac{(X_i - X_j)^2}{\sigma_{ij}}, \quad (3.7)$$

onde σ_{ij} é a distância euclidiana média entre a observação i e a observação j . O número de vizinhos próximos nNB é usado para construir os pares de vizinhos (i, j_t) , onde $t = 1, 2, \dots, nNB$. Aqui, as distâncias são usadas apenas para selecionar vizinhos.

O segundo tipo consiste em $N \times nNB$ pares intermediários selecionados ao amostrar aleatoriamente N observações adicionais, usando a segunda menor distância euclidiana delas para o par intermediário.

Finalmente, o terceiro tipo é composto por uma seleção aleatória de nFP pontos distantes de cada observação. Para conveniência, o número de pares intermediários e de pontos distantes é determinado pelos parâmetros MN ratio e FP ratio, que especificam a razão dessas quantidades em relação ao número de vizinhos mais próximos, ou seja

$$nMN = MNratio \times nNB, \quad (3.8)$$

$$nFP = FP ratio \times nNB, \quad (3.9)$$

onde nMN e nFP representam o número de pares intermediários e de pontos distantes, respectivamente; e nNB é o número de vizinhos mais próximos.

Esses parâmetros permitem ajustar a quantidade de pontos intermediários e pontos distantes em relação à vizinhança de cada observação, o que pode ser útil em diferentes cenários de aplicação. Uma vez que nNB é tipicamente uma ordem de grandeza menor que o número total de observações, a amostragem aleatória efetivamente escolhe vizinhos mais distantes entre si como pares intermediários e distantes. Ressalta-se que a decisão de escolher pares aleatoriamente, em vez de deterministicamente (por exemplo, quantis fixos), visa reduzir a carga computacional.

Para definição da localização dos pontos no espaço reduzido, o PaCMAP utiliza três funções de perda distintas, uma para cada tipo de par. As Equações (3.10), (3.11) e (3.12) representam, respectivamente, a função de perda dos pares de vizinhos, intermediários e distantes:

$$Loss_{NB} = \frac{\tilde{d}_{ij}}{10 + \tilde{d}_{ij}}, \quad (3.10)$$

$$Loss_{MN} = \frac{\tilde{d}_{ik}}{1000 + \tilde{d}_{ik}}, \quad (3.11)$$

$$Loss_{FP} = \frac{1}{1 + \tilde{d}_{il}}. \quad (3.12)$$

onde \tilde{d}_{ij} é a distância euclidiana na dimensão reduzida entre duas observações vizinhas i e j ; \tilde{d}_{ik} representa a distancia euclidiana na dimensão reduzida entre duas observações intermediárias i e k e \tilde{d}_{il} é a distancia euclidiana na dimensão reduzida entre duas observações distantes i e l .

O processo de otimização consiste em três fases, projetadas para evitar ótimos locais. O objetivo dessa otimização é encontrar uma representação de baixa dimensionalidade dos dados que preserve tanto as estruturas globais quanto as estruturas locais. Isso é crucial para garantir que a representação final capture de forma precisa as relações entre os pontos no espaço de alta dimensionalidade original. Na primeira fase, o objetivo é aprimorar a colocação inicial dos pontos na dimensão reduzida de forma a preservar tanto as estruturas globais quanto as locais, mas principalmente a estrutura global. Isso é alcançado ao atribuir pesos significativos aos pares intermediários em (3.11). Ao longo da primeira fase, são diminuídos gradualmente os pesos dos pares intermediários, permitindo que o algoritmo se concentre na estrutura local. Na segunda fase, o objetivo é aprimorar a estrutura local mantendo a estrutura global capturada durante a primeira fase, atribuindo um peso pequeno (mas não zero) aos pares intermediários.

Em conjunto, as duas primeiras fases tentam evitar ótimos locais usando um processo que apresenta semelhanças com o arrefecimento simulado e a técnica de *early exaggeration* usada no algoritmo de redução de dimensionalidade t-SNE (MAATEN; HINTON, 2008). No entanto, o *early exaggeration* enfatiza mais os vizinhos, em vez dos pontos intermediários, enquanto o PaCMAP se concentra primeiro nos pares intermediários e nos vizinhos posteriormente. Os principais obstáculos que essas fases tentam evitar são: primeiro, negligenciar a aplicação de forças em pontos distantes não próximos, o que ignora a estrutura global; e segundo, posicionar vizinhos no espaço de baixa dimensão muito distantes entre si nas primeiras iterações, causando saturação nas derivadas, e tornando difícil para esses vizinhos se aproximarem novamente. Isso levaria a *clusters* falsos na incorporação de baixa dimensão.

Finalmente, na terceira fase, o foco é aprimorar a estrutura local, reduzindo o peso dos pares intermediários para zero e o dos vizinhos para um valor menor, enfatizando o papel da força repulsiva para separar possíveis *clusters* e tornar suas fronteiras mais claras. O algoritmo utiliza o algoritmo de descida gradiente estocástica Adam (KINGMA; BA, 2017).

3.2 Algoritmo de agrupamento de dados *k-means*

Técnicas de agrupamento de dados buscam a categorização de objetos ou dados em grupos distintos, chamados de *clusters* (agrupamentos), com base em suas similaridades. O objetivo principal é identificar padrões inerentes nos dados e organizar informações de maneira significativa.

O algoritmo de agrupamento *k-means* (LLOYD, 1982; MACQUEEN et al., 1967) se destaca por sua simplicidade conceitual e eficiência computacional. O funcionamento do algoritmo envolve a atribuição de dados a agrupamentos de maneira iterativa, minimizando a variância intra-agrupamento e maximizando a variância inter-agrupamento. A principal limitação do *k-means* é a necessidade de especificar previamente a quantidade desejada de agrupamentos, representada pelo parâmetro k .

Dado um conjunto inicial de n observações (x_1, x_2, \dots, x_n) , o algoritmo *k-means* procede da seguinte forma: são determinados k agrupamentos, sendo cada observação atribuída a um agrupamento S_i ($i = 1, 2, \dots, k$) de forma a minimizar soma dos quadrados intra-clusters (WCSS, do inglês *within cluster sum of squares*), i.e.,

$$\operatorname{argmin}_S \left(\sum_{i=1}^k \sum_{x \in S_i} \|x - \mu_i\|^2 \right), \quad (3.13)$$

onde $S = \{S_1, S_2, \dots, S_k\}$ é o conjunto de k *clusters*, S_i corresponde ao i -ésimo *cluster* do conjunto, μ_i representa o centroide de S_i e $\|x - \mu_i\|^2$ corresponde à distância euclidiana entre uma observação x e o centroide de S_i .

O algoritmo é interrompido quando as atribuições não mudam mais. No entanto, não há garantia de convergência à solução ótima. Frequentemente, o algoritmo é apresentado como a atribuição de objetos ao agrupamento mais próximo em termos de distância. O método de *partição aleatória* (HAMERLY; ELKAN, 2002) atribui aleatoriamente um agrupamento a cada observação e, em seguida, procede ao passo de atualização, calculando a média inicial como o centroide dos pontos atribuídos aleatoriamente ao *cluster*.

3.3 Redes neurais artificiais

3.3.1 Histórico

O desenvolvimento das redes neurais artificiais (ANN, do inglês *artificial neural networks*) tem sua origem na elaboração dos primeiros modelos para representar o funcionamento dos neurônios humanos.

Redigido em 1943 por Warren McCulloch, um neuro-fisiólogo, e Walter Pitts, um matemático, o trabalho “A Logical Calculus of the Ideas Immament in Nervous Activity” (MC-

CULLOCH; PITTS, 1943) apresenta um modelo feito com circuitos elétricos para representar artificialmente o comportamento e o armazenamento de memória de um neurônio biológico.

Esse estudo avançou ainda mais em 1949 quando Donald Hebb publicou seu trabalho “The Organization of Behavior” (HEBB, 1949), quando foi criada uma hipótese sobre o aprendizado não supervisionado baseada na plasticidade neural, onde os caminhos neurais são fortalecidos com o uso.

Em 1958, Frank Rosenblatt (ROSENBLATT, 1958) propôs um novo modelo que poderia ser treinado para classificar padrões, denominado de *perceptron*. Este modelo propõe uma relação simples de entrada e saída, onde esta é descrita por um limiar aplicado à soma ponderada das entradas, enquanto são ajustados os pesos de ponderação para minimizar a diferença entre o resultado desejado e saída real da rede. Porém, foi demonstrado que esse modelo só é capaz de gerar soluções para problemas de solução linearmente separável.

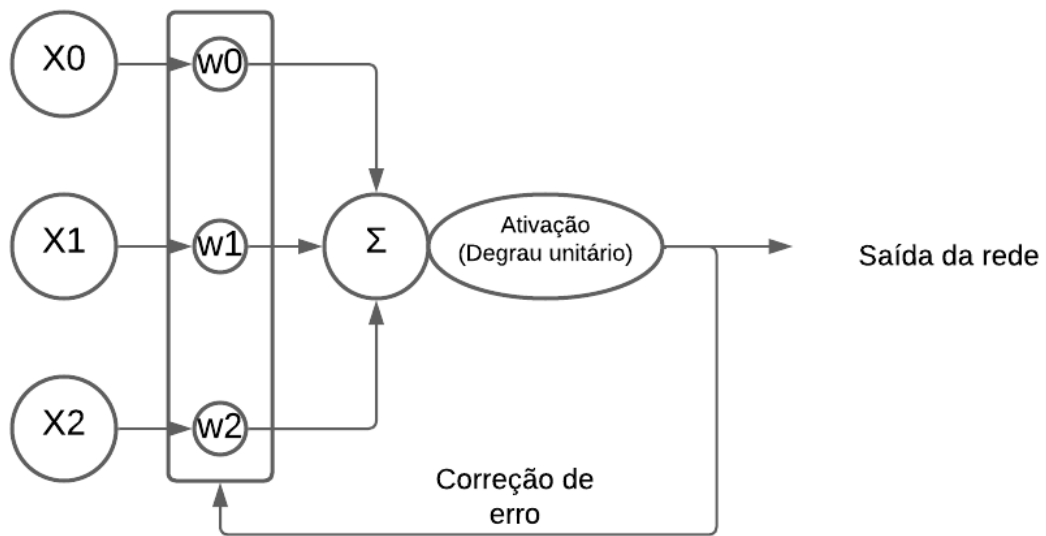
Em 1959, Bernard Widrow e Marcian Hoff (WIDROW, 1960) desenvolveram os modelos MADALINE (do inglês, *multiple adaptative linear elements*) e ADALINE (do inglês, *adaptative linear elements*) para aplicação na remoção de ruído em telefones. O interesse nesse campo de estudo seria revitalizado em 1986 com o algoritmo *backpropagation*, desenvolvido por David Rumelhart, Geoffrey Hinton, e Ronald Williams (WERBOS, 1990), que possibilitou o treinamento de redes com múltiplas camadas. Atualmente, esse campo de estudo ainda está se desenvolvendo com várias técnicas especializadas para solução de tipos específicos de problemas, assim como implementações desses modelos para soluções de problemas no mundo real (RABUNAL; DORADO, 2006).

3.3.2 Perceptron

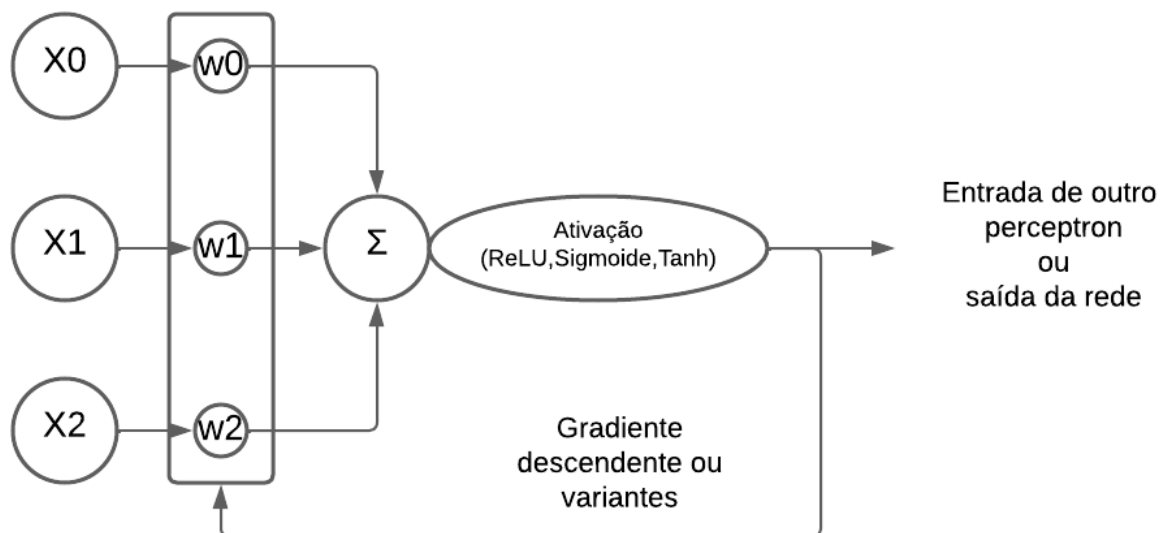
- **Arquitetura:** O Perceptron de Rosenblatt é um modelo de rede neural de uma única camada, composto por um conjunto de entradas, pesos associados a essas entradas e uma função de ativação. Ele foi concebido originalmente para resolver problemas de classificação binária linearmente separáveis. Por outro lado, as redes neurais modernas são geralmente compostas por várias camadas de neurônios, incluindo uma camada de entrada, uma ou mais camadas ocultas e uma camada de saída. Essas redes podem ter arquiteturas profundas e complexas, permitindo a modelagem de funções mais complexas e o aprendizado de representações hierárquicas dos dados.

- **Treinamento:** O treinamento do Perceptron de Rosenblatt é baseado no algoritmo de aprendizado supervisionado de correção de erro, onde os pesos são ajustados iterativamente para minimizar o erro entre as previsões do modelo e os rótulos verdadeiros. Já as redes neurais modernas são geralmente treinadas usando algoritmos de otimização como o gradiente descendente estocástico e suas variantes. O treinamento pode ser realizado usando técnicas avançadas como retropropagação (backpropagation) para calcular gradientes e ajustar os pesos em todas as camadas da rede.
- **Funções de Ativação:** Tradicionalmente, o Perceptron de Rosenblatt utiliza uma função de ativação degrau (step function) para produzir saídas binárias (0 ou 1) com base em um limiar. Por outro lado, as redes neurais modernas geralmente empregam funções de ativação não-lineares, como ReLU (Rectified Linear Unit), sigmoid, tanh, entre outras. Essas funções permitem que as redes aprendam representações não-lineares dos dados, tornando-as capazes de modelar relações complexas.

A Fig. 11 ilustra uma comparação entre os dois modelos de perceptron.



