

UNIVERSIDADE DO ESTADO DE SANTA CATARINA
CENTRO DE CIÊNCIAS TECNOLÓGICAS – CCT
DEPARTAMENTO DE ENGENHARIA ELÉTRICA – DEE
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA – PPGEEL

STÉFANO FRIZZO STEFENON

**CLASSIFICAÇÃO E PREVISÃO DE FALHAS EM ISOLADORES DE DISTRIBUIÇÃO
POR MEIO DO PROCESSAMENTO DE SINAL ULTRASSÔNICO**

JOINVILLE, SC

2021

STÉFANO FRIZZO STEFENON

**CLASSIFICAÇÃO E PREVISÃO DE FALHAS EM ISOLADORES DE DISTRIBUIÇÃO
POR MEIO DO PROCESSAMENTO DE SINAL ULTRASSÔNICO**

Tese apresentada ao Programa de Pós-Graduação
em Engenharia Elétrica, da Universidade do Estado
de Santa Catarina, como requisito parcial para
obtenção do título de Doutor em Engenharia
Elétrica.

Orientador: Prof. Dr. Ademir Nied

JOINVILLE, SC

2021

**Ficha catalográfica elaborada pelo programa de geração automática da
Biblioteca Setorial do CCT/UDESC,
com os dados fornecidos pelo(a) autor(a)**

Stefenon, Stéfano Frizzo
CLASSEFAÇÃO E PREVISÃO DE FALHAS EM
ISOLADORES DÉ DISTRIBUIÇÃO POR MEIO DO
PROCESSAMENTO DE SINAL ULTRASSÔNICO / Stéfano
Frizzo Stefenon. -- 2021.
150 p.

Orientador: Ademir Nied
Tese (doutorado) -- Universidade do Estado de Santa Catarina,
Centro de Ciências Tecnológicas, Programa de Pós-Graduação em
Engenharia Elétrica, Joinville, 2021.

1. Classificação de isoladores. 2. Máquina de aprendizagem
extrema. 3. inteligência artificial. 4. redes neurais artificiais. I. Nied,
Ademir. II. Universidade do Estado de Santa Catarina, Centro de
Ciências Tecnológicas, Programa de Pós-Graduação em Engenharia
Elétrica. III. Título.

**CLASSIFICAÇÃO E PREVISÃO DE FALHAS EM ISOLADORES DE
DISTRIBUIÇÃO POR MEIO DO PROCESSAMENTO DE SINAL ULTRASSÔNICO**

por

STÉFANO FRIZZO STEFENON

Esta tese foi julgada adequada para obtenção do título de

DOUTOR EM ENGENHARIA ELÉTRICA

Área de concentração em “Sistemas Eletroeletrônicos”
e aprovada em sua forma final pelo

**CURSO DE DOUTORADO EM ENGENHARIA ELÉTRICA
DO CENTRO DE CIÊNCIAS TECNOLÓGICAS DA
UNIVERSIDADE DO ESTADO DE SANTA CATARINA**

Banca Examinadora:

Prof. Dr. Ademir Nied
CCT/UDESC (Presidente/Orientador)

Prof. Dr. Luiz Henrique Meyer
FURB

Profa. Dra. Andreza Sartori
FURB

Prof. Dr. Roberto Zanetti Freire
PUC/PR

Prof. Dr. Douglas Wildgrube Bertol
CCT/UDESC

Joinville, SC, 26 de novembro de 2021.

Dedico este trabalho a minha esposa Anne Carolina Rodrigues Klaar que renunciou a seus sonhos para que o meu fosse possível. Me apoiando e me incentivando nesta trajetória.

AGRADECIMENTOS

Agradeço a minha família que me deu o apoio e suporte para que este trabalho fosse possível, além de toda a motivação necessária. Agradeço em especial a minha esposa Anne Carolina Rodrigues Klaar que renunciou a várias oportunidades para continuar ao meu lado nesta jornada. Agradeço a minha irmã Stéfani Frizzo Stefenon que tem nos dado muito orgulho pela sua garra e força de vontade em crescer profissionalmente, que meu deu suporte financeiro nas horas difíceis e sempre me apoiou, agradeço a minha mãe, Rosangela de Fátima Frizzo que sempre esteve torcendo pelos meus resultados.

Agradeço ao professor Ademir Nied, meu orientador, que aceitou darmos continuidade a minha pesquisa sobre isoladores e sempre esteve a disposição para ajudar tanto no desenvolvimento do trabalho como em orientação pessoal. Agradeço ao professor Kin-Choong Yow que supervisionou minha pesquisa como estudante internacional na *Faculty of Engineering and Applied Science, University of Regina*.

Agradeço ao professor Luiz Henrique Meyer que além da orientação do mestrado se tornou um parceiro para pesquisa científica e amigo, além dos professores Ivo Barbi e Sérgio Vidal Garcia Oliveira que são minha inspiração pessoal no meio acadêmico. Agradeço aos meus amigos Rafael Bartnik Grebogi, Marcelo Campos Silva, Roberto Zanetti Freire e Jonatas Policarpo Américo que foram além de amigos grandes incentivadores e motivadores ao desenvolvimento de minha produção científica.

Agradeço a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela bolsa de doutorado. Agradeço ao *Emerging Leaders in the Americas Program (ELAP)*, *Canadian Bureau for International Education (CBIE)*, Governo do Canadá, que providenciou bolsa de estudo como pesquisador visitante de pós-graduação na *University of Regina*, em Regina, Canadá.

“A percepção do desconhecido é a mais fascinante das experiências. O homem que não tem os olhos abertos para o misterioso passará pela vida sem ver nada.”

Albert Einstein

RESUMO

O presente trabalho apresenta um estudo sobre a aplicação de técnicas de inteligência artificial para analisar a resposta do sinal ultrassônico emitido por isoladores de distribuição do tipo pino de corpo cerâmico, a fim de automatizar a classificação das condições deste componente. Isoladores de distribuição são fundamentais para manter o sistema elétrico em funcionamento, e a identificação prévia de falha pode auxiliar as concessionárias a manter a continuidade do fornecimento de energia elétrica. O objetivo geral deste trabalho é avaliar o desempenho de algoritmos baseados em redes neurais artificiais para melhorar a identificação de condições adversas em isoladores. Para isso, foram avaliados os modelos *adaptive neuro fuzzy inference system*, *long short-term memory*, *group method of data handling*, *nonlinear autoregressive*, e *multilayer perceptron*. Os objetivos específicos deste trabalho estão relacionados com a avaliação de técnicas combinadas com a máquina de aprendizagem extrema para obtenção do algoritmo o mais rápido e preciso o quanto possível. Desta forma, é proposto nesta tese o modelo *optimized ensemble extreme learning machine* (OEN-ELM) para avaliação das condições dos isoladores. O algoritmo proposto é baseado na combinação de técnicas como a transformada wavelet e a *principal component analysis* para extração de características com o *particle swarm optimization*, para otimização do classificador *ensemble extreme learning machine*. A partir do modelo OEN-ELM é possível identificar falhas antes que estas ocorram e se tornem irreversíveis, mantendo assim o sistema elétrico em funcionamento. A metodologia utilizada neste trabalho é do tipo quantitativa. Serão analisados os resultados do processamento dos sinais de ultrassom que foram captados em isoladores energizados, sob diversas condições controladas em laboratório. A aplicação da otimização no algoritmo de máquina de aprendizagem extrema se mostrou uma solução promissora, resultando em uma precisão de até 98,02% para classificação das condições dos isoladores.

Palavras-chave: Classificação de isoladores, máquina de aprendizagem extrema, inteligência artificial, redes neurais artificiais.

ABSTRACT

This work presents a study on the application of artificial intelligence techniques to analyze the response of the ultrasonic signal emitted by pin type distribution insulators of ceramic material to automate the classification of these components' conditions. Distribution insulators are essential to keep the electrical system in operation. The prior identification of failure can help utilities to maintain the continuity of electricity supply. The general aim is to evaluate the performance of algorithms based on artificial neural networks to improve the identification of adverse conditions in insulators. For this, the adaptive neuro-fuzzy inference system, long short-term memory, group method of data handling, nonlinear autoregressive, and the multilayer perceptron models were evaluated. The specific objectives of this work are related to the evaluation of techniques combined with the extreme learning machine to obtain the algorithm as fast and accurate as possible. This way, in this thesis, the optimized ensemble extreme learning machine (OEN-ELM) model is proposed for evaluating the insulators conditions. The proposed algorithm is based on the combination of techniques such as wavelet transform and principal component analysis for feature extraction with particle swarm optimization, to optimize the ensemble extreme learning machine classifier. From the OEN-ELM model, it is possible to identify failures before they occur and become irreversible, thus keeping the electrical system in operation. The methodology used in this work is quantitative. The results of processing the ultrasound signals that were captured in energized insulators, under various laboratory-controlled conditions will be analyzed. The application of optimization in the extreme learning machine algorithm proved to be an efficient solution, resulting in accuracy up to 98.02% for the classification of insulator conditions.

Keywords: Classification of insulators, extreme learning machine, artificial intelligence, artificial neural networks.

LISTA DE ILUSTRAÇÕES

Figura 1 - Potencial elétrico em isolador do tipo pino em boas condições de uso.....	33
Figura 2 - Perfuração em isolador de compósito.....	34
Figura 3 - Trilhamento em isolador polimérico.....	35
Figura 4 - Isolador com superfície com contaminação incrustada.....	36
Figura 5 - Isolador parcialmente quebrado por provável vandalismo.....	37
Figura 6 - Inspeção do sistema de distribuição por meio do detector de ultrassom.....	41
Figura 7 - Inspeção do sistema de distribuição através da câmera termográfica.....	42
Figura 8 - Isolador sendo avaliado em laboratório por meio da câmera ultravioleta.....	43
Figura 9 - Transformada rápida de Fourier do sinal gerado pelo equipamento de ultrassom	47
Figura 10 - Árvore de pacotes wavelet.....	49
Figura 11 - Árvore ótima de pacotes wavelet.....	51
Figura 12 - Comparaçao do sinal original e sinal reconstruido utilizando WPT.....	51
Figura 13 - PCA aplicado sobre 4 condições de isoladores.....	53
Figura 14 - Método k -fold de validação cruzada para $k = 4$	54
Figura 15 - Perceptron de camada única com peso de entradas.....	56
Figura 16 - Estrutura de um Perceptron multicamadas.....	58
Figura 17 - Evolução da posição da partícula no método PSO.....	76
Figura 18 - Formatos para topologias de enxame: A) melhor global, B) topologia em anel, C) topologia circular, D) topologia em pirâmide, E) topologia Von Neumann.....	76
Figura 19 - Funções de ativação para redes neurais artificiais.....	78
Figura 20 - Estrutura do ANFIS para avaliação de séries temporais.....	82
Figura 21 - Fluxo de uma série temporal para LSTM.....	87
Figura 22 - Diagrama do fluxo de dados para LSTM.....	88
Figura 23 - Arquitetura do <i>group method of data handling</i>	91
Figura 24 - Isoladores de perfil corpo simples: A) limpo, B) contaminado, C) furo lateral, D) furo superior.....	96
Figura 25 - Isoladores de perfil corpo duplo: A) limpo, B) contaminado, C) furo superior, D) quebrado.....	98

Figura 26 - Configuração para análise em laboratório.....	98
Figura 27 - Desenho esquemático da análise em laboratório.....	99
Figura 28 - Equipamento de ultrassom utilizado em laboratório.....	100
Figura 29 - Fluxograma da configuração de laboratório.....	101
Figura 30 - Sinal original gerado pelo equipamento de ultrassom.....	102
Figura 31 - Fluxograma do algoritmo para treinamento da rede OEN-ELM.....	103
Figura 32 - Evolução do erro da rede Perceptron camada única.....	107
Figura 33 - Evolução do erro quadrático médio da rede MLP.....	109
Figura 34 - Avaliação da convergência do algoritmo em 50 experimentos.....	121
Figura 35 - Comparaçao do sinal original (objetivo) e previsão de sinal (resposta).....	123
Figura 36 - Regressão de rede para treinamento usando o GMDH.....	127
Figura 37 - Regressão de rede para teste para o GMDH.....	128
Figura 38 - Sinal com aplicação de WEC em janela de 480 pontos.....	130
Figura 39 - Resultado do teste final utilizando o WEC com o GMDH.....	131

LISTA DE TABELAS

Tabela 1 - Descrição da configuração de dados para as simulações.	104
Tabela 2 - Espectro harmônico do sinal do equipamento de ultrassom (ganho 0,5).....	106
Tabela 3 - Espectro harmônico do sinal do equipamento de ultrassom (ganho 1,0).....	107
Tabela 4 - Validação de resultados FFT da rede Perceptron camada única (10/250/0,5).	108
Tabela 5 - Treinamento Utilizando o Método BFGS quasi-Newton.....	110
Tabela 6 - Treinamento Utilizando o Método LM.	110
Tabela 7 - Treinamento Utilizando o Método Gradiente Descendente.	111
Tabela 8 - Treinamento Utilizando o Método Gradiente com <i>Momentum</i> e Taxa Adaptativa. ..	112
Tabela 9 - Treinamento com reinicializações de Powell-Beale.	112
Tabela 10 - Treinamento com atualizações de Polak-Ribiére.	113
Tabela 11 - Treinamento com atualizações de Fletcher-Reeves.	113
Tabela 12 - Treinamento do gradiente conjugado em escala.	114
Tabela 13 - Treinamento de retropropagação resiliente.	114
Tabela 14 - Avaliação da função de ativação sobre o treinamento.	115
Tabela 15 - Resultados do treinamento em relação ao número de neurônios ocultos.....	116
Tabela 16 - Resultados de precisão na classificação com FFT da rede MLP (10/250/1,0).....	116
Tabela 17 - Resultados de precisão na classificação com sinal bruto da rede MLP.	117
Tabela 18 - Resultado da precisão na classificação para o teste das amostras.	118
Tabela 19 - Tempo médio necessário para o treinamento da RNA.	118
Tabela 20 - Tempo médio necessário para o teste da RNA.	119
Tabela 21 - Detalhes da otimização do algoritmo (OEN-ELM) utilizando 2 <i>Fitness Functions</i> .120	
Tabela 22 - Resultado usando diferentes estruturas ANFIS.	122
Tabela 23 - Resultado usando diferentes números de clusters.	123
Tabela 24 - Resultado para treinamento do FCM <i>clustering</i>	124
Tabela 25 - Resultado para teste do FCM <i>clustering</i>	124
Tabela 26 - Resultado para validação do FCM <i>clustering</i>	125
Tabela 27 - Resultado estatístico para validação do FCM.	125
Tabela 28 - Precisão do método autorregressivo não linear.	126

Tabela 29 - Erro do método autorregressivo não linear.	126
Tabela 30 - Resultado para teste da rede usando o GMDH.	127
Tabela 31 - Resultados da variação no número de neurônios.	128
Tabela 32 - Resultados para variação no número de camadas.	129
Tabela 33 - Resultados para aplicação do WEC.....	130
Tabela 34 - Análise estatística utilizando o WEC com o GMDH.	131

LISTA DE ABREVIATURAS E SIGLAS

ADAM	<i>adaptive moment estimation</i> (estimativa adaptativa do momento)
ANFIS	<i>adaptive neuro fuzzy inference system</i> (sistema de inferência neuro-fuzzy adaptativo)
ANN, RNA	<i>artificial neural network</i> (rede neural artificial)
APSO	<i>adaptive particle swarm optimization</i> (optimização por enxame de partículas adaptativa)
ARIMA	auto-regressive integrated moving average (média móvel integrada auto-regressiva)
BFGS	método de otimização proposto por Broyden, Fletcher, Goldfarb e Shanno
BP	<i>backpropagation</i> (retropropagação)
BR	<i>bayesian regularization</i> (regularização bayesiana)
CELESC	Centrais Elétricas de Santa Catarina
CNN	<i>convolutional neural network</i> (rede neural convolucional)
CMNN	<i>cerebellar model neural network</i> (modelo de rede neural cerebelar)
DWT	<i>discrete wavelet transform</i> (transformada discreta wavelet)
EI-ELM	<i>enhanced incremental extreme learning machine</i> (máquina incremental de aprendizagem extrema melhorada)
ELM	<i>extreme learning machine</i> (máquina de aprendizagem extrema)
EN	<i>elastic net</i> (rede elástica)
EN-GMDH	<i>elastic net to group method of data handling</i> (rede elástica para o método de grupo de tratamento de dados)
EN-ELM, Ens- ELM, E ² LM	<i>ensemble extreme learning machine</i> (conjunto para máquina de aprendizagem extrema)
EOS-ELM	<i>ensemble online sequencial extreme learning machine</i> (conjunto para máquina sequencial de aprendizagem extrema online)
ER ² -ELM	<i>ensemble based reactivated regularization extreme learning machine</i> (conjunto para máquina de aprendizagem extrema baseado na regularização reativada)
FCM	fuzzy c-means clustering
FFT	<i>fast Fourier transform</i> (transformada rápida de Fourier)
FN	<i>false negative</i> (falso negativo)
FOS-ELM	<i>fuzziness-based online sequential extreme learning machine</i> (lógica fuzzy baseada na máquina sequencial de aprendizagem extrema online)

FP	<i>false positive</i> (falso positivo)
FLVQ	<i>fuzzy learning vector quantization</i> (quantização vetorial de aprendizagem fuzzy)
GMDH	<i>group method of data handling</i> (método de grupo de tratamento de dados)
HFWNN	hybrid fuzzy wavelet neural network (sistema híbrido fuzzy wavelet neural network)
I-ELM	<i>incremental extreme learning machine</i> (máquina incremental de aprendizagem extrema)
JCR	<i>journal citation reports</i> (relatórios de citações de periódicos)
LM	algoritmo de Levenberg-Marquardt
LMS	<i>least-mean-square</i> (mínimo quadrado médio)
LSE	<i>least square error</i> (mínimo erro quadrático)
LS-SVM	<i>least square support vector machine</i> (máquina de vetores de suporte de mínimos quadrados)
LSTM	<i>long short-term memory</i> (memória longa de curto prazo)
MAE	<i>mean absolute error</i> (erro médio absoluto)
MLP	<i>multilayer perceptron</i> (rede neural de múltiplas camadas)
MAPE	<i>mean absolute percentage error</i> (erro médio percentual absoluto)
MSE	<i>mean square error</i> (erro quadrático médio)
NAR	<i>nonlinear autoregressive</i> (autoregressivo não linear)
NARX	<i>nonlinear autoregressive exogenous model</i> (modelo exógeno autorregressivo não linear)
NIO	<i>nonlinear input-output</i> (entrada-saída não linear)
N1	nomenclatura para estrutura do tipo normal com uma cruzeta para rede convencional de distribuição no padrão CELESC
OEN-ELM	<i>optimized ensemble extreme learning machine</i> (aprendizado extremo de conjunto otimizado)
OS-ELM	<i>online sequential extreme learning machine</i> (máquina sequencial de aprendizagem extrema online)
OR-KGLRT	<i>online reduced kernel generalized likelihood ratio test</i> (teste de razão de verossimilhança generalizada de kernel reduzido online)
PCA	<i>principal component analysis</i> (análise de componente principal)
PCA-GMDH	<i>principal component analysis to group method of data handling</i> (análise de componente principal para o método de grupo de tratamento de dados)
PDs	<i>partial discharges</i> (descargas parciais)

PSO	<i>particle swarm optimization</i> (optimização por enxame de partículas)
PSVM	<i>proximal support vector machine</i> (máquina proporcional de vetores de suporte)
RBF	<i>radial basis function</i> (função de base radial)
ReLU	unidade linear retificada
RF	<i>rotation forest</i> (floresta rotacional)
RF-ELM	<i>rotation forest extreme learning machine</i> (máquina de aprendizagem extrema por floresta rotacional)
RI	rádio interferência
RMS	<i>root mean square</i> (raiz do valor quadrático médio)
RMSE	<i>root mean square error</i> (raiz do erro quadrático médio)
RMSProp	<i>root mean square propagation</i> (propagação de raiz quadrada média)
RNN	<i>recurrent neural network</i> (rede neural recorrente)
SCI	<i>science journal impact factor</i> (fator de impacto de periódico científico)
SCG	<i>scaled conjugate gradient</i> (gradiente conjugado em escala)
SGDM	<i>stochastic gradient descent with momentum</i> (descida gradiente estocástica com momentum)
SLFN	<i>single-hidden layer feedforward neural network</i> (rede neural feedforward de camada oculta única)
SLP	<i>single layer perceptron</i> (perceptron camada única)
SPSO	<i>accelerated particle swarm optimization</i> (optimização por enxame de partículas acelerado)
ST	série temporal
SVM	<i>support vector machine</i> (máquina de vetores de suporte)
TanH	tangente hiperbólica
TP	<i>true positive</i> (verdadeiro)
UV	<i>ultraviolet</i> (ultravioleta)
WEC	<i>wavelet energy coefficient</i> (coeficiente de energia wavelet)
WP	<i>wavelet packets</i> (pacotes wavelet)
WPC	<i>wavelet packet coefficients</i> (coeficientes de pacotes wavelet)
WPD	<i>wavelet packet decomposition</i> (decomposição de pacotes wavelet)
WPT	<i>wavelet packets transform</i> (transformada de pacotes wavelet)

LISTA DE SÍMBOLOS E DEFINIÇÕES

a	parâmetro de mudança de escala da transformada wavelet
b	parâmetro de translação da transformada wavelet
bi	bias
c	cluster
c_1, c_2	constantes de aceleração
d	valor em um ponto no passado de uma série temporal
e_i	polarização da camada oculta
ens	conjunto de partículas
$ensPar$	vetor das funções de ativação
f	portão de esquecimento do controle de reconfiguração de estado
f^T	transposta da função
f'	derivada da função
$fuzzy$	camada de saída fuzzy
g	célula candidata a adicionar informações
gP	previsão da melhor posição da partícula
h_t, c_t	saída e o estado da célula na etapa de tempo
i	valor da iteração
j	valor do neurônio
k	variável de iteração
$klin$	parâmetros lineares da função de ativação
knl	parâmetros não lineares da função de ativação
lP_q	melhor posição da partícula
m	objeto da iteração
$maxEns$	conjuntos de partículas necessários
n	amostra de um conjunto de dados
o	portão de saída que controla o nível de estado da célula
p	direção de descida
pos	posição de cada particular
pr	precisão
q	partícula de um enxame

r	peso sináptico recorrentes
re	recordação
r_1, r_2	números aleatórios
s	comprimento através de uma camada
t	tempo em uma série temporal
$TanH$	função de ativação tangente hiperbólica
vb	vetor de posição da partícula
w	peso sináptico de entrada
w_{out}	resultado do cálculo de um critério externo
x	valor de entrada da rede
x^j	valor de x com potência complexa
y_{R_j}	conjunto de treinamento
z	porta de entrada para controlar o nível de atualização do estado da célula
C	números complexos
H	matriz H de Moore-Penrose
I	matriz inversa
N	exemplo de treinamento
\tilde{N}	neurônios ocultos
Nt	números naturais
P	valor em um ponto no futuro de uma série temporal
P_j	número de conjuntos de teste
$P_k(\lambda)$	conjunto fuzzy
R	números reais
R^2	coeficiente de determinação
$ReLU$	função de ativação de unidade linear retificada
α	valor do passo
α_g	valor do passo da iteração anterior para à atual
β	peso sináptico de saída
β_1, β_2	taxa de decaimento da média móvel
γ	saída da rede neural

γt	trade-off entre precisão e recordação
$\delta, \xi, \tau, v, \chi$	função de ativação genérica
δ_c	função de ativação de célula em uma camada LSTM
δ_g	função de ativação de gate em uma camada LSTM
δ_{TAF}	função de ativação variável
ε	valor pequeno e positivo
ζ	fator de mapeamento da função de ativação
η	deslocamento vertical da função de ativação
θ	inclinação da função da rede
μ	fator do ângulo de inclinação da função de ativação
ς	parâmetro de posicionamento horizontal
σ	função de ativação sigmoide
φ	função de escala de transformada wavelet
ψ	função principal de transformada wavelet
ω	momento de inércia
ω_n	frequência angular
ϑ_q	velocidade da partícula
$\varrho(b)$	filtro passa-baixo em uma transformada wavelet
$\varpi(b)$	filtro passa-alto em uma transformada wavelet
Δ	unidades no tempo
\emptyset	curva da função
\varOmega	coeficiente de pacotes wavelet
$ x $	valor absoluto de x
$\ x\ $	norma de x
\in	pertence a um conjunto
\forall	quantificação universal (é verdadeiro para todos)
Σ	somatório
∇	gradiente
$\stackrel{\text{def}}{=}$	igual por definição

SUMÁRIO

1	INTRODUÇÃO.....	23
1.1	PROBLEMÁTICA	23
1.2	OBJETIVO GERAL	24
1.3	OBJETIVOS ESPECÍFICOS	24
1.4	MOTIVAÇÃO E JUSTIFICATIVA	24
1.5	ARTIGOS PUBLICADOS EM PERIÓDICOS	25
1.6	ORGANIZAÇÃO DO TRABALHO	31
2	TÉCNICAS DE INSPEÇÃO DE REDE ELÉTRICA	32
2.1	PRINCIPAIS FALHAS	33
2.1.1	Perfurações	34
2.1.2	Descarga em banda seca.....	34
2.1.3	Erosão e trilhamento	35
2.1.4	Contaminação	35
2.1.5	Isoladores quebrados.....	37
2.2	INSPEÇÃO DE CAMPO DA REDE DE DISTRIBUIÇÃO	38
2.2.1	Rádio interferência	39
2.2.2	Ultrassom.....	40
2.2.3	Termografia.....	41
2.2.4	Radiação ultravioleta.....	43
2.3	CONSIDERAÇÕES FINAIS DO CAPÍTULO.....	44
3	CLASSIFICAÇÃO DE PADRÕES.....	45
3.1	CRITÉRIOS DE AVALIAÇÃO DE DESEMPENHO PARA CLASSIFICAÇÃO	45
3.2	SEGMENTAÇÃO E EXTRAÇÃO DE CARACTERÍSTICAS	46
3.2.1	Espectro harmônico.....	46
3.2.2	Transformada wavelet.....	48
3.2.3	Redução das dimensões	52
3.3	VALIDAÇÃO CRUZADA	54
3.4	REDES NEURAIS ARTIFICIAIS	55

3.4.1	Adaline	56
3.4.2	Perceptron multicamadas	57
3.4.3	Máquina de aprendizagem extrema.....	61
3.5	OTIMIZAÇÃO	67
3.5.1	Gradiente	67
3.5.2	Newton	70
3.5.3	Levenberg-Marquardt	71
3.5.4	Estimativa do momento adaptativa	72
3.5.5	Otimização por enxame de partículas.....	74
3.6	FUNÇÃO DE ATIVAÇÃO.....	77
3.7	CONSIDERAÇÕES FINAIS DO CAPÍTULO	79
4	PREVISÃO DE SÉRIES TEMPORAIS.....	80
4.1	CRITÉRIOS DE AVALIAÇÃO DE DESEMPENHO PARA PREVISÃO	80
4.2	SISTEMA DE INFERÊNCIA NEURO-FUZZY ADAPTATIVO	81
4.3	SISTEMAS AUTOREGRESSIVOS NÃO LINEARES	86
4.4	MEMÓRIA LONGA DE CURTO PRAZO	87
4.5	GRUPO DE TRATAMENTO DE DADOS	90
4.6	CONSIDERAÇÕES FINAIS DO CAPÍTULO	95
5	OBTENÇÃO DE DADOS E MÉTODO PROPOSTO	96
5.1	AMOSTRAS.....	96
5.2	CONFIGURAÇÃO DE LABORATÓRIO.....	98
5.2.1	Calibração do equipamento de ultrassom	99
5.2.2	Configuração de software	101
5.3	MÉTODO PROPOSTO PARA CLASSIFICAÇÃO	102
5.3.1	Configuração do conjunto.....	103
5.3.2	Metodologia adotada	105
5.4	CONSIDERAÇÕES FINAIS DO CAPÍTULO	105
6	ANÁLISE DOS RESULTADOS	106
6.1	RESPOSTA DO ESPECTRO HARMÔNICO	106
6.2	PERCEPTRON CAMADA ÚNICA	107

6.3	PERCEPTRON MULTIPLAS CAMADAS	108
6.3.1	BFGS quasi-Newton	109
6.3.2	Levenberg-Marquardt	110
6.3.3	Gradiente descendente	111
6.3.4	Gradiente descendente com <i>momentum</i> e taxa adaptativa	111
6.3.5	Variações do método gradiente	112
6.3.6	Retropropagação resiliente	114
6.3.7	Função de ativação	115
6.3.8	Resultados do teste da rede MLP usando a FFT e o sinal sem tratamento.....	115
6.4	MÁQUINA DE APRENDIZAGEM EXTREMA OTIMIZADA	117
6.4.1	Avaliação da configuração do conjunto.....	117
6.4.2	Análise estatística.....	119
6.5	ANÁLISE DE RESULTADOS ANFIS	121
6.5.1	Avaliação da estrutura	121
6.5.2	Número de clusters	122
6.5.3	Utilização da transformada wavelet.....	123
6.6	AVALIAÇÃO DOS MÉTODOS AUTOREGRESSIVOS	125
6.7	ANÁLISE DO MÉTODO DE GRUPO DE TRATAMENTO DE DADOS	127
6.7.1	Número de neurônios e camadas.....	128
6.7.2	Utilização da transformada wavelet.....	129
6.8	CONSIDERAÇÕES FINAIS DO CAPÍTULO.....	131
7	CONCLUSÕES.....	132
7.1	TRABALHOS FUTUROS	133
	REFERÊNCIAS	135

1 INTRODUÇÃO

A energia elétrica se torna cada dia mais necessária no cotidiano da sociedade, possibilitando o desenvolvimento social e econômico, e trazendo conforto para as pessoas. Partindo deste pressuposto, manter o sistema elétrico em funcionamento com qualidade e continuidade é necessário. Desta forma, as concessionárias, que são responsáveis por transmitir e distribuir energia elétrica dentro de limites de qualidade aceitáveis, estabelecidos pelas agências regulamentadoras, têm procurado equipamentos e técnicas para antecipar a ocorrência de falhas e garantir a continuidade no fornecimento de energia elétrica. A substituição de todos os componentes que estão há muito tempo em uso não é opção viável, considerando que é possível realizar manutenções preventivas e preditivas na rede evitando possíveis interrupções de fornecimento de energia.