(A)



(B)

Figura 11 – Comparação entre: (A) perceptron de Roseblatt e (B) perceptron moderno. Fonte: o autor

Este modelo de perceptron compõe uma pequena célula que é normalmente utilizada em múltiplas camadas conectadas.

3.3.3 Redes neurais do tipo *feedforward*

São modelos de rede nos quais os sinais de entrada são propagados em um único sentido, não havendo, portanto, realimentação do sinal em nenhuma camada.

Redes neurais artificiais normalmente possuem múltiplas camadas: a primeira camada é denominada *camada de entrada*; a última é referida como *camada de saída*; e as camadas intermediárias (entre a de entrada e a de saída) são chamadas de *camadas ocultas*. Em uma rede *feedforward* multicamadas, a saída de uma camada é a entrada da outra imediatamente posterior. Um exemplo de rede *feedforward* pode ser visto na Fig. 12.

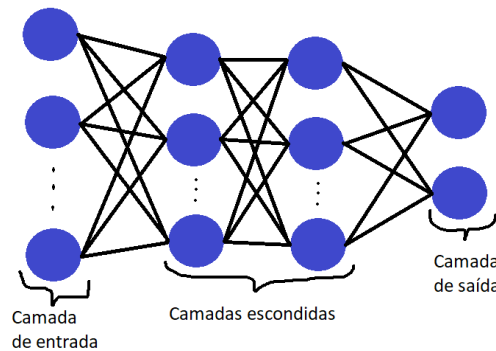


Figura 12 – Exemplo de ANN do tipo *feedforward*. Fonte: o autor

3.3.4 Função de custo

As ANNs aprendem através do processo de ajuste dos pesos em cada camada, sendo necessário definir uma forma de calcular o quão próximo um resultado está do desejado. Tal proximidade é caracterizada pela função de custo. Exemplos de funções de custo incluem o erro quadrático médio, bastante utilizado em problemas de regressão e a função de entropia cruzada, utilizada em problemas de classificação com múltiplas classes. Algumas dessas funções estão descritas a seguir:

1. Erro médio

$$C = \frac{1}{N} \sum_{i=1}^N (y_i - Y_i) \quad (3.14)$$

2. Erro quadrático médio

$$C = \frac{1}{N} \sum_{i=1}^N (y_i - Y_i)^2 \quad (3.15)$$

3. Erro absoluto médio

$$C = \frac{1}{N} \sum_{i=1}^N |y_i - Y_i| \quad (3.16)$$

4. Entropia binária cruzada

$$C = \frac{1}{N} \sum_{i=1}^N [y_i \log(Y_i) + (1 - y_i) \log(1 - Y_i)] \quad (3.17)$$

onde y_i e Y_i denotam, respectivamente, a i -ésima saída da ANN e o resultado desejado correspondente; e N é o número total de exemplos de treinamento.

Durante o treinamento é possível que o modelo gerado não tenha um grande alinhamento com os padrões de dados observados na entrada, situação em que o modelo gerado possui baixo viés e alta variância de seus resultados, e que é denominada de *underfitting*. Outra situação que pode ocorrer durante o processo de treinamento é quando há a geração de um modelo de erro que se alinhe demais aos padrões presentes nos dados de entrada. Como resultado, este modelo é útil apenas em referência ao conjunto de dados inicial e não a outros conjuntos, gerando um modelo enviesado e com baixa variância. Tal situação é denominada *overfitting* e denota baixa capacidade de generalização do modelo treinado (Fig. 13- 15).

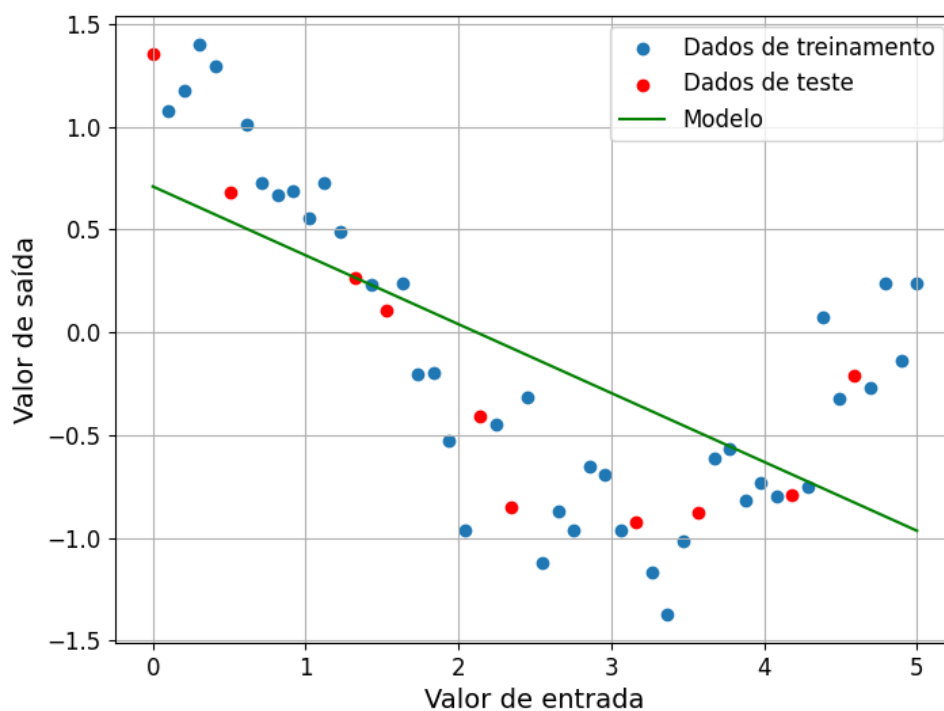


Figura 13 – Exemplo de *underfitting*. Fonte: o autor

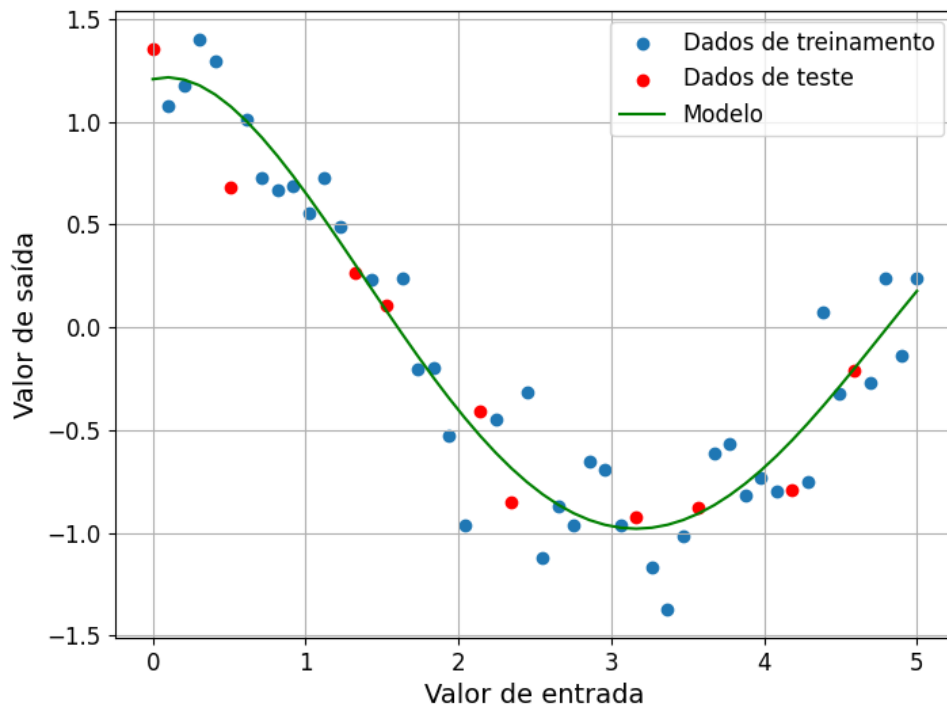


Figura 14 – Exemplo de boa generalização. Fonte: o autor

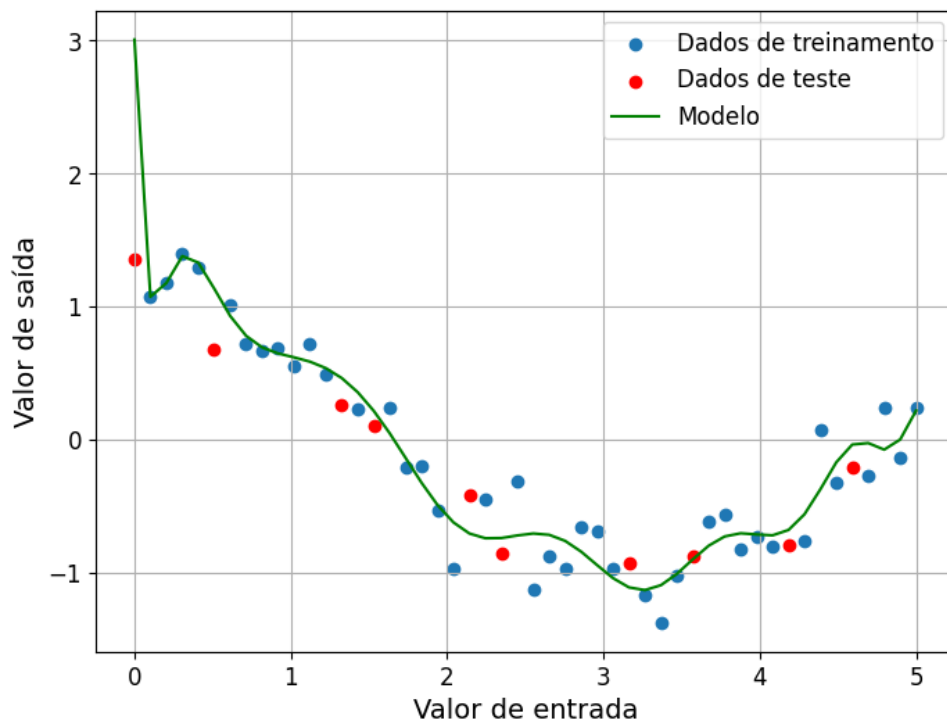


Figura 15 – Exemplo de *overfitting*. Fonte: o autor

3.3.5 Algoritmo de *backpropagation*

É necessário ajustar os pesos da ANN para atingir o resultado desejado a partir de uma entradas específicas. Uma das estratégias mais utilizadas para realizar tal tarefa, é o algoritmo

backpropagation em conjunto com gradiente descendente. Em termos gerais, o *backpropagation* utiliza a regra da cadeia para o cálculo de derivadas parciais e opera no sentido inverso ao da entrada de dados, camada a camada da rede neural. Com o auxílio da técnica do gradiente descendente, as derivadas parciais são usadas para ajustar os valores dos pesos de cada camada considerada. O objetivo dessa técnica é minimizar localmente a função de custo. É necessário cuidado na escolha da taxa de aprendizado do algoritmo, pois ela impacta diretamente no tempo necessário para se chegar ao mínimo da superfície de desempenho. Outro componente fundamental no treinamento de redes neurais feedforward é o algoritmo de *backpropagation*. Esse algoritmo é usado para calcular o gradiente da função de custo em relação aos pesos da rede, permitindo ajustar esses pesos para reduzir o erro entre a saída prevista pelo modelo e os valores reais associados (GOODFELLOW; BENGIO; COURVILLE, 2016).

3.3.6 Gradiente descendente

O gradiente descendente é um algoritmo utilizado para encontrar o parâmetros que minimizem uma função. Isto é feito de maneira iterativa dando pequenos passos na direção oposta do vetor gradiente. Consideremos, como exemplo, a função de custo.

$$f(\mathbf{w}, \mathbf{b}) = \frac{1}{N} \sum_{i=1}^N [\mathbf{y}_i - (\mathbf{w}_i \cdot \mathbf{x}_i + \mathbf{b}_i)]^2, \quad (3.18)$$

onde i é um índice que representa cada exemplo nos dados, \mathbf{w} é o vetor de pesos, \mathbf{x}_i é o vetor de entrada, \mathbf{b}_i é o vetor de vieses, e N é a quantidade de exemplos. Calcula-se, então, o gradiente da função custo.

$$\begin{bmatrix} \frac{\partial f}{\partial \mathbf{w}} \\ \frac{\partial f}{\partial \mathbf{b}} \end{bmatrix} = \begin{bmatrix} -\frac{1}{N} \sum_{i=1}^N 2\mathbf{x}_i((\mathbf{w}_i \cdot \mathbf{x}_i + \mathbf{b}_i) - \mathbf{y}_i) \\ -\frac{1}{N} \sum_{i=1}^N 2((\mathbf{w}_i \cdot \mathbf{x}_i + \mathbf{b}_i) - \mathbf{y}_i) \end{bmatrix}. \quad (3.19)$$

A solução ótima é determinada iterativamente ajustando-se os valores dos pesos \mathbf{w} e dos vieses \mathbf{b} em pequenos passos feitos na direção oposta do gradiente, determinados pela taxa de aprendizado μ . Assim, novos valores de \mathbf{w} e \mathbf{b} são calculados, i é um índice que representa cada exemplo nos dados. A equação (3.20) representa esse procedimento.

$$\mathbf{b}_{i+1} = \mathbf{b}_i - \mu \frac{\partial f}{\partial \mathbf{b}} \quad (3.20)$$

$$\mathbf{w}_{i+1} = \mathbf{w}_i - \mu \frac{\partial f}{\partial \mathbf{w}} \quad (3.21)$$

Este processo é repetido por um determinado número de iterações. A Fig. 16 ilustra de maneira simplificada este processo.

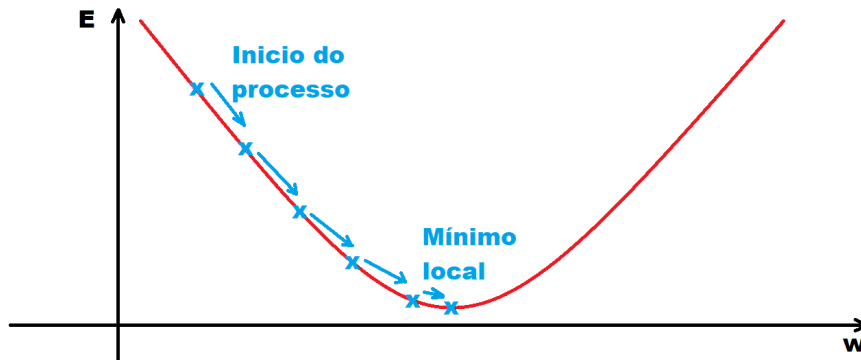


Figura 16 – Esquema simplificado do gradiente descendente. Fonte: o autor

3.3.7 Funções de ativação

São responsáveis pela geração de saída do neurônio a partir da combinação linear de suas entradas, havendo uma variedade de funções utilizadas para este fim. A seguir, estão descritas algumas das mais utilizadas.

1. Sigmoide

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (3.22)$$

2. Tangente hiperbólica

$$\sigma(x) = \frac{2}{1 + e^{-2x}} - 1 \quad (3.23)$$

3. Linear

$$\sigma(x) = ax \quad (3.24)$$

4. ReLU

$$\sigma(x) = \max(0, x) \quad (3.25)$$

5. Softmax

$$\sigma(z_j) = \frac{e^{z_j}}{\sum_{k=1}^K e^{z_k}}, \quad \text{para } j = 1, 2, \dots, K \quad (3.26)$$

z_j refere-se à entrada para a função Softmax associada à classe j

3.3.8 Regularização em redes neurais

São estratégias utilizadas para reduzir o erro de teste, sem melhorar, necessariamente, o desempenho do treinamento (GOODFELLOW; BENGIO; COURVILLE, 2016). Existem vários tipos de métodos para regularização que adicionam algum tipo de penalidade ou restrições em parâmetros da rede. Quando escolhidas corretamente, essas restrições podem melhorar o desempenho do sistema com o conjunto de teste. Dentro do contexto das redes neurais artificiais, a regularização tenta reduzir a variância o máximo possível. Existem várias estratégias de regularização, incluindo o *dropout*, *data augmentation* e as técnicas de regularização L1 e L2. O *dropout* consiste em realizar múltiplos treinamentos, aleatoriamente desativando neurônios durante os treinamentos, o *data augmentation* envolve a geração de dados adicionais através de modificações nos dados originais, o que é útil quando o conjunto de dados é limitado. As regularizações L1 e L2 adicionam termos de penalidade à função de custo, com o objetivo de controlar a complexidade do modelo. A regularização L1 tende a produzir pesos esparsos, sendo útil para seleção de *features*, enquanto a L2 favorece soluções com pesos menores em magnitude, contribuindo para evitar o *overfitting*.

3.4 Redes neurais convolucionais

Redes neurais convolucionais (CNN, do inglês *convolutional neural networks*) são um tipo específico de redes neurais artificiais, proposta pelo pesquisador francês Yann LeCun (LE-CUN et al., 1998). As CNNs mostram-se muito eficazes para resolver problemas de classificação envolvendo imagens, sendo uma alternativa viável aos métodos tradicionais para esse tipo de problema. Uma desvantagem desse tipo de rede é o alto custo computacional necessário para realizar os cálculos de convolução, cuja complexidade aumenta com o tamanho das imagens utilizadas. Outra desvantagem das CNNs é a necessidade de uma grande quantidade de dados rotulados para seus treinamentos. Essas redes são compostas de três componentes básicos: camada de convolução, *pooling* e camada completamente conectada. Uma ilustração da rede criada por Yann LeCun, a LeNet, pode ser vista na Fig. 17.

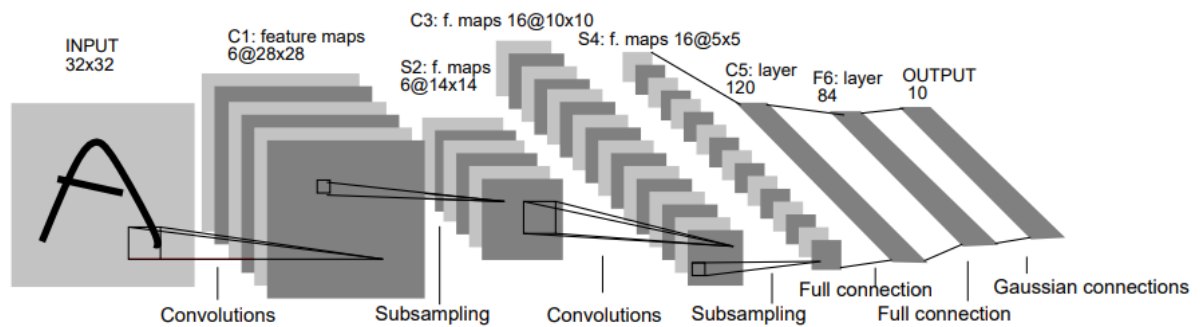


Figura 17 – LeNet5. Fonte: (LECUN et al., 1998)

3.4.1 Camada convolucional

A convolução é uma operação linear que tem duas funções como argumento de entrada e produz uma terceira função, dada por (3.27):

$$(f * g)[n] = \sum_{k=-\infty}^{\infty} f[k]g[n-k], \quad (3.27)$$

onde $f[n]$ e $g[n]$ são sinais discretos e a fórmula fornece o valor do n -ésimo elemento do resultado $(f * g)$.

Em processamento de sinais, esse tipo de equação é, em sua forma discreta, comumente utilizada para descrever o relacionamento entre o sinal de entrada e o sinal de saída. A camada convolucional em uma CNN tem por função extrair os atributos (*features*) existentes na entrada de maneira automática. O processo de extração se dá por meio de filtros convolucionais de tamanhos reduzidos. Esses filtros percorrem, de maneira iterativa, os dados de entrada em largura, altura e profundidade realizando a operação de convolução sobre os dados. No treinamento da rede, os filtros são ajustados de tal modo a indicar quando a entrada contiver uma determinada característica de interesse para o problema. No decorrer do treinamento, os filtros vão aprendendo estruturas cada vez mais complexas, e quanto mais filtros são incluídos, mais *features* podem ser obtidas. Porém, isso tem um custo de memória e processamento, o que precisa ser levado em conta na hora de definir a arquitetura.

3.4.2 Camada de *pooling*

A camada de *pooling* é destinada a reduzir a dimensão dos dados de entrada. Normalmente, essa camada é aplicada após uma camada convolucional. Uma técnica bastante utilizada é a *maxpooling*. Nessa técnica, substituímos todos os elementos de uma dada região de observação

por um único elemento, justamente o de maior valor, reduzindo com isso o tamanho da imagem como ilustrado na Fig. 18.

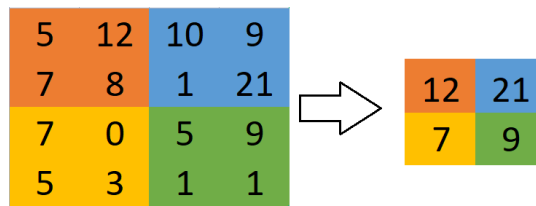


Figura 18 – Exemplo de *maxpooling* com filtro 2×2 em uma matriz 4×4 . Fonte: o autor

A técnica de *pooling* tem benefícios variados para rede neural: na medida que a quantidade de dados é reduzida de camada para camada, há melhora da regularização, bem como redução do custo associado à memória e ao processamento.

3.4.3 Camada completamente conectada

A camada completamente conectada se situa ao final da rede. Nessas camadas as *features* extraídas nas camadas de convolução anteriores são utilizadas para classificar os dados de entrada, gerando uma saída para a rede. Um exemplo de arquitetura com as três camadas básicas é apresentado na Fig. 19.

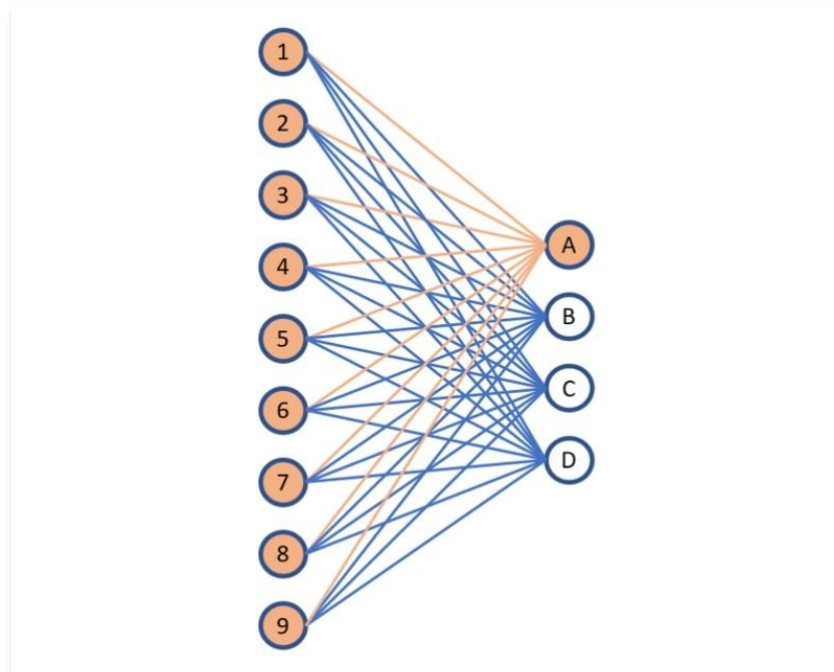


Figura 19 – Exemplo de camada completamente conectada. Fonte (IN, 2022)

3.5 Generative artificial network (GAN)

As *redes generativas adversariais* (GANs) representam uma classe revolucionária de modelos de aprendizado de máquina que tem transformado significativamente a paisagem da geração de dados. Propostas por Ian Goodfellow e seus colaboradores em 2014 (GOODFELLOW et al., 2014), as GANs introduzem um paradigma inovador ao combinar duas redes neurais, uma geradora e outra discriminadora, em um processo iterativo de aprendizado. Esta abordagem única permite a criação de dados sintéticos que se assemelham de maneira notável aos dados de treinamento originais, sendo aplicável em diversas áreas, desde a geração de imagens até a síntese de texto. A Fig. 20 ilustra este modelo

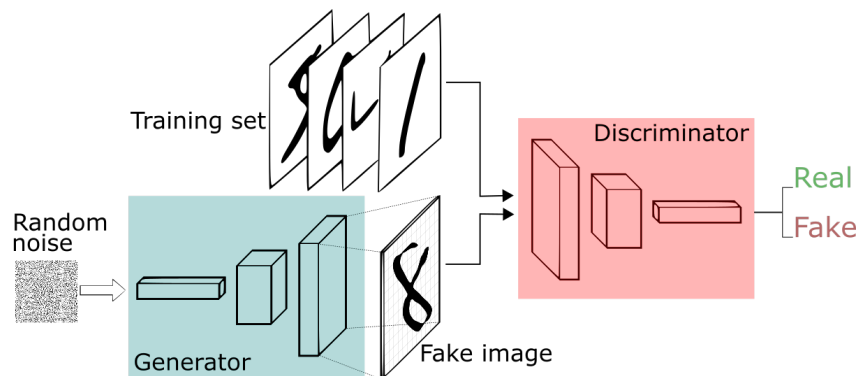


Figura 20 – Exemplo do modelo de GAN. Fonte: (GOODFELLOW; BENGIO; COURVILLE, 2016)

Uma evolução notável das GANs é o modelo DCGAN (Deep Convolutional GAN) (RADFORD; METZ; CHINTALA, 2015), que incorpora camadas convolucionais profundas para melhorar a capacidade de representação e aprendizado de características complexas em dados de alta dimensão, como imagens. Essa arquitetura DCGAN tem sido fundamental para o sucesso das GANs em aplicações práticas, impulsionando avanços significativos em campos como visão computacional e geração de imagens realistas. A Fig. 21 exemplifica a arquitetura do modelo DCGAN

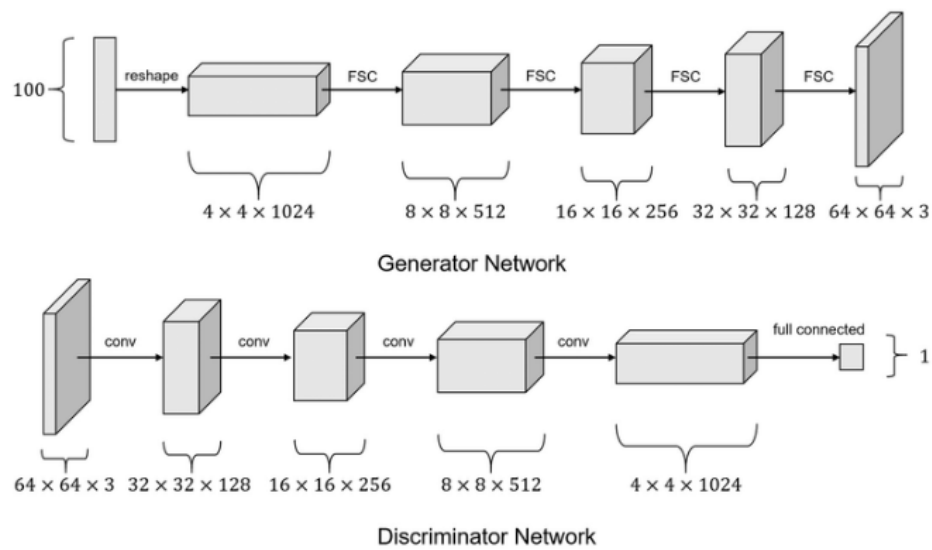


Figura 21 – Arquitetura do modelo DCGAN. Fonte: (RADFORD; METZ; CHINTALA, 2015)

A aplicação de DCGAN na geração de padrões de descarga parcial tem ganhado destaque em áreas relacionadas à análise de dados elétricos. A capacidade da DCGAN em capturar padrões intrincados e gerar artificialmente descargas parciais similares a dados reais oferece uma ferramenta valiosa para pesquisadores e profissionais envolvidos na monitorização e diagnóstico de sistemas elétricos. A obra de referência (ARDILA-REY et al., 2020) aborda esta aplicação específica, destacando a relevância do DCGAN na geração de dados realistas para estudos relacionados à descarga parcial.

Neste contexto, as GANs e, em particular, os avanços proporcionados pelo modelo DCGAN, continuam a desempenhar um papel de destaque na inovação e no desenvolvimento de técnicas de geração de dados em diversas disciplinas.

4 SISTEMA DE ROTULAGEM E USO DE GAN PARA TREINAMENTO DE CLASSIFICADORES DE PD

Este capítulo apresenta em detalhes os sistemas de rotulagem automática e geração artificial de PRPDs propostos, bem como aborda os procedimentos adotados para a realização de testes de desempenho. Os dados de PRPD utilizados neste trabalho foram obtidos a partir de medições realizadas para o monitoramento de descargas parciais em hidrogeradores. Foram acompanhadas 24 máquinas em um período de aproximadamente quatro anos, entre março de 2017 e fevereiro de 2021. Neste conjunto de dados, existe um total de 3.662 arquivos, um para cada medição realizada.