Para realizar a inspeção do sistema elétrico e localizar dispositivos que possam vir a falhar, a concessionária realiza inspeções periódicas em suas redes elétricas. Durante a inspeção são utilizados equipamentos de alta tecnologia, como o detector de ultrassom, que é um equipamento que gera um ruído audível quando identifica a presença de ruído ultrassônico. Apesar do equipamento de ultrassom ser eficiente na localização de falhas, é cada vez mais difícil se dispor de uma equipe específica em uma concessionária para realizar as inspeções no sistema elétrico, pois os equipamentos utilizados requerem pessoal com treinamento avançado. A automatização desta tarefa pode ser uma alternativa que facilitará o desenvolvimento da inspeção, além de ser uma forma mais rápida e precisa, que a desempenhada pelo ouvido humano.

1.1 PROBLEMÁTICA

A fim de manter o sistema de distribuição em funcionamento é necessário identificar componentes que podem estar danificados, para que estes sejam substituídos e não ocorram desligamentos. A identificação de falhas é uma tarefa difícil que exige experiência do operador, tendo em vista que os equipamentos utilizados para inspeção do sistema elétrico requerem interpretação das medições. Em algumas situações a diferença de sinal entre componentes em boas condições e componentes avariados pode ser pequena, dificultando a classificação da condição do isolador.

1.2 OBJETIVO GERAL

Realizar um estudo sobre algoritmos de inteligência artificial para avaliação das condições de isoladores a partir dos sinais do detector de ultrassom utilizado.

1.3 OBJETIVOS ESPECÍFICOS

- Executar uma avaliação sobre as capacidades de otimização da configuração dos parâmetros do modelo de uma rede neural artificial;
- Comparar algoritmos classificadores e de previsão de falhas, com base no mesmo conjunto de dados para avaliação de falhas em isoladores de distribuição;
- Otimizar o algoritmo máquina de aprendizado extremo aplicado a detecção de falhas.

1.4 MOTIVAÇÃO E JUSTIFICATIVA

A aplicação de inteligência artificial tem sido uma excelente alternativa para melhorar a análise de dados e automatizar tarefas. Termos como *machine learning*, *big data* e *deep learning* estão cada vez mais presentes no dia a dia do engenheiro. Conforme Kanellopoulos e Wilkinson (1997) as redes neurais artificiais (RNAs) têm sido utilizadas com sucesso para classificação de padrões em diversas áreas de estudo. No sistema elétrico, os isoladores são componentes importantes da rede que podem, mesmo operando, estar danificados. A aplicação de RNAs para classificação de isoladores e previsão de possíveis falhas é um assunto promissor e será apresentado neste trabalho (STEFENON *et al.*, 2020b).

Devido às características não lineares do sinal ultrassônico, algoritmos mais simples não são eficientes neste tipo de análise, considerando que os dados, geralmente, não são linearmente separáveis (STEFENON *et al.*, 2017). Dentre os algoritmos mais avançados de classificação destaca-se o *extreme learning machine* (ELM), proposto por Huang, Zhu e Siew (2005). O ELM será apresentado neste trabalho como uma solução para classificação de isoladores. Além deste algoritmo, algumas modificações e comparações com algoritmos clássicos serão apresentadas com o objetivo principal de avaliar o quanto uma técnica é melhor/pior que outra para este problema.

A combinação do ELM com técnicas baseadas em wavelet (MALANGE, 2010), *principal component analysis* (PCA) (HOSSEINPOUR; AGHBASHLO; TABATABAEI, 2018) e *particle swarm optimization* (PSO) (MIRANDA; FONSECA, 2002) é um assunto inédito para aplicação de detecção de falhas no sistema elétrico de potência. Desta forma, a combinação destas técnicas é nomeada neste trabalho como *optimized ensemble extreme learning machine* (OEN-ELM).

1.5 ARTIGOS PUBLICADOS EM PERIÓDICOS

Nesta seção são apresentados os artigos publicados em periódicos e que foram resultado da aplicação e comparação entre as técnicas abordadas nesta tese de doutorado. Para cada artigo publicado, foi incluída a descrição da contribuição realizada pelo trabalho e o propósito de sua aplicação. As métricas *journal citation reports* (JCR) da Web of Science Group (Clarivate) e *science journal impact factor* (SCI) baseada no Scopus (fator de impacto quadrienal) referentes ao ano de 2020 são apresentadas.

O projeto de doutorado iniciou no ano de 2018, no qual foram realizadas as disciplinas necessárias para aprimorar os conhecimentos sobre o sistema elétrico. Inicialmente os artigos desenvolvidos neste ano foram relacionados ao método de elementos finitos. Este método foi abordado neste trabalho por ser promissor para melhorar a compreensão dos efeitos do campo elétrico no sistema elétrico de potência.

- **STEFENON, S. F.; AMÉRICO, J. P.; MEYER, L. H.; GREBOGI, R. B.; NIED, A.** Analysis of the Electric Field in Porcelain Pin-Type Insulators via Finite Elements Software. **IEEE Latin America Transactions** (IEEE), v. 16, n. 10, p. 2505-2512, 2018. (SCI = **0,99**. JCR 2020 = **0,73**). Neste trabalho foi avaliada a contaminação de isoladores utilizados em redes de distribuição. Através do método de elementos finitos foram simuladas variações na condição da superfície do isolador com o objetivo de verificar o impacto da contaminação que se acumula sobre estes componentes.

- **CORSO, M. P.; STEFENON, S. F.; COUTO, V. F.; CABRAL, S. H. L.; NIED, A.** Evaluation of Methods for Electric Field Calculation in Transmission Lines. **IEEE Latin America Transactions** (IEEE), v. 16, n. 12, p. 2970-2976, 2018. (SCI = **0,99**. JCR 2020 = **0,73**). Neste artigo foram comparados os métodos utilizados para realizar o cálculo do campo elétrico em

comparação com o método de elementos finitos. Assim, foi possível validar a utilização do método para avaliação da influência do campo elétrico no sistema elétrico de potência.

No ano de 2019 foi dada sequência no aprimoramento dos conhecimentos relacionados a aplicação de inteligência artificial para avaliar o sistema elétrico. Durante este ano foram realizadas as primeiras avaliações envolvendo inteligência artificial para classificar as condições de isoladores de distribuição. Além disto, se deu início à otimização dos modelos para melhorar o design de equipamentos elétricos.

- **STEFENON, S. F.; NIED, A.** FEM Applied to Evaluation of the Influence of Electric Field on Design of the Stator Slots in PMSM. **IEEE Latin America Transactions** (IEEE), v. 17, n. 4, p. 590-596, 2019. (SCI = 0,99. JCR 2020 = 0,73). O foco deste artigo foi o melhoramento do projeto de equipamentos elétricos. O método de elementos finitos foi utilizado para avaliar a influência na alteração de parâmetros de projeto em relação a distribuição de campo elétrico.

- **LESSE, C. S.; STEFENON, S. F.; AMÉRICO, J. P.; CORSO, M. P.; NIED, A.** Avaliação de Falhas em Redes de Distribuição de Energia Elétrica Através da Corrente de Curto-Círcuito. **Revista de Engenharia e Tecnologia**, v. 11, n. 1, p. 127-139, 2019. Neste trabalho foi avaliada a possibilidade de se identificar uma falha em redes de distribuição através da corrente de curto-círcuito, que é registrada pelo sistema supervisório da concessionária de energia elétrica.

- **STEFENON, S. F.; SILVA, M. C.; BERTOL, D. W.; MEYER, L. H.; NIED, A.** Fault Diagnosis of Insulators from Ultrasound Detection Using Neural Networks. **Journal of Intelligent & Fuzzy Systems** (IOS Press), v. 37, n. 5, p. 6655-6664, 2019. (SCI = 1,78. JCR 2020 = 1,85). Este trabalho foi o primeiro artigo que envolveu a utilização do ultrassom para detecção de falhas no sistema elétrico de distribuição, que é o foco desta tese. Nesta análise foi utilizado o modelo clássico de inteligência artificial *multilayer perceptron* para classificar condições adversas em componentes retirados do sistema.

No ano de 2020 os resultados mais importantes desta tese foram encontrados. Os assuntos abordados nas análises foram em relação à otimização através do enxame de partículas e aplicação da transformada wavelet para redução do ruído. Diversos modelos foram testados a fim de melhorar a capacidade de previsão e classificação.

- **STEFENON, S. F.; SEMAN, L. O.; FURTADO NETO, C. S.; NIED, A.; SEGANFREDO, D. M.; LUZ, F. G.; SABINO, P. H.; GONZALEZ, J. T.; LEITHARDT, V. R. Q.** Electric Field Evaluation Using the Finite Element Method and Proxy Models for the Design of

Stator Slots in a Permanent Magnet Synchronous Motor. **Electronics** (MDPI), v. 9, n. 11, p. 1975, 2020. (SCI = **1,70**. JCR 2020 = **2,40**). Neste trabalho diversas estratégias foram avaliadas para melhoria do projeto de equipamentos elétricos a partir da análise do método de elementos finitos.

- STEFENON, S. F.; BRANCO, N. W.; NIED, A.; BERTOL, D. W.; FINARDI, E. C.; SARTORI, A.; MEYER, L. H.; GREBOGI, R. B. Analysis of Training Techniques of ANN for Classification of Insulators in Electrical Power Systems. **IET Generation Transmission & Distribution** (Wiley), v. 14, n. 8, p. 1591-1597, 2020. (SCI = **4,41**. JCR 2020 = **2,86**). Como a alteração de parâmetros resultou em variação dos resultados do modelo classificador, neste artigo foi avaliado o impacto da alteração da configuração da rede sobre a acurácia e a velocidade para a convergência.

- CABRAL, S. H. L.; BERTOLI, S. L.; MEDEIROS, A.; HILLESHEIM, C. R.; DE SOUZA, C. K.; STEFENON, S. F.; NIED, A.; LEITHARDT, V. R. Q.; GONZALEZ, G. V. Practical Aspects of the Skin Effect in Low Frequencies in Rectangular Conductors. **IEEE Access** (IEEE), v. 9, p. 49424-49433, 2021. (SCI = **4,93**. JCR 2020 = **3,37**). Neste artigo o efeito pelicular foi analisado com o objetivo de avaliar a influência do design dos condutores.

- STEFENON, S. F.; GREBOGI, R. B.; NIED, A.; MEYER, L. H.; FREIRE, R. Z. Optimized Ensemble Extreme Learning Machine for Classification of Electrical Insulators Conditions. **IEEE Transactions on Industrial Electronics** (IEEE), v. 67, n. 6, p. 5170-5178, 2020. (SCI = **10,49**. JCR 2020 = **8,24**). Este trabalho traz a combinação das técnicas mais avançadas, que foram encontradas durante esta pesquisa, para solucionar o problema de classificação de isoladores retirados de campo. Nesta avaliação se utilizou um modelo com baixo esforço computacional que resulta em um alto desempenho quando combinado com o otimizador de enxame de partículas.

- STEFENON, S. F.; RIBEIRO, M. H. D. M.; NIED, A.; MARIANI, V. C.; COELHO, L. S.; ROCHA, D. F. M.; GREBOGI, R. B.; RUANO, A. E. B. Wavelet Group Method of Data Handling for Fault Prediction in Electrical Power Insulators. **International Journal of Electrical Power & Energy Systems** (Elsevier), v. 123, p. 106269, 2020. (SCI = **5,98**. JCR 2020 = **4,63**). Como a aplicação da transformada wavelet foi consideravelmente promissora na análise de classificação, esta técnica foi utilizada para melhorar a capacidade de previsão de uma série temporal caótica e os resultados se mostraram superiores também para esta avaliação.

- STEFENON, S. F.; FREIRE, R. Z.; MEYER, L. H.; CORSO, M. P.; SARTORI, A.; NIED, A.; KLAAR, A. C. R.; YOW, K-C. Fault Detection in Insulators Based on Ultrasonic Signal Processing Using a Hybrid Deep Learning Technique. **IET Science Measurement & Technology** (Wiley), v. 14, no. 10, p. 953-961, 2020. (SCI = 2,46. JCR 2020 = 1,91). Tendo em vista o sucesso na utilização da transformada wavelet, o modelo LSTM foi aplicado para previsão da série temporal. Esta avaliação foi realizada utilizando este modelo por ser popular para este tipo de aplicação.
- STEFENON, S. F.; KASBURG, C.; NIED, A.; KLAAR, A. C. R.; FERREIRA, F. C. S.; BRANCO, N. W. Hybrid Deep Learning for Power Generation Forecasting in Active Solar Trackers. **IET Generation Transmission & Distribution** (Wiley), v. 14, n. 23, p. 5667-5674, 2020. (SCI = 4,41. JCR 2020 = 2,86). Aplicando o mesmo modelo que combina o LSTM com a transformada wavelet, a análise foi realizada para a previsão de uma série temporal de geração de energia por meio de painéis fotovoltaicos. Esta análise foi realizada para verificar se o modelo é eficiente para prever outros tipos de sinais.
- RIBEIRO, M. H. D. M.; STEFENON, S. F.; LIMA, J. D.; NIED, A.; MARIANI, V. C.; COELHO, L. S. Electricity Price Forecasting Based on Self-Adaptive Decomposition and Heterogeneous Ensemble Learning. **Energies** (MDPI), v. 13, n. 19, p. 5190, 2020. (SCI = 3,45. JCR 2020 = 3,00). Para um problema semelhante relacionado à questão de sistemas de energia, neste artigo foram comparados os modelos *ensemble* para previsão da série temporal.
- STEFENON, S. F.; FREIRE, R. Z.; COELHO, L. S.; MEYER, L. H.; GREBOGI, R. B.; BURATTO, W. G.; NIED, A. Electrical Insulator Fault Forecasting Based on a Wavelet Neuro-Fuzzy System. **Energies** (MDPI), v. 13, n. 2, p. 484, 2020. (SCI = 3,45. JCR 2020 = 3,00). A partir da utilização das redes neurais e sistemas de inferência fuzzy, este trabalho mostrou que combinando estas técnicas com a transformada wavelet se obtém um modelo promissor para predição.
- MUNIZ, R. N.; STEFENON, S. F.; BURATTO, W. G.; NIED, A.; MEYER, L. H.; FINARDI, E. C.; KUHL, R. M.; SA, J. A. S.; ROCHA, B. R. P. Tools for Measuring Energy Sustainability: A Comparative Review. **Energies** (MDPI), v. 13, n. 9, p. 2366, 2020. (SCI = 3,45. JCR 2020 = 3,00). Dando sequência na análise energética, de forma paralela foram avaliadas as métricas utilizadas para medir sustentabilidade.

No ano de 2021 foi dada sequência na análise da previsão de séries temporais utilizando os modelos que tiveram melhores resultados apresentados nos trabalhos anteriores. Neste ano foi avaliada a utilização de outros tipos de sinais de entrada para a classificação das condições dos isoladores, como o sinal ultrassônico sem pré-processamento e a imagem.

- STEFENON, S. F.; FURTADO NETO, C. S.; COELHO, T. S.; NIED, A.; YAMAGUCHI, C. K.; YOW, K-C. Particle Swarm Optimization for Design of Insulators of Distribution Power System Based on Finite Element Method. **Electrical Engineering** (Springer), aceito, 2021. (SCI = 2,13. JCR 2020 = 1,84). A partir das análises iniciais em relação à aplicação do método de elementos finitos para a melhoria do projeto de componentes do sistema elétrico, neste artigo o método de otimização de enxame de partículas foi aplicado para a melhoria do projeto do isolador.
- STEFENON, S. F.; KASBURG, C.; FREIRE, R. Z.; FERREIRA, F. C. S.; BERTOL, D. W.; NIED, A. Photovoltaic Power Forecasting Using Wavelet Neuro-Fuzzy for Active Solar Trackers. **Journal of Intelligent & Fuzzy Systems** (IOS Press), v. 40, n. 1, p. 1083-1096, 2021. (SCI = 1,78. JCR 2020 = 1,85). Este artigo foi uma extensão da análise sobre a geração de energia em painéis fotovoltaicos. Neste caso, a análise foi realizada utilizando a combinação da transformada wavelet com redes *neuro-fuzzy*, assim como já realizado para geração de energia.
- STEFENON, S. F.; RIBEIRO, M. H. D. M.; NIED, A.; MARIANI, V. C.; COELHO, L. S.; LEITHARDT, V. R. Q.; SILVA, L. A.; SEMAN, L. O. Hybrid Wavelet Stacking Ensemble Model for Insulators Contamination Forecasting. **IEEE Access** (IEEE), v. 9, p. 66387-66397, 2021. (SCI = 4,41. JCR 2020 = 3,37). Neste artigo foi avaliado o modelo *stacking ensemble* para previsão de uma série temporal caótica. Este modelo se destacou para a previsão de séries temporais por exigir menor esforço computacional, sendo promissor para análise em questão.
- SOPELSA NETO, N. F.; STEFENON, S. F.; MEYER, L. H.; BRUNS, R.; SEMAN, L. O.; GONZALEZ, G. V.; LEITHARDT, V. R. Q.; YOW, K-C. A Study of Multilayer Perceptron Networks Applied to Classification of Ceramic Insulators Using Ultrasound. **Applied Sciences** (MDPI), v. 11, n. 4, p. 1592, 2021. (SCI = 2,15. JCR 2020 = 2,68). Neste trabalho a análise foi realizada a partir de um equipamento de ultrassom que capta especificamente o sinal ultrassônico, sendo que o modelo clássico *multilayer perceptron* foi utilizado nesta avaliação.
- ITAJIBA, J. A.; VARNIER, C. A. C.; CABRAL, S. H. L.; STEFENON, S. F.; LEITHARDT, V. R. Q.; OVEJERO, R. G.; NIED, A.; YOW, K-C. Experimental Comparison of

Preferential vs. Common Delta Connections for the Star-Delta Starting of Induction Motors. **Energies** (MDPI), v. 14, no. 5, p. 1318, 2021. (SCI = 3,45. JCR 2020 = 3,00). Este trabalho foi uma avaliação paralela para variar a influência da forma de ligação de motores elétricos. A partir de equipamentos conectados no sistema elétrico com maior eficiência se obtém o menor consumo de energia.

Os trabalhos com data de publicação no ano de 2022 foram consequência das submissões realizadas nos anos anteriores. Estes trabalhos tiveram o foco de melhorar a avaliação das condições específicas dos modelos utilizados anteriormente.

- STEFENON, S. F.; SEMAN, L. O.; SOPELSA NETO, N. F.; MEYER, L. H.; NIED, A.; Yow, K-C. Echo State Network Applied for Classification of Medium Voltage Insulators. **International Journal of Electrical Power & Energy Systems** (Elsevier), vol. 134, p. 107336, 2022. (SCI = 5,98. JCR 2020 = 4,63). A definição de parâmetros da rede gerou grande influência nos resultados das análises dos modelos classificadores, por este motivo neste trabalho foi analisado um método que avalia os parâmetros de forma dinâmica e define a melhor configuração possível para a rede de forma automática.

- STEFENON, S. F.; RIBEIRO, M. H. D. M.; NIED, A.; YOW, K-C.; MARIANI, V. C.; COELHO, L. S., *et al.* Time Series Forecasting Using Ensemble Learning Methods for Emergency Prevention in Hydroelectricpower Plants with Dam. **Electric Power Systems Research** (Elsevier), vol. 202, p. 107584, 2022. (SCI = 4,41. JCR 2020 = 3,41). Considerando que a estrutura *ensemble* teve resultados promissores para a previsão de séries temporais, neste artigo diferentes formas de organização desta estrutura são avaliadas.

- STEFENON, S. F.; CORSO, M. P.; NIED, A.; PEREZ, F. L.; YOW, K-C.; GONZALEZ, G. V.; LEITHARDT, V. R. Q. Classification of Insulators Using Neural Network Based on Computer Vision. **IET Generation Transmission & Distribution** (Wiley), aceito, 2022. (SCI = 4,41. JCR 2020 = 2,86). Este trabalho foi resultado do estudo sobre visão computacional realizado durante o intercâmbio na *Faculty of Engineering and Applied Science, University of Regina* (Canada).

- FERNANDES, F.; STEFENON, S. F.; SEMAN, L. O.; NIED, A.; FERREIRA, F. C. S.; SUBTIL, M. C. M.; KLAAR, A. C. R.; LEITHARDT, V. R. Q. Long Short-Term Memory Stacking Model to Predict the Number of Cases and Deaths Caused by COVID-19. **Journal of Intelligent & Fuzzy Systems** (IOS Press), aceito, 2022. (SCI = 1,78. JCR 2020 = 1,85). Neste

artigo foi avaliado uma nova forma de organização das camadas em uma rede de camadas profundas para previsão de séries temporais.

- MEDEIROS, A.; SARTORI, A.; STEFENON, S. F.; MEYER, L. H.; NIED, A. Comparison of Artificial Intelligence Techniques to Failure Prediction in Contaminated Insulators Based on Leakage Current. **Journal of Intelligent & Fuzzy Systems** (IOS Press), aceito, 2022. (SCI = 1,78. JCR 2020 = 1,85). Este trabalho apresentou uma avaliação da variação do tipo do modelo para realização da previsão da série temporal, especialmente neste caso em relação a a corrente de fuga que pode sinalizar o aumento da contaminação em isoladores.

1.6 ORGANIZAÇÃO DO TRABALHO

O Capítulo 1 apresenta o problema sobre a identificação de falhas em isoladores de distribuição. Com base neste problema são definidos os objetivos deste trabalho, justificativa e motivação para realização desta pesquisa.

No Capítulo 2 são apresentadas as técnicas utilizadas para inspeção de redes de distribuição e localização de falhas em isoladores. Neste capítulo são descritos os principais equipamentos utilizados na inspeção do sistema elétrico, com foco no ultrassom.

O Capítulo 3 apresenta uma revisão de literatura sobre a aplicação de inteligência artificial para classificação de padrões. Esta descrição é apresentada de acordo com a evolução dos algoritmos utilizados para este fim. Ao fim deste capítulo, é realizada uma revisão sobre as capacidades de previsibilidade do sinal.

No Capítulo 4 são apresentadas as amostras utilizadas para análise em laboratório e as configurações necessárias para o desenvolvimento da análise. Além disso, neste capítulo são discutidos os métodos propostos e utilizados para este tipo de avaliação.

No Capítulo 5 todos os resultados serão comparados e discutidos pontualmente. Inicialmente são apresentados os resultados dos métodos clássicos baseados em Perceptron, e em seguida são apresentados os resultados da máquina de aprendizagem extrema.

O Capítulo 6 apresenta uma avaliação em relação a previsão de séries temporais para prognóstico de defeitos de isoladores contaminados. O Capítulo 7 apresenta as conclusões do trabalho desenvolvido e sugestões para trabalhos futuros.

2 TÉCNICAS DE INSPEÇÃO DE REDE ELÉTRICA

O sistema elétrico de distribuição convencional, geralmente opera em ambiente externo e desta forma, está exposto às condições do meio ambiente, como contaminação por poeira de ruas não pavimentadas, resíduos industriais e salinidade (GHOSH; CHATTERJEE; CHAKRAVORTI, 2018). Quando a superfície de um isolador tem suas propriedades alteradas devido ao acúmulo destes resíduos, esta superfície pode se tornar mais condutiva e com isto suscetível a falhas, que podem levar a desligamentos (PARK; MOTAI; YOON, 2017).

A localização de isoladores com possíveis avarias e a classificação do tipo de problema é difícil de ser realizada em uma inspeção do sistema elétrico. Quando ocorre um desligamento do sistema devido à descarga em um isolador, é difícil, para a equipe de campo, identificar o isolador defeituoso, devido à extensão dos ramais de distribuição. Além disso, na grande maioria das vezes, após o religamento do sistema, este volta a funcionar normalmente, até que as condições ambientais proporcionem o desenvolvimento de uma nova falha, no mesmo, ou em outro isolador. Desta forma, o sistema pode se tornar vulnerável e perder sua confiabilidade (NGUYEN; JENSSEN; ROVERSO, 2018).

Uma das formas de manifestações iniciais de falhas no sistema elétrico de potência são as descargas parciais. Estas descargas são de pequena intensidade e ocorrem em dielétricos sujeitos a um campo elétrico. As descargas parciais podem ocorrer de diversas formas, principalmente sendo manifestadas internamente ou de forma superficial (STEFENON; MEYER, 2015).

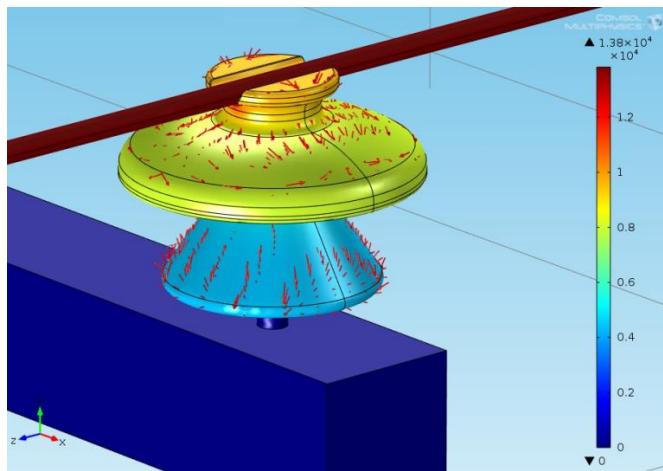
No sistema de transmissão, as descargas parciais são conhecidas como efeito corona, que ocorrem após a ionização do ar devido a alta intensidade do campo elétrico nas proximidades do condutor. O efeito corona pode gerar interferências eletromagnéticas, ruído sonoro, formação de ozônio, emissão de radiação ultravioleta e geração de calor nas proximidades das linhas de alta tensão (KUFFEL; ZAENGL; KUFFEL, 2000).

A combinação destes efeitos pode causar perdas de energia elétrica, expressas em watts por quilômetro de linha, que podem ser insignificantes em boas condições climáticas. Contudo, podem chegar a muitos quilowatts em más condições meteorológicas, como chuva ou alta umidade relativa do ar (CORSO *et al.*, 2018).

Como a distribuição de campo está associada a geometria do local, a variação do design deve ser analisada. Uma avaliação do projeto é importante para determinar uma possível condição

de vulnerabilidade (STEFENON *et al.*, 2018). Por exemplo, uma amarração de isolador realizada de forma inadequada, pode aumentar a influência do campo elétrico e acelerar o processo de degradação nas proximidades da rede. Uma avaliação da influência do campo elétrico sobre um isolador em perfeitas condições de utilização é apresentada na Figura 1, utilizando o método de elementos finitos.

Figura 1 - Potencial elétrico em isolador do tipo pino em boas condições de uso.



Fonte: Stefenon *et al.* (2018, p. 2509).

Através da simulação de elementos finitos é perceptível que um componente em boas condições de uso apresente uma distribuição linear de potencial elétrico e linhas de campo. Quando uma falha no componente é simulada, se acentuam as linhas de campo no local que indicam que existe maior chance de ocorrer uma descarga elétrica (STEFENON; NIED, 2019).

2.1 PRINCIPAIS FALHAS

Os isoladores podem apresentar danos físicos que podem comprometer sua capacidade de isolação e com isto, seu funcionamento. Grande parte das redes de distribuição ficam instaladas em ambiente externo e sem cobertura, com isto ficam expostas a variações ambientais que podem causar a degradação destes componentes (ZOELLER *et al.*, 2017).

A desgradação pode ocorrer ao longo de vários anos, podem ser encontradas rachaduras, corrosão de suportes metálicos, contaminação na superfície de componentes isolantes, e criação de fungos e liquens nos isoladores e estruturas de suporte da rede (STEFENON; MEYER, 2015).