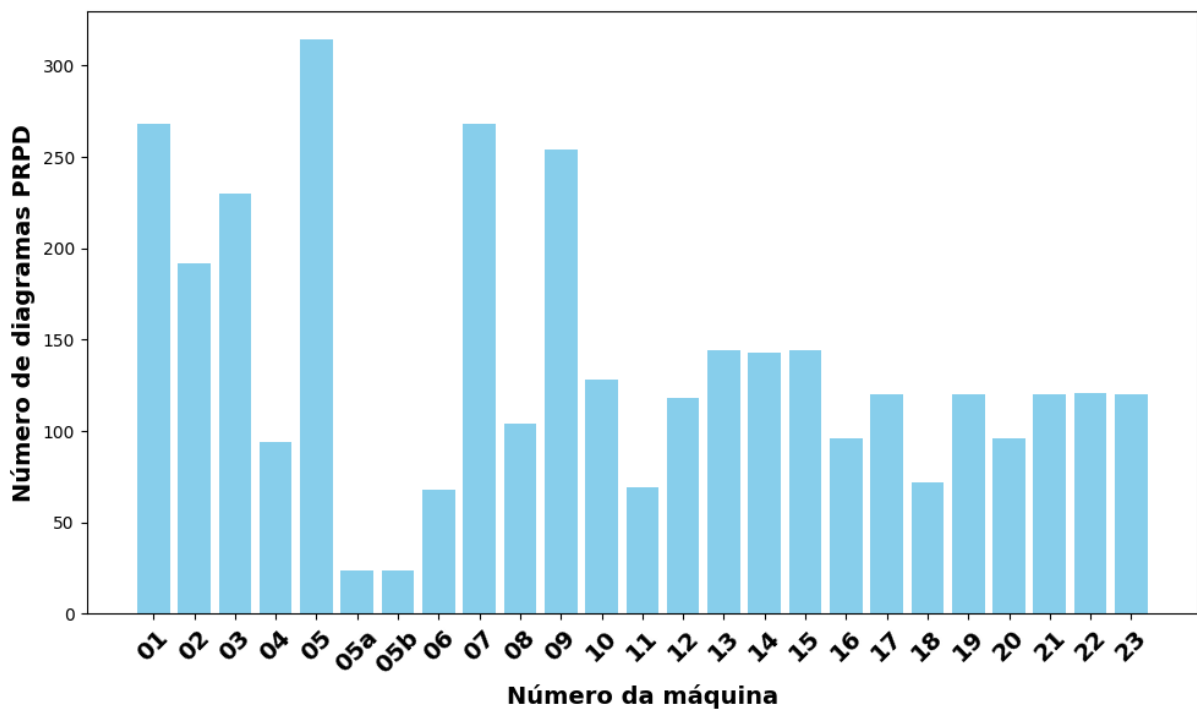


Figura 22 – Número de diagramas PRPD de cada máquina do conjunto de dados

4.1 Técnica semiautomática de rotulagem de PRPDs

Esta seção apresenta uma proposta de algoritmo para propagação de rótulos em conjuntos de dados PRPD com o objetivo de reduzir o tempo necessário para rotulagem manual. Este algoritmo consiste em uma técnica semi-supervisionada que combina operações de redução de dimensionalidade e clusterização. Deve-se ressaltar que a técnica propagação de rótulos proposta é testada em conjuntos de dados PRPD reais.

O algoritmo de propagação de rótulos é composto por etapas de agrupamento de dados PRPD, rotulagem manual de centroides, e propagação de rótulos (Fig. 23).

- **Agrupamento de Dados PRPD:** Nesta etapa, os dados PRPD são agrupados usando algum algoritmo de clustering, como o K-Means mencionado anteriormente. Esse agrupamento tem como objetivo identificar padrões semelhantes nos dados, formando clusters que representam diferentes tipos de descargas parciais.
- **Rotulagem Manual de Centroides:** Após o agrupamento, os centroides de cada cluster são identificados. Os centroides são pontos representativos de cada cluster e são geralmente escolhidos como os pontos de dados mais centrais em termos de distância euclidiana. Nesta etapa, os centroides são rotulados manualmente com base na interpretação dos padrões de descarga parcial.
- **Propagação de Rótulos:** Com os centroides rotulados, os rótulos são propagados para os demais pontos do mesmo cluster. Isso é feito assumindo que os pontos dentro de um mesmo cluster possuem características semelhantes e, portanto, devem ser rotulados da mesma forma que o seu centroide. Essa propagação de rótulos automatiza o processo de rotulagem para o restante dos pontos.

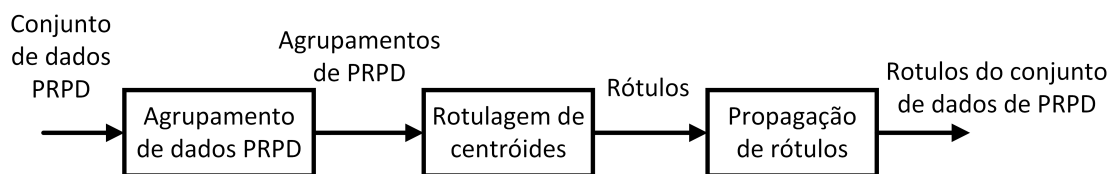


Figura 23 – Estratégia de propagação de rótulos para conjunto de dados de PRPD.

A etapa de agrupamento de dados PRPD é subdividida em pré-processamento, redução de dimensionalidade e clusterização. A Fig. 24 representa essa subdivisão da etapa de agrupamento de dados PRPD, que será detalhada a seguir.

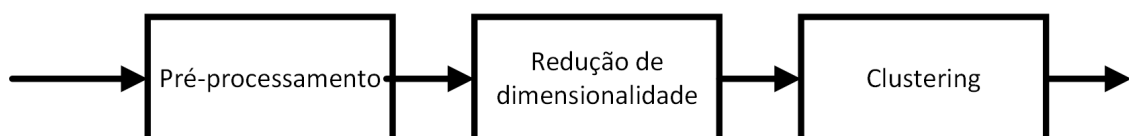


Figura 24 – Diagrama em blocos, detalhando a estratégia de agrupamento de dados PRPD.

4.1.1 Pré-processamento

Como a configuração de um equipamento de aquisição de dados deve ser ajustada para maximizar a qualidade do sinal, a faixa de amplitude dos PRPDs, mesmo quando obtidos de uma mesma máquina, pode ser diferente. Essa diferença pode potencialmente levar a problemas de classificação incorreta na fase de clusterização. Por essa razão, é introduzido um procedimento de escalonamento para colocar todos os PRPDs na mesma faixa de amplitude. Na primeira etapa do procedimento, são lidas todas as faixas de amplitude dos PRPDs que serão analisados. Em seguida, encontra-se o k -ésimo percentil, valor que especificará o intervalo para o qual todas as amplitudes dos PRPDs serão escalonadas. Empiricamente, definiu-se o valor de k como 75.

Se a faixa de amplitude original for maior do que a faixa após o escalonamento, o PRPD resultante exibirá um padrão de linhas escuras (Fig. 25b) que pode potencialmente perturbar a fase de clusterização. Assim, para mitigar o padrão de linhas escuras, utilizou-se o operador morfológico de fechamento para imagens em tons de cinza (SERRA, 1982) com um elemento estruturante de dimensões 3×3 . Um exemplo de saída da operação de fechamento aplicada a um PRPD escalonado é mostrado na Fig. 25c.

4.1.2 Redução de dimensionalidade

A etapa seguinte do procedimento de agrupamento de PRPDs visa simplificar a representação dos dados, preservando ao máximo suas características essenciais. Foram empregadas as seguintes técnicas de redução de dimensionalidade para lidar com conjuntos de dados complexos: PCA; *kernel* PCA; e PaCMAP (Seção 3.1). Em relação ao *kernel* PCA, foram testados os *kernels*: função de base radial (RBF); cosseno; e linear.

A determinação do número ideal de dimensões do espaço latente – espaço de menor dimensão que o original, mas que preserva as características mais importantes do conjunto de dados – em técnicas de redução de dimensionalidade é uma etapa crucial e não trivial, demandando uma análise criteriosa. Uma abordagem eficaz envolve a aplicação de métodos objetivos para determinar esse número.

Os critérios utilizados nesse trabalho para definir o número de dimensões do espaço latente para cada uma das técnicas de redução de dimensionalidade são os seguintes:

- PCA: a dimensão do espaço latente foi determinada pelo valor associado ao pico do perfil

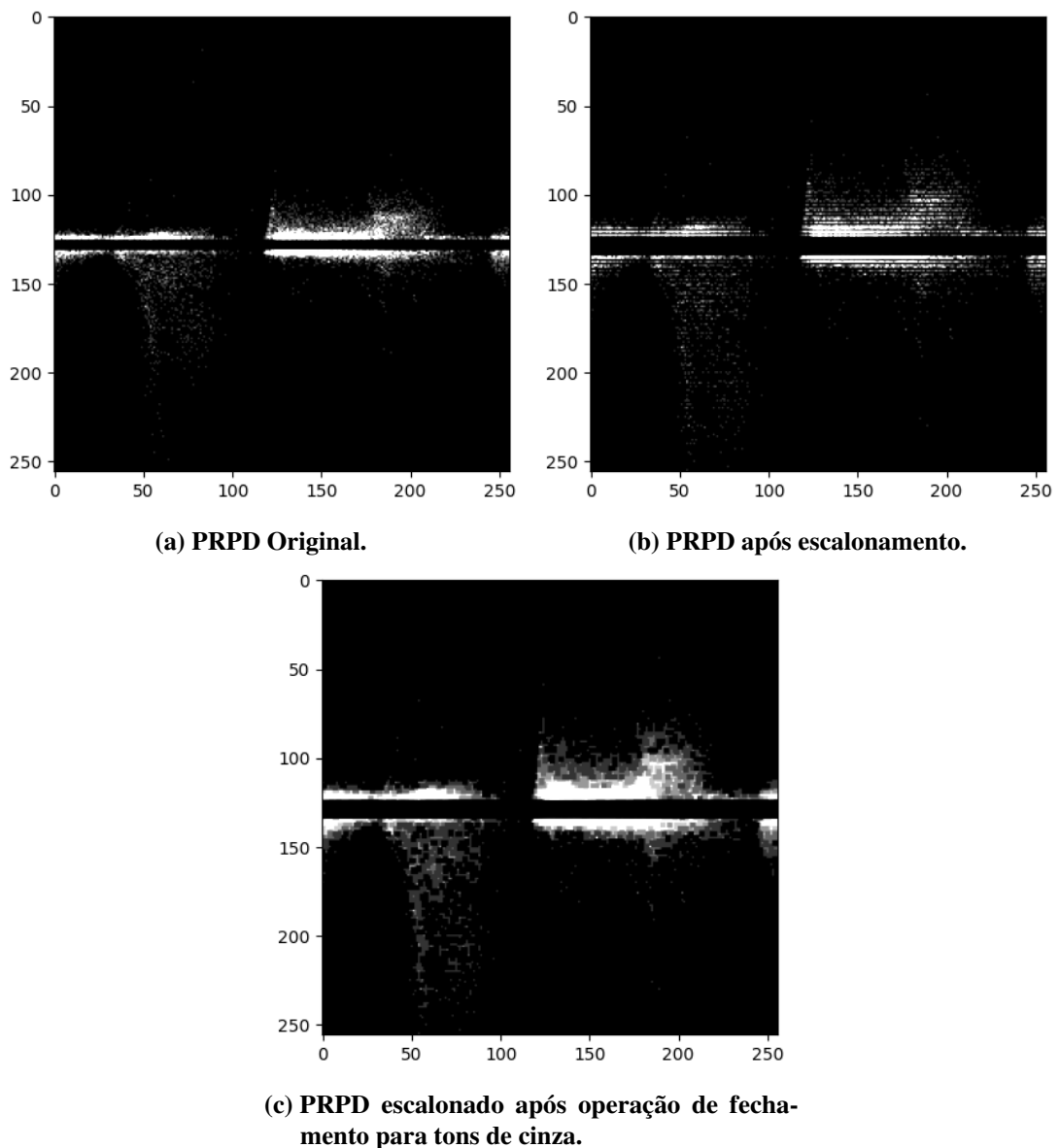


Figura 25 – Efeitos do pré-processamento nos PRPDs.

de probabilidade logarítmica (PLL, do inglês *profile loglikelihood*) (ZHU; GHODSI, 2006) aplicado aos valores singulares do PCA;

- Kernel-PCA: a determinação do número de dimensões do espaço latente é calculado o pico do PLL do MSE de reconstrução dos dados;
- PaCMAP: esta técnica não permite a reconstrução dos dados a partir da dimensão reduzida; além disso, a técnica requer a definição de 5 parâmetros: o número de componentes (dimensões do espaço latente); o número de vizinhos (nNB); a razão do número de pares intermediários para o número de vizinhos (MN ratio); e a razão do número de pares mais distantes para o número de vizinhos (FP ratio). A estratégia escolhida para otimizar esses

parâmetros foi a busca em grade. As faixas para cada parâmetro foram definidas a partir dos valores padrão ou conhecimento prévio.

4.1.3 Clustering

No trabalho, o algoritmo K-Means foi empregado para agrupar os padrões de descarga parcial (PRPDs) em clusters distintos. Após a inicialização dos centroides e a execução do algoritmo, os PRPDs foram atribuídos aos clusters correspondentes. Esses clusters foram então analisados para identificar padrões com características semelhantes, o que facilitou a interpretação dos dados de descarga parcial.

4.1.4 Escalonamento padrão de características

Às vezes, devido aos seus valores absolutos relativamente maiores, algumas características dominam outras no processo de treinamento de uma estratégia de aprendizado de máquina. Para evitar essa dominância, é realizado o escalonamento de características que consiste em normalizar cada atributo ao longo de todo o conjunto de dados, ou seja, cada característica é transformada para ter média zero e desvio padrão unitário. Dois blocos opcionais de escalonamento de características estão localizados antes da redução de dimensionalidade e antes da operação de agrupamento, como ilustrado na Fig. 26.

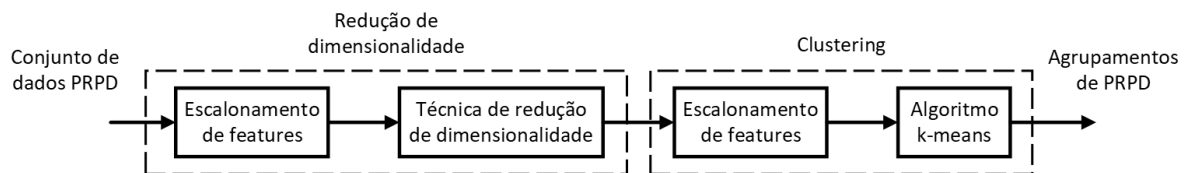


Figura 26 – Diagrama detalhado da etapa do agrupamento de PRPDs, incluindo as operações de escalonamento de características (*features*).

4.1.5 Avaliação de desempenho do sistema de agrupamento de PRPDs

1. Coeficiente de silhueta : o coeficiente de silhueta é definido como

$$s = \frac{b - a}{\max(a, b)}, \quad (4.1)$$

onde a e b são as médias das distâncias intra e interclusters, respectivamente.

Especificamente, a é a média da distância entre uma amostra e todas as outras amostras no mesmo cluster para todos os clusters; e b é a média da distância entre uma amostra e todas as outras amostras no cluster mais próximo para todas as amostras e clusters.

O coeficiente de silhueta varia de -1 a 1. Atribuições incorretas são indicadas por índices próximos a -1, enquanto valores próximos a 0 indicam clusters sobrepostos. Um índice alto implica clusters bem separados e densos.

2. Índice de Calinski-Harabasz:

O índice de Calinski-Harabasz, também conhecido como critério de razão de variância, pode ser usado para avaliar o desempenho de um procedimento de agrupamento quando a verdade fundamental não é conhecida. Tal métrica é a razão entre a soma da dispersão entre clusters e a soma da dispersão dentro de clusters.

$$s = \frac{\text{tr}(B_k) N - k}{\text{tr}(W_k) k - 1}, \quad (4.2)$$

onde k é o número de clusters, N denota o número total de amostras em um conjunto de dados.

Por sua vez, B_k e W_k são dados por

$$B_k = \sum_{q=1}^k n_q (c_q - c_E) (c_q - c_E)^T, \quad (4.3)$$

$$W_k = \sum_{q=1}^k \sum_{x \in C_q} (x - c_q) (x - c_q)^T, \quad (4.4)$$

onde C_q é o conjunto de pontos no q -ésimo cluster, c_q é o ponto central do cluster q , c_E é o ponto central do conjunto de dados, n_q é o número de pontos no cluster q .

Quanto maior o índice, clusters mais bem definidos.

3. Índice de Davies-Bouldin:

Esta métrica é uma medida de similaridade de clusters definida como a razão entre distâncias dentro de clusters e distâncias entre clusters. O índice de Davies-Bouldin é definido como

$$s = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} R_{ij}, \quad (4.5)$$

onde k é o número de clusters, e R_{ij} é dado por

$$R_{ij} = \frac{s_i + s_j}{d_{i,j}}, \quad (4.6)$$

onde s_i é a distância média entre cada ponto do cluster i e o centroide desse cluster; e d_{ij} é a distância entre os centroides dos clusters i e j .

Índices mais altos indicam clusters mais próximos entre si e mais dispersos. Assim, um melhor agrupamento está associado a índices mais baixos (o índice mínimo é zero).

4.2 Treinamento de classificadores usando PRPDs gerados por GAN

4.2.1 Geração de PRPDs artificiais

Os classificadores usados recebem como entrada diagramas PRPD. Dessa forma, foi necessário criar um *script* em linguagem Python para automatizar a leitura dos arquivos de dados de descargas parciais e colocá-los no formato de PRPD. O *script* permite ainda exibir e salvar os gráficos PRPD. A Fig. 27 ilustra exemplos de PRPD gerado com esse *script*.

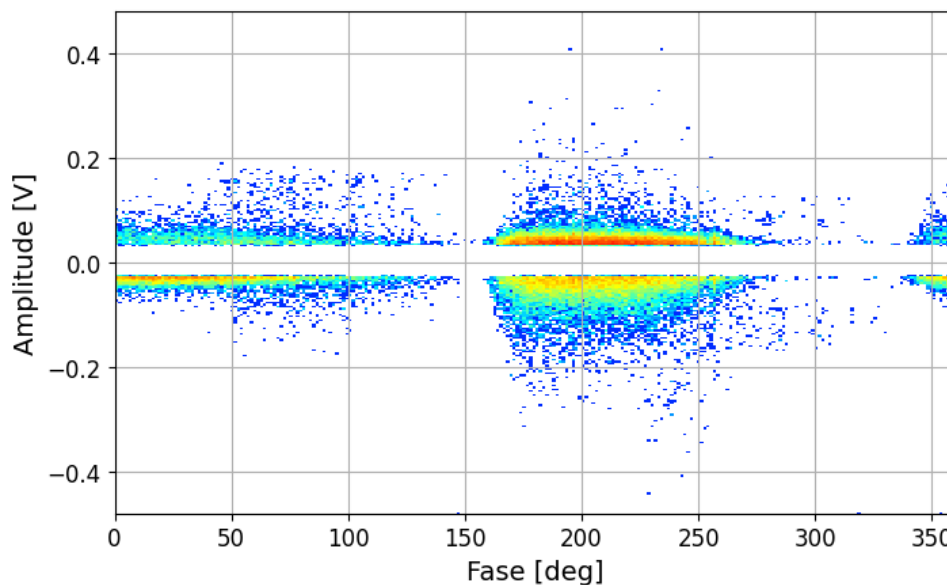


Figura 27 – Exemplo de PRPD obtido do conjunto de dados. Fonte: o autor

O processo de rotulagem do conjunto de dados desempenha um papel crucial na preparação de informações fundamentais para o treinamento de classificadores de descargas parciais em hidrogeradores. Para garantir uma abordagem padronizada e conforme as normas IEC (IEC-60034-27-2, 2012), os dados são classificados de acordo com diretrizes específicas. A estratégia adota utiliza técnicas de redução de dimensionalidade e agrupamento, para classificação em lote. O processo inicia com a divisão do conjunto de dados por máquina, os dados de cada máquina

Tabela 1 – Número de exemplos para cada classe no conjunto de dados original

Tipo	Número
Gap	413
Ranhura	949
Superfície	344
Lacuna	94
Corona	8
Não identificado	1854
Total	3662

passam por um processo de redução de dimensionalidade utilizando a técnica PaCMAP (WANG et al., 2021) que visa projetar os diagramas de PRPD para um plano bidimensional.

Os dados de cada agrupamento são inspecionados manualmente e são rotulados em lote de acordo com a norma IEC 60034-27-2. (IEC-60034-27-2, 2012).

Um total de 3.662 exemplos foram rotulados, cujas classes estão numericamente distribuídas como mostra a Tabela 1.

Observa-se um grande número de situações em que não foi possível associar os exemplos a nenhum dos padrões presentes na norma. Dentre as razões para essa dificuldade, cita-se: baixa contagem de descargas na amostra, presença elevada de ruído e ocorrência combinada de diferentes tipos de descarga parcial. Para avaliar de maneira imparcial o impacto do aumento do conjunto de dados no desempenho do classificador, o conjunto original (sem aumento de dados) foi dividido em quatro partes. Três dessas partes foram utilizadas no processo de aumento de dados, enquanto a quarta foi reservada para teste. Essa abordagem garante que o classificador não seja treinado com exemplos derivados de PRPDs que serão apresentados na fase de teste. Os resultados desta etapa são denominadas de rede com data augmentation. Uma estratégia adotada para aprimorar a qualidade do conjunto de treinamento foi a implementação da técnica de *data augmentation*, utilizando modelos de rede GAN. A arquitetura escolhida para a geração de dados sintéticos foi a DCGAN (RADFORD; METZ; CHINTALA, 2015), que emprega CNNs. Essa escolha foi motivada pela capacidade comprovada da DCGAN em aprender e sintetizar características de descargas parciais, (ARDILA-REY et al., 2020) utilizou DCGAN para gerar artificialmente descargas parciais, tornando-a particularmente eficaz na criação de dados realistas. Foram implementadas quatro GANs independentes. Cada uma dessas redes é dedicada à representação de um tipo específico de defeito. O treinamento foi feito com os parâmetros presentes na Tabela 2

Tabela 2 – Parâmetros de treinamento parcial da rede base.

Parâmetro	Valor/opção
Algoritmo de otimização	Adam
Número de épocas	100.000
Tamanho do <i>batch</i>	16
Taxa de aprendizado do gerador	2×10^{-4}
Taxa de aprendizado do discriminador	1×10^{-6}
Função de perda do gerador	Entropia binária cruzada
Função de perda do discriminador	Entropia binária cruzada

Ao longo de 100.000 épocas e com um *batch size* de 16, ajustou-se os hiperparâmetros para otimizar o desempenho dos modelos. A taxa de aprendizagem para treinamento do gerador foi definida em 2×10^{-4} , enquanto para o discriminador foi ajustada para 1×10^{-6} , ambos usando o otimizador Adam. Este processo foi realizado utilizando o conjunto de dados fornecido pela Eletronorte, garantindo que as GANs fossem treinadas com exemplos relevantes e representativos dos defeitos encontrados em hidrogeradores. A partir deste processo foi criado um novo conjunto de dados para treinamento contendo 4500 exemplos artificiais de PRPD, com 1500 exemplos por tipo de defeito. A inclusão desses dados sintéticos gerados pela DCGAN no conjunto de treinamento visa mitigar a escassez de exemplos em algumas classes, contribuindo para um treinamento mais equilibrado. Essa estratégia visa capacitar as redes a lidar de maneira mais eficaz com exemplos menos representados, resultando em um melhor desempenho geral do classificador. O treinamento foi feito de maneira similar com Em síntese, esta abordagem metodológica foi concebida não apenas para superar desafios relacionados à distribuição desigual de dados, mas também para explorar a capacidade das GANs, e em particular da DCGAN, em enriquecer conjuntos de treinamento, com o intuito de aprimorar a capacidade de generalização e classificação dos modelos propostos. Os resultados desta etapa são denominadas de rede com GAN.

4.2.2 Arquitetura do classificador

O modelo utilizado neste trabalho foi baseado no estudo (MANTACH et al., 2021), que apresentou um modelo de rede baseado em CNN para classificação *multi-label* composto por múltiplos classificadores binários independentes nas camadas finais, compartilhando camadas convolucionais para extração de *features*. O classificador foi concebido para classificar descargas no interior de recipientes com óleo e os dados de teste e treinamento foram obtidos em um ambiente controlado. A arquitetura é formada por duas camadas convolucionais de 36 filtros

3×3 com *maxpooling*, seguidas por uma camada densa de 128 filtros, outra camada densa com 64 filtros e um camada de classificação com 7 nós, um para cada classe ilustrada na Fig. 28.

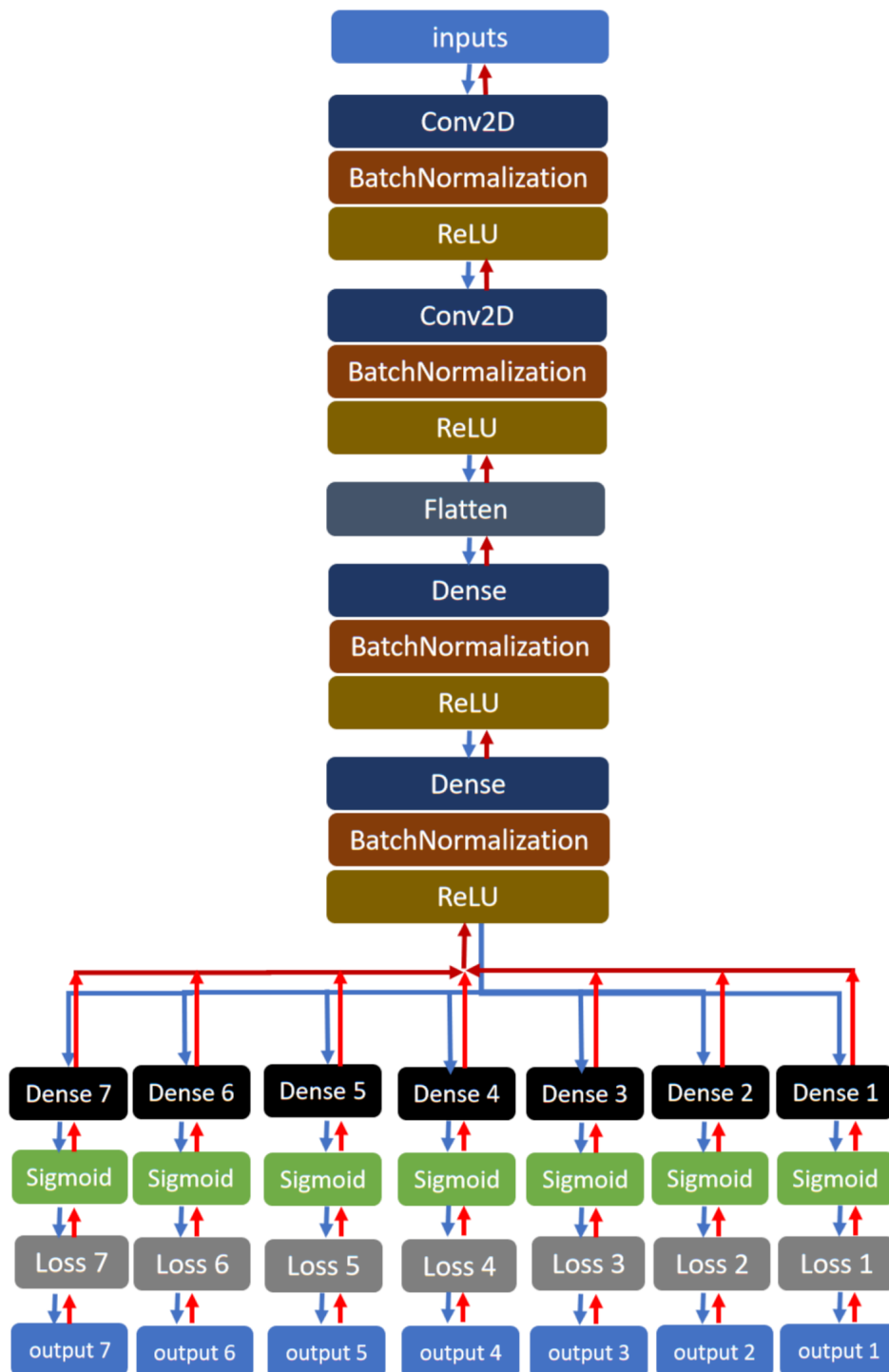


Figura 28 – Arquitetura da rede base. Fonte: (MANTACH et al., 2021)

Este modelo de rede usa como entrada PRPDs binarizados de dimensões 100×100 e originalmente trabalhou com um problema de classificação *multi-label* com 7 classes distintas. Os resultados referentes a este modelo serão referidos como rede base, já os resultados com os

dados não binarizados são denominados como rede não binarizada

4.2.3 Metodologia de treinamento do classificador

Os pesos otimizados para o modelo de rede 1 foram obtidos junto aos autores da proposta original (OLIVEIRA et al., 2017). Nesse trabalho, é relatado que no treinamento foi usada estratégia de validação cruzada *k-folds* (WITTEN; FRANK; HALL, 2011) com $k = 4$, sendo o procedimento repetido 10 vezes. Em cada iteração do algoritmo de validação cruzada, são escolhidos de 2 *folds* para formar o conjunto de treinamento, 1 *fold* para ser usado na etapa de validação, e 1 *fold* para servir como conjunto de teste. Os pesos e os *bias* foram ajustados interativamente, utilizando o algoritmo *scaled conjugate gradient backpropagation* (SCG) (MOLLER, 1993).