2.1.1 Perfurações

Uma das falhas que podem ser encontradas em isoladores são as perfurações (ver Figura 2), que podem ser geradas por frentes de sobretensão muito rápidas, de ordem de décimos de microssegundos. Caso ocorram descargas de contorno, que aumentam a temperatura da superfície do isolador, a umidade dentro das fissuras pode se transformar em vapor, aumentando a pressão interna dos isoladores e podendo acarretar até em explosões (STEFENON *et al.*, 2017; FERRARINI *et al.*, 2016).

Figura 2 - Perfuração em isolador de compósito.



Fonte: Brito (2017, p. 9).

Esta falha é difícil de ser identificada em uma inspeção visual, porque os furos geralmente são pequenos e podem ficar escondidos sob o cabo e/ou amarrações. Este defeito pode causar falhas esporádicas no sistema elétrico de distribuição, levando a atuação da proteção, desligando o sistema (KALLA *et al.*, 2016).

2.1.2 Descarga em banda seca

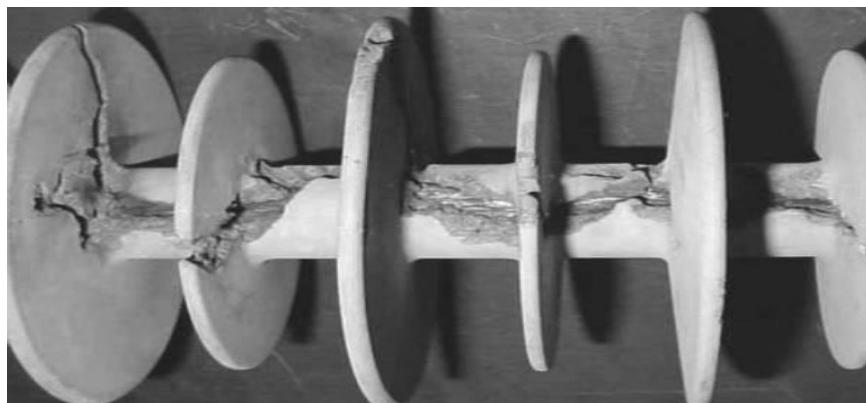
A descarga em banda seca ou arco de banda seca, ocorre na superfície de isoladores quando há corrente de fuga sobre uma superfície molhada. A corrente de fuga produz calor por efeito Joule, que pode fazer com que a água evapore da superfície molhada. O aquecimento da superfície cria partes secas que interrompem a corrente criando uma banda seca (YAMASHITA *et al.*, 2019).

Quando a banda seca está em um local com potencial elétrico alto, podem ocorrer arcos elétricos gerando assim descargas de banda seca. A ocorrência das descargas de banda seca na superfície de isoladores poliméricos pode causar a erosão e trilhamento da sua superfície, reduzindo assim a vida útil destes componentes (ALBANO *et al.*, 2016).

2.1.3 Erosão e trilhamento

O trilhamento é o processo que produz um caminho parcialmente condutivo através da deterioração da superfície isolante, conforme apresentado na Figura 3. Isoladores poliméricos suportam temperaturas inferiores que isoladores de vidro ou porcelana, desta forma as descargas elétricas causam uma limitação em sua utilização em ambientes externos com alta contaminação. A perda progressiva de material isolante pela ação de descargas elétricas é denominada erosão (KHAN; AMIN; AHMAD, 2017).

Figura 3 - Trilhamento em isolador polimérico.



Fonte: Costa, Ferreira e Neri (2007, p. 2).

2.1.4 Contaminação

Com o passar do tempo a contaminação fica incrustada (fortemente grudada) no isolador e mesmo a ação da chuva não é suficiente para remover estes resíduos da superfície isolante. A contaminação pode ser causada por resíduos industriais, salinidade ou poeira de ruas não

pavimentadas, que ocorre principalmente em áreas rurais (LI *et al.*, 2017). Esta condição pode ser visualizada na Figura 4.

Figura 4 - Isolador com superfície com contaminação incrustada.



Fonte: Elaborada pelo autor (2014).

Conforme apresentado por Stefenon *et al.* (2018), isoladores que possuem contaminação, tendem a ter maior condutividade superficial. Com o aumento da umidade a corrente de fuga pode atingir valores elevados, e desta forma podem ocorrer descargas disruptivas, que danificam os isoladores. A corrente de fuga pode ocorrer na superfície do isolador e internamente.

A contaminação de componentes isolantes não necessariamente representa uma falha na qual os isoladores devam ser substituídos, porém a permanência desta característica pode levar ao desenvolvimento de outras falhas irreversíveis, nas quais os sistemas isolantes perdem suas propriedades de isolamento e necessitam ser substituídos (MAVRIKAKIS *et al.*, 2017).

Um isolador de porcelana ou vidro pode suportar altas temperaturas, com isto geralmente pode suportar descargas disruptivas. Esta propriedade faz com que os isoladores cerâmicos sejam muito utilizados em redes de distribuição e transmissão de energia elétrica. No isolador de material polimérico, quando ocorrem descargas, pode ocorrer erosão, que faz com que o componente perca suas características isolantes (STEFENON; MEYER, 2015).

A contaminação varia de acordo com o local onde a rede está instalada. Em regiões litorâneas há níveis mais altos de salinidade; em regiões industrializadas pode ter maior presença de resíduos químicos na atmosfera. Em áreas rurais pode haver resíduos orgânicos e até resíduos químicos como agrotóxicos ou fertilizantes, além disso, a poeira do solo pode acumular sobre a superfície dos isoladores em locais que possuem ruas não pavimentadas (STEFENON; MEYER; MOLINA, 2014).

A contaminação de isoladores gera descargas de banda seca, que ocorrem de forma aleatória quando um filme aquoso gera uma pequena descarga que aquece o local. O aquecimento faz com que haja uma pequena área seca e devido a diferença de potencial ocorrem mais descargas neste local (STEFENON *et al.*, 2017).

No sul do Brasil, outro tipo de contaminação comum no sistema de distribuição é a presença de ninhos de pássaros próximas aos isoladores, especificamente o *Furnarius Rufus* (João de Barro). Estes pássaros constroem seus ninhos sobre as cruzetas e próximas aos condutores elétricos de média tensão. A aplicação de espaçadores, que não permitem que pássaros construam suas casas próximas aos isoladores, tem se mostrado uma ótima alternativa, visto que estes são de fácil instalação e não agridem o meio ambiente (STEFENON; MEYER, 2015).

2.1.5 Isoladores quebrados

Os isoladores instalados em locais abertos também ficam suscetíveis ao vandalismo. Tiros de armas de fogo podem destruir a superfície de isoladores vidro ou porcelana e com isto sua capacidade de isolação. Um isolador com sua saia superior quebrada pode ser visualizado na Figura 5 à direita. Nesta figura o local da saia quebrada está circulado em vermelho, de forma comparativa um isolador em boas condições de uso é apresentado na mesma figura à esquerda.

Figura 5 - Isolador parcialmente quebrado por provável vandalismo.



Fonte: Elaborada pelo autor (2014).

Para localização das falhas, é realizada uma inspeção nos isoladores da rede elétrica. Quando ocorre uma falha no sistema de distribuição de energia, a rede é percorrida do início ao fim procurando identificar, de forma visual, o local de ocorrência do defeito. Se o defeito não for localizado, é efetuado o religamento para restabelecimento da energia. É gerado um relatório de inspeção que é entregue ao centro de operação, que sistematiza a análise de defeitos para otimizar a manutenção da rede elétrica. A partir da inspeção é possível determinar quais componentes têm maior chance de desenvolver problemas, que podem prejudicar a continuidade no fornecimento de energia elétrica (VELÁSQUEZ; LARA, 2017).

2.2 INSPEÇÃO DE CAMPO DA REDE DE DISTRIBUIÇÃO

Para melhorar a confiabilidade no sistema de distribuição de energia elétrica, uma das boas práticas da concessionária é realizar inspeções periódicas em sua rede, tanto na parte ativa quanto da parte de isolamento. Na inspeção de campo do sistema elétrico, uma equipe percorre o ramal problemático, utilizando equipamentos e técnicas específicas como o ultrassom (MAHMOUDI; MOUSSAVI; SHARIFIAN, 2017), processamento de imagem (LV *et al.*, 2017), câmera ultravioleta (LI *et al.*, 2017) e infravermelha (SHEN-PEI *et al.*, 2017) visando encontrar indícios de falhas (WANG *et al.*, 2015).

Dentre os equipamentos de inspeção destaca-se o equipamento de ultrassom por ser direcionável, indicando qual isolador pode estar avariado. O equipamento gera um ruído audível, em que a interpretação depende da experiência do operador. Além disto, a habilidade do ouvido humano depende de pessoa para pessoa, e isto torna essa análise um tanto quanto subjetiva (STEFENON *et al.*, 2017).

A inspeção do sistema elétrico vem sendo aprimorada ao longo dos anos, através de técnicas (HUI *et al.*, 2018) e equipamentos (FREER; MANAHAN, 2018) que automatizam as tarefas do operador (NASLE, 2017). Além da utilização de equipamentos de alta tecnologia tem-se estudado a aplicação do processamento de dados nos sinais gerados pelos equipamentos de inspeção para identificação de falhas (EIBAR *et al.*, 2018). Sistemas adaptativos que utilizam lógica fuzzy (KHAN; MOKHTAR; HUSSIN, 2015) ou redes neurais (FU *et al.*, 2006; MAS'UD *et al.*, 2017c) podem aprimorar ainda mais a classificação das condições dos isoladores.

As descargas parciais, do inglês *partial discharges* (PDs), que ocorrem no sistema de energia elétrica, são indícios de que pode haver falhas de isolação (GÓMEZ *et al.*, 2018). As PDs podem ocorrer sem que haja uma falha que necessite de ação corretiva, como por exemplo, em componentes com contaminação sobre sua superfície, ou em amarrações de isoladores realizadas de forma inadequada. Entretanto, antes que ocorra uma falha disruptiva, as PDs podem ser um forte indício de que existe uma condição anormal no local (KHAN *et al.*, 2018). As descargas parciais geram ruído ultrassônico, rádio interferência e efeito corona que podem ser detectadas por equipamentos específicos.

Durante a inspeção de campo são utilizados diversos equipamentos para localização de falhas. Estes equipamentos têm suas particularidades e podem ser usados em conjunto durante uma inspeção elétrica (DONG *et al.*, 2018). Dentre os mais comumente utilizados destacam-se os detectores de rádio interferência e ultrassom, e as câmeras infravermelha e ultravioleta. Atualmente, esses equipamentos têm evoluído para facilitar a realização de inspeções, existindo equipamentos específicos que transformam a informação ultrassônica em imagem, auxiliando assim o entendimento da resposta do sinal.

2.2.1 Rádio interferência

A rádio interferência (RI) pode ser detectada por equipamentos semelhantes aos utilizados para sintonia de sinais de radiodifusão. Contudo, a capacidade de detecção e sensibilidade destes equipamentos é muito superior aos rádios comuns, tornando possível identificar falhas no sistema elétrico com grande acurácia (ALBARRACIN *et al.*, 2015).

As RIs podem ser divididas em interferências irradiadas, que se propagam através de campos eletromagnéticos e interferências conduzidas, que se propagam por meio de ligações físicas como cabos de alimentação ou dispositivos elétricos. No contexto deste trabalho, as fontes de RI de interesse são as oriundas de descargas parciais que, quando ocorrem no ar atmosférico, são chamadas de descargas corona (ROCHA *et al.*, 2019).

O medidor de RI é um receptor de rádio que funciona como um voltímetro e está ligado a uma antena apropriada para captação do sinal a ser medido. As conexões devem ser feitas com cabos apropriados, para o casamento adequado de impedâncias. O medidor possui um atenuador em sua entrada, para proteger o equipamento de níveis de sinais mais elevados. O sinal é filtrado

após passar pelo atenuador, assim são bloqueadas as frequências fora da faixa de interesse. Após passar pelo atenuador o sinal é alterado para uma frequência intermediária e amplificado para que fique visível na tela do equipamento. Por fim, o sinal passa por um circuito de amostragem e após isto é apresentado para o operador (STEFENON; MEYER, 2015).

A grande dificuldade na utilização desta técnica, é que normalmente a antena do equipamento que capta a rádio interferência não é direcional, o que dificulta a localização exata da falha. Geralmente a capacidade de detecção de falhas fica em torno de 30 a 50 metros do equipamento, dependendo das características do local e da presença de outros ruídos que podem dificultar a interpretação do sinal captado (MAS'UD *et al.*, 2016).

A técnica de detecção de RI é normalmente utilizada em conjunto com outras técnicas durante inspeções no sistema elétrico. Outras técnicas de detecção de PDs têm sido estudadas para localização de falhas no sistema elétrico e estas evoluíram juntamente com a capacidade do processamento digital dos sinais (MAS'UD *et al.*, 2017a).

2.2.2 Ultrassom

A inspeção por ultrassom é uma forma não destrutiva de localizar descargas parciais e pode ser utilizada mesmo na presença de interferência eletromagnética. Geralmente são utilizados sensores de 20 kHz até 100 kHz, em conjunto com amplificadores e filtros. Através da utilização do ultrassom é possível identificar defeitos de fabricação em isoladores, como rachaduras internas ou materiais estranhos, ou falhas que ficam ocultas após sua fabricação (STEFENON *et al.*, 2019).

O equipamento de ultrassom tem sido vastamente utilizado em inspeções do sistema elétrico de potência devido sua direcionalidade, que torna possível identificar pontos emissores de ultrassom nas redes elétricas, de forma segura, a partir do solo. A direcionalidade é obtida pelo uso de uma antena parabólica adequada, que converge os sinais obtidos para o microfone ultrassônico. O ruído ultrassônico emitido por um componente em falha é traduzido para um sinal audível que é interpretado pelo técnico de inspeção (STEFENON; MEYER; MOLINA, 2014).

A desvantagem na utilização deste equipamento é que se necessita de um operador com treinamento especializado e com experiência. Além disto, o ruído ultrassônico captado pelo equipamento pode vir de outras fontes, como motores a combustão ou elétricos com escovas, o que

torna necessário maior atenção do operador a fim de desconsiderar estes ruídos (STEFENON; MEYER; MOLINA, 2015).

Durante a inspeção a equipe especializada passa pelo ramal problemático e ao encontrar indícios de uma falha faz as medições, conforme pode ser visualizado na Figura 6. Nesta inspeção um técnico está avaliando o sinal ultrassônico de isoladores de distribuição em uma estrutura tipo *NI* que possui acúmulo de contaminação orgânica, como fungos e liquens.

Figura 6 - Inspeção do sistema de distribuição por meio do detector de ultrassom.



Fonte: Radar Engineers, *datasheet of model 251* (2021).

2.2.3 Termografia

A câmera termográfica funciona a partir da medição de emissividade de radiação eletromagnética na faixa infravermelha. A câmera termográfica é utilizada com frequência para inspeção de quadros elétricos de baixa tensão, considerando que a sobrecarga ou o mau contato pode causar pontos quentes nas instalações elétricas, que são facilmente identificados utilizando este equipamento (DONG *et al.*, 2018).

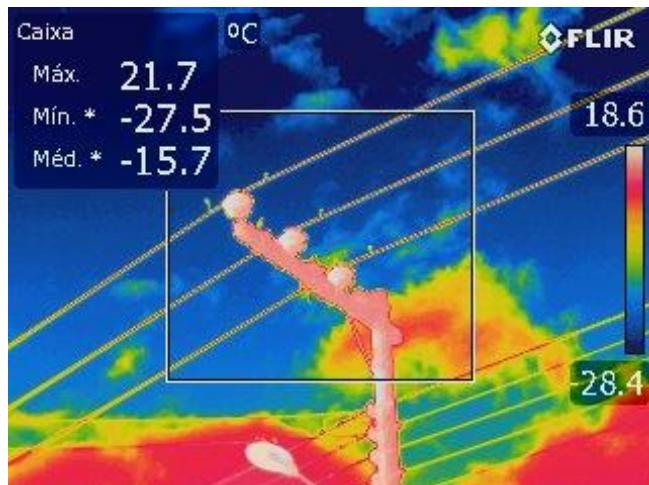
A câmera também pode ser utilizada para localizar problemas de conexão e emendas de condutores, o que pode facilitar a identificação de problemas em inspeções no sistema elétrico.

Para localização de pontos quentes são avaliadas frequências abaixo da capacidade de visualização do ser humano, que têm espectro correspondente ao infravermelho (FERRARINI *et al.*, 2016).

A câmera capta a radiação infravermelha que é emitida por objetos sob inspeção e transforma, em um sinal elétrico que é apresentado na tela do equipamento e mensurado através da variação de cor da imagem. Com base na imagem gerada pela câmera termográfica, pode ser realizada a prevenção de falhas com substituição de componentes ou manutenção em locais que apresentam variações de temperatura desconformes dos padrões (LIU *et al.*, 2017a).

Durante inspeções no sistema elétrico de distribuição utilizando a câmera termográfica, como pode ser visto na Figura 7, percebe-se dificuldade de localizar problemas em isoladores pois nem sempre esses problemas se manifestam em forma de calor. Este equipamento é utilizado em inspeções do sistema elétrico principalmente para avaliação de transformadores e conexões elétricas (STEFENON; MEYER, 2015).

Figura 7 - Inspeção do sistema de distribuição através da câmera termográfica.



Fonte: Elaborada pelo autor (2014).

O aquecimento dos componentes oculta as falhas e como estas não são facilmente diferenciáveis, a utilização da câmera termográfica se mostra pouco eficiente na identificação de defeitos em isoladores (STEFENON; MEYER; MOLINA, 2014).

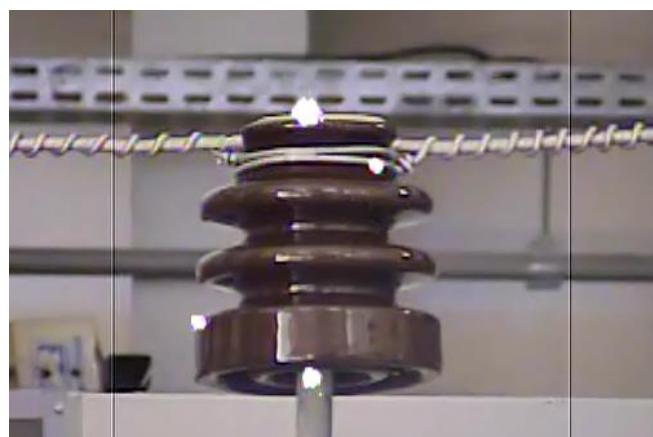
2.2.4 Radiação ultravioleta

A câmera ultravioleta (UV) que também é conhecida como câmera corona, têm a capacidade de detectar radiação na faixa ultravioleta. A câmera corona é baseada na detecção de radiação ultravioleta que é gerada pelo processo de ionização/de-ionização do ar em proximidades dos sistemas elétricos. Inicialmente as câmeras ultravioletas tinham sua imagem prejudicada por estarem suscetíveis a interferência de radiação da luz solar.

Atualmente são utilizadas técnicas para evitar a interferência da radiação solar, utilizando uma largura de onda entre 240 e 280 nm, tornando possível a visualização da luz ultravioleta advinda de outras fontes, como as provenientes de pequenas descargas elétricas, que tipicamente ocorrem onde há maior intensidade de campo elétrico. Utilizando os filtros adequados existem câmeras que podem ser utilizadas durante o dia e durante a noite (ZHANG; MENG, 2019).

A imagem é gerada no equipamento com uma sobreposição de pontos brancos sobre o espectro visível, indicando onde ocorreu a emissão da radiação ultravioleta naquele local (STEFENON; MEYER, 2015), como pode ser visto na Figura 8.

Figura 8 - Isolador sendo avaliado em laboratório por meio da câmera ultravioleta.



Fonte: Elaborada pelo autor (2016).

Existe interesse sobre o estudo do efeito corona pelas concessionárias, tendo em vista que esta pode causar a deterioração em equipamentos e com isto descontinuidade no fornecimento de energia elétrica. A detecção de corona através de UV é uma forma de inspeção indireta, pois não é necessário o contato com a rede. O equipamento pode ser utilizado em redes de distribuição de

média tensão ou no sistema de transmissão, sendo necessário utilizar lentes adequadas para cada necessidade.

Um ponto importante a ser mencionado em relação as câmeras é que um componente que possui alta intensidade de campo elétrico não necessariamente caracteriza uma falha. Os componentes do sistema elétrico são projetados para trabalhar em níveis altos de campo elétrico. Os eletrodos de sacrifício, por exemplo, são projetados para receber alta intensidade de campo elétrico e não representam defeitos (COETZER; LEUSCHNER, 2016).

2.3 CONSIDERAÇÕES FINAIS DO CAPÍTULO

Neste capítulo foram apresentados os equipamentos utilizados em inspeções do sistema elétrico de potência. Foi discutido sobre as falhas encontradas em ramais de distribuição que podem resultar em desligamentos da rede elétrica. Tipicamente o processo de inspeção é “humano”, e a automação da classificação do que o instrumento mede pode auxiliar na identificação de falhas. No próximo capítulo serão apresentados os modelos utilizados para classificação de padrões.

3 CLASSIFICAÇÃO DE PADRÕES

Neste capítulo será apresentado o estado da arte em relação a classificações de padrões. Inicialmente é discorrido sobre as técnicas utilizadas para extração de características do sinal avaliados nesta tese. Após a extração de características, são apresentados os modelos de inteligência artificial aplicados para classificação das condições dos isoladores analisados. Por fim, os métodos de otimização e configuração do modelo são comentados.

3.1 CRITÉRIOS DE AVALIAÇÃO DE DESEMPENHO PARA CLASSIFICAÇÃO

Existem várias maneiras de avaliar o desempenho de um algoritmo para classificação. Geralmente, as métricas são derivadas de informações obtidas da matriz de confusão. Nesta matriz se leva em consideração o *true negative* (*TN*), *true positive* (*TP*), *false negative* (*FN*) e *false positive* (*FP*). Com isso, a acurácia (*acc*) é calculada conforme:

$$acc = \frac{TP + TN}{TP + FN + FP + TN}. \quad (1)$$

A precisão (*pr*) é a fração de instâncias relevantes entre as instâncias recuperadas (SOKOLOVA, LAPALME, 2009), que pode ser estimada por:

$$pr = \frac{TP}{TP + FP}. \quad (2)$$

A recordação (*re*) também é conhecida como sensibilidade, sendo a fração de instâncias relevantes que foram recuperadas, calculada por:

$$re = \frac{TP}{TP + FN}. \quad (3)$$

A medida adotada para avaliação da precisão da classificação do algoritmo proposto é a *f-measure* que é definida como a média harmônica ponderada das métricas, chamada precisão e recordação (BALLABIO; GRISONI; TODESCHINI, 2018). A partir da precisão e da recuperação, o *f-measure* (*F_{acc}*) é dado pela equação:

$$F_{acc} = \frac{1 + \gamma t^2}{\frac{1}{pr} + \gamma t^2 \frac{1}{re}}, \quad (4)$$

em que $\gamma t \in [0, \infty]$ determina o *trade-off* entre *pr* e *re*. Neste trabalho, o *trade-off* utilizado foi $\gamma t = 2$.

3.2 SEGMENTAÇÃO E EXTRAÇÃO DE CARACTERÍSTICAS

Considerando que os dados que são irrelevantes para análise aumentam o tempo de treinamento e reduzem a precisão no modelo, a seleção correta dos dados de entrada torna a construção da rede eficiente e otimiza o modelo. Muitas abordagens têm sido estudadas para encontrar as características de entrada mais influentes como a *partial least squares regression*, *genetic algorithm correlation analysis*. Contudo, estes modelos não são adequados para aplicações mais complexas (JIA *et al.*, 2016).

Para divisão do sinal em segmentos menores, o método de segmentação *bottom-up* avalia quando o sinal começa a melhorar a aproximação e atende o critério de parada. Nesta tese, o critério de parada para o treinamento dos modelos analisados, foi que o algoritmo atingisse o erro mínimo desejado ou o máximo de iterações definido. A partir da segmentação, cada conjunto de informações que for extraído é considerado uma amostra de sinal.

3.2.1 Espectro harmônico

A transformada de Fourier converte um sinal $f(x)$ de seu domínio original (tempo) para o domínio da frequência f . Dentro desta perspectiva, a transformada de Fourier no domínio do tempo é obtida através de (5):

$$\hat{f}(\omega_n) = \int_{-\infty}^{\infty} f(t) e^{-j\omega_n t} dt, \quad (5)$$

em que,

$$\omega_n = 2\pi f. \quad (6)$$

Neste trabalho, a característica de frequência do sinal é utilizada para identificação de uma condição, seja esta normal ou em falha. O algoritmo utilizado para este fim é obtido a partir da *fast Fourier transform* (FFT), obtida por:

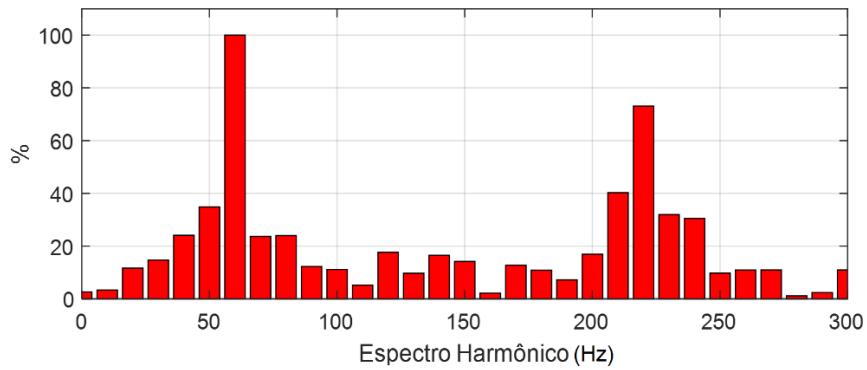
$$FFT = \frac{1}{n} \sum_{x=0}^{n-1} f(x) \cdot e^{-j2\pi x/n}. \quad (7)$$

Este tipo de análise técnica de compactação do sinal como *downsample* reduz a qualidade do sinal e compromete sua análise harmônica. A média móvel também não pode ser utilizada pois exclui as frequências mais elevadas e com isto reduz a qualidade do sinal. Para padronização de

análises é realizada uma ponderação em porcentagem em relação a frequência fundamental do sinal audível gerado pelo equipamento de ultrassom, conforme apresentado na Figura 9.

A fim de padronizar a análise e ter divisões de tempo homogêneas entre as comparações o sinal precisa ser segmentado. Neste trabalho, os dados foram segmentados a cada 4.800 pontos, correspondente a 0,1 segundo de aquisição, com uma taxa de aquisição de dados de 48 kHz. Optou-se por esta segmentação, em vista que dentro desta janela de tempo de aquisição, é possível avaliar toda a variação do sinal.

Figura 9 - Transformada rápida de Fourier do sinal gerado pelo equipamento de ultrassom



Fonte: Elaborada pelo autor (2021).

Para validação de resultados, foram avaliadas 250 respostas de FFT deslocadas 10 pontos de medição em cada sinal, em relação ao treinamento da rede. Devido a característica não linear do sinal, seu espectro harmônico também é variável. Entretanto, de forma global dentro de uma faixa, em relação a frequência fundamental, é possível realizar a análise destas relações de frequência para cada condição do isolador.

Polisetty, Jayaram e El-Hag (2018) realizaram uma análise para classificação de PDs baseadas na FFT a partir das frequências de 60 Hz, 120 Hz e 180 Hz. A partir da utilização da FFT se obteve uma precisão global de 91%. Analisando os trabalhos de Stefenon, Meyer e Molina (2014; 2015) os resultados se mostram promissores para análise de falha utilizando FFT.

Uma análise global do sinal produzido por isoladores de porcelana em estado de falha é apresentada no trabalho de Park, Motai e Yoon (2017). Este trabalho comprova que a técnica acústica com rede neural pode ser usada com alta precisão para diagnosticar o estado dos isoladores. A partir dessa premissa, um estudo de técnicas avançadas pode ser uma excelente maneira de automatizar inspeções elétricas.

3.2.2 Transformada wavelet

Para extração de características a transformada wavelet utiliza janelas deslizantes ao longo de intervalos de tempo e decompõe uma função com base em uma série temporal, através da decomposição em diferentes níveis de resolução e escala. A transformada wavelet pode ser definida como:

$$W(a, b) = \int_{-\infty}^{\infty} f(t) \frac{1}{\sqrt{|a|}} \psi * \left(\frac{t-b}{a} \right) dt, \quad (8)$$

sendo que a e b são parâmetros reais da função:

$$\psi_{a, b}(t) = \frac{1}{\sqrt{|a|}} \psi * \left(\frac{t-b}{a} \right). \quad (9)$$

Esta transformada pode ser reescrita como o produto interno das funções $f(t)$ e $\psi_{a, b}(t)$:

$$W(a, b) = \int_{-\infty}^{\infty} f(t) \psi_{a, b}(t) dt. \quad (10)$$

onde a função ψ é a função principal e a função $\psi_{a, b}(t)$ é a função secundária. O parâmetro b indica que a função foi transladada no eixo de t . O parâmetro a gera uma mudança de escala, que pode aumentar quando $a > 1$, ou diminuir quando $a < 1$. Uma das grandes vantagens na utilização da transformada wavelet é sua capacidade de manipular dados e extrair recursos ou características compactadas da função (MALANGE, 2010).