Também foram obtidos pesos pré-treinados para a rede base. No entanto, como esta rede foi projetada para um tipo de classificação diferente da original, foi necessário fazer um retreinamento da camada final do classificador para adaptá-lo às classes que estão sendo usadas neste trabalho. Esta abordagem foi escolhida baseando-se na premissa de que as *features* extraídas nas camadas convolucionais são igualmente boas para caracterização de PRPDs de diferentes aplicações, além de reduzir substancialmente o tempo de treinamento. O treinamento para este modelo foi feito utilizando os seguintes parâmetros, presentes na Tabela 3.

Tabela 3 – Parâmetros de treinamento parcial da rede base

Parâmetro	Valor/opção
Algoritmo de otimização	Adam
Estratégia de otimização	<i>Cross-validation k-folds</i>
Número de <i>folds</i>	10
Número de épocas	1000
Tamanho do <i>mini-batch</i>	80
Função de perda	Entropia binária cruzada
Métrica de desempenho	Acurácia

O número de *folds* (k) não estava explícito no trabalho original, sendo escolhido empiricamente, mediante comparação dos resultados para k entre 4 e 12 (os valores dos outros parâmetros foram mantidos iguais aos da Tabela 3). A Fig. 29 mostra os resultados deste processo.

O número de épocas utilizado no trabalho original foi de 4.000. Já para este trabalho, foram utilizados 1.000 épocas. A função de perda é a mesma do trabalho original que é geralmente escolhida para classificação binária. Por fim, a métrica de análise escolhida foi a mesma utilizada

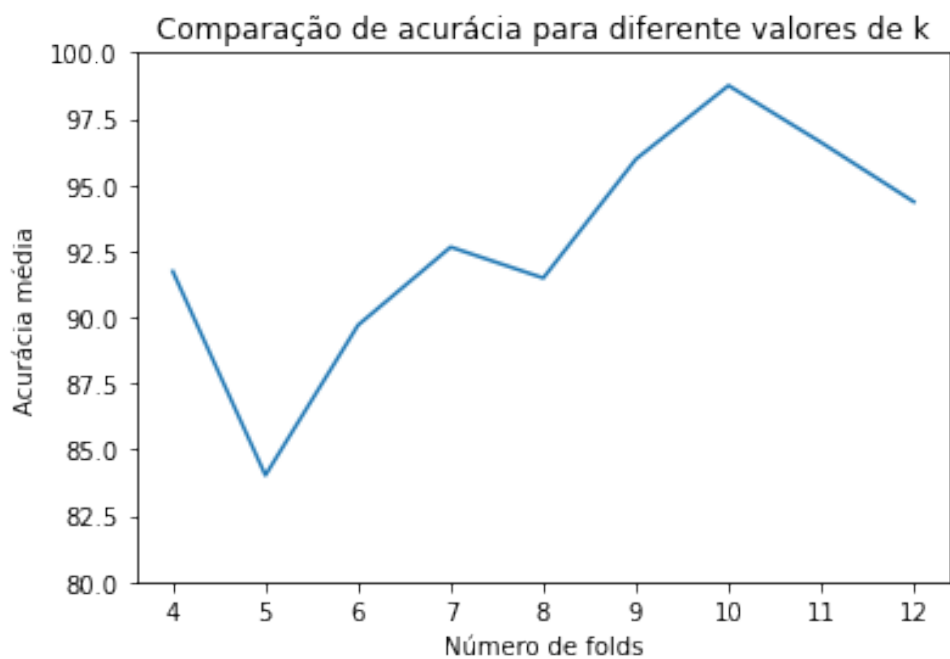


Figura 29 – Gráfico de acurácia para vários valores de k . Fonte: o autor

no trabalho original, a acurácia. Os resultados desta etapa são denominadas de rede 10 folds.

5 RESULTADOS

Neste capítulo, são comparados os desempenhos das redes neurais utilizadas na classificação de descargas parciais a partir de diagramas PRPD calculados de dados medidos em hidrogeradores. Foram comparados os desempenhos obtidos com diferentes configurações de treinamento.

5.1 Avaliação do sistema de agrupamento automático de PRPDs

Foi avaliado o desempenho de várias configurações de nossa estratégia de agrupamento do conjunto de dados PRPD de acordo com três métricas - *silhouette score*, Calinski-Harabasz e Davies-Bouldin. Buscaram-se configurações com um bom equilíbrio entre desempenho da operação de agrupamento e complexidade computacional. Começando com o *silhouette score*, a Fig. 30 mostra a média ponderada e o desvio padrão para cada configuração. Deve-se lembrar que quanto maior o *silhouette score*, melhor o desempenho de agrupamento de PRPDs. Observa-se que a média ponderada varia significativamente entre as configurações. As 10 melhores configurações de acordo com o *silhouette score* são apresentadas na Tabela 4. Na primeira coluna, tem-se a média ponderada, enquanto a segunda coluna mostra o desvio padrão correspondente. As terceira e quarta colunas exibem o número máximo de dimensões do espaço latente e o número máximo de *clusters* resultante, respectivamente, obtidos seguindo os procedimentos no Capítulo 4. A última coluna mostra o número identificador da configuração.

Observa-se que todas as configurações têm 10 como número máximo de dimensões latentes, e o número máximo de *clusters* permanece no intervalo entre 6 e 9.

Agora, considerando a métrica de Calinski-Harabasz, o desempenho relativo para as mesmas configurações muda. A Fig. 31 exibe a média ponderada e o desvio padrão para cada configuração. Como anteriormente, as 10 melhores configurações de acordo com essa métrica na Tabela 5.

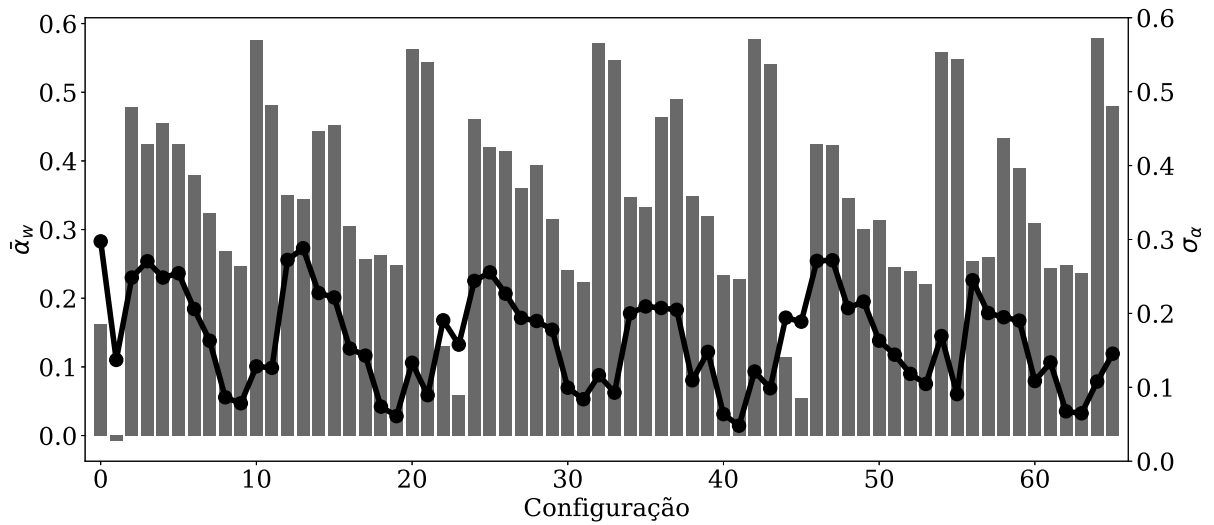


Figura 30 – *Silhouette score* para cada configuração. $\bar{\alpha}_w$: média ponderada (barras em cinza); σ_α : desvio padrão (curva em preto).

Tabela 4 – As 10 melhores configurações para o *silhouette score* (α). $\bar{\alpha}_w$: média ponderada de α ; σ_α : desvio padrão de α ; d_{\max} : número máximo de dimensões no espaço latente; e n_{\max} : número máximo de *clusters* resultantes

$\bar{\alpha}_w$	σ_α	d_{\max}	n_{\max}	Configuração
0,5791	0,1078	10	7	64
0,5775	0,1213	10	8	42
0,5766	0,1283	10	8	10
0,5711	0,1162	10	9	32
0,5626	0,1331	10	7	20
0,5580	0,1693	10	9	54
0,5481	0,0908	10	9	55
0,5464	0,0927	10	6	33
0,5441	0,0893	10	7	21
0,5411	0,0986	10	7	43

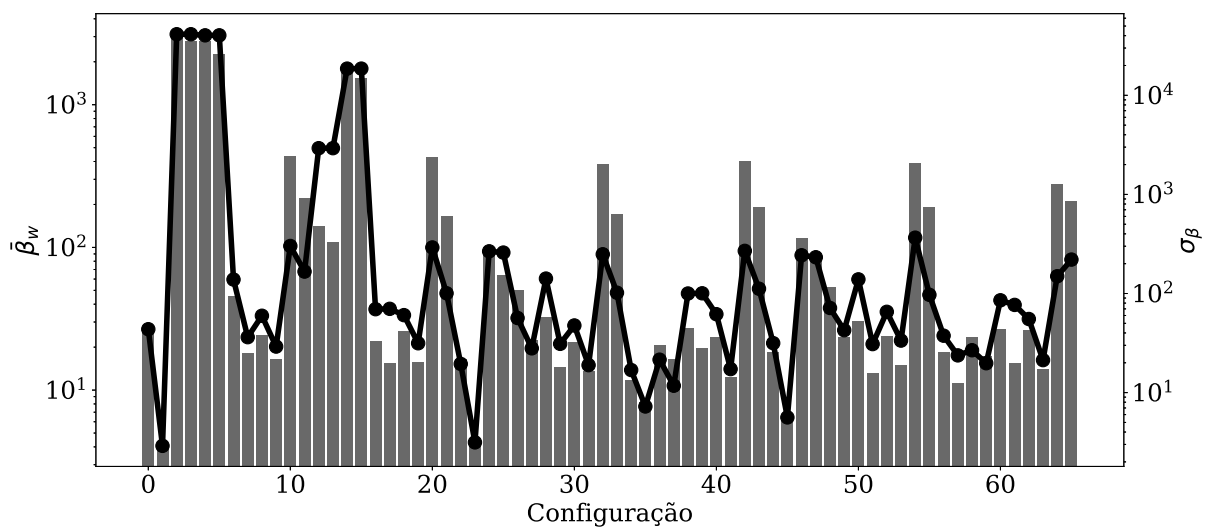


Figura 31 – Métrica de Calinski-Harabasz para cada configuração. $\bar{\beta}_w$: média ponderada (barras em cinza); σ_β : desvio padrão (curva em preto).

Tabela 5 – As 10 melhores configurações para a métrica de Calinski-Harabasz (β). $\bar{\beta}_w$: média ponderada de β ; σ_β : desvio padrão de β ; d_{\max} : número máximo de dimensões no espaço latente; e n_{\max} : número máximo de *clusters* resultantes

$\bar{\beta}_w$	σ_β	d_{\max}	n_{\max}	Configuração
31184×10^3	24845×10^{-1}	35	19	2
27782×10^3	27054×10^{-1}	35	25	3
27754×10^3	24850×10^{-1}	19	14	4
22416×10^3	25429×10^{-1}	19	15	5
16682×10^3	22756×10^{-1}	25	17	14
15380×10^3	22156×10^{-1}	25	15	15
43822×10^2	12832×10^{-1}	10	8	10
42722×10^2	1.3319×10^{-1}	10	7	20

Da mesma forma, a Fig. 32 e a Tabela 6 mostram os números para a métrica de Davies-Bouldin (neste caso, quanto menor o *score*, melhor o desempenho).

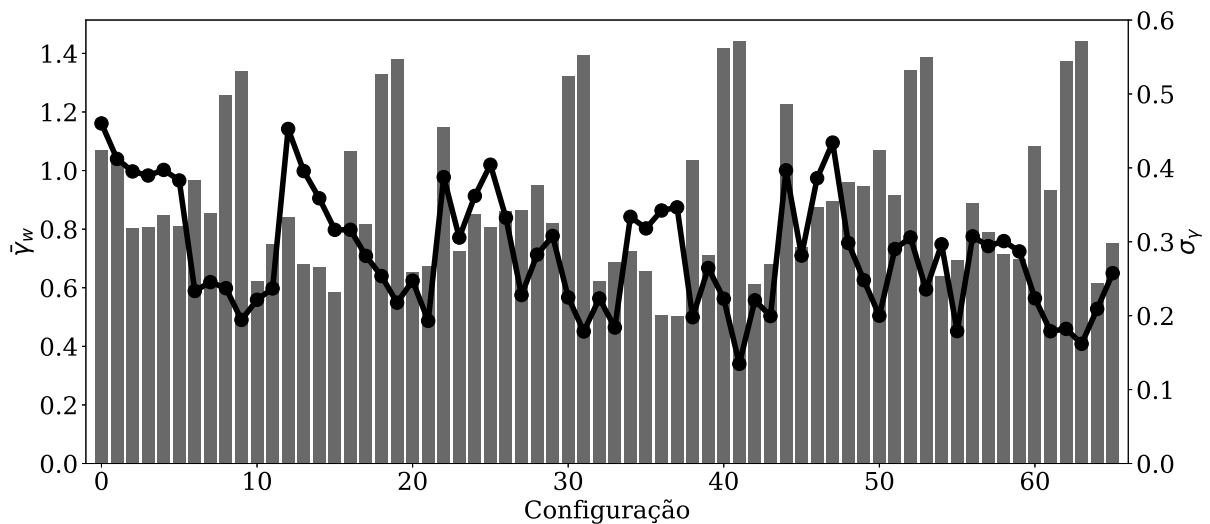


Figura 32 – Métrica de Davies-Bouldin para cada configuração. $\bar{\gamma}_w$: média ponderada (barras em cinza); σ_γ : desvio padrão (curva em preto).

Ao comparar as Tabelas 4 - 6, nota-se que o *silhouette score* seleciona configurações com um número máximo menor de dimensões no espaço latente do que as configurações selecionadas pelas métricas de Calinski-Harabasz e Davies-Bouldin. Um número menor de dimensões no espaço latente para o conjunto de dados potencialmente levará a uma etapa de *clustering* menos complexa. O mesmo ocorre com o número máximo de *clusters*. Isso afeta o tempo necessário para rotular manualmente os dados (quanto menor o número de *clusters*, menor o tempo necessário para rotular o conjunto de dados).

A Tabela 7 detalha as melhores configurações apontadas por cada uma das três métricas. Observa-se que a melhor configuração selecionada pela métrica de Calinski-Harabasz também é

a menos complexa, onde o PCA (a estratégia mais simples quando comparada com *kernel* PCA e PaCMAP) é a abordagem de redução de dimensionalidade, com ausência de escalonamento de amplitude PRPD, operação de fechamento e escalonamento de características padrão. Por outro lado, as configurações do *silhouette score* são as mais elaboradas, requerendo a realização de escalonamento de amplitude dos PRPD, operação de fechamento, bem como o escalonamento de características antes da aplicação do algoritmo PaCMAP para redução de dimensionalidade.

Tabela 7 – Melhor configuração de acordo com cada métrica. SIL: *silhouette score*; CHA: métrica de Calinski-Harabasz; DBO: métrica de Davies-Bouldin; EA: escalonamento de amplitude; RD: redução de dimensionalidade; EC-RD: escalonamento de características antes da redução de dimensionalidade; EC-CL: escalonamento de características antes do clustering

Itens	SIL	CHA	DBO
Número	64	2	37
EA	Sim	Não	Sim
Fechamento	Sim	Não	Não
RD	PaCMAP	PCA	K-PCA (linear)
Clustering	kmeans++	kmeans++	kmeans++
EC-RD	Sim	Não	Sim
EC-CL	Não	Não	Sim

Na Tabela 8, é apresentado as características das configurações mais comuns, ou seja, as configurações que aparecem nas três tabelas dos dez melhores. Foi observado uma alternância no uso de escalonamento de amplitude PRPD e escalonamento de características antes da operação de redução de dimensionalidade. Em relação ao escalonamento de características antes do clustering, vemos que a ausência dessa operação é verificada em todas as configurações relatadas, o mesmo acontecendo com o algoritmo PaCMAP na redução de dimensionalidade. A operação

Tabela 6 – As 10 melhores configurações para a métrica de Davies-Bouldin (γ). $\tilde{\gamma}_w$: média ponderada de γ ; σ_γ : desvio padrão de γ ; d_{\max} : número máximo de dimensões no espaço latente; e n_{\max} : número máximo de *clusters* resultantes

$\tilde{\gamma}_w$	σ_γ	d_{\max}	n_{\max}	Configuração
0,5023	0,2048	27	17	37
0,5078	0,2073	27	16	36
0,5837	0,2215	25	15	15
0,6126	0,1213	10	8	42
0,6165	0,1078	10	7	64
0,6220	0,1283	10	8	10
0,6241	0,1162	10	9	32
0,6396	0,1693	10	9	54
0,6522	0,1331	10	7	20
0,6560	0,2094	96	36	35

de fechamento é a opção mais comum entre as configurações relatadas.

Tabela 8 – Configurações mais comuns. EA: Escalonamento de Amplitude; RD: Redução de Dimensionalidade; EC-RD: Escalonamento de Características antes da Redução de Dimensionalidade; EC-CL: Escalonamento de Características antes do Agrupamento

Número	EA	Fechamento	DR	Agrupamento	FS-DR	FS-CL
42	Sim	Não	PaCMAP	kmeans++	Sim	Não
10	Não	Não	PaCMAP	kmeans++	Não	Não
20	Não	Não	PaCMAP	kmeans++	Sim	Não
54	Sim	Sim	PaCMAP	kmeans++	Não	Não

5.2 Resultados da geração artificial com rede GAN

No contexto específico deste estudo, as GANs foram utilizadas para gerar novos exemplos diagramas PRPDs binarizados, com o objetivo de equilibrar as classes de maneira mais robusta. As Figs. 33, 34, 35 e 36 ilustram os resultados desse processo, evidenciando a capacidade das GANs em criar amostras que preservam as características essenciais dos dados originais.

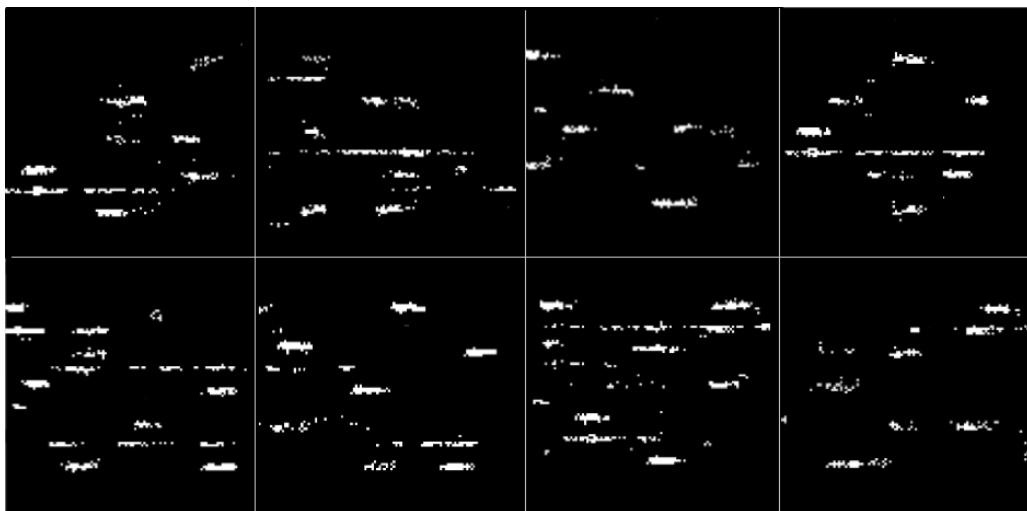


Figura 33 – Exemplo de PRPDs artificiais da classe gap gerados com GAN. Fonte: o autor

5.3 Resultados da classificação de PRPDs

5.3.1 Classificador treinado com PRPDs não binarizados

Foi realizado um experimento adicional em que foram utilizados PRPDs não binarizados, em contraste com o método empregado por (MANTACH et al., 2021). O experimento foi conduzido de maneira análoga à da seção anterior, com a única distinção sendo a exclusão do

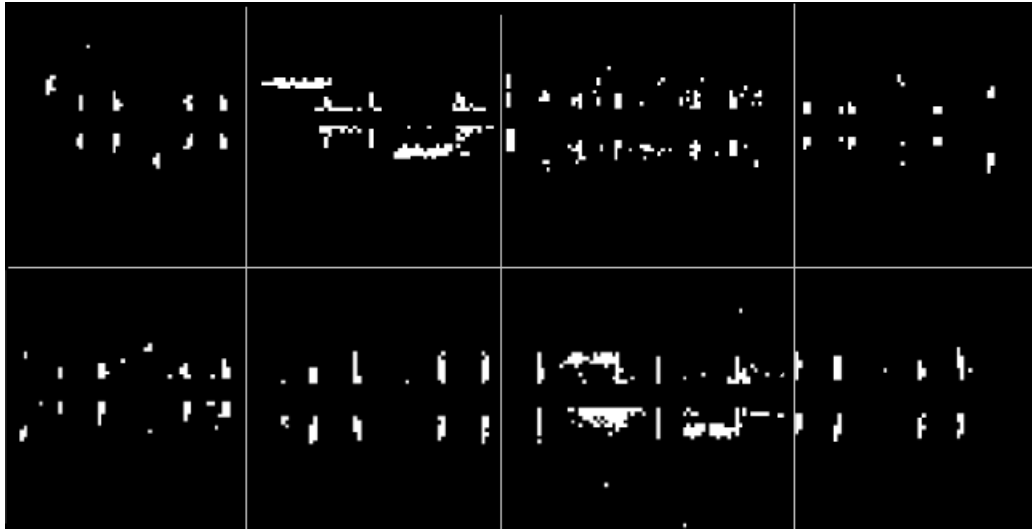


Figura 34 – Exemplo de PRPDs artificiais da classe superficial gerados com GAN. Fonte: o autor

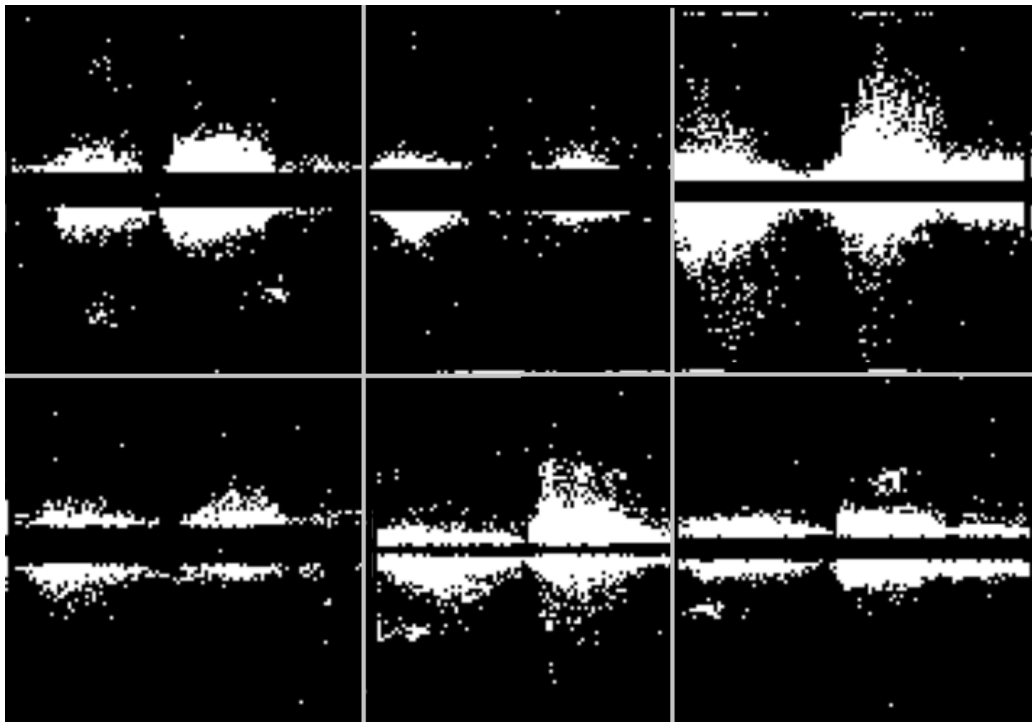


Figura 35 – Exemplo de PRPDs artificiais da classe cavidade gerados com GAN. Fonte: o autor

processo de binarização dos PRPDs. Esse procedimento preservou a dimensão de contagem de descargas parciais nos dados de entrada do classificador. A Tabela 9 mostra a acurácia de cada *fold*.

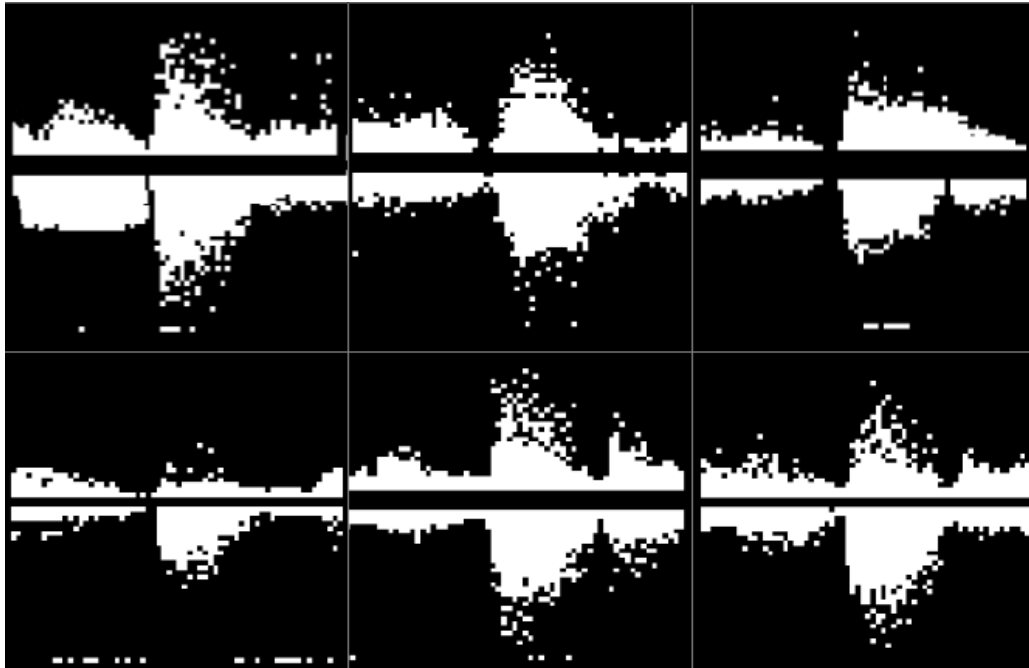


Figura 36 – Exemplo de PRPDs artificiais da classe ranhura gerados com GAN. Fonte: o autor



Tabela 9 – Acurácia dos testes com dados não binarizados: média de acurácia de 85,16% ($\pm 3,21\%$)

Fold	Acurácia do teste
1	83,72%
2	86,90%
3	85,20%
4	87,16%
5	81,98%
6	83,45%
7	84,09%
8	84,02%
9	82,76%
10	83,68%

Por fim, foi realizado mais um teste com uma parte do conjunto de dados que não foi utilizada no treinamento. A matriz de confusão é apresentada na Tabela 10.

Tabela 10 – Matriz de Confusão do Teste com a Rede Base (Dados não binarizados)

		Previsão			
		Gap	Superficial	Ranhura/Corona	Cavidade
Verdadeiro	Gap	90	10	1	0
	Superficial	10	74	1	1
	Ranhura/Corona	12	11	210	5
	Cavidade	2	1	0	5

 Verdadeiro Positivo
 Falso Positivo

A Tabela 11 ilustra a acurácia por classe obtida nesse teste.

Tabela 11 – Acurácia por classe feito com a rede após treinamento com dados não binarizados

	gap	superficial	ranhura/corona	cavidade	Geral
Acurácia	89,10%	86,04%	88,23%	62,50%	87,53%

5.3.2 Classificador treinado mediante *data augmentation* usando GAN

A Tabela 12 mostra os resultados obtidos do treinamento.


Tabela 12 – Acurácia dos testes feitos com *data augmentation* Não-treinamento da rede base: média de acurácia de 91,22% ($\pm 1,66\%$)

Fold	Acurácia do teste
1	90,12%
2	92,34%
3	88,98%
4	94,84%
5	89,56%
6	93,45%
7	91,09%
8	94,02%
9	88,76%
10	92,67%

O teste com o dataset original, feito após o treinamento gerou a matriz de confusão na Tabela 13

Tabela 13 – Matriz de Confusão do Teste com a Rede Base (*Data augmentation com GAN*)

		Previsão			
		Gap	Superficial	Ranhura/Corona	Cavidade
Verdadeiro	Gap	92	5	0	1
	Superficial	4	82	0	0
	Ranhura/Corona	10	0	226	2
	Cavidade	2	1	0	5

 Verdadeiro Positivo
 Falso Positivo

A Tabela 14 mostra a acurácia por classe para este teste.

Tabela 14 – Acurácia por classe da rede base

Classe	gap	superficial	ranhura	cavidade	Geral
Acurácia	94,23%	95,34%	95,16%	87,50%	94,84%

5.3.3 Comparação de resultados

Dessa forma, pode-se estabelecer uma comparação entre todos os testes realizados, incluindo resultados de (LOPES et al., 2022) e (LOPES, 2022), presente na Tabela 15.