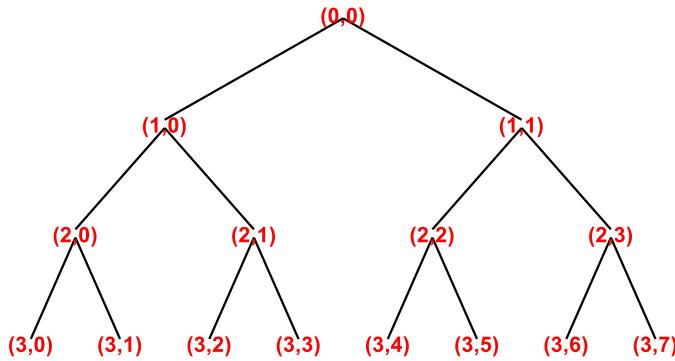
A partir da discretização da transformada wavelet, o *wavelet energy coefficient* (WEC) leva em consideração as informações de ambos os lados do espectro. O WEC é obtido após o sinal ser decomposto pela *discrete wavelet transform* (DWT) ou pelo *wavelet packets transform* (WPT). O WPT é uma generalização da transformada wavelet que é superior ao DWT por oferecer uma melhor resolução na região de alta frequência. Conforme Tiwari e Shukla (2002, *apud* MALANGE, 2010, p. 80), usando o DWT “é possível uma análise muito mais eficiente, pois se escolhe apenas um conjunto de escalas e posições para realização dos cálculos, sendo que as escalas e as posições são escolhidas em potências de dois”.

O WPT realiza em cada iteração uma nova decomposição de toda abordagem e detalha os coeficientes das iterações anteriores. Com isto o número de coeficientes obtidos depende de quantas decomposições foram realizadas. No WPT é realizado uma decomposição da wavelet

ortogonal em vários níveis de nós *wavelet packet* (WP) formando uma estrutura de árvore através da divisão dos coeficientes de aproximação (KUMAR *et al.*, 2018).

Após esta divisão, um vetor de coeficientes de aproximação e um vetor de coeficientes de detalhe é obtido. A informação perdida nas aproximações é capturada nestes coeficientes e um novo vetor é gerado, sendo que detalhes sucessivos não são reanalisados (LAINE; FAN, 1993). Inicialmente é realizado o *wavelet packet decomposition* (WPD), que retorna a uma árvore binária completa de pacotes, neste caso, para três níveis conforme Figura 10.

Figura 10 - Árvore de pacotes wavelet.



Fonte: Elaborada pelo autor (2021).

A função *wavelet packet* pode ser descrita da seguinte forma:

$$W_{a,b}^n(t) = 2^{a/2} W^n(2^a t - b), \quad (11)$$

onde a é o parâmetro de escala e b é o parâmetro de translação e n representa a oscilação do parâmetro. As duas primeiras funções WP para $n = 0$ e $n = 1$ são:

$$W_{0,0}^0(t) = \varphi(t), \quad (12)$$

$$W_{0,0}^1(t) = \psi(t), \quad (13)$$

representando, respectivamente, a função de escala e função principal. Outras funções para $n = 2, 3, \dots, N$ são definidas pelas relações:

$$W_{0,0}^{2n}(t) = \sqrt{2} \sum_b \varrho(b) W_{1,b}^n(2t - b), \quad (14)$$

$$W_{0,0}^{2n+1}(t) = \sqrt{2} \sum_b \varpi(b) W_{1,b}^n(2t - b), \quad (15)$$

onde $\varrho(b)$ é um filtro passa-baixo e $\varpi(b)$ é um filtro passa-alta, que estão associados à função de escala predefinida e à função principal. Os coeficientes WP $\Omega_a^n(k)$ podem ser calculados usando o produto da função $x(t)$ para cada função de WP, conforme a equação:

$$\Omega_a^n(k) = \int_{-\infty}^{\infty} x(t) W_{a,b}^n dt. \quad (16)$$

Cada coeficiente WP é específico com um nível de resolução de frequência, associado com um parâmetro de escala e um parâmetro de oscilação. A transformada wavelet decompõe somente os componentes de baixa frequência, enquanto o WPT decompõe todos os componentes em todos os níveis e nós. Desta forma, o WPT resulta em componentes de alta e baixa frequência, também chamados de aproximação passa-baixa e passa-alta (GAN *et al.*, 2016).

Para realização de classificação utilizando o WPT, alguns conceitos são necessários como entropia, energia e variação, e serão utilizados no cálculo dos coeficientes do WP. Dentre estes conceitos, a energia é utilizada para discriminar diferentes classes, em vista que contêm a informação da falha e sua flutuação em um componente específico que corresponde a ocorrência de uma falha específica. O sinal é decomposto em a níveis, resultando em subespaços ortogonais, e desta forma, a componente de frequência pode ser calculada a partir da seguinte equação:

$$E_a^n = \sum_b [\Omega_a^n(b)]^2. \quad (17)$$

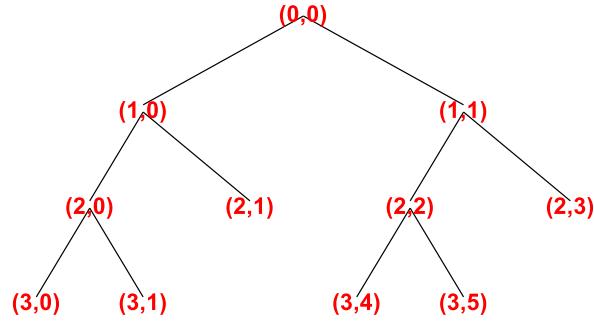
Para normalização da energia em cada banda de frequência, o percentual de distribuição do componente de energia é definido pela seguinte equação:

$$e_a^n = \frac{E_a^n}{E_{total}} = \frac{E_a^n}{\sum_{n=1}^a E_a^n}. \quad (18)$$

A energia relativa do vetor demonstra o desenvolvimento da frequência pelo tempo em subespaços da baixa para alta frequência. Cada nó pode ser considerado um recipiente e a mudança de padrão na distribuição formará o fluxo de energia, e esta distribuição de energia revela o padrão a ser identificado (GAN *et al.*, 2016).

Para que o melhor valor desta árvore de pacotes seja obtido é determinado o valor binário ótimo. Desta forma, é calculada uma sub-árvore ótima em relação a árvore inicial, conforme pode ser visualizado na Figura 11, com base no critério de entropia. Dependendo da aplicação, a árvore resultante pode ser muito menor que a árvore inicial. Esta técnica é aplicada considerando que o objetivo é encontrar um critério mínimo para que o algoritmo seja computável e eficiente (HAO, *et al.*, 2018).

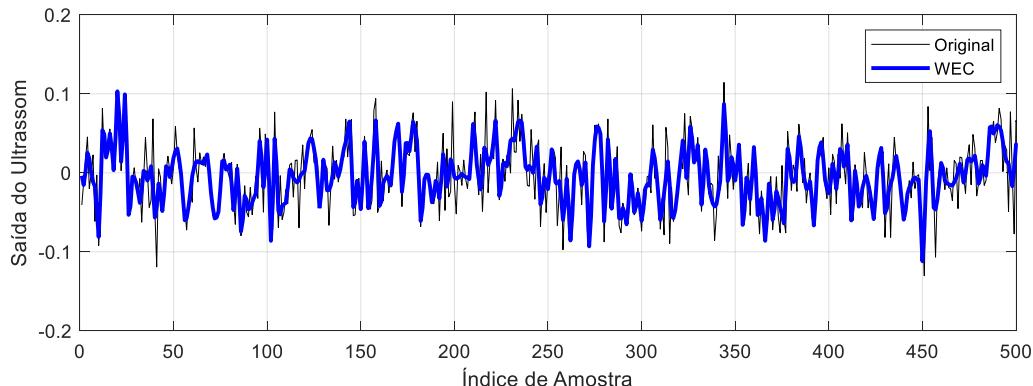
Figura 11 - Árvore ótima de pacotes wavelet.



Fonte: Elaborada pelo autor (2021).

Os coeficientes utilizando a *wavelet packet coefficients* (WPC) são armazenados a partir da entropia de Shannon, e por fim são reconstruídos para gerar o sinal filtrado, conforme apresentado na Figura 12. A utilização do WPT permite uma redução do sinal de entrada em relação ao tempo e frequência, contudo este método não é recomendado para análise de harmônica, pois reduz a capacidade de analisar as altas frequências (ZHAO, *et al.*, 2019).

Figura 12 - Comparação do sinal original e sinal reconstruído utilizando WPT.



Fonte: Elaborada pelo autor (2021).

O WPC pode ser utilizado para representar quantitativamente distribuições dos coeficientes obtidos pelo WPT baseado na entropia de Shannon, e combina ambas as características do sinal (GRAPS, 1995). Com isto, estas características podem ser utilizadas de forma eficaz como base para o treinamento do problema de classificação.

A entropia de Shannon descreve o conteúdo de energia de um sinal através da distribuição dos níveis de amplitude de um sinal. A definição da incerteza probabilística, proposta por Shannon (1948), utiliza o tratamento probabilístico através da função logarítmica,

$$T(p_1, \dots, p_n) = - \sum_{i=1}^n p_i \log(p_i), \quad (19)$$

em que p_i é a probabilidade de ocorrência de um evento i .

A equação pode ser vista como uma extensão da entropia estatística de Boltzmann-Gibbs, e desta forma, em um sistema macroscópico, leva-se em consideração sua parte microscópica. Com isto, a entropia passa a indicar a incerteza probabilística de uma distribuição de probabilidade, sendo denominada entropia de Shannon (SHANNON, 1948, *apud* NASCIMENTO, 2013, p. 1).

A extração de características é parte central de um processo de classificação bem sucedido. Em Lasalvia *et al.* (2015) a utilização do *wavelet energy coefficients* gera uma média de precisão de resultados de apenas de 82,14%, enquanto utilizando o *spectral subband centroid energy vectors* o valor de precisão é de 98,63%. Por este motivo, além da extração de características, métodos que reduzem as dimensões e selecionam os parâmetros ótimos melhoram as capacidades de classificação do modelo.

3.2.3 Redução das dimensões

Utilizando a *principal component analysis* (PCA), as características de entrada do algoritmo são extraídas com a equalização de histograma. PCA é um procedimento matemático que reduz variáveis não correlacionadas e encontra variáveis que representam a maior variabilidade possível nos dados. Através do método PCA, é realizada uma redução nas dimensões do conjunto de dados, mantendo as principais características comparadas aos conjuntos maiores, o que é uma maneira eficiente de reduzir o custo computacional (AÏT-SAHALIA, 2018).

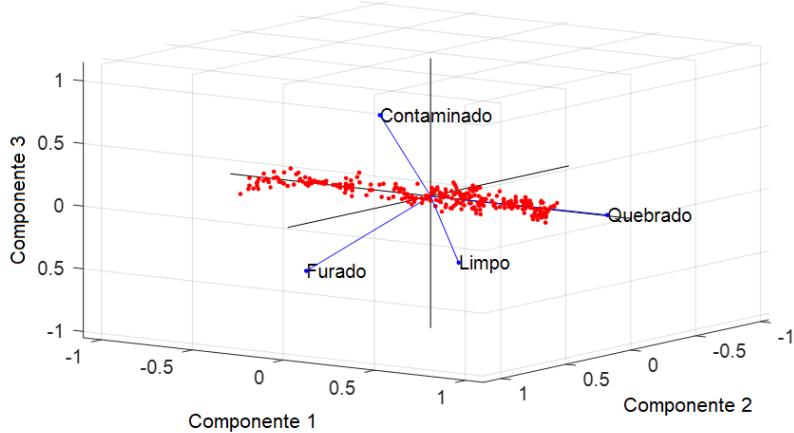
O PCA gera um número de variáveis igual ou inferior ao conjunto original. O primeiro componente desta transformação tem a maior variância possível e os componentes seguintes têm as máximas variâncias sendo ortogonais as componentes anteriores, sendo que os componentes principais são independentes somente se os dados foram normalmente distribuídos (GAJJAR; KULAHCI; PALAZOGLU, 2018).

O PCA pode ser aplicado para avaliar a variância dos dados e desta forma extrair as características fundamentais de um sinal. Para um conjunto de dados multivariados, o PCA pode ser visto como uma representação do sinal original de seu ponto mais informativo, de forma que sua dimensionalidade é reduzida. O cálculo do PCA é realizado pela decomposição em autovalores

de uma matriz de covariância, após centralizar os dados em cada atributo (FEBRERO-BANDE; GALEANO; GONZÁLEZ-MANTEIGA, 2017).

Os resultados do cálculo do PCA são obtidos em termos de pontuação de componentes e carregamento, sendo que o peso de cada variável deve ser multiplicado para se obter a pontuação do componente (BRO; SMILDE, 2014). Para exemplificar esta aplicação, a Figura 13 apresenta o PCA aplicado a 4 condições de isoladores (limpo, contaminado, furado e quebrado). As três componentes apresentadas neste caso têm 98,96% da variância total do sinal.

Figura 13 - PCA aplicado sobre 4 condições de isoladores.



Fonte: Elaborada pelo autor (2021).

Estas componentes representam as características extraídas do sinal. Para a resolução do PCA, um conjunto de dados é selecionado e os valores de média, variância e covariância são calculados (VALPOLA, 2015). A variância V_i de cada variável é definida como o desvio quadrático médio de seus valores em torno da média e pode ser expressa como:

$$V_i = \frac{1}{n-1} \sum_{m=1}^n (y_{im} - \bar{y}_i)^2. \quad (20)$$

A covariância C_{ik} representa o grau de correlação linear entre as variáveis, expresso por:

$$C_{ik} = \frac{1}{n-1} \sum_{m=1}^n (y_{im} - \bar{y}_i)(y_{km} - \bar{y}_k), \quad (21)$$

sendo y_{im} o valor da variável no objeto m , \bar{y}_i a média do valor da variável, y_{km} o valor da variável k no objeto m , \bar{y}_k a média do valor da variável k .

A partir da covariância, os autovalores e os correspondentes autovetores dessa matriz são calculados, com isto os componentes principais são obtidos baseados na variabilidade percentual

cumulativa. Os fatores com valores próprios mais altos são selecionados e calcula-se os indicadores de avaliação de cada fator. Por fim, as características influentes são escolhidas com base na avaliação de indicadores considerando os fatores mais significativos (LEDOIT; WOLF, 2015). Após a extração de características, a validação cruzada é utilizada para validar a capacidade de generalização do modelo com base em seu conjunto de dados.

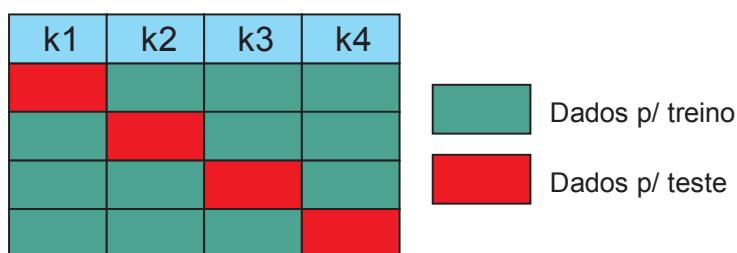
3.3 VALIDAÇÃO CRUZADA

A validação cruzada é amplamente utilizada para problemas de predição, onde se avalia uma previsão com base em um conjunto de dados distintos ao conjunto utilizado para o treinamento da rede. A aplicação desta técnica visa estimar o quanto preciso é o modelo, ou seja, avalia o desempenho do modelo com um novo conjunto de dados (GRIMM; MAZZA; DAVOUDZADEH, 2017).

Na validação cruzada os dados são particionados em subconjuntos mutuamente exclusivos, onde parte dos subconjuntos são usados para a estimação dos parâmetros do modelo (treinamento da rede) e os subconjuntos restantes são utilizados para validação ou teste do modelo (ROBERTS *et al.*, 2017).

O método *k-fold* de validação cruzada consiste em dividir o conjunto total de dados em *k* subconjuntos mutuamente exclusivos do mesmo tamanho. Um subconjunto é utilizado para teste e os conjuntos restantes são utilizados para estimação dos parâmetros. O procedimento de avaliação da acurácia é realizado *k* vezes de forma circular para o subconjunto de teste (BISGIN *et al.*, 2011). A Figura 14 mostra o esquema de particionamento e execução do método *k-fold* com *k* = 4.

Figura 14 - Método *k-fold* de validação cruzada para *k* = 4.



Fonte: Elaborada pelo autor (2021).

A partir da extração de características e validação do conjunto de dados, são utilizados modelos classificadores baseados em redes neurais artificiais. Estes modelos são empregados com o intuito de determinar quais isoladores estão avariados e precisão ser retirados do sistema elétrico de potência.

3.4 REDES NEURAIS ARTIFICIAIS

A identificação de padrões e utilização de algoritmos com implementação matemática vem avançando com o passar dos anos, desde quando McCulloch e Pitts (1943) desenvolveram um modelo matemático de um neurônio biológico que revolucionou este tipo de análise. Este modelo computacional é inspirado no sistema nervoso central de animais, que é capaz de reconhecer padrões. Em seu trabalho “*the general and logical theory of automata*”, Von Neumann (1951) aplicou a técnica e conseguiu simular um organismo vivo que foi um grande avanço na época.

Com o intuito de melhorar as RNAs, Hebb (1957) criou uma variação no peso das entradas dos neurônios que melhora a eficiência do algoritmo. Esta técnica ficou conhecida como regra de aprendizagem de Hebb ou aprendizagem Hebbiana. Partindo destes princípios surgiu a inteligência artificial que tenta simular o comportamento de um ser humano (MINSKY, 1958).

Rosenblatt (1958) criou o sistema Perceptron, que foi baseado na necessidade da máquina ter capacidade de interpretar informações sobre o mundo físico, de forma que a informação seja lembrada e o comportamento seja reconhecido. Em teoria criou-se um sistema nervoso hipotético.

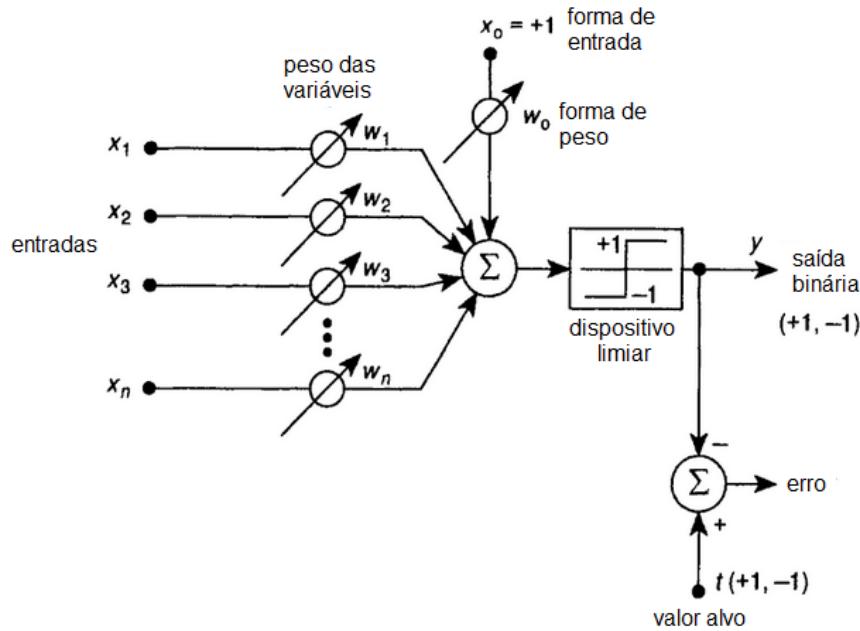
Na rede Perceptron é possível prever curvas de aprendizado, onde a compreensão do sistema cognitivo é baseada em uma abordagem estatística quantitativa, podendo classificar padrões linearmente separáveis. Esta técnica é um tipo de rede neural que funciona como classificador linear, mapeando uma entrada x para um valor de saída através de uma matriz conforme equação (22), para saída unipolar.

$$f(\text{net}_j) = f\left(\sum_i x_i \cdot w_{ij} + \theta_i\right) = \begin{cases} 1 & \text{net}_j > 0, \\ 0 & \text{net}_j \leq 0. \end{cases} \quad (22)$$

Em (22) w_{ij} é um vetor de peso real de cada entrada, $x_i \cdot w_{ij}$ é um produto escalar de uma soma com os pesos e θ_i é a inclinação, ou seja, o *bias* que não depende do valor de entrada. Com isto o somatório é a combinação linear das entradas ponderada pelos pesos (ROSENBLATT, 1958), conforme pode ser visualizado na Figura 15, em que x_0 representa o *bias*.

Minsky e Papert (2017) mostraram que inicialmente a técnica de Perceptron não era capaz de detectar paridade, conectividade e simetria, com isto a rede não seria capaz de resolver problemas “difíceis”. Foi provado o teorema mostrando que assim como uma camada de entrada se torna grande suficiente para apresentar classes geométricas importantes, o número de nós do Perceptron também se torna casualmente grande (GARSON, 1998).

Figura 15 - Perceptron de camada única com peso de entradas.



Fonte: Traduzido de Tarassenko (1998, p. 10).

Para Tarassenko (1998) o sucesso na utilização da rede neural depende de quanto conhecimento sobre o problema pode ser utilizado para realizar o treinamento da rede. Um estudo detalhado sobre as técnicas de aprendizagem da máquina e formas de organização e memória foi apresentado por Stracuzzi (2006) em seu trabalho “*scalable knowledge acquisition through cumulative learning and memory organization*”.

3.4.1 Adaline

O *adaptive linear element* (ADALINE) é uma generalização do algoritmo de treinamento do Perceptron. Este procedimento se dá pelo cálculo do *least-mean-square* (LMS) também

conhecido como a regra delta. A regra delta usa um custo ou função de erro com base nas diferenças do erro para ajustar os pesos (KRÖSE; VAN DER SMAGT, 1996).

O LMS contém os valores de todos os pesos que minimizam a função de erro por um método chamado gradiente descendente, que será explicada ao decorrer deste trabalho. A idéia é fazer uma mudança no peso proporcional ao negativo da derivada do erro, conforme medido no padrão atual em relação a cada peso (XU, 1993).

A principal diferença da regra de treinamento do Perceptron é a forma como a saída do sistema é usada na regra de aprendizado. O Perceptron gera diversas retas para separar linearmente os dados, enquanto através do Adaline é encontrada apenas uma reta ótima para esta divisão. Desta forma o ADALINE é menos suscetível a ruídos (KRÖSE; VAN DER SMAGT, 1996).

Rumelhart, Hinton e Williams (1985) propuseram um algoritmo com novo procedimento de aprendizagem através de retropropagação, que será explicado em detalhes na próxima seção. Através deste algoritmo, é possível treinar redes alimentadas com múltiplas camadas. O algoritmo é baseado na correção do erro e pode ser visto como generalização do algoritmo LMS, que serve como base para filtragem adaptativa linear.

3.4.2 Perceptron multicamadas

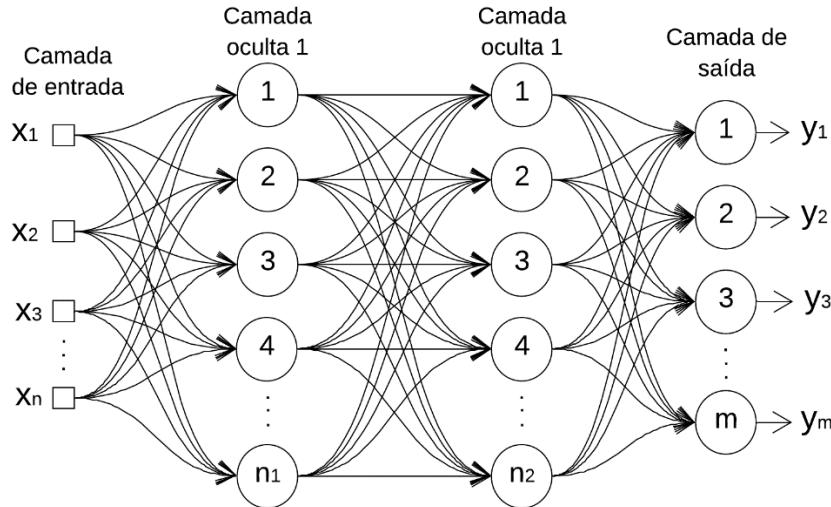
O *multilayer perceptron* (MLP) é dividido em camada de entrada, camada intermediária ou “oculta” e camada de saída, conforme pode ser visto na Figura 16. Os neurônios tipicamente têm como suas entradas os sinais das saídas precedentes, constituindo a característica de *feedforward* (WIDROW; LEHR, 1990).

A partir desta lógica, a rede MLP possui função de ativação não linear e com isto pode ser diferenciável em qualquer ponto. Sua implementação pode ser dada através da não linearidade sigmóide¹ (RUMELHART; HINTON; WILLIAMS, 1986). A escolha da função tem motivação biológica, tendo em vista que procura levar em conta a fase refratária de neurônios reais. Caso não houvesse existência de não linearidade a relação da entrada da rede poderia ser reduzida a um Perceptron camada única (CYBENKO, 1989).

¹As funções sigmóides são chamadas desta forma, pois seus gráficos são em forma de “s”.

O cálculo para o neurônio do MLP requer o conhecimento da derivada da função de ativação. Para esta existir é necessário que a função seja contínua, ou seja, a diferencialidade é a única exigência que a função de ativação deve satisfazer. Uma função de ativação não linear, deve ser continuamente diferenciável (NIED, 2007).

Figura 16 - Estrutura de um Perceptron multicamadas.



Fonte: Elaborada pelo autor (2021).

A rede MLP tem camadas de neurônios escondidos que capacitam a rede a aprender tarefas complexas extraindo progressivamente as características significativas dos padrões de entrada e possui um alto grau de conectividade determinado pelas suas sinapses (elos de conexões). É possível modificar a rede apenas alterando sua conectividade através da mudança do número de sinapses e de seus pesos (NIED, 2007).

A aprendizagem da rede MLP por retropropagação consiste em duas etapas, a programação e a reprogramação. Na programação é aplicado um padrão de entrada de ativação nos nós das camadas da rede e seu efeito se propaga através das camadas. Na última camada, um conjunto de saídas é produzido como resposta real da rede. Na etapa de programação todos os pesos sinápticos da rede são fixos, e na reprogramação (passo reverso), os pesos sinápticos são ajustados de acordo com a regra de supervisão (treinamento) (RUMELHART; HINTON; WILLIAMS, 1985).

O algoritmo *backpropagation* (BP) ajusta os pesos das conexões repetidamente para minimizar a diferença entre o vetor de saída real da rede \hat{y}_i e a saída desejada y_i e com isto produz um sinal de erro, apresentado em (23).

$$e_i(n) = \gamma_i(n) - \hat{\gamma}_i(n). \quad (23)$$

O sinal de erro é propagado para trás através da rede, contra a direção das conexões sinápticas, com isso os pesos são ajustados para que a resposta real se aproxime da resposta desejada de forma estatística. Desta forma, as unidades ocultas que não fazem parte da entrada ou da saída se tornam características importantes do algoritmo e a regularidade da tarefa é obtida pelas interações dessas unidades. A partir desta capacidade, a programação reversa supera métodos simples de convergência (RUMELHART; HINTON; WILLIAMS, 1986).

O algoritmo BP utiliza uma aproximação semelhante em raciocínio à usada na derivação do algoritmo LMS para minimizar erro, ou seja, os pesos são atualizados de padrão em padrão até formar uma estimativa da alteração real que resultaria na modificação dos pesos baseada na minimização da função custo. De forma resumida, se busca a direção da minimização do erro para o ajuste do peso na saída do neurônio.

O uso do algoritmo BP implica na determinação de um ganho ou taxa de aprendizagem para o treinamento. A escolha do ganho deve ser feita com cuidado. Se o ganho for grande o aprendizado pode ser rápido, mas pode se tornar instável. Por outro lado, se o ganho for pequeno, o tempo de treinamento pode se tornar muito longo. Este problema se torna ainda mais complexo se considerar inicializações diferentes para os pesos da rede e diferentes topologias de MLP. Um bom ganho para uma rede não é necessariamente bom para outra.

O algoritmo BP fornece uma aproximação para a trajetória no espaço dos pesos que pode ser calculada pelo método da descida mais íngrime. Assim, uma pequena taxa de aprendizagem causa pequenas variações nos pesos sinápticos e com isto torna a aprendizagem lenta. Uma forma de resolver este impasse é modificar a regra delta, incluindo o termo momento (RUMELHART; HINTON; WILLIAMS, 1986). A aprendizagem por retropropagação pode proceder de duas formas:

- De modo sequencial, a partir de uma iteração, constituída de N exemplos de treinamento, onde a atualização dos pesos é realizada após a apresentação de cada exemplo de treinamento. O modo sequencial é popular para uso na aprendizagem por retropropagação por ser simples de implementar e fornecer soluções efetivas para problemas complexos.
- Por lote, onde o ajuste dos pesos é realizado após a apresentação de todos os exemplos de treinamento que constituem uma iteração.

Uma característica do algoritmo BP baseado no gradiente é a possibilidade dele parar o processo de treinamento em mínimos locais onde os gradientes são próximos a zero ou nulos. Estas características são devido à formulação do método de gradiente descendente ou regra de cadeia (LEUNG; HAYKIN, 1991).

Uma função real no domínio de x tem um ponto x^* de mínimo global se:

$$f(x^*) \leq f(x) \quad \text{para } \forall x. \quad (24)$$

Um ponto x^* é um mínimo local (fraco) se existe uma vizinhança Nt de x^* tal que:

$$f(x^*) \leq f(x) \quad \text{para } x \in Nt. \quad (25)$$

Um ponto x^* é um mínimo local estrito (forte) se existe uma vizinhança Nt de x^* tal que:

$$f(x^*) < f(x) \quad \text{para } x \in Nt, \quad \text{com } x \neq x^*. \quad (26)$$

Em vista que o algoritmo BP possui uma convergência lenta que depende das condições iniciais e há possibilidade de o processo convergir para regiões de mínimos locais, foram desenvolvidas variações deste algoritmo com o objetivo de corrigir ou minimizar suas deficiências. Alguns algoritmos implementam técnicas de ajuste da taxa de aprendizagem, como no trabalho de Silva e Almeida (1990) que é utilizada uma taxa de aprendizagem para cada conexão. Com isto, a taxa é adaptada em funções do sinal de gradiente do erro no instante atual e no anterior.

A técnica MLP é aplicada no trabalho de Mas'ud *et al.* (2017b) e a precisão do algoritmo chega até 96%. No trabalho de Mas'ud *et al.* (2017c) a eficiência do algoritmo para localização de PDs chega a 98% em um problema equivalente. Em Mas'ud *et al.* (2017a) aplicando o algoritmo *ensemble-boosting* (EBA) a detecção de padrões em PDs chega a 97% e prova-se que este algoritmo é superior aos algoritmos clássicos baseados em inteligência artificial. Neste trabalho são usados 20%, 50% e 80% do total de dados adquiridos para o treinamento utilizando a função de ativação tangente hiperbólica.

Outra técnica para aprendizagem de redes neurais é o treinamento da rede MLP baseado no *extended kalman filter* que garante um aumento na taxa de convergência, pois fornece uma estimativa de variação mínima dos pontos de ligação da rede (IIGUNI; SAKAI; TOKUMARU, 1992). Porém na utilização desta técnica não há garantia de estabilidade numérica podendo degradar a convergência do aprendizado tornando a implementação prática deste algoritmo questionável.

Yu, Efe e Kaynak (2002) desenvolveram uma proposta para generalização do algoritmo BP, utilizando o método de Lyapunov² para análise da convergência dos pesos. Neste caso a convergência dos pesos para um mínimo global é obtida somente se os pesos iniciais estiverem próximos do mínimo global ou se a distribuição dos pesos permitir a convergência para o mínimo global.

Behera, Kumar e Patnaik (2006) propuseram um algoritmo para treinamento de redes MLP com taxa de aprendizagem adaptativa que garante a convergência global. Esta taxa de aprendizado também é obtida utilizando o método de Lyapunov. Os resultados desta análise mostraram que o algoritmo proposto pode ser mais rápido que o BP tradicional.

No trabalho de Junior *et al.* (2018) o estudo foi realizado com *machine learning* a partir do sinal ultrassônico, em que o classificador *support vector machine* (SVM) teve um resultado de 89%, enquanto o MLP de 85% para o mesmo sinal analisado. Isso mostra que é necessário realizar uma avaliação completa dos modelos de classificação para definir o melhor algoritmo a ser utilizado no problema avaliado.

3.4.3 Máquina de aprendizagem extrema

Para utilização da rede em aplicações reais é necessário que o conjunto de treinamento seja finito. Huang e Babri (1998) provaram que as *single-hidden layer feedforward neural networks* (SLFNs) com o máximo de \tilde{N} neurônios ocultos e com uma função de ativação não linear limitada, pode aprender amostras distintas com erro igual a zero. Com isto, mostraram que anteriormente a esta técnica não era viável escolher os pesos arbitrariamente para qualquer SLFNs.

A partir desta técnica é possível encontrar diretamente os pesos das SLFNs, contrário aos algoritmos de treinamento iterativos convencionais apresentados na literatura. Uma SLFN que possui \tilde{N} neurônios ocultos e considerando um conjunto de treinamento com amostras distintas (x_j , y_{R_j}), onde

$$x_j = [x_1, x_2, \dots, x_N] \quad (27)$$

pode ser modelada por (28).

² Método que permite investigar a estabilidade local de um sistema não-linear através do seu modelo linearizado.

$$y_{Rj} = \sum_{i=1}^{\tilde{N}} \beta_i g(w_i \cdot x_j + e_i), \quad (28)$$

$$j = 1, \dots, N.$$

onde β_i e w_i são os pesos de saída do i -ésimo neurônio e entrada respectivamente, e_i é a polarização da camada oculta e a função de ativação é denotada por $\delta(\cdot)$. A saída real é dada por y_{Rj} , onde

$$y_{Rj} = [y_{Rj1}, y_{Rj2}, \dots, y_{Rjm}]^T \in R^m. \quad (29)$$

A equação (28) pode ser compactada e rescrita como (30).

$$H\beta = Y_R \quad (30)$$

onde,

$$H \begin{pmatrix} w_1, \dots, w_N, \\ b_1, \dots, b_{\tilde{N}}, x_1, \dots, x_N \end{pmatrix} = \begin{bmatrix} \delta(w_1 \cdot x_1 + b_1) & \dots & \delta(w_{\tilde{N}} \cdot x_1 + b_{\tilde{N}}) \\ \vdots & \ddots & \vdots \\ \delta(w_1 \cdot x_N + b_1) & \dots & \delta(w_{\tilde{N}} \cdot x_N + b_{\tilde{N}}) \end{bmatrix}_{N \times \tilde{N}} \quad (31)$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_{\tilde{N}}^T \end{bmatrix}_{\tilde{N} \times m} \quad \text{e} \quad Y_R = \begin{bmatrix} y_{R1}^T \\ \vdots \\ y_{RN}^T \end{bmatrix}_{N \times m}.$$

O teorema apresentado por Huang, Zhu e Siew (2005) afirma que uma SLFN com uma função de ativação infinitamente diferenciável e \tilde{N} neurônios ocultos, pode aproximar essas N amostras usando pesos de entrada inicializados aleatoriamente determinando a matriz de saída. Este procedimento fornecerá uma aproximação dos valores desejados.

Significa que

$$\|H_{N \times \tilde{N}} \beta_{\tilde{N} \times m} - Y_{RN \times m}\| < \varepsilon \quad (32)$$

onde ε é um valor pequeno e positivo.

Para resolver o sistema linear (30), a menor solução padrão de mínimos quadrados é

$$\hat{\beta} = H^{\dagger} Y_R \quad (33)$$

onde (33) é uma das soluções de mínimos quadrados e H^{\dagger} denota o inverso generalizado da matriz H de Moore-Penrose.

A máquina de aprendizagem extrema, do inglês *extreme learning machine* (ELM), foi proposta por Huang, Zhu e Siew (2005) para SLFNs que escolhem aleatoriamente os pesos de entrada e determinam de forma analítica os pesos de saída. Teoricamente, o algoritmo tende a

proporcionar um melhor desempenho para a velocidade de aprendizado de uma RNA. Este algoritmo provou ser eficiente quando analisado em problemas de classificação, incluindo aplicações complexas, aprendendo muito mais rápido do que algoritmos tradicionais para redes neurais *feedforward*.

Conforme Huang, Zhu e Siew (2006), a pesquisa sobre a capacidade de aproximação do *feedforward* na rede neural tem dois aspectos importantes que são: a aproximação universal do conjunto de entrada compacto e a aproximação de um conjunto de amostras finito. O ELM pode ser utilizado com uma ou várias camadas de nós ocultos, onde estes nós podem ser atribuídos aleatoriamente e não atualizados ou herdados de valores anteriores sem serem alterados (HUANG; ZHU; SIEW, 2006). Geralmente os pesos da saída dos nós ocultos são aprendidos em uma única etapa, o que torna equivalente a aprender um modelo linear.

Devido à facilidade de implementação da *least square support vector machine* (LS-SVM) e da *proximal support vector machine* (PSVM) estas técnicas também são muito utilizadas para classificação binária. Porém, sua forma convencional não pode ser utilizada diretamente para regressão e classificação multilassses. Com isto, conforme apresentado por Huang *et al.* (2012) uma simplificação e unificação do LS-SVM e PSVM podem ser aplicadas para construção do ELM.

A partir disto, esta técnica pode ser utilizada para o mapeamento de recursos e aplicada diretamente na classificação multilassada. Se comparado aos SVMs o ELM possui restrições de otimização mais suaves, requerendo menor esforço computacional (HUANG *et al.*, 2012).

Conforme proposto por Huang, Ding e Zhou (2010), o ELM resulta em um melhor desempenho quando comparado com classes de regressão binária e um desempenho ainda melhor do que as técnicas tradicionais baseadas no gradiente. Novas teorias foram propostas por Huang e Chen (2007) para mostrar que as SLFNs com aditivo gerado aleatoriamente e *radial basis function* (RBF) com nós ocultos, podem funcionar como aproximadores universais e resultar no *incremental* ELM (I-ELM) que supera os algoritmos de aprendizagem mais populares.

O I-ELM gera nós ocultos aleatoriamente e calcula analiticamente os pesos da saída, porém não recalcula todos os pesos da saída quando um novo nó é adicionado. A taxa de convergência do I-ELM pode ser melhorada ao se recalcular o peso das saídas dos nós baseados no método *convex optimization* quando um novo nó é adicionado (HUANG; CHEN, 2007). Para escolher os nós de

forma aleatória é necessário ajustar os pesos de saída que ligam a camada oculta e a camada de saída.

Verificou-se que alguns nós ocultos podem desempenhar um papel menor na saída da rede e podem aumentar a complexidade da rede. Para evitar este problema, Huang e Chen (2008) desenvolveram um método ainda mais aprimorado para obter uma arquitetura mais compacta. Este método é denominado como *enhanced* para o I-ELM (EI-ELM). Nesta técnica, dentre os vários nós ocultos que são gerados aleatoriamente, o nó que tiver o maior erro residual será adicionado à rede existente. Deste modo, o método denominado EI-ELM funciona de forma generalizada para nós ocultos contínuos por partes.

No trabalho de Sun *et al.* (2008), uma interação entre vários ELMs foi aplicada com sucesso para prever o valor de vendas, no qual várias redes foram conectadas em paralelo e a média dos resultados foi usada como valor final para previsão de vendas. O resultado aplicado utilizando esta técnica mostrou ter melhor desempenho que apenas uma rede única. Liu e Wang (2010) mostram em seu trabalho que a validação cruzada no treinamento da rede pode melhorar sua estabilidade, além de que o algoritmo é robusto e eficiente para classificação.

Recentemente, o *ensemble* ELM (EN-ELM) tem sido vastamente aplicado para classificação (VINAY *et al.*, 2018). O algoritmo de EN-ELM foi inicialmente proposto por Hansen e Salamon (1990). Conforme apresentado por Cao *et al.* (2018b), algumas variações têm sido utilizadas com alta acurácia para classificação de padrões. Atualmente o EN-ELM também pode ser encontrado na literatura como E²LM (SONG *et al.*, 2018) ou Ens-ELM (KUMAR; SAVITHA; AL MAMUN, 2018).

O *ensemble* consiste em algumas redes únicas que podem ter adaptabilidade diferente para novos dados. Algumas redes podem se adaptar mais rapidamente que as outras na fase de treinamento, o que pode fazer que o conjunto de redes supere o problema em que somente uma rede é utilizada (HUANG; WANG; LAN, 2011).

Van Heeswijk *et al.* (2009) avaliaram os modelos de conjuntos para aplicação de previsão de séries temporais. Os estudos empíricos provaram que o EN-ELM tem um erro aceitável e uma boa adaptabilidade. Além disto, Van Heeswijk *et al.* (2011) utilizaram o EN-ELM para grandes aplicações em regressão e Zhai, Zhang e Wang (2017) e Zhai *et al.* (2018) utilizaram o EN-ELM com dados desequilibrados.

Zhou *et al.* (2017) aplicaram o EN-ELM utilizando *rotation forest* (RF) para classificação de imagem. Em seu trabalho, a árvore de decisão foi utilizada como base para o classificador original. Para obtenção dos componentes principais foi utilizado o PCA, assim foi proposto o método *ensemble* aplicado juntamente com o ELM e com RF (RF-ELM). Esta técnica tem sido amplamente explorada e comparada, como pode ser visto nos trabalhos de Abuassba *et al.* (2017) e Chai, Xu e Jia (2018).

No trabalho de Lv, Han e Qiu (2017) os algoritmos *bagging*, *adaboost*, *random forest*, *rotate forest* e ELM são utilizados e comparados com o *ensemble* ELM otimizado para resolução de uma classificação de imagem. Através de uma *stacked autoencoder* é adotada uma característica de segmentação, com isto é criada uma diversidade na base de classificação que melhora a precisão do algoritmo. A segmentação é característica importante para melhorar o desempenho do método e nesta tese está apresentada na seção 3.2.

A técnica ELM pode ser utilizada inclusive através de aprendizagem *online sequencial* (OS-ELM) que pode aprender os dados de forma unitária ou pedaço por pedaço com tamanho de bloco fixo ou variável (LIANG *et al.*, 2006). O trabalho de Lan, Soh e Huang (2009) mostra que o OS-ELM pode ser melhorado ainda mais, propondo um *ensemble online sequencial* ELM (EOS-ELM) baseado no OS-ELM, onde os resultados mostram uma melhora de estabilidade em relação ao OS-ELM original.

O trabalho de Cao *et al.* (2018a) também mostra uma nova forma de aplicação do OS-ELM na perspectiva da imprecisão, o *fuzziness-based online sequential* ELM (FOS-ELM). A partir desta técnica as novas amostras com dificuldade de alto rendimento são selecionadas e usadas em aprendizagem sequencial. Outras variações também podem ser encontradas na literatura como o *ensemble based reactivated regularization extreme learning machine* (ER²-ELM).

Algumas técnicas sofisticadas foram propostas para localização de falhas no sistema elétrico como o trabalho de Pan *et al.* (2018) que utilizam *deep learning network*, Cui e Tu (2015) a partir de decomposição e regressão da rede neural, além dos trabalhos de Ünal e Ekici (2017) e de Saini e Beniwal (2018) que aplicam o ELM para detecção e classificação de falhas no sistema elétrico. A aplicação do EN-ELM para classificação de condições no sistema elétrico, que é foco deste trabalho é mais rara. Neste contexto, Khamis *et al.* (2018) apresentaram um esquema para detecção rápida de eventos de ilhamento em uma microrrede.

A classificação das condições dos isoladores de redes elétricas é uma tarefa difícil (LASALVIA *et al.* 2015), e por conta disso, a aplicação de técnicas sofisticadas, como as apresentadas neste trabalho, se mostram bastante promissoras (ZHANG *et al.* 2018). Alguns trabalhos utilizando técnicas inovadoras relacionadas à localização de falhas em isoladores do sistema elétrico já foram explorados (MARAABA; AL-HAMOUZ; AL-DUWAISH, 2014; BELHOUCHET; BAYADI; BENDIB, 2016), no entanto, não utilizaram o EN-ELM que se destaca entre os algoritmos classificadores.

No trabalho de Belhouchet, Bayadi e Bendib (2016), as classificações foram baseadas em estruturas tradicionais de RNA. Pode-se observar que os resultados de aplicações anteriores têm uma baixa taxa de acertos. O estudo proposto por Cui e Tu (2015) mostrou que através da RNA foi possível melhorar a localização de falhas em linhas de transmissão, mas assim como outros trabalhos aqui mencionados, esse tipo de aplicação poderia ser melhorada, como mostrado por Ünal e Ekici (2017) onde o ELM foi aplicado com sucesso.

O trabalho apresentado por Pan *et al.* (2018) mostrou avanços na resolução do problema de identificação de falhas, através da rede de aprendizagem profunda com aprendizado de recursos em camadas, onde resultados consistentes foram relatados. No entanto, a validação deste método é realizada em comparação com SVMs clássicos.

Uma abordagem inicial para a aplicação de ELM e PSO foi realizada por Mišák *et al.* (2016) para identificar descargas parciais em redes isoladas de média tensão em ambiente externo. O trabalho mostra que a técnica é precisa para indicação de descargas parciais, que geralmente estão associadas a falhas nos componentes isolantes do sistema. No entanto, neste trabalho em específico não há comparação direta e clara entre componentes em falha e em boas condições, além de que a precisão do algoritmo é baixa para esta classificação.

O trabalho de Chen, Shang e Wu (2016) propõe o uso de ELM e PSO com transformada wavelet e PCA combinados para extrair a característica da informação das amostras. Embora o excelente resultado e a comparação sejam em relação às técnicas avançadas de ELM, foram realizadas poucas análises e variações da técnica para validação do algoritmo, o que não deixa claro sua robustez para análise de diferentes condições de falha.

3.5 OTIMIZAÇÃO

Para minimização do erro no processo de treinamento do modelo, métodos de otimização são utilizados. Dependendo do método de otimização, existe um maior esforço computacional. Desta forma, um método adequado deve ser utilizado para que torne possível processar as informações e treinar o modelo.

3.5.1 Gradiente

O gradiente é utilizado para encontrar um mínimo (local) de uma função de forma interativa em que cada passo que se realiza na direção negativa do gradiente corresponde à direção de maior decréscimo ou declive máximo (NOCEDAL; WRIGHT, 2006). Para isto, x_i é o valor atual e a pesquisa linear é dada pela direção de descida p_i que é o oposto do gradiente da função objetivo, conforme (34).

$$p_i = -[I]^{-1} \nabla f(x_i). \quad (34)$$

Após o cálculo da direção de descida o valor é atualizado conforme (35) em busca do mínimo.

$$x_{i+1} = x_i + \alpha_i p_i. \quad (35)$$

Neste método há um *trade-off* na escolha do passo α_i , pois este pode reduzir substancialmente $f(x_i)$ versus optar por utilizar o mínimo de esforço nesta escolha. Com isto, uma solução para este problema é realizar uma busca linear inexata que reduza $f(x_i)$ adequadamente a um mínimo custo computacional, descobrindo um intervalo que possua valores desejáveis para o passo α_i (FINARDI, 2017). Para avaliar α_i algumas condições são impostas como o decréscimo suficiente (Condição de Armijo) apresentada em (36).

$$f(x_i + \alpha_i p_i) \leq f(x_i) + c_1 \alpha_i \nabla f_i^T p_i \quad e \quad c_1 \in (0,1). \quad (36)$$

Nesta condição não há garantia que o algoritmo tenha um processo de resolução razoável, pois pode ser atendido por valores muito pequenos de passo. A condição de curvatura também é utilizada para melhorar a forma de encontrar um passo que seja aceitável.

$$\nabla f(x_i + \alpha p_i)^T p_i \geq c_2 \nabla f_i^T p_i \quad e \quad c_2 \in (c_1, 1). \quad (37)$$

Se as condições de decréscimo suficiente e curvatura forem respeitadas se atende à condição de Wolfe (FINARDI, 2017). Aumentando mais as restrições de curvatura para atendimento das condições fortes de Wolfe, tem-se:

$$f(x_i + \alpha_i p_i) \leq f(x_i) + c_1 \alpha_i \nabla f_i^T p_i \quad e \quad |\nabla f(x_i + \alpha p_i)^T p_i| \geq |c_2 \nabla f_i^T p_i|. \quad (38)$$

O valor de α pode ser calculado através de *backtracking line search* ou interpolação quadrática, por exemplo. No *backtracking line search* a atualização de α é dada por um valor fixo informado pelo operador, no entanto, uma escolha ruim resulta em uma convergência demorada. A interpolação quadrática é baseada em valores da função e de suas derivadas (FINARDI, 2017). Reescrevendo as variáveis para a curva da função em relação ao passo se obtém:

$$\emptyset(\alpha_i) = f(x_i + \alpha_i p_i) \quad \emptyset(0) = f(x_i) \emptyset'(0) = \nabla f_i^T p_i. \quad (39)$$

E substituindo as variáveis de (39) para a condição (36) tem-se:

$$\emptyset(\alpha_i) \leq \emptyset(0) + c_1 \alpha_i \emptyset'(0). \quad (40)$$

Com isto constrói-se uma sequência que determina α_i que atenda a condição (40). Partindo de um passo que tenha um intervalo aceitável utilizando interpolação quadrática e então calcula-se o novo passo que atenda as condições, sem ser muito pequeno.

$$\emptyset_m(\alpha) = \left(\frac{\emptyset(\alpha_i) - \emptyset(0) - \alpha_i \emptyset'(0)}{\alpha_i^2} \right) \alpha^2 + \emptyset'(0) \alpha + \emptyset'(0). \quad (41)$$

Cujo mínimo é dado por (42).

$$\alpha_{i+1} = -\frac{\emptyset'(0) \alpha_1^2}{2[\emptyset(\alpha_1) - \emptyset(0) - \emptyset'(0) \alpha_1]}. \quad (42)$$

Se as condições não forem atendidas deve-se calcular novamente um novo alfa, entre α_i e α_{i+1} e realiza-se nova interpolação quadrática. Após encontrar um novo alfa que atende a condição (40), é avaliada a condição (37) em módulo. Com isto são atendidas as condições fortes de Wolfe apresentadas em (38) e o algoritmo segue em busca do mínimo até que o erro, ou seja, a diferença do valor anterior para o valor atual seja aceitável (FINARDI, 2017).

Para aplicações em redes neurais, na descida do gradiente o otimizador é estocástico porque as atualizações de parâmetro calculadas são uma estimativa ruidosa da atualização de parâmetro que resultaria do uso do conjunto de dados completo (MURPHY, 2012). O algoritmo de descida de gradiente estocástico pode oscilar ao longo do caminho de descida mais íngreme em direção ao resultado ótimo. Adicionar um termo de *momentum* à atualização de parâmetros é uma maneira de

reduzir essa oscilação (MANDT; HOFFMAN; BLEI, 2017). A atualização *stochastic gradient descent with momentum* (SGDM) é dada por:

$$x_{i+1} = x_i - \alpha \nabla f(x_i) + \alpha_g (x_i - x_{i-1}), \quad (43)$$

onde α_g determina a contribuição do passo do gradiente anterior para a iteração atual. O cálculo do SGDM usa uma única taxa de aprendizado para todos os parâmetros.

O método proposto por Powell baseado em Beale reinicia o treinamento se houver pouca ortogonalidade entre o gradiente atual g_i e o gradiente anterior g_{i-1} . Os requisitos de armazenamento para o algoritmo de Powell-Beale (seis vetores) são ligeiramente maiores do que para o algoritmo de Polak-Ribiere (quatro vetores) (ABBO; KHUDHUR, 2017).

Esses requisitos podem ser testados por:

$$|g_{i-1}^T g_i| \geq 0,2 \|g_i\|^2. \quad (44)$$

O procedimento geral para determinar a nova direção de busca é combinar a nova direção de descida mais íngreme com a direção de busca anterior. Para a atualização de Fletcher-Reeves, o procedimento é:

$$B_i = \frac{g_i^T g_i}{g_{i-1}^T g_{i-1}}. \quad (45)$$

Esta é a razão entre a norma quadrada do gradiente atual e a norma quadrada do gradiente anterior. Os algoritmos de gradiente conjugado requerem um pouco mais de armazenamento do que os algoritmos mais simples, por isso são bons para redes com um grande número de pesos. O algoritmo de gradiente conjugado escalado é baseado em direções conjugadas, e o algoritmo não realiza uma busca de linha em cada iteração (AHMAD *et al.*, 2018).

O algoritmo de treinamento de retropropagação resiliente permite reduzir os efeitos das diferenças das magnitudes das derivadas parciais, já que apenas o sinal derivativo pode determinar a direção da atualização de peso e não a magnitude da derivada (MANSOURIAN; GHANIZADEH; GOLCHIN, 2019).

Cybenko (1989) mostrou que combinações lineares finitas podem aproximar uniformemente qualquer função contínua de variáveis reais, com isso é possível uma universalidade na aproximação geral de funções. Bishop (1995) explorou a relação do número de parâmetros a serem ajustados em função do número de variáveis que se deseja aproximar. A partir destes conceitos comprovou-se que há viabilidade no uso de redes neurais artificiais como aproximadores de funções universais.

No trabalho de Tollenaere (1990) é utilizado um algoritmo semelhante ao utilizado por Silva e Almeida (1990), porém não são feitas as atualizações nos pesos que causam alterações no sinal do gradiente. Estes algoritmos propostos apesar de não dependerem das condições iniciais, como acontece com o algoritmo BP, necessitam da determinação de três parâmetros ao invés de um. O uso destes algoritmos é bastante trabalhoso, em vista que não existe uma metodologia eficaz para escolha dos parâmetros podendo ocorrer instabilidade no treinamento caso um parâmetro seja inadequado.

Apesar do trabalho de Maraaba, Al-hamouz e Al-duwaish (2014) ter apenas 86,67% de acurácia, a avaliação é baseada na detecção de contaminantes que podem indicar a possibilidade do desenvolvimento de uma falha antes que esta se torne irreversível. A variação do formato da RNA tem apresentado sucesso em resultados aplicados para classificação de isoladores. Utilizando convergência rápida Liu *et al.* (2017b) em um algoritmo baseado no método Gradiente Descendente teve um resultado de precisão de 93%.

3.5.2 Newton

Para melhorar o ajuste dos pesos foram utilizadas técnicas avançadas de otimização como o método de Newton. Assim como o método de gradiente, o método de Newton constrói-se através de iterações em busca da solução do problema (MARTINEZ; SANTOS, 1995). No método de Newton a direção p_i depende da matriz hessiana, conforme (46).

$$p_i = -[\nabla^2 f(x_i)]^{-1} \nabla f(x_i). \quad (46)$$

O método de Newton pode ser mais eficiente para convergir que o método de gradiente, no entanto, exige maior esforço computacional. Com isto, o método de quase-Newton pode ser utilizado, sendo que possui uma boa convergência e exige menor esforço computacional que o método original de Newton (NOCEDAL; WRIGHT, 2006).

$$B_{i+1} s_i = y_i \quad (47)$$

em que,

$$s_i = x_{i+1} - x_i y_i = \nabla f(x_{i+1}) - \nabla f(x_i) \quad (48)$$

com isto,

$$B_{i+1} = B_i - \frac{(y_i - B_i s_i)(y_i - B_i s_i)^T}{(y_i - B_i s_i)^T s_i}. \quad (49)$$

Reorganizando a fórmula conhecida como BFGS (de seus inventores Broyden, Fletcher, Goldfarb e Shanno) tem-se (50).

$$B_{i+1} = B_i - \frac{B_i s_i s_i^T B_i}{s_i^T B_i s_i} + \frac{y_i y_i^T}{y_i^T s_i}. \quad (50)$$

Desta forma, se obtém a direção para o método quase-Newton sem precisar encontrar a hessiana da função. Aplicando estes métodos há um ganho significativo em relação ao número de interações, porém, geralmente apresentam maior custo computacional. Para melhorar as propriedades de convergência e reduzir o custo computacional algumas alternativas foram propostas. Zhou e Si (1998) desenvolveram um método de treinamento avançado para redes neurais que é baseado na deficiência de classificação da matriz jacobiana.

Conforme Zhao (1996), as técnicas de otimização irrestritas para treinamento de RNAs *feedforward* não podem ser utilizadas em casos *online* em que as saídas e entradas são constantemente atualizadas, pois é possível realizar a otimização somente quando as amostras estiverem disponíveis. A partir desta necessidade, em seu trabalho Zhao (1996) propôs uma nova técnica para rastreamento de mínimos que é dependente do tempo, podendo desta forma ser utilizada para classificação de componentes em tempo real, não necessitando de dados de entradas estáticos.

Novos algoritmos foram propostos em que a camada é decomposta em uma parte linear e uma parte não linear. Inicialmente é analisado o passo em direção ao mínimo através de estratégias de otimização, após isto, o bloco linear é otimizado separadamente usando um critério de mínimos quadrados. Esta abordagem torna a convergência deste algoritmo mais rápida e com menos custo computacional do que os métodos tradicionais de Newton, gradiente conjugado ou quase-Newton (PARISI *et al.*, 1996).

3.5.3 Levenberg-Marquardt

O algoritmo de Levenberg-Marquardt (LM) vem sendo utilizado para o treinamento de redes neurais com sucesso para solução de problemas de mínimos quadrados não lineares, superando os algoritmos de retropropagação com taxa de aprendizagem variável (ZHOU; SI, 1998). Levenberg (1944) inicialmente desenvolveu um estudo para solução de problemas não lineares e Marquardt (1963) deu sequência no estudo para soluções numéricas complexas. Os

métodos tradicionais de mínimos quadrados de equações não lineares para redução de resíduos dependem de aproximações de Taylor³ e produzem novos valores para os parâmetros que não estão suficientemente próximos aos valores iniciais.

No trabalho de Moré (2006) é apresentado como o método LM pode ser robusto e eficiente tendo propriedades de convergência fortes, em que as principais características desta implementação são o uso adequado de variáveis escalonadas implicitamente. Sendo uma função continuamente diferenciável para o problema de mínimos quadrados não lineares, será possível encontrar o minimizador para (51).

$$\emptyset(x) = \frac{1}{2} \sum_{i=1}^m f_i^2(x) = \frac{1}{2} \|f(x)\|^2. \quad (51)$$

No trabalho de Yam e Chow (1997) é mostrado que um algoritmo baseado em mínimos quadrados possui uma convergência superior comparando-se ao método de LM e o número de operações por interação é inferior ao algoritmo padrão de retropropagação melhorando assim sua eficiência. Apesar destas características, este método tem que lidar com a possibilidade da camada escolhida não ser linearmente separável em relação a uma especificação, tornando impossível a redução dos erros de mínimos quadrados das camadas de saída e escondida.

No trabalho de Moreira *et al.* (2006) foi aplicado o método de Levenberg-Marquardt (LM) com sucesso, com um volume inferior de dados. Outros métodos de otimização, que exigem menor esforço computacional, podem ser utilizados tendo resultados de precisão adequados para os problemas considerados.

3.5.4 Estimativa do momento adaptativa

Outros algoritmos de otimização buscam melhorar o treinamento da rede usando taxas de aprendizado que diferem por parâmetro e podem se adaptar automaticamente à função de perda que está sendo otimizada, como a *root mean square propagation* (RMSProp), que mantém uma média móvel dos quadrados elementares dos gradientes de parâmetro (TIELEMAN; HINTON, 2012). Assim, o algoritmo pode ser calculado como:

³ Fórmula que fornece uma aproximação polinomial local cada vez melhor para uma função (muitas vezes) diferenciável à medida que se calculam as suas derivadas.

$$v_i = \beta_2 v_{i-1} + (1 - \beta_2)[\nabla f(x_i)]^2. \quad (52)$$

Dessa forma, β_2 é a taxa de decaimento da média móvel. Os comprimentos médios correspondentes dos gradientes quadrados são iguais ao inverso de $(1 - \beta_2)$. Esse algoritmo usa essa média móvel para normalizar as atualizações de cada parâmetro individualmente,

$$x_{i+1} = x_i - \frac{\alpha \nabla f(x_i)}{\sqrt{v_i + \epsilon}}. \quad (53)$$

O uso do RMSProp reduz efetivamente as taxas de aprendizado para parâmetros de gradiente grande e aumenta as taxas de aprendizado para parâmetros de gradiente pequeno. O ϵ é uma pequena constante usada para evitar a divisão por zero.

O algoritmo de estimação *adaptive moment estimation* (ADAM), mantém uma média móvel elementar de ambos os gradientes de parâmetro e seus valores quadrados. O ADAM usa uma atualização de parâmetro semelhante ao RMSProp, mas com um termo de momento adicional, desta forma mantém uma média móvel em termos de elementos, tanto dos gradientes dos parâmetros quanto dos seus valores ao quadrado.

$$m_i = \beta_1 m_{i-1} + (1 - \beta_1)\nabla f(x_i). \quad (54)$$

Se os gradientes forem semelhantes em muitas iterações, o uso de uma média móvel do gradiente permite que as atualizações de parâmetros captem o momento em uma determinada direção (KINGMA; BA, 2015).

$$v_i = \beta_2 v_{i-1} + (1 - \beta_2)[\nabla f(x_i)]^2. \quad (55)$$

O ADAM usa as médias móveis para atualizar os parâmetros de rede conforme:

$$x_{i+1} = x_i - \frac{\alpha m_i}{\sqrt{v_i + \epsilon}}. \quad (56)$$

Além desses métodos de otimização, atualmente está se tornando popular utilizar métodos de otimização meta-heurísticos, pois fazem poucas ou nenhuma suposição sobre a resolução do problema e podem pesquisar um grande espaço de soluções candidatas. Modelos que simulam o comportamento social de animais, como *particle swarm optimization* (PSO), vem sendo vastamente utilizados.

Por estes algoritmos não utilizarem o gradiente para resolução do problema, não exigem que o problema seja diferenciável, se sobressaindo em relação a métodos clássicos de otimização. Contudo, por serem métodos meta-heurísticos, não garantem uma solução ótima (SINGH; DHILLON, 2016).

3.5.5 Otimização por enxame de partículas

A otimização por enxame de partículas simula o comportamento social das aves em um bando, sendo possível descrevê-lo como um sistema em evolução automática. Cada ave é considerada uma partícula no enxame, onde cada partícula é composta por um conjunto de variáveis a serem otimizadas. Neste caso, as variáveis são os parâmetros de cada conjunto do ELM (KENNEDY; EBERHART, 1995). O movimento das partículas é influenciado por sua posição conhecida e é orientado para uma posição no espaço de busca (VLACHOGIANNIS; LEE, 2009).

As partículas se movem com certa velocidade e encontram uma nova posição após cada iteração, que é avaliada para obter seu valor de adequação. Estas posições são atualizadas conforme as melhores posições são encontradas por outras partículas, e com isto, o enxame se move para as melhores soluções. O processo é repetido até que um critério de parada seja atendido (XU *et al.*, 2015). A velocidade é calculada com base na influência das partículas em sua melhor posição (lP_q), da previsão de sua melhor posição (gP) e do peso do momento de inércia (ω).

$$\begin{aligned} \vartheta_q(i+1) = \omega \cdot \vartheta_q(i) + r_1 \cdot c_1 \cdot (lP_q(i) - pos_q(i)) + \\ r_2 \cdot c_2 \cdot (gP(i) - pos_q(i)). \end{aligned} \quad (57)$$

A velocidade de cada partícula q é expressa por ϑ_q , i é a iteração atual do algoritmo, pos é a posição de cada partícula q -th, c_1 e c_2 são as constantes de aceleração, r_1 e r_2 são os números aleatórios entre 0 e 1 com distribuição uniforme.

$$pos_q(i+1) = pos_q(i) + \vartheta_q(i+1). \quad (58)$$

Para aumentar o desempenho do algoritmo, implementa-se o peso de inércia dinamicamente. Os parâmetros podem reduzir gradativamente conforme o aumento de cada iteração em que a inércia atual é apresentada como,

$$\omega(i) = \frac{\omega_2 - \omega_1}{\omega_e - 1} (i - 1) + \omega_2, \quad (59)$$

o peso de inércia inicial como ω_2 , e o valor final do peso de inércia como ω_1 . Utilizando esta técnica é evitada uma convergência prematura para um mínimo local.

O vetor de posição da partícula tem um valor variável dependendo de quantos *ensembles* são necessários (*maxEns*). Para tornar a explicação mais fácil, o vetor é dividido em partes. Inicialmente o vetor possui peso de entrada w e bias b para cada conjunto de partículas (*ens*), estes

parâmetros criam o vetor wb . Após isto, todos os parâmetros da função de ativação são ajustados para todos os conjuntos de partículas. Os parâmetros lineares da função de ativação $klin$, e não lineares knl são representados pelo vetor $ensPar$.

$$wb_{ens} = \begin{bmatrix} w_{ens_{11}}, w_{ens_{12}}, \dots, w_{ens_{1N}} \\ w_{ens_{21}}, w_{ens_{22}}, \dots, w_{ens_{\tilde{N}N}} \\ b_{ens_1}, b_{ens_2}, \dots, b_{ens_{\tilde{N}}} \end{bmatrix},$$

$$ensPar_{ens} = \begin{bmatrix} \delta_{ens_1}, \delta_{ens_2}, \dots, \delta_{ens_{klin}} \\ \xi_{ens_1}, \xi_{ens_2}, \dots, \xi_{ens_{knl}} \\ \tau_{ens_1}, \tau_{ens_2}, \dots, \tau_{ens_{knl}} \\ v_{ens_1}, v_{ens_2}, \dots, v_{ens_{knl}} \\ \chi_{ens_1}, \chi_{ens_2}, \dots, \chi_{ens_{knl}} \end{bmatrix}, \quad (60)$$

$$ens = 1, \dots, maxEns.$$

A posição do vetor pode ser descrita como:

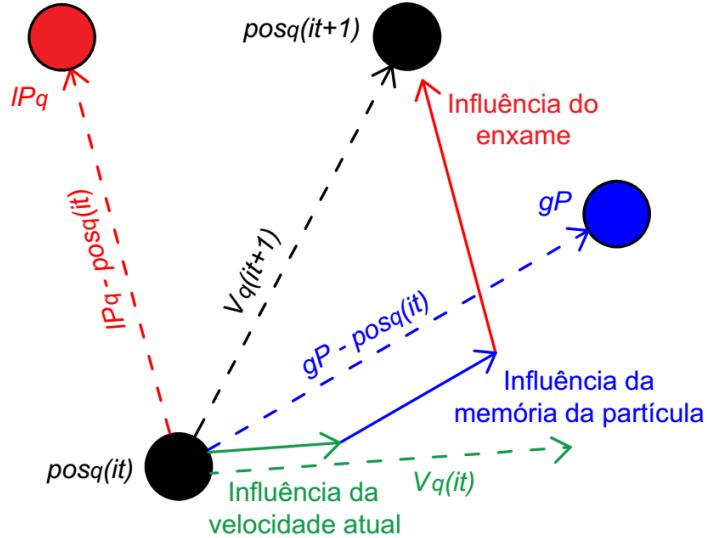
$$pos_q = \begin{bmatrix} wb_1, wb_2, \dots, wb_{maxEns} \\ ensPar_1, ensPar_2, \dots \\ ensPar_{maxEns} \end{bmatrix}. \quad (61)$$

Com base no cálculo da influência da velocidade atual da partícula, de sua memória e da influência do enxame, uma nova posição é obtida $pos_q(i + 1)$, conforme pode ser visualizado na Figura 17, com base nas equações já apresentadas. A posição tende a ser atualizada em direção ao máximo global, pois à medida que as partículas se ajustam, tendem a encontrar posições melhores e convergir para o mesmo local.

Para resolver o problema de otimização e selecionar os parâmetros ótimos, o algoritmo de PSO modifica os pesos de entrada, melhorando o desempenho da generalização e o condicionamento do SLFN. Esta otimização tem o objetivo de resolver um problema procurando encontrar, de forma interativa, a melhor solução em relação a uma medida de qualidade (SHI; EBERHART, 1999).

A topologia do enxame define com quais partículas um subconjunto pode trocar informações, permitindo que todas as partículas possam se comunicar, e desta forma, todo o enxame tenda a convergir para uma mesma posição. Entretanto, existe a possibilidade do algoritmo de otimização tender a um mínimo local, e por isso outras técnicas têm sido utilizadas para controlar o fluxo de informações das partículas (DEL VALLE *et al.*, 2008).

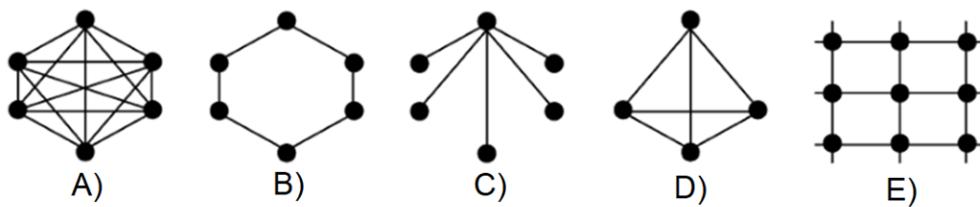
Figura 17 - Evolução da posição da partícula no método PSO.



Fonte: Elaborada pelo autor (2021).

Na topologia local, as partículas compartilham informações apenas com um subconjunto. Este pode ser geométrico em que a comunicação se dá por proximidade, ou social em que o conjunto não depende da distância. Além disso, o formato da topologia do enxame tem influência sobre o resultado da otimização. Alguns formatos possíveis são apresentados na Figura 18. Alguns autores têm procurado melhorar estas topologias como proposto por Yang, Deb, Fong (2011) para o *accelerated particle swarm optimization* (SPSO), por Zhan *et al.* (2009) para o *adaptive particle swarm optimization* (APSO) e por Yin *et al.* (2011) com *cyber* enxame.

Figura 18 - Formatos para topologias de enxame: A) melhor global, B) topologia em anel, C) topologia circular, D) topologia em pirâmide, E) topologia Von Neumann.



Fonte: Adaptado de Del Valle *et al.* (2008, p. 174).

Conforme Zhan *et al.* (2009), a otimização do enxame de partículas de forma adaptativa apresenta uma melhor eficiência para realizar a busca pelo mínimo global que o PSO padrão, com maior velocidade de convergência. Através do APSO, o controle dos coeficientes de aceleração, peso de inércia e outros parâmetros é realizado de forma automática. Desta forma, o APSO pode

gerar maior influência na busca pelo mínimo global, evitando ótimos locais, introduzindo novos parâmetros ao algoritmo sem aumentar a complexividade em sua implementação. Nesta tese foi utilizado um PSO adaptativo a partir da topologia em anel.

Além do otimizador, para uma boa aplicação dos algoritmos apresentados é necessária a configuração adequada da rede. Neste trabalho, a configuração da rede foi realizada com base na avaliação da resposta à convergência conforme alteração dos parâmetros fundamentais de configuração da rede, como sua função de ativação, número de neurônios e taxa de aprendizado.

3.6 FUNÇÃO DE ATIVAÇÃO

Nas redes neurais artificiais a função de ativação de um neurônio define sua saída, para uma entrada ou conjunto de entradas. A saída do neurônio é usada como entrada do próximo e assim sucessivamente, até que se encontre uma saída desejada para o problema. Com isto, a função de ativação representa a taxa de ativação do potencial de ação do neurônio.

A função de ativação tem a capacidade de introduzir um componente de não linearidade nas redes artificiais. Entretanto, ao incorporar um componente não linear na rede, a superfície de custo da rede deixa de ser convexa, tornando o problema de otimização mais complicado e exigindo maior custo computacional.

Quando não existe função de ativação na rede, os pesos e *bias* fazem uma transformação linear que é simples de resolver, porém é limitada para resolver problemas complexos. Com isto, uma rede neural sem função de ativação é apenas um modelo de regressão linear. A não linearidade da função de ativação faz com que esta tenha capacidade de resolver problemas complexos, como tradução de idiomas, processamento de imagem e classificação de dados (NIED, 2007), que é o foco deste trabalho.

Na sua forma mais simples a função de ativação pode ser representada por uma função Degrau, sendo a saída possível ser sim ou não. Por esta função não possuir derivada em zero, não é aplicável para problemas complexos. Grande parte dos problemas de classificação pode ser resolvida com a função de ativação Sigmoíde, apresentada em (62), que também é conhecida como função logística. Esta função é considerada suave e possui boa aplicação por tender os valores para os extremos. Inicialmente a função tem um crescimento exponencial e quando a função se aproxima da saturação o crescimento diminui para um modelo linear até sua convergência.

$$\sigma(x) = \frac{1}{1 + e^x} \quad \sigma'(x) = \sigma(x)(1 - \sigma(x)). \quad (62)$$

Outra função de ativação usada para multiclassificação é a tangente hiperbólica (*TanH*), apresentada em (63), que conforme De Mello Canalli (2017) esta função se aproxima de um modelo realista, geralmente sendo mais interessante do que a Sigmoide.

$$\text{TanH}(x) = 2\sigma(2x) - 1 \quad \text{TanH}'(x) = 1 - \tanh^2(x). \quad (63)$$

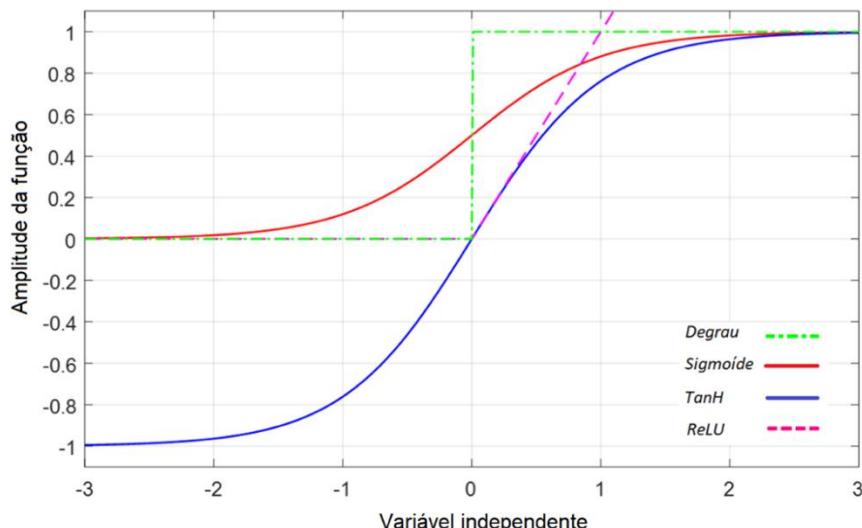
Além destas, outras funções também têm sido utilizadas em redes neurais como a função unidade linear retificada (*ReLU*), apresentada em (64). A função *ReLU* é facilmente otimizada, sendo uma função semelhante a função identidade que possui vasta aplicação para *deep learning*.

A principal vantagem nesta função é que ela não ativa todos os neurônios ao mesmo tempo, ou seja, se uma entrada for negativa ela será considerada zero e o neurônio não será ativado, e com isto a rede se torna mais eficiente computacionalmente.

$$\text{ReLU}(x) = \max\{0, x\} \quad \text{ReLU}'(x) = \begin{cases} 1, & \text{se } x \geq 0, \\ 0, & \text{c.c..} \end{cases} \quad (64)$$

As funções descritas nesta seção são apresentadas na Figura 19. Além destas funções de ativação tradicionais, pesquisadores vem avaliando a influência de outras funções de ativação em relação ao comportamento das redes neurais e tem proposto variações (DE MELLO CANALLI, 2017), como a função linear com saturação apresentada em (65).

Figura 19 - Funções de ativação para redes neurais artificiais.



Fonte: Elaborada pelo autor (2021).

$$f(x) = \begin{cases} 0, & n \leq 0, \\ n, & 0 \leq n \leq 1, \\ 1, & 1 \leq n. \end{cases} \quad (65)$$

Vários pesquisadores mostraram que a função de ativação pode atuar como aproximador universal em redes neurais multicamada. Leshno *et al.* (1993) conseguiu melhorar esta análise provando que uma rede multicamada padrão com função de ativação contínua, que não for um polinômio, pode aproximar uma função contínua para qualquer grau de precisão.

Hornik (1991) desenvolveu um estudo sobre as capacidades de aproximação das RNAs, no qual mostrou que as redes convencionais de múltiplas camadas, com somente uma camada oculta e função de ativação arbitrária limitada e não constante, são aproximadores universais em relação aos critérios de desempenho. Se a função de ativação for contínua, limitada e não constante, então mapeamentos contínuos podem ser aprendidos uniformemente em conjuntos de entradas compactas.

3.7 CONSIDERAÇÕES FINAIS DO CAPÍTULO

Neste capítulo os modelos usados para classificação de padrões foram apresentados. Dentro deste contexto, foram apresentadas técnicas para extração de características e formas de otimização do modelo. Considerando que o sinal ultrassônico avaliado nesta tese é uma série temporal, no próximo capítulo serão apresentados os algoritmos utilizados para análise deste sinal em relação à esta característica.

4 PREVISÃO DE SÉRIES TEMPORAIS

Uma série temporal (ST) pode ser definida com um conjunto de dados obtidos com uma determinada frequência de tempo e pode ser utilizada para previsão, pois fornece um meio de ponderação dos dados para o cálculo de valores futuros. Além disso, as STs permitem a identificação de modelos estruturais com tendências. Inicialmente, os modelos são construídos a partir de distribuições de probabilidade que geralmente são utilizados para descrever o comportamento da ST (YANG; LIU, 2018).

As observações das STs caóticas são realizadas em períodos equidistantes. Neste trabalho os períodos são iguais a $2,08 \times 10^{-5}$ s que correspondem ao período de amostragem em relação a taxa de aquisição de 48 kHz. A ST é considerada estacionária se não há grandes variações na variância dos valores observados ao longo do tempo, para que desta forma a sequência dos valores seja estável e tenha um comportamento regular. Se o período de análise não for suficientemente longo, não é possível avaliar tendências de dados, sazonalidade e irregularidade (YANG; LIU, 2018).

Na predição, valores conhecidos da série temporal são utilizados até o ponto no tempo t , para prever o valor em algum ponto no futuro, $t + P$ (YANG; LIU, 2018). Com isto, se cria um mapeamento a partir dos pontos de dados da amostra n , amostrados a cada Δ unidades no tempo,

$$x(t - (n - 1)\Delta), \dots, x(t - \Delta), x(t) \quad (66)$$

para um valor futuro previsto

$$x = (t + P). \quad (67)$$

4.1 CRITÉRIOS DE AVALIAÇÃO DE DESEMPENHO PARA PREVISÃO

Além do cálculo do erro apresentado em (2), neste trabalho outras métricas foram utilizadas para avaliação das séries temporais, como o *mean square error* (MSE), que pode ser calculado por:

$$MSE = \frac{1}{n} \sum_{i=1}^n (\gamma_i - \hat{\gamma}_i)^2. \quad (68)$$

O MSE é muito usado por penalizar os grandes erros de predição, contudo, o *root mean square error* (RMSE) é geralmente mais utilizado para avaliar o desempenho do modelo de regressão. Para realizar uma avaliação do erro o RMSE é calculado por:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\gamma_i - \hat{\gamma}_i)^2}. \quad (69)$$

O *mean absolute error* (MAE) denota a média da diferença absoluta entre o valor observado e o previsto, sendo promissor quando há *outliers*. O MAE é calculado conforme:

$$MAE = \frac{1}{n} \sum_{i=1}^n |\gamma_i - \hat{\gamma}_i|. \quad (70)$$

O *mean absolute percentage error* (MAPE) calcula a taxa de erro média para os valores corretos, onde:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{\gamma_i - \hat{\gamma}_i}{\gamma_i} \right|. \quad (71)$$

O coeficiente de determinação também é uma métrica muito utilizada para avaliação do erro em algoritmos de previsão. Este coeficiente pode ser calculado através da equação:

$$R^2 = 1 - \frac{\sum_{i=1}^n (\gamma_i - \hat{\gamma}_i)^2}{\sum_{i=1}^n (\gamma_i - \bar{\gamma})^2}. \quad (72)$$

O coeficiente de determinação é usado para estabelecer o quanto bem as variáveis independentes, no modelo de regressão linear, explicam a variabilidade na variável dependente. O valor quadrático de R aumenta com a adição das variáveis independentes, o que pode levar à adição de variáveis redundantes no modelo (STEFENON *et al.*, 2020b).

4.2 SISTEMA DE INFERÊNCIA NEURO-FUZZY ADAPTATIVO

Um *adaptive neuro fuzzy inference system* (ANFIS) é um tipo de rede neural artificial baseada em Takagi-Sugeno. Como o ANFIS integra nas redes neurais princípios de lógica fuzzy, tem potencial para utilizar as vantagens de ambas as técnicas em um mesmo *framework* (YANG; FU; WANG, 2017).

A principal função dos conjuntos fuzzy é fornecer uma maneira sistemática de aproximação de fenômenos complexos ou mal definidos. Ao invés de utilizar valores numéricos, a lógica fuzzy utiliza expressões de regras para representação das operações matemáticas. A lógica fuzzy é baseada em funções de pertinência que determinam o grau que um determinado elemento pertence

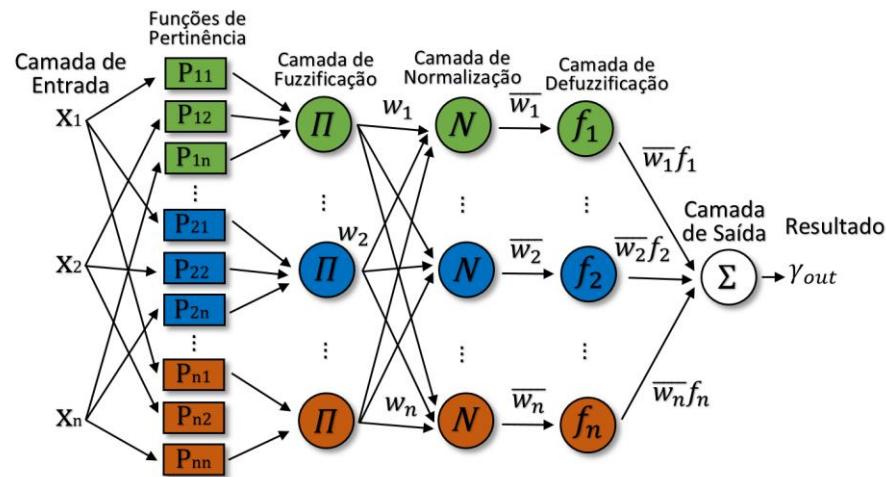
a um conjunto, permitindo assim uma transição gradual da falsidade para a verdade (YANG; CHAO; SHEN, 2016).

Existem diversas formas de função de pertinência, como a função triangular, trapezoidal, gaussiana, entre outras. São aplicáveis conceitos de conjuntos para estas funções de pertinência como, por exemplo, intersecção, união e complemento. No sistema de inferência fuzzy, inicialmente é realizada a fuzzificação. Nesta etapa são definidas as funções de membro, também conhecidas como funções de pertinência (MENDEL, 2017).

Após a fuzzificação e definição das regras das funções de pertinência é realizada a defuzzificação, etapa que as regiões resultantes são convertidas em valores para as variáveis do sistema. As principais técnicas de defuzzificação utilizadas são centroide, critério máximo, mínimo do máximo e primeiro do máximo. A partir da defuzzificação se obtém o resultado da função (WANG *et al.*, 2017).

A integração das redes neurais e dos sistemas fuzzy é possível por ambos serem aproximadores universais e desta forma, a rede *neuro-fuzzy* também mantém essa característica. O resumo dessa arquitetura é apresentado na Figura 20. Com base nas características de ambos os sistemas, a rede *neuro-fuzzy* tem a capacidade de manipular sistemas com dados imprecisos, complexos e não-lineares (GIL; OLIVEIRA; PALMA, 2018).

Figura 20 - Estrutura do ANFIS para avaliação de séries temporais.



Fonte: Elaborada pelo autor (2021).

As redes *neuro-fuzzy* herdam da rede neural sua capacidade de aprendizagem, classificação, robustez, adaptação e agrupamento de dados comuns em *clusters*. O comportamento deste modelo

pode ser entendido através da observação das variáveis das funções de pertinência, dos relacionamentos entre entrada e saída e das regras fuzzy, que podem explicar o funcionamento do sistema de forma simplificada devido a proximidade com a linguagem humana. A partir destas características, o ANFIS pode ser utilizado para previsão de séries temporais caóticas (PARSAPOOR; BILSTRUP; SVENSSON, 2018).

A partir do sinal filtrado aplicando o modelo ANFIS, são mapeadas as características de entrada para as funções de participação, e estas funções se transformam em regras de entrada. As regras geram um conjunto de características de saída para funções de saída, com uma decisão associada. As funções de associação são escolhidas arbitrariamente e a estrutura de regras é predeterminada com base nas características das variáveis do modelo (FU *et al.*, 2018).

Após a normalização das variáveis de entrada do sinal, na camada fuzzy calcula-se o grau de pertinência com que as entradas satisfazem aos conjuntos fuzzy associados a cada entrada. Nas camadas de regra, calcula-se o nível de disparo correspondente a cada regra. Por fim, na camada de saída calcula-se o valor da função da rede $f(\text{net})$ em relação aos conjuntos fuzzy $P_k(\lambda_i)$ e pesos associados w .

$$\text{fuzzy}_i = f\left(\frac{\sum_{k=1}^m P_k(\lambda_i) \cdot \omega_k}{\sum_{k=1}^m P_k(\lambda_i)}\right). \quad (73)$$

A estrutura de inferência fuzzy baseada em participação de grade gera uma modelo do tipo Sugeno usado como condição inicial para o treinamento do ANFIS. O método de partição em grade possui características que melhoram o processamento em paralelo, pois garante que cada processador tenha uma carga igual. Para este tipo de *cluster* existe uma regra para cada combinação de função de participação de entrada, em consequência, cada regra corresponde a uma função de associação de saída diferente (VERSTRAEETE, 2017).

A estrutura que utiliza o armazenamento em *cluster* subtrativo requer um conjunto separado de dados de entrada e saída assim como argumentos. A partir desta técnica são extraídos os conjuntos de regras que modelam o comportamento dos dados. Neste tipo de *cluster* há uma regra para cada *fuzzy cluster* (REZAEIAN; ESMAEILI; FADAEINEDJAD, 2018).

O sistema de inferência *fuzzy c-means clustering* (FCM) escolhe o número de *clusters* e atribui os coeficientes aleatoriamente a cada ponto de dados. O algoritmo repete esta lógica até que tenha convergido. Para que haja convergência, a centroide de cada *cluster* é calculada para cada

ponto. A centroide de um *cluster* é a média de todos os pontos, ponderada pelo seu grau de pertencimento ao *cluster*, expressa por,

$$c_i = \frac{\sum_i w_i(x)^m x}{\sum_i w_i(x)^m} \quad (74)$$

em que qualquer ponto de x tem um conjunto de coeficientes em relação ao grau k_{th} do *cluster* $w_i(x)$.

O FCM tenta particionar uma coleção finita de elementos x , em uma coleção de clusters em relação a um critério determinado (YANG; NATALIANI, 2018). A partir de um conjunto de dados o algoritmo retorna para o centro dos *clusters*,

$$W = w_{i,j} \in [0, 1], \quad i = 1, \dots, n, \quad j = 1, \dots, c, \quad (75)$$

em que cada elemento w_{ij} determina o grau para cada x_i pertencente ao *cluster* c_j (Havens *et al.*, 2012). Desta forma a função objetivo a ser minimizada pode ser expressa por,

$$\arg \min_C \sum_{i=1}^n \sum_{j=1}^c w_{ij}^m \|x_i - c_j\|^2, \quad (76)$$

em que,

$$w_{ij} = \frac{1}{\sum_{k=1}^c \left(\frac{\|x_i - c_k\|}{\|x_i - c_j\|} \right)^{\frac{2}{m-1}}}. \quad (77)$$

Inicialmente foi aplicado um filtro de escala correspondente e após isto o sinal foi decomposto em pacotes wavelet do terceiro até o quinto nível, visto que a utilização de mais níveis de decomposição resultou em sinais equivalentes. Foram utilizados dois e três nós nesta decomposição, pois nesta análise com um número maior de nós o sinal perde suas características fundamentais.

Para estrutura de inferência fuzzy baseada em participação de grade especificou-se duas funções de associação para cada entrada. Para isto, utilizou-se a função de associação gaussiana em relação a entrada,

$$f(x; \sigma, c) = e^{\frac{-(x-c)^2}{2\sigma^2}}. \quad (78)$$

e a função de associação linear em relação a saída.

A combinação das técnicas wavelet e redes neurais baseadas em lógica fuzzy tem tido grande visibilidade entre os pesquisadores para diversas aplicações. O trabalho de Zhao e Lin (2018) propõe uma nova estrutura fuzzy para sistemas não lineares. Esta rede é baseada na função

wavelet e no *fuzzy cerebellar model neural network* (CMNN). Este trabalho combina as vantagens da função wavelet, do CMNN e do sistema de inferência fuzzy Takagi–Sugeno–Kang, com resultados de maior eficiência para resolução de problemas não lineares.

No trabalho de Huang, Oh e Pedrycz (2017) um sistema *hybrid fuzzy wavelet neural network* (HFWNN) é apresentado. Os parâmetros do algoritmo são inicializados a partir do método *fuzzy c-means clustering* (FCM), assim como o realizado em uma das análises neste trabalho. A primeira camada da rede é utilizada para refletir a incerteza dos dados enquanto a segunda, que possui um alto nível de flexibilidade, realiza combinações lineares da função wavelet. Os parâmetros utilizados no projeto do HFWNN são ajustados através de *genetic optimization*.

Outra aplicação de fuzzy e wavelet é apresentada no trabalho de Huang, Oh e Pedrycz (2016) para uma rede neural polinomial. Nesta análise também é aplicado FCM na premissa e na consequência e a função wavelet estimada pelo método dos mínimos quadrados. Esta análise apresenta vantagens em relação a habilidade de descrever relações não lineares entre as variáveis de entrada e saída. Uma aplicação de ANFIS baseada na extração de características utilizando a transformada wavelet e transformada de Fourier é apresentada no artigo de Alkhraijah e Abido (2017) para classificação de um problema de qualidade de energia do sistema elétrico.

Outras avaliações do ANFIS e da transformada wavelet para identificação e classificação de falhas no sistema elétrico são apresentadas nos trabalhos de Babayomi *et al.* (2017) e Basantak *et al.* (2018). Uma comparação entre o método *fuzzy learning vector quantization* (FLVQ) e o ANFIS, baseado em uma entrada resultante de transformada wavelet é apresentado por Gan *et al.* (2016) em comparação ao método clássico de Levenberg-Marquardt.

Uma aplicação para previsão de preços de eletricidade baseada na wavelet e ANFIS é apresentada por Karri, Durgam e Raghuram (2018). Assim como nos trabalhos mencionados anteriormente, esta técnica apresenta um bom resultado para resolução deste problema não linear. Uma comparação completa com base no RMSE, entre diversas variações de algoritmos, para um problema de previsão é apresentada no trabalho de Fu *et al.* (2018).

Fu *et al.* (2018) mostram que o ANFIS tem uma melhor capacidade de previsão que técnicas clássicas e aplicando uma melhoria no wavelet-ANFIS os resultados são ainda superiores, utilizando três funções de associação. Damayanti (2017) realiza uma comparação entre ANFIS e FLVQ e mostra que a utilização de FLVQ tem um resultado superior na maioria das análises em relação a precisão para classificação de imagem com a transformada wavelet.

No trabalho de Alkhraijah e Abido (2017) é aplicado o ANFIS a partir de duas funções de associação gaussiana juntamente com o WPT, para avaliação da estrutura de inferência fuzzy baseada em participação de grade. Foram obtidos resultados promissores, com precisão de até 99,56%, contudo se utilizou um volume de dados consideravelmente pequeno e não se realizou grandes variações para avaliar a variabilidade do algoritmo. Desta forma, apesar do resultado promissor, não se tem segurança em relação a precisão do algoritmo.

Basantak *et al.* (2018) apresentaram um trabalho equivalente com um resultado de precisão de 92%. Babayomi *et al.* (2017) obtiverem uma precisão de apenas 81% para localização de defeitos e previsão de 10 tipos de faltas no sistema elétrico, aplicando somente o ANFIS também baseado em participação de grade.

Huang, Oh e Pedrycz apresentaram em (2016) e (2017) trabalhos completos comparando diversas técnicas com o *neuro-fuzzy* baseado em FCM e wavelet. Nestas avaliações se obtiveram resultados de erro baixo comparativamente com outros algoritmos. Também baseado em FCM Yang e Liu (2018) apresentaram uma boa aplicação para previsão de séries temporais.

4.3 SISTEMAS AUTOREGRESSIVOS NÃO LINEARES

Os sistemas autoregressivos não lineares são modelos bem consolidados para previsão de séries temporais. Para valores passados de uma série x_t , a *nonlinear input-output* (NIO) é dada por:

$$f(\mathbf{x}) = f(x(t-1), \dots, x(t-d),) + \varepsilon_t. \quad (79)$$

Os resultados de $f(\mathbf{x})$ dados d valores passados de $f(\mathbf{x})$, podem ser apresentados para prever séries de $f(\mathbf{x})$ através do modelo *nonlinear autoregressive* (NAR).

$$f(\mathbf{x}) = f(y(t-1), \dots, y(t-d),) + \varepsilon_t. \quad (80)$$

Em séries temporais, o *nonlinear autoregressive exogenous model* (NARX) é um modelo autorregressivo não linear que possui entradas exógenas. O NARX relaciona o valor atual de uma série temporal com valores passados da mesma série e os valores atuais e passados de séries exógenas. O modelo NARX pode ser declarado como:

$$f(\mathbf{x}) = f(x(t-1), \dots, x(t-d), y(t-1), \dots, y(t-d),) + \varepsilon_t. \quad (81)$$

Neste caso, o modelo é utilizado para prever séries $f(\mathbf{x})$ dados d valores anteriores de $f(\mathbf{x})$ e outra série $x(t)$.

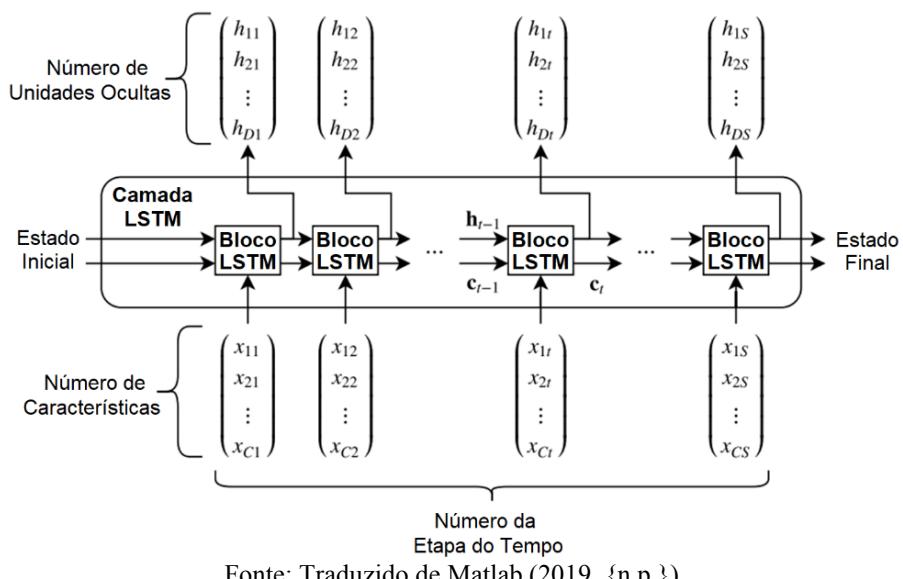
4.4 MEMÓRIA LONGA DE CURTO PRAZO

O *long short-term memory* (LSTM) é uma *recurrent neural network* (RNN) utilizada na aprendizagem profunda. O LSTM possui conexões de *feedback* diferentes das redes neurais *feedforward*. Uma unidade LSTM geralmente é composta de uma célula, uma porta de entrada, uma porta de saída e uma porta de esquecimento (GREFF *et al.*, 2016).

Com base em dados de séries temporais, as LSTMs são adequadas para fazer previsões, sendo que foram desenvolvidas para lidar com problemas graduais do treinamento de RNNs. A insensibilidade relativa é uma vantagem do LSTM em relação a outros modelos e métodos de aprendizado (ZHAO *et al.*, 2017).

As LSTMs podem lembrar o estado da rede entre as previsões, sendo útil quando não se tem a série temporal completa antecipadamente ou se deseja fazer várias previsões em uma série temporal longa. O fluxo de uma série temporal x de comprimento s através de uma camada é apresentado na Figura 21 para o LSTM.

Figura 21 - Fluxo de uma série temporal para LSTM.

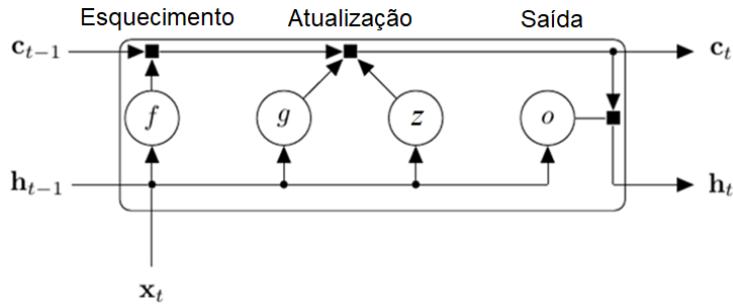


Fonte: Traduzido de Matlab (2019, {n.p.}).

Neste caso, h_t e c_t denotam a saída e o estado da célula na etapa de tempo t , respectivamente. O primeiro bloco LSTM utiliza o estado inicial da rede e o primeiro passo de tempo da sequência para calcular a primeira saída e o estado da célula atualizado. Em um intervalo de tempo t , o bloco usa o estado atual da rede (c_{t-1}, h_{t-1}) e o próximo passo de tempo da sequência

para calcular a saída e o estado da célula atualizado c_t (HOCHREITER; SCHMIDHUBER, 1997). O diagrama da Figura 22 ilustra o fluxo de dados na etapa de tempo t . O diagrama destaca como os portões esquecem, atualizam e produzem a célula e os estados ocultos.

Figura 22 - Diagrama do fluxo de dados para LSTM.



Fonte: Adaptado de Matlab (2019, {n.p.}).

O estado da camada está baseado no estado oculto e no estado da célula. O estado oculto no passo de tempo t contém a saída da camada LSTM para este intervalo de tempo. O estado da célula contém informações aprendidas nas etapas de tempo anteriores. Em cada etapa de tempo, a camada adiciona informações ou remove informações do estado da célula.

Os pesos aprendidos de uma camada LSTM são os pesos de entrada w , os pesos recorrentes r e o bias bi (PASCANU; MIKOLOV; BENGIO, 2013). As matrizes w , r e bi são, respectivamente, as concatenações dos pesos de entrada, dos pesos recorrentes e das tendências de cada componente, conforme:

$$w = \begin{bmatrix} w_z \\ w_f \\ w_g \\ w_o \end{bmatrix}, \quad r = \begin{bmatrix} r_z \\ r_f \\ r_g \\ r_o \end{bmatrix}, \quad bi = \begin{bmatrix} bi_z \\ bi_f \\ bi_g \\ bi_o \end{bmatrix}. \quad (82)$$

No qual z denota a porta de entrada para controlar o nível de atualização de estado da célula, f é o portão de esquecimento do controle de reconfiguração de estado, g é a célula candidata a adicionar informações e o é o portão de saída que controla o nível de estado da célula adicionado ao estado oculto. As equações a seguir descrevem os componentes na etapa de tempo t .

$$\begin{aligned} z_t &= \delta_g(w_z x_t + r_z h_{t-1} + bi_z), \\ f_t &= \delta_f(w_f x_t + r_f h_{t-1} + bi_f), \\ g_t &= \delta_g(w_g x_t + r_g h_{t-1} + bi_g), \end{aligned} \quad (83)$$

$$o_t = \delta_g(w_o x_t + r_o h_{t-1} + b_i o).$$

Nessas equações, δ_g denota a função de ativação de gate e a função de ativação de célula δ_c é usada para atualizar a célula e o estado oculto. Para prever os valores das etapas de tempo futuro de uma sequência, as respostas das sequências de treinamento são deslocadas por uma etapa de tempo. Com isto, a cada passo de tempo da sequência de entrada, a rede LSTM aprende a prever o valor do próximo passo de tempo (HOCHREITER; SCHMIDHUBER, 1997).

A detecção precoce de falhas é crucial para reduzir o tempo de inatividade de um processo ou sistema. A LSTM é capaz de descobrir dependências temporais em dados sequenciais, o que permite ao estimador utilizar informações de contexto anteriores e realizar a análise ao longo do eixo do tempo (LU *et al.*, 2018).

Com base na necessidade de maior confiabilidade do sistema elétrico, Bhattacharya e Sinha (2017) criaram um sistema inteligente artificial para analisar as informações da rede de energia e determinar sua condição. Para isso, são simuladas várias condições da rede elétrica. Segundo os autores, o LSTM tem excelente precisão para esse problema e a técnica pode ser usada para aplicações mais complexas.

Para previsão de falhas em sistemas de energia, Zhang, Wang, Liu e Bao (2018) desenvolveram um trabalho equivalente. Nesta aplicação, o LSTM tem bom desempenho ao extrair recursos de séries temporais por um longo período. O trabalho é baseado em dados da subestação de Wanjiang na *China Southern Power Grid*.

Uma apresentação prática mostra a viabilidade da aplicação do LSTM em cenários reais, principalmente devido à capacidade do algoritmo de lidar com grandes séries temporais (XIA; LU; LI; XU, 2018). Contudo, aplicações de LSTM para predição ou identificação de falhas especificamente em isoladores são raras (WANG; JIN; LU; GAO, 2018).

A pesquisa desenvolvida por Cheng *et al.* (2016) apresenta uma análise da expectativa de vida de isoladores compostos em regiões de alta temperatura e umidade. A vida útil dos isoladores amostrados é determinada pelo processo de oxidação térmica da borracha de silicone e de acordo com os resultados dos testes de carga mecânica e espectroscopia. Este trabalho é baseado em um grande volume de amostras retiradas do sistema de transmissão de energia.

As descargas nos isoladores de redes de transmissão de energia, são um problema crescente no sistema elétrico. Uma previsão de probabilidade de eventos de alto risco é apresentada por Ali (2016) para prever os valores de corrente de fuga em isoladores de alta tensão. Nesta avaliação, o

cálculo da densidade de probabilidade da corrente de fuga pelo estimador de densidade do Kernel apresenta um resultado relevante para garantir a operação segura do sistema de potência.

A previsão de falhas nos isoladores é uma questão importante, considerando que o funcionamento adequado desses componentes contribui para uma maior confiabilidade no fornecimento de eletricidade. Atualmente, o LSTM é uma das técnicas mais avançadas que podem ser usadas para previsão de falhas com base em séries temporais caóticas. A aplicação de técnicas wavelet para extração de características de sinal pode melhorar ainda mais essas análises, reduzindo o ruído e aumentando a precisão do algoritmo.

4.5 GRUPO DE TRATAMENTO DE DADOS

O *group method of data handling* (GMDH) apresenta uma otimização automática estrutural e paramétrica de sua estrutura. O GMDH é um algoritmo indutivo para modelagem matemática de conjuntos de dados multiparamétricos. Esse modelo é caracterizado por ser um procedimento que realiza a classificação dos modelos polinomiais de forma gradativa e realiza a seleção da melhor solução por critérios externos (MO *et al.*, 2018). Para encontrar a melhor solução, o GMDH considera vários subconjuntos de componentes da função base chamados de modelos parciais. Os coeficientes desses modelos são estimados pelo método dos mínimos quadrados.

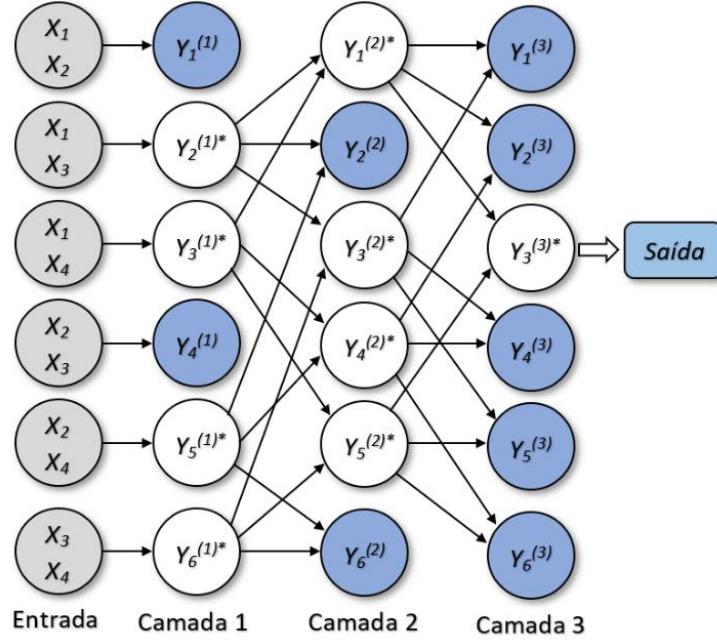
No GMDH, a regressão é utilizada para minimizar a soma dos quadrados dos resíduos feitos nos resultados das equações (RIBEIRO; COELHO, 2020). O modelo aumenta gradativamente o número de componentes parciais e encontra uma estrutura com a complexidade ideal indicada pelo valor mínimo de um critério externo. Li, Fong e Chong (2017) mostram que o GMDH pode ter um desempenho melhor do que os algoritmos de previsão clássicos, como *auto-regressive integrated moving average* (ARIMA), rede neural de retropropagação e outros.

Os valores observados são chamados de variáveis de entrada. Inicialmente, os dados de entrada são divididos em um conjunto de treinamento e conjunto de teste de acordo com uma proporção definida. O conjunto de treinamento é usado para estimar os coeficientes e o conjunto de teste é usado para verificar o erro. O neurônio é avaliado e testado por um critério externo, e os neurônios que têm a pior previsão são eliminados (LIU *et al.*, 2018a).

Todo o processo de reestruturação, treinamento, teste e seleção é realizado novamente na próxima camada e a rede não é concluída até que o erro de previsão do neurônio pare de diminuir.

A arquitetura do GMDH pode ser vista na Figura 23, onde neurônios de cor azul são eliminados por terem a pior previsão (LIU *et al.*, 2018b).

Figura 23 - Arquitetura do *group method of data handling*.



Fonte: Elaborada pelo autor (2021).

Considerando o uso de séries temporais, o algoritmo aprende a relação entre as defasagens. Após aprender as relações, o GMDH seleciona automaticamente o caminho a seguir no algoritmo (EBTEHAJ, Isa *et al.*, 2015). O mapeamento entre as variáveis de entrada e saída usando uma rede neural GMDH é uma função não linear apresentada na forma de equação. Os coeficientes são resolvidos com métodos de regressão para cada par de x_i e x_j variáveis de entrada, conforme apresentado nas equações:

$$\hat{y}(x_1, \dots, x_n) = a_0 + \sum_{i=1}^n a_i x_i + \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j + \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n a_{ijk} x_i x_j x_k + \dots \quad (84)$$

$$G(x_i, x_j) = a_0 + a_1 x_i + a_2 x_j + a_3 x_i^2 + a_4 x_j^2 + a_5 x_i x_j. \quad (85)$$

Os critérios de regularidade podem ser expressos por:

$$w_{out} = \frac{\sum_{n=1}^{P_j} (\hat{y} - y)^2}{\sum_{n=1}^P (y)^2}, \quad (86)$$

em que, y é o valor observado, \hat{y} é o valor previsto de y e w_{out} é o resultado do cálculo do critério externo.

O w_{out} mínimo da camada é registrado. Quando w_{out} deixa de ser reduzido em relação à camada anterior, o erro de previsão da rede deixa de diminuir. Então, a rede para de se expandir e o resultado da camada anterior é gerado. P_j é o número de conjuntos de teste.

O coeficiente é estimado pelo *least square error* (LSE) e o processo de cálculo é o seguinte:

$$\begin{cases} \hat{y}(x_1, \dots, x_n) = G(x_i, x_j) \\ e = \sum_{n=1}^N (y - \hat{y})^2 \\ \frac{de}{da_k} = 0, \quad k = 1, 2, 3, 4, 5. \end{cases} \quad (87)$$

Para facilitar a análise, os resultados são expressos em forma de matriz:

$$a = (X^T X)^{-1} X^T y \quad (88)$$

em que,

$$X = \begin{pmatrix} 1 & x_{i1} & x_{j1} & x_{i1}x_{j1} & x_{i1}^2 & x_{j1}^2 \\ 1 & x_{i2} & x_{j2} & x_{i2}x_{j2} & x_{i2}^2 & x_{j2}^2 \\ 1 & x_{i3} & x_{j3} & x_{i3}x_{j3} & x_{i3}^2 & x_{j3}^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{in} & x_{jn} & x_{in}x_{jn} & x_{in}^2 & x_{jn}^2 \end{pmatrix}. \quad (89)$$

Comparando as aplicações de GMDH para a previsão de séries temporais, o trabalho de Dag e Yozgatligil (2016) atinge um MSE de 4,98 para a previsão. Como R^2 não é apresentado neste trabalho, não é possível fazer uma comparação global do método. Em EBTEHAJ *et al.* (2015) os seguintes valores foram obtidos, 5,263, 0,038, 0,779, para RMSE, MAPE e R^2 , respectivamente. Esses resultados são considerados satisfatórios em comparação com outros modelos de previsão para aplicações equivalentes.

Outros autores pesquisaram a aplicação do GMDH para previsões em diversas áreas da engenharia. No trabalho de Ebtehaj *et al.* (2015) o GMDH foi aplicado para prever o coeficiente de descarga em uma barragem lateral. Como neste trabalho, dois conjuntos de dados foram utilizados para a realização do treinamento e teste da rede. Assim, a validação cruzada foi utilizada para os resultados apresentados. Os resultados mostram que o GMDH é mais preciso do que o modelo de rede neural feed-forward e as equações de regressão não linear existentes para previsão.

Com base nos diferentes graus de dados históricos, Liu *et al.* (2018b) avalia o processamento do peso variável nos dados de entrada da previsão de carga. Nessa avaliação, é utilizado um algoritmo híbrido baseado em *elastic net* (EN), definido como EN-GMDH. O EN-GMDH é usado para estabelecer a relação entre as variáveis e a carga, e uma previsão de carga de curto prazo é estabelecida. Liu *et al.* (2018b) mostra que a aplicação híbrida de GMDH é superior à sua aplicação tradicional. Conforme apresentado nesta pesquisa para uma das aplicações, o MAE reduz de 2,744 para 2,223, e o MAPE reduz de 3,657 para 2,698.

Também usando um GMDH híbrido, Shen *et al.* (2019) une o GMDH à PCA. Um pré-processamento para redução de dimensão é executado e a análise de wavelet é aplicada. Mostra-se que a aplicação do algoritmo híbrido, definido como PCA-GMDH, aumenta a precisão e distingue todas as classes de litofácies presentes no estudo.

Com o objetivo de tentar desenvolver modelos eficientes de detecção de falhas, diversos autores desenvolveram modelos híbridos. Fezai *et al.* (2019) propôs uma técnica eficaz chamada *online reduced kernel generalized likelihood ratio test* (OR-KGLRT) para detecção de falhas em sistemas fotovoltaicos. Por meio do uso de estruturas kernel para mapear as não linearidades dos dados, a abordagem proposta torna-se interessante. Dados simulados foram adotados para avaliar a eficácia do modelo desenvolvido. Segundo os autores, o método proposto reduz a complexidade computacional enquanto conserva a estrutura dos dados no espaço de recursos, especialmente para grandes conjuntos de dados.

Wang *et al.* (2019a) desenvolveu a modelagem probabilística para intervalo de saída de matrizes fotovoltaicas para sua aplicação no diagnóstico de falhas. Os problemas de incerteza e não linearidades podem ser resolvidos por meio de análises probabilísticas. Jufri, Oh e Jung (2019) propuseram um modelo em cascata para detecção de condição fotovoltaica anormal. Primeiramente, o modelo de regressão é usado para estimar a geração de energia esperada sob a respectiva irradiância solar, que é usada como entrada para o modelo de máquina de vetor de suporte. Os resultados mostraram que a estrutura proposta pode distinguir com sucesso a condição normal, bem como identificar a condição anormal de um sistema fotovoltaico usando as medidas básicas.

Eom *et al.* (2019) avaliaram o uso de uma *convolutional neural network* (CNN) para detecção de falha de carga em uma bomba de ar para fonte de calor. Dois modelos preditivos para classificação e regressão foram propostos para prever a variação do sistema tanto no resfriamento

quanto no aquecimento. Os hiperparâmetros CNN foram definidos por uma grade de pesquisa, atingindo resultados 99% de precisão e um erro de 3,1% para a tarefa proposta.

Alguns autores realizaram trabalhos específicos para avaliar o sistema de distribuição elétrica. Com o objetivo de detectar falhas de alta impedância que causam degradação do isolamento, Bakar *et al.* (2014) apresentam um estudo baseado na transformada wavelet para extrair as características da falha das tensões medidas. Este método é aplicado para detectar e localizar falhas em isoladores contaminados. Amini e Sedighi (2014) apresentam um método para determinar o nível de contaminação da superfície do isolador. A transformada wavelet e o PCA são usados para extrair características da linha de corrente com base na corrente de fuga.

Considerando a aplicação específica apresentada por Stefenon *et al.* (2020), os resultados também são inferiores a aplicação do wavelet GMHD. Os valores de erro calculados por RMSE, MAE, MAPE são maiores e o valor de precisão é menor. De todo modo, a aplicação da transformada wavelet é muito promissora em conjunto com o sistema neuro-fuzzy, melhorando o R^2 de 0,9669 para 0,9700 no teste da rede para o melhor conjunto de dados.

Nos trabalhos de Bakar *et al.* (2014) e Amini e Sedighi (2014) a análise é baseada apenas na extração de características; nenhum método melhorado para avaliação de sinal é usado. Como o sinal de ultrassom não é linear, apenas a extração das características não é suficiente para realizar uma análise de previsão adequada. Assim, é necessário utilizar algoritmos baseados em inteligência artificial para realizar uma análise do sinal. A utilização do ultrassom para inspeções no sistema elétrico é muito conveniente, pois pode ser realizada sem contato com a rede elétrica e a uma distância segura, possibilitando, portanto, uma análise não destrutiva.

Como nos trabalhos de Bakar *et al.* (2014) e Amini e Sedighi (2014) os valores são apenas para a função da transformada wavelet, não são apresentadas métricas para avaliar a assertividade na previsão. Em Velasco, Amaris e Alonso (2019) a precisão no melhor caso foi de 0,9365, sendo inferior ao GMDH. Algoritmos que usam muito processamento não necessariamente resultam em avaliações mais precisas, e dependendo das características do sinal, uma classe de algoritmos pode ser mais apropriada que outra (KASBURG; STEFENON, 2019). Muitos autores têm usado técnicas de aprendizado profundo para prever séries temporais, como Wang *et al.* (2019b).

A escolha do algoritmo adequado depende muito das características do sinal a ser analisado, considerando que sinais com não linearidades requerem algoritmos apropriados para avaliação (VELASCO; AMARIS; ALONSO, 2019). De acordo com Li *et al.* (2020) e Paiva *et al.* (2020), os

algoritmos híbridos resultaram em um desempenho superior aos algoritmos clássicos. Apenas a escolha do algoritmo indicado para o problema proposto não necessariamente resulta em sucesso na aplicação, uma vez que é necessário configurar o algoritmo adequadamente (ALHASSAN *et al.*, 2020).

4.6 CONSIDERAÇÕES FINAIS DO CAPÍTULO

Neste capítulo uma revisão sobre os algoritmos para previsão de séries temporais foi apresentada. O LSTM se mostrou uma arquitetura relevante para este tipo de análise, sendo aplicada por diversos autores. O GMDH se mostrou uma técnica promissora por desconsiderar neurônios que não resultam em uma melhoria na capacidade de previsão do sinal.

O próximo capítulo terá o foco na descrição da configuração para análise de laboratório para aplicação dos modelos apresentados nesta tese. Com base nos fundamentados apresentados nos capítulos anteriores será proposto um modelo otimizado para classificação das condições dos isoladores.

5 OBTENÇÃO DE DADOS E MÉTODO PROPOSTO

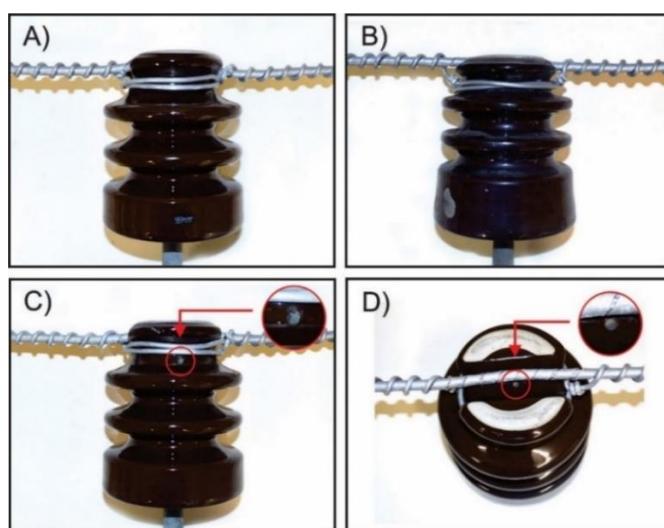
Nesse capítulo serão apresentadas as amostras consideradas neste estudo, assim como a configuração do laboratório de alta tensão para análise das condições aos quais os isoladores serão avaliados. O método proposto com base na combinação das melhores técnicas consideradas para resolução do problema em questão será apresentado.

5.1 AMOSTRAS

Nas redes de distribuição localizadas no sul do Brasil, os principais isoladores utilizados são do tipo pino, que podem ser de um corpo simples ou duplo, dependendo sua classe de tensão e projeto. Este perfil de isolador tem seu pino de suporte conectado na parte isolante e a base do pino fixada na cruzeta. Para o estudo, foram selecionadas amostras de isoladores tipo pino, com características representativas das principais condições que são encontrados em campo.

As amostras de isoladores de perfil corpo simples são apresentadas na Figura 24, para as condições: A) novo e limpo, B) contaminado, C) perfurado com furo lateral e D) perfurado com furo superior. No perfil corpo simples as perfurações foram realizadas, na lateral próximo da amaração do condutor e do topo até o pino de suporte.

Figura 24 - Isoladores de perfil corpo simples: A) limpo, B) contaminado, C) furo lateral, D) furo superior.



Fonte: Elaborada pelo autor (2021).

Os isoladores contaminados foram retirados da rede de distribuição da cidade de Pomerode em Santa Catarina. Estes isoladores apresentam contaminação biológica, além de possuírem acúmulo de poeira típica de ruas não pavimentadas. A contaminação biológica destas amostras está incrustada na superfície isolante, e isto ocorre devido a permanência de contaminantes durante um longo período em contato com o isolador, em conjunto com a ocorrência de descargas parciais.

Para simular uma perfuração nos isoladores, causada por descargas disruptivas, foram feitos furos utilizando uma furadeira de bancada com broca de 5 mm de diâmetro. Este tipo de característica é dificilmente identificada em inspeções do sistema elétrico pois as perfurações são pequenas e podem ficar escondidas sob o cabo ou amarrações.

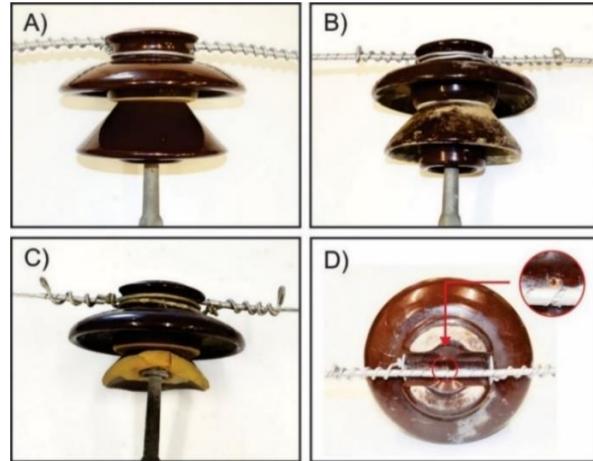
Devido as redes de distribuição estarem em ambientes abertos, ficam suscetíveis a vandalismo e variações ambientais que podem reduzir a capacidade de isolação destes componentes. Um isolador com saia quebrada também foi retirado do sistema durante a inspeção. Esta condição se deu por provável vandalismo, em que este componente ficou parcialmente destruído. Foi percebido que além de estar quebrado, o isolador possuía marcas de descargas, o que deixa claro que este tipo de isolador deve ser substituído.

Outro problema comum que é percebido em redes rurais é a presença de ninho de pássaros sobre as cruzetas, que podem reduzir significativamente a distância entre o potencial (cabo) e o terra (cruzeta). Como esta característica é facilmente visualizada e não é especificamente uma condição do isolador, não foi considerada como amostra nas análises (utilizando inteligência artificial).

As amostras de isoladores de perfil corpo duplo são apresentadas na Figura 25, para as condições: A) novo e limpo, B) contaminado, C), saia inferior quebrada D) perfurado com furo superior. O isolador quebrado, além de estar visivelmente danificado, possuía marcas de descargas próximas a sua amarração.

Os isoladores utilizados nesta tese, foram selecionados com base em inspeções de campo com o objetivo de ter na avaliação de condições representativas equivalentes as encontradas nas redes de distribuição do sul do Brasil. Os isoladores quebrados e perfurados, também possuem contaminação incrustada em sua superfície, por serem componentes retirados do sistema elétrico.

Figura 25 - Isoladores de perfil corpo duplo: A) limpo, B) contaminado, C) furo superior, D) quebrado.



Fonte: Elaborada pelo autor (2021).

5.2 CONFIGURAÇÃO DE LABORATÓRIO

As amostras foram montadas sobre um poste a 1,5 m do solo, em uma cruzeta de 2 m de comprimento, conforme apresentado na Figura 26. Foram utilizadas condições de instalação que simulam as condições em campo. A cruzeta foi instalada conforme a estrutura tipo *NI* e as amarrações também adotaram o padrão da concessionária. Foi aplicada a tensão de 13,8 kV *root mean square* (RMS), correspondente à tensão de fase da rede de 23,8 kV.

Figura 26 - Configuração para análise em laboratório.



Fonte: Elaborada pelo autor (2014).

Para análise ultrassônica foi utilizado o detector de ultrassom modelo 250 do fabricante Radar Engineers, que é um equipamento próprio para análise de emissão de ultrassom em redes elétricas (STEFENON *et al.*, 2017). Este equipamento possui uma saída de áudio, que indica a presença de ruído ultrassônico para o operador através de um sinal audível. Esta saída foi conectada a um computador que fez o registro do sinal gerado nos experimentos e que, posteriormente, foi utilizado nas análises.

O detector de ultrassom foi instalado sobre um tripé à 2,2 m de distância das amostras, sendo esta distância fixada para todas as medições, como pode ser observado na Figura 27. A fim de garantir que o sinal gravado não tivesse a interferência de outras fontes, todos os equipamentos elétricos que não faziam parte do experimento foram desligados durante a análise, inclusive a iluminação.

Figura 27 - Desenho esquemático da análise em laboratório.



Fonte: Elaborada pelo autor (2021).

5.2.1 Calibração do equipamento de ultrassom

O equipamento de ultrassom modelo 250 da Radar Engineers, apresentado na Figura 28, possui uma escala de ganho de 0 até 8, que define sua sensibilidade durante a medição. Para

determinar qual o ganho adequado para uma análise é necessário haver um operador experiente, visto que um alto ganho satura a medição e um baixo ganho pode não ser suficiente para evidenciar uma falha.

Figura 28 - Equipamento de ultrassom utilizado em laboratório.

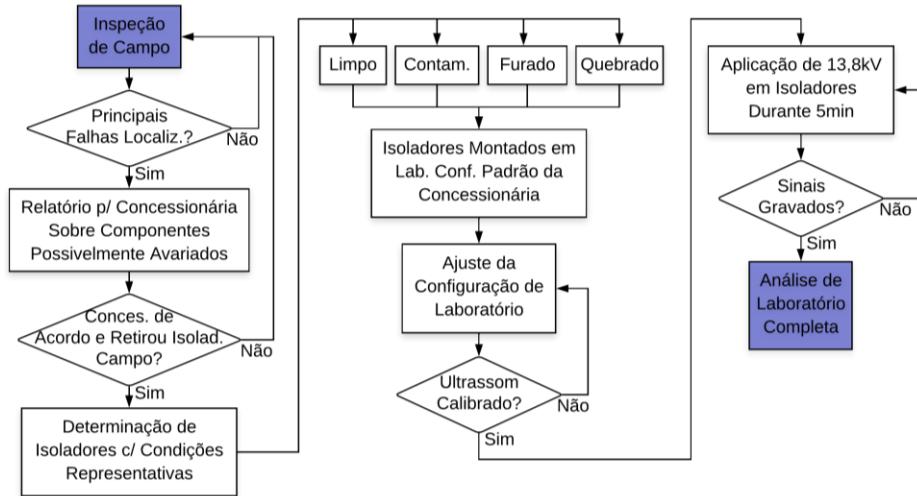


Fonte: Elaborada pelo autor (2021).

A partir de várias inspeções de campo e testes em laboratório, percebeu-se que o ganho adequado nesta análise deve ser entre 0,5 e 2,0. Com isto, em todas as análises o sinal foi gravado com o ganho de 0,5, 1,0 e 2,0. Mesmo com o ganho ajustado dentro deste nível, se o ambiente estiver úmido pode ocorrer saturação do equipamento devido as descargas parciais presentes no local. O desenvolvimento completo da análise em laboratório é apresentado na Figura 29.

Por haver um grande volume de dados a serem analisados, o sinal foi gravado com frequência de amostragem de 48 kHz durante 5 minutos, com isto o sinal bruto completo possui 14.4×10^6 pontos a serem analisados. Esta frequência foi definida em função da capacidade de captura de áudio da placa de som do computador utilizado durante está análise em laboratório. Considerando que o equipamento foi projetado para emitir um sinal audível para o operador, que atinge um máximo de 20 kHz, a frequência de amostragem garante a captura de toda a informação necessária para análise.

Figura 29 - Fluxograma da configuração de laboratório.



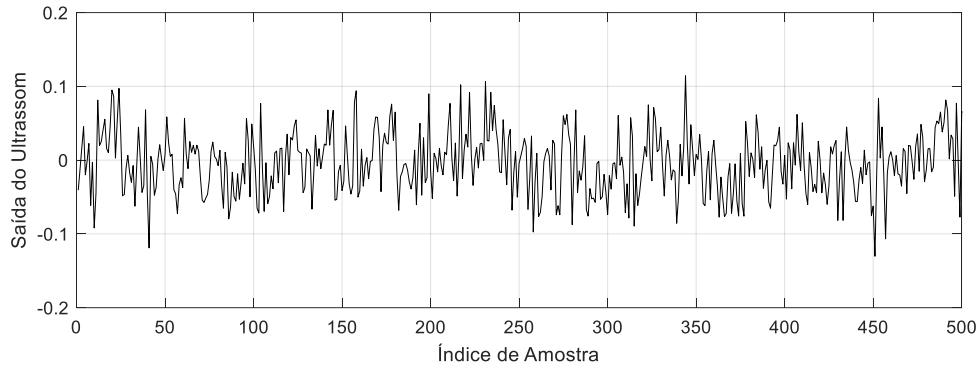
Fonte: Elaborada pelo autor (2021).

5.2.2 Configuração de software

O sinal gerado pelo equipamento de ultrassom foi transmitido para o computador através de uma saída de áudio do tipo philips P2, do detector ultrassônico para entrada de áudio do computador. O software LabVIEW foi utilizado para gravar o sinal. Após ser gravado o sinal foi processado através do software MATLAB, em que se realizaram as análises considerando todas as medições. Todas as análises desta tese foram realizadas considerando a utilização de um computador pessoal, Intel Core i7-3520M com 8 GB de *random-access memory* (RAM) e uma placa de vídeo dedicada GeForce GT 640M NVIDIA.

A partir das condições de falha das amostras foram desenvolvidos procedimentos de processamento de dados para identificação do padrão de falhas, ou seja, a assinatura de uma condição de defeito foi obtida. Um sinal típico de uma medição ultrassônica para um isolador novo e limpo é apresentado na Figura 30, com 500 pontos, correspondentes a 0,1 segundos de medição. Este sinal está normalizado para as análises comparativas em relação as outras condições.

Figura 30 - Sinal original gerado pelo equipamento de ultrassom.



Fonte: Elaborada pelo autor (2021).

5.3 MÉTODO PROPOSTO PARA CLASSIFICAÇÃO

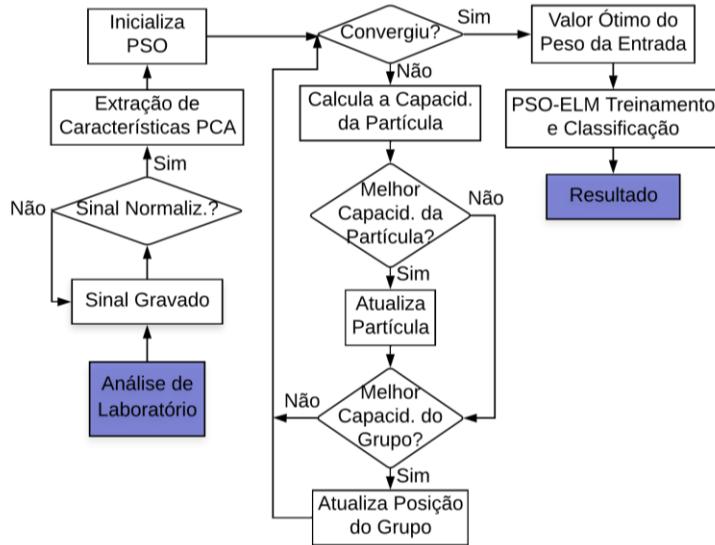
Com base na necessidade de ter um modelo robusto e eficiente para classificação de condições adversas em isoladores de distribuição, foi proposto nesta tese o *optimized ensemble extreme learning machine* (OEN-ELM). No algoritmo proposto, inicialmente se realiza a extração das características, especificamente é aplicada a transformada wavelet (3.2.2) para redução do ruído, em seguida é utilizada a análise da componente principal (3.2.3) para redução das dimensões do sinal. Após este tratamento inicial dos dados, é realizada a valiação cruzada a fim de ter uma análise mais completa do problema em relação a variabilidade dos dados.

Para a valiação cruzada foi utilizada a técnica *k-fold cross-validation* (3.3). A média das precisões resultantes da valiação cruzada de *k-fold* resulta na avaliação de desempenho de classificação do algoritmo. Esse método é usado para intensificar o estágio de treinamento da classificação, garantindo mais exposições a exemplos desconhecidos, apesar das amostras serem pertencentes ao conjunto de treinamento. Para uma padronização nas análises se utilizou um $k=10$ na valiação cruzada.

Para realizar a classificação foi utilizado o modelo EN-ELM (3.4.3) que se destaca por ser um modelo que exige menor esforço computacional. A ideia geral do classificador *ensemble* é melhorar o desempenho geral da rede. Nesta tese foram aplicados dois classificadores estimados pela saída e obtidos por processo de votação simples. No caso de empate ao usar um número par de classificadores, foi decidido que a classe vencedora é a “classe verdadeira”. Para otimizar o

modelo foi utilizado o *particle swarm optimization* (3.5.5). Um resumo deste algoritmo pode ser visualizado na Figura 31.

Figura 31 - Fluxograma do algoritmo para treinamento da rede OEN-ELM.



Fonte: Elaborada pelo autor (2021).

5.3.1 Configuração do conjunto

Cada simulação testou uma combinação de variáveis distintas considerando: i) número de neurônios ocultos; ii) número de neurônios não lineares; e iii) número de neurônios lineares. O conjunto selecionado para este trabalho assume dois ELMs. Durante os experimentos computacionais, a validação cruzada (*10-fold*) foi utilizada para avaliar o desempenho dos classificadores e conjuntos de dados. No total 4 conjuntos de dados foram utilizados para as simulações.

Os conjuntos foram divididos em 2 grupos de 2 conjuntos de dados, sendo no grupo 1, 40% das amostras foram usadas para treinar o classificador e 60% usadas para teste. No grupo 2, 60% das amostras foram usadas para treinamento do classificador e 40% usadas para testes. A Tabela 1 apresenta as características dos 4 conjuntos de dados criados para este trabalho, onde 2 deles tiveram o método PCA aplicado para redução de dimensionalidade.

Tabela 1 - Descrição da configuração de dados para as simulações.

	Dados #	Utilização de PCA	Amostras p/ treinamento	Amostras p/ teste	Total de amostras
Group 1	1	Sim	800	1200	2000
	2	Não	800	1200	2000
Group 2	3	Sim	1200	800	2000
	4	Não	1200	800	2000

Fonte: Elaborada pelo autor (2021).

Conforme Li, Li e Rong (2013), o uso de mais de um tipo de função de ativação pode melhorar a capacidade da rede neural de capturar as características do sistema analisado, então foram utilizadas duas funções de ativação. A primeira função de ativação utilizada no EN-ELM, foi proposta por Wang, Li e Cao (2012) e é chamada de função de ativação variável apresentada em (90).

$$\delta_{TAF}(n) = \frac{\zeta}{1 + e^{-\frac{n-\zeta}{\mu}}} + \eta, \quad (90)$$

em que ζ é o fator de mapeamento que determina o comprimento e o alcance do mapeamento, η é o deslocamento vertical, ζ é o parâmetro de posicionamento horizontal e μ é o fator do ângulo de inclinação, que define a forma do $\delta_{TAF}(.)$. Se for utilizado o parâmetro $\eta = \zeta = 0$ e $\zeta = \mu = 1$, irá resultar em uma função sigmoide.

A segunda função de ativação utilizada nesta tese é descrita através do mapeamento linear, em que k_{lin} é um número constante positivo, calculada pela equação:

$$\delta_{lin}(n) = nk_{lin}. \quad (91)$$

A função objetivo utilizada neste modelo, também chamada de *fitness function*, é composta pelo valor médio das medidas da *F-measure* obtidas a partir dos *k-folds* e dos valores médios do módulo β para o peso de saída, descrito a seguir:

$$Fitness \#1 = \frac{k1}{kFold} * \sum_{i=1}^{kFold} Facc_i + \frac{(1 - k1)}{\tilde{N}} * \sum_{k=1}^{\tilde{N}} \sqrt{\beta_k^2}, \quad (92)$$

em que $k1 \in [0, 1]$ controla a influência da precisão e do peso de saída no valor da função. Após realizar 10 experimentos diferentes, alterando o valor de $k1$ entre 0 e 1 com um passo de 0,1, o melhor desempenho foi alcançado adotando-se $k1 = 0,2$.

Esta função foi projetada para maximizar a precisão da classificação e reduzir o valor absoluto do peso de saída. Ao procurar valores menores, é possível minimizar a variação da amplitude do sinal durante sua propagação através do neurônio artificial. Além disso, também pode

diminuir o consumo de energia em uma possível implementação em um sistema embarcado. Outra *fitness function* foi avaliada nesta tese apresentada em (93), sendo que esta função não utiliza os pesos de saída da rede neural artificial.

$$Fitness \#2 = \frac{1}{kFold} * \sum_{i=1}^{kFold} Facc_i . \quad (93)$$

5.3.2 Metodologia adotada

Nesta tese se adotou uma metodologia quantitativa, baseada na comparação entre os modelos aplicados para classificação e predição. Nesta pesquisa a análise é realizada com base na comparação numérica de desempenho dos modelos avaliados (NEVES, 1996).

Os modelos de previsão de séries temporais foram avaliados para melhorar a capacidade de lidar com dados em uma série temporal, e a partir das melhores técnicas de tratamento dos dados um modelo classificador foi proposto. Os resultados do modelo proposto foram comparados com variações de modelos bem consolidados para avaliar suas capacidades de classificação.

5.4 CONSIDERAÇÕES FINAIS DO CAPÍTULO

Neste capítulo os métodos utilizados para análise em laboratório foram apresentados em detalhes, além do modelo proposto OEN-ELM. Com base nos capítulos anteriores as técnicas combinadas tornam o modelo proposto promissor, por ser mais rápido e eficiente que técnicas baseadas em *deep learning*. No próximo capítulo os resultados das análises realizadas nesta tese serão apresentados e discutidos.

6 ANÁLISE DOS RESULTADOS

Os resultados são apresentados e discutidos neste capítulo seguindo a sequência de apresentação das técnicas descritas nas seções anteriores. Para padronização das comparações, os valores sublinhados representam os melhores resultados de cada configuração e os valores em negrito e sublinhado são os melhores resultados gerais da análise.

6.1 RESPOSTA DO ESPECTRO HARMÔNICO

Inicialmente, para realizar a análise do espectro harmônico por meio da FFT, são utilizados dois ganhos no equipamento de ultrassom, os quais representam a sensibilidade a qual o equipamento está configurado. Os valores gravados em análise são apresentados na Tabela 2 para o ganho do equipamento regulado em 0,5 e na Tabela 3 com ganho de 1,0 de forma percentual, em relação a frequência fundamental de 60 Hz para as amostras consideradas.

Tabela 2 - Espectro harmônico do sinal do equipamento de ultrassom (ganho 0,5).

Condição	120Hz (%)	180Hz (%)	240Hz (%)
Isolador limpo e novo	66,4704	10,5623	112,3341
Isolador contaminado	55,7486	53,8046	41,8069
Isolador perfurado	57,3024	15,0279	87,7243
Isolador quebrado	54,8266	13,4764	80,9988

Fonte: Elaborada pelo autor (2021).

O número de entradas da rede se deu pela característica do sinal gravado, sendo que foram utilizadas 4 entradas na análise FFT, considerando as respostas em frequência de 60 Hz, 120 Hz, 180 Hz e 240 Hz. Na apresentação do resultado das FFTs não é considerada a frequência de 60 Hz, tendo em vista que esta é a fundamental e desta forma foi utilizada para normalização da resposta em frequência tendo seu resultado fixado em 100% em todas as análises para fins de comparação.

Nesta perspectiva a resposta em frequência é comparada em relação a frequência fundamental. Para realização da análise, o sinal bruto foi segmentado em 4.800 pontos, correspondente à um período de 100 ms. No algoritmo utilizado sem as características da FFT, foi avaliado o sinal bruto.

Tabela 3 - Espectro harmônico do sinal do equipamento de ultrassom (ganho 1,0).

Condição	120 Hz (%)	180 Hz (%)	240 Hz (%)
Isolador limpo e novo	19,3685	40,5249	66,0062
Isolador contaminado	33,9793	6,0272	145,6062
Isolador perfurado	25,7262	4,8721	115,0087
Isolador quebrado	84,4424	91,7504	652,8794

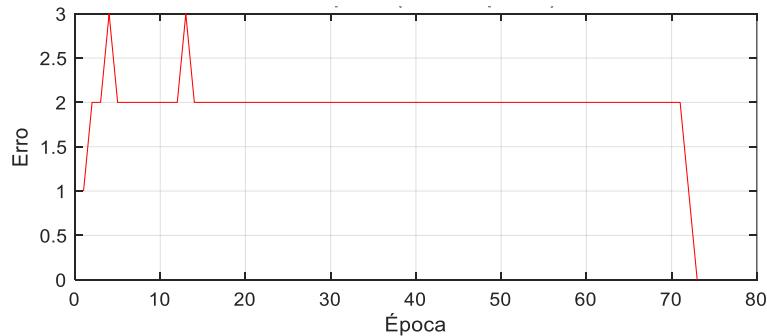
Fonte: Elaborada pelo autor (2021).

6.2 PERCEPTRON CAMADA ÚNICA

Devido as características não lineares do sinal, em diversas configurações o algoritmo *single layer perceptron* (SLP) não convergiu. Neste trabalho, o SLP é baseado em uma saída binária, e para esta aplicação, 1 representa um componente danificado que precisa ser substituído e 0 representa um componente que está em boas condições.

O tempo necessário para convergência do algoritmo variou entre 0,06 e 0,09 segundos e a evolução do cálculo do erro em relação as iterações é apresentada na Figura 32. Na rede SLP em poucas iterações há convergência do algoritmo, contudo isto não representa que o padrão foi suficientemente aprendido, apenas gera um algoritmo rápido para o problema analisado.

Figura 32 - Evolução do erro da rede Perceptron camada única.



Fonte: Elaborada pelo autor (2021).

O treinamento na rede Perceptron camada única foi realizado com ganho do equipamento igual a 0,5, pois com ganho igual a 1,0 não houve convergência. A validação dos resultados foi realizada com passo de 10 pontos com base na característica extraída em cada cálculo de FFT. O resultado desta análise é apresentado na Tabela 4.

Nesta análise foi utilizada a FFT pois os resultados de classificação utilizando o sinal bruto foram inferiores. Isto ocorre porque este sinal possui não linearidades. A rede Perceptron camada única não é adequada para classificar padrões não lineares.

Tabela 4 - Validação de resultados FFT da rede Perceptron camada única (10/250/0,5)⁴.

Condição	Precisão (%)
Isolador limpo e novo	57,60
Isolador contaminado	36,80
Isolador perfurado	87,20
Isolador quebrado	8,00

Fonte: Elaborada pelo autor (2021).

A alteração no peso das entradas ou função de ativação resultou em diferentes velocidades de convergência, entretanto, a validação dos resultados se manteve baixa utilizando o Perceptron camada única. Nesta avaliação o isolador quebrado teve uma precisão de classificação consideravelmente inferior aos demais.

6.3 PERCEPTRON MULTIPLAS CAMADAS

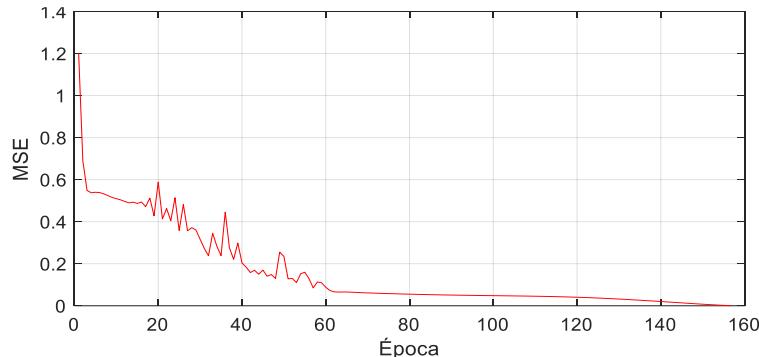
A rede de múltiplas camadas tem a capacidade de classificar dados não linearmente separáveis por uma reta. Neste modelo, a incialização dos pesos da camada oculta ocorre de forma aleatória. Para uma avaliação inicial do método, a função de ativação utilizada para a rede MLP foi a tangente hiperbólica e para saída da rede foi utilizada a função de ativação linear. O critério de convergência da rede MLP foi baseado na utilização do MSE. A evolução do MSE, em relação as iterações, está apresentada na Figura 33.

Para comparar os métodos de treinamento foram utilizadas 15 configurações de redes neurais, sendo avaliadas as variações de 1, 2 e 3 camadas ocultas e de 5, 10 e 15 neurônios em cada camada oculta. Para uma avaliação comparativa entre os métodos de otimização foi utilizada a função de ativação tangente hiperbólica nas camadas ocultas e na saída da rede foi utilizada a função linear. Para esta avaliação comparativa foi definido o critério de parada de tempo de treinamento de 600 segundos (10 min) ou 1000 iterações.

⁴ Configuração da análise (passo para o cálculo da FFT/janela de amostragem considerada/ganho do equipamento).

O valor de 300.000 pontos foi definido com o total do sinal para o treinamento, considerando que dentro deste volume de dados é possível capturar plenamente a característica