Tabela 15 – Relação geral das acurácias de todos os testes

Classificador	gap	superficial	ranhura	cavidade	Geral
Rede base (MANTACH et al., 2021)	76,47%	69,81%	95,36%	0,00%	84,29%
Rede 10 <i>folds</i> (LOPES, 2022)	77,35%	83,87%	89,72%	80,00%	86,22%
Rede <i>data augmentation</i> (LOPES, 2022)	90,38%	90,69%	89,83%	75,00%	89,44%
Rede PRPD não binarizado	89,10%	86,04%	88,23%	62,50%	87,53%
Rede com GAN	94,23%	92,55%	94,95%	87,50%	94,72%

Particularmente, ao analisar o desempenho da rede base com dados não binarizados, foi evidenciado um aumento geral na acurácia, com relação aos resultados de (LOPES et al., 2022) e (LOPES, 2022), destacando uma melhoria substancial em várias classes. Por exemplo, a classe gap teve sua acurácia aumentada de 76,47% para 90,38%, enquanto a classe superficial experimentou um aumento de 69,81% para 90,69%. Além disso, a classe cavidade, que inicialmente tinha uma acurácia de 0%, obteve um aumento significativo, alcançando 75% de acurácia.

Já com a utilização de GANs, o desempenho da rede base alcançou níveis ainda maiores. Houve melhorias notáveis em todas as classes, resultando em uma acurácia geral ainda mais alta. Por exemplo, a classe gap atingiu uma acurácia de 94,23%, enquanto a classe superficial obteve 92,55%. Além disso, a classe cavidade também teve um aumento significativo em sua acurácia, chegando a 87,50%.

Esses resultados destacam a eficácia das GANs em gerar dados sintéticos de alta qualidade, que contribuíram para um desempenho ainda mais robusto e preciso da Rede neural 2 em comparação com os dados não binarizados.

6 CONCLUSÕES

Neste trabalho, realizou-se a implementação de *label propagation*, *data augmentation* com redes GAN e redes neurais aplicados a problema de classificação multi-classe de descargas parciais obtidas em hidrogeradores. Foi feita uma análise comparativa dos resultados da classificação de PRPDs a partir de cada técnica implementada.

Os resultados mostraram não apenas um aumento geral na acurácia da classificação, mas também melhorias substanciais em classes individuais, incluindo aquelas que inicialmente apresentavam desafios significativos de classificação. Em particular, a classe Void, que antes tinha uma acurácia de 0%, experimentou um aumento impressionante para 75% com dados não binarizados e um aumento ainda mais substancial para 87,50% com a inclusão de dados gerados por GANs.

Esses resultados não apenas validam a eficácia das GANs na geração de dados sintéticos de alta qualidade, mas também destacam sua capacidade de melhorar significativamente o desempenho das redes neurais em tarefas complexas de classificação. Portanto, este estudo não só contribui para o avanço do campo da análise de descargas parciais em hidrogeradores, mas também ressalta a importância de abordagens inovadoras na melhoria da precisão e confiabilidade dos sistemas de classificação. O funcionamento adequado dos hidrogeradores é de suma importância, tanto para as empresas de distribuição de energia quanto para a população que depende desses serviços. Para garantir a operação contínua desses equipamentos, é crucial monitorá-los de forma contínua, visando identificar e corrigir possíveis falhas antes que causem danos graves. Um método eficaz para realizar esse monitoramento é através da classificação de descargas parciais. Este trabalho representa uma contribuição significativa para o avanço nesse campo, ao explorar classificadores baseados em redes neurais artificiais, considerando uma variedade de fatores que impactam seu desempenho, examinando o efeito do procedimento de *label propagation* e do uso de técnicas de *data augmentation*, incluindo a geração de dados sintéticos por meio de GANs.

Ao integrar essas abordagens, foi aprimorada a capacidade dos classificadores em identificar e categorizar descargas parciais em hidrogeradores. A inclusão de técnicas de *data augmentation*, especialmente com o uso da GAN, permitiu melhorar a robustez e precisão dos modelos de classificação. Isso é crucial para garantir um monitoramento eficaz e antecipar ações corretivas, contribuindo para a segurança e a confiabilidade dos sistemas de energia elétrica.

Como planos futuros de trabalho, tem-se a implementação de testes para detecção de múltiplas classes de descargas parciais ocorrendo simultaneamente. Além disso, desenvolver um sistema para localização de descargas parciais, utilizando sensores de detecção de ozônio, e adicionar a análise dados de concentração de ozônio, temperatura e umidade dentro dos hidrogeradores. Por fim, busca-se desenvolver um sistema de análise de dados para monitorar a evolução do estado de uma máquina a partir das medidas de descargas parciais.

REFERÊNCIAS

- ANDRADE, M. C.; PINTO, L. C. M. Segmentação de imagens com base em regiões utilizando histogramas e resíduos. In: **Anais do IV Workshop em Tratamento de Imagens**. NPDI/DCC/ICEx/UFGM: [s.n.], 2003. p. 54–61.
- ARDILA-REY, J. A. et al. Artificial generation of partial discharge sources through an algorithm based on deep convolutional generative adversarial networks. **IEEE Access**, v. 8, p. 24561–24575, 2020.
- CHEN, X. et al. Electrical treeing behavior at high temperature in xlpe cable insulation samples. **IEEE Transactions on Dielectrics and Electrical Insulation**, v. 22, p. 2841–2851, 2015.
- CRUZ, J. d. S. et al. Partial discharges monitoring for electric machines diagnosis: A review. **Energies**, v. 15, n. 21, 2022. ISSN 1996-1073. Disponível em: <<https://www.mdpi.com/1996-1073/15/21/7966>>.
- DION, M.; MARTOFF, C.; HOSACK, M. On the mechanism of townsend avalanche for negative molecular ions. **Astroparticle Physics**, v. 33, n. 4, p. 216–220, 2010.
- Empresa de Pesquisa Energética. **Matriz Energética e Elétrica**. Brasília, 2022. Acesso em: 10 dezembro 2023. Disponível em: <<https://www.epe.gov.br/pt/abcdenergia/matriz-energetica-e-eletrica>>.
- FRANCO, J. R.; RICHARDS, M.; SELTZER-GRANT, M. 8 years experience in large-scale remote partial discharge monitoring of hv motors in an oil and gas environment. In: **2019 IEEE Petroleum and Chemical Industry Committee Conference (PCIC)**. [S.l.: s.n.], 2019. p. 23–32.
- GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. MIT Press, 2016. Disponível em: <<http://www.deeplearningbook.org>>.
- GOODFELLOW, I. J. et al. **Generative Adversarial Networks**. 2014.
- HAMERLY, G.; ELKAN, C. Alternatives to the k-means algorithm that find better clusterings. In: **Proceedings of the Eleventh International Conference on Information and Knowledge Management**. New York, NY, USA: Association for Computing Machinery, 2002. (CIKM '02), p. 600–607. ISBN 1581134924. Disponível em: <<https://doi.org/10.1145/584792.584890>>.
- HASSAN, W. et al. Effects of environmental factors on partial discharge activity and estimation of insulation lifetime in electrical machines. **IEEE Access**, v. 8, p. 108491–108502, 2020. [CrossRef]. Disponível em: <<https://ieeexplore.ieee.org/document/9094443>>.
- HEBB, D. O. **The organization of behavior: a neuropsychological theory**. [S.l.]: John Wiley & Sons, 1949.
- HOTELLING, H. Relations between two sets of variates. **Biometrika**, [Oxford University Press, Biometrika Trust], v. 28, n. 3/4, p. 321–377, 1936. ISSN 00063444. Disponível em: <<http://www.jstor.org/stable/2333955>>.
- HUDON, C.; BÉLEC, M. Partial discharge signal interpretation for generator diagnostics. **Dielectrics and Electrical Insulation, IEEE Transactions on**, v. 12, p. 297 – 319, 05 2005.

IEC-60034-27-2. [S.l.], 2012.

IEC-60270. [S.l.], 2000.

IEEE. **IEEE Std 1434-2014 (Revision of IEEE Std 1434-2000); IEEE Guide for the Measurement of Partial Discharges in AC Electric Machinery**. Piscataway, NJ, USA, 2014. 1–89 p.

IN, B. **What Is a Fully Connected Layer?** 2022. <<https://builtin.com/machine-learning/fully-connected-layer>>. [Accessed: 2024].

Iris Power. **Partial Discharge (PD) Monitoring**. 2016. <<https://irispower.com/monitoring/partial-discharge-pd-monitoring/>>. Accessed: February 23, 2024.

JAVED, E. H.; LI, K. Online monitoring of partial discharge by measuring air decomposition by-products under low and high humidity. In: . [S.l.: s.n.], 2017.

JOLLIFFE, I. T.; CADIMA, J. Principal component analysis: a review and recent developments. **Phil. Trans. R. Soc. A.**, v. 374, p. 20150202, 2016. Disponível em: <<http://doi.org/10.1098/rsta.2015.0202>>.

KHAN, Q. et al. Partial discharge modeling of internal discharge in electrical machine stator winding. In: **Proceedings of the 2020 IEEE Kansas Power and Energy Conference (KPEC)**. Manhattan, KS, USA: [s.n.], 2020. p. 1–6. [CrossRef].

KINGMA, D. P.; BA, J. **Adam: A Method for Stochastic Optimization**. 2017.

LECUN, Y. et al. Gradient-based learning applied to document recognition. **Proceedings of the IEEE**, v. 86, n. 11, p. 2278–2324, 1998.

LLOYD, S. Least squares quantization in pcm. **IEEE Transactions on Information Theory**, v. 28, n. 2, p. 129–137, 1982.

LOPES, F. H. d. R. **Classificação de descargas parciais em hidrogeradores baseada em redes neurais profundas**. 2022. Trabalho de Conclusão de Curso (TCC).

LOPES, F. H. R. et al. Evaluation of transfer learning approaches for partial discharge classification in hydrogenerators. In: **2022 Workshop on Communication Networks and Power Systems (WCNPS)**. [S.l.: s.n.], 2022. p. 1–6.

MAATEN, L. van der; HINTON, G. Visualizing data using t-sne. **Journal of machine learning research**, v. 9, n. Nov, p. 2579–2605, 2008.

MACQUEEN, J. et al. Some methods for classification and analysis of multivariate observations. In: OAKLAND, CA, USA. **Proceedings of the fifth Berkeley symposium on mathematical statistics and probability**. [S.l.], 1967. v. 1, n. 14, p. 281–297.

MANTACH, S. et al. A convolutional neural network-based model for multi-source and single-source partial discharge pattern classification using only single-source training set. **Energies**, v. 14, n. 5, p. 1355, 2021.

MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. **The Bulletin of Mathematical Biophysics**, v. 5, n. 4, p. 115–133, 1943.

MITICHE, I. et al. Classification of partial discharge signals by combining adaptive local iterative filtering and entropy features. **Sensors**, v. 18, n. 2, 2018. ISSN 1424-8220. Disponível em: <<https://www.mdpi.com/1424-8220/18/2/406>>.

MOLLER, M. F. A scaled conjugate gradient algorithm for fast supervised learning. **Neural Networks**, v. 6, n. 4, p. 525–533, 1993.

NAILEN, R. L. Corona: What it does; how to detect it – part 1. **Electrical Apparatus Magazine**, January 1999.

OLIVEIRA, R. M. et al. A system based on artificial neural networks for automatic classification of hydro-generator stator windings partial discharges. **Journal of Microwaves, Optoelectronics and Electromagnetic Applications**, v. 16, n. 3, p. 628–645, 2017.

PEARSON, K. Liii. on lines and planes of closest fit to systems of points in space. **The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science**, Taylor & Francis, v. 2, n. 11, p. 559–572, 1901. Disponível em: <<https://doi.org/10.1080/14786440109462720>>.

RABUNAL, J. R.; DORADO, J. **Artificial neural networks in real-life applications**. [S.l.]: Idea Group Pub., 2006.

RADFORD, A.; METZ, L.; CHINTALA, S. Unsupervised representation learning with deep convolutional generative adversarial networks. 11 2015.

ROSENBLATT, F. The perceptron: A probabilistic model for information storage and organization in the brain. **Psychological Review**, v. 65, n. 6, p. 386–408, 1958.

SERRA, J. Image analysis and mathematical morphology. **Academic Press Inc.**, Elsevier, v. 1, n. 1, p. 1–22, 1982.

SOOMRO, I. A.; RAMDON, N. Study on different techniques of partial discharge (pd) detection in power transformers winding: simulation between paper and epoxy resin using uhf method. In: . [S.l.: s.n.], 2014.

STONE, G.; WU, R. Examples of stator winding insulation deterioration in new generators. In: **2009 IEEE 9th International Conference on the Properties and Applications of Dielectric Materials**. [S.l.: s.n.], 2009. p. 180–185.

STONE, G. C. et al. Impact of slot discharges and vibration sparking on stator winding life in large generators. **IEEE Electrical Insulation Magazine**, v. 24, n. 5, p. 14–21, 2008.

THAKOR, K.; SHAH, P. Partial discharge detection techniques: A review. **International Research Journal of Engineering and Technology (IRJET)**, v. 07, p. 885, 2020. ISSN 2395-0056. Impact Factor value: 7.529 | ISO 9001:2008 Certified Journal. Disponível em: <<https://www.irjet.net>>.

WANG, Y. et al. **Understanding How Dimension Reduction Tools Work: An Empirical Approach to Deciphering t-SNE, UMAP, TriMAP, and PaCMAP for Data Visualization**. 2021.

WERBOS, P. Backpropagation through time: what it does and how to do it. **Proceedings of the IEEE**, v. 78, n. 10, p. 1550–1560, 1990.

WIDROW, B. **Adaptive "adaline" neuron using chemical "memistors."** [S.l.: s.n.], 1960.

WITTEN, I. H.; FRANK, E.; HALL, M. A. **Data Mining: Practical Machine Learning Tools and Techniques**. [S.l.]: Elsevier, 2011.

YAACOB, M. M. et al. Partial discharge signal detection using ultra high frequency method in high voltage power equipments: a review. In: . [S.l.: s.n.], 2013.

YAACOB, M. M. et al. Review on partial discharge detection techniques related to high voltage power equipment using different sensors. **Photonic Sensors**, v. 4, n. 4, p. 325–337, 2014.

ZHU, M.; GHODSI, A. Automatic dimensionality selection from the scree plot via the use of profile likelihood. **Computational Statistics & Data Analysis**, v. 51, n. 2, p. 918–930, nov. 2006. ISSN 01679473. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0167947305002343>>.

ZNIDARICH, M. Hydro generator high voltage stator windings: Part 1 – essential characteristics and degradation mechanisms. **Australian Journal of Electrical and Electronics Engineering**, v. 5, n. 1, p. 1–17, 2008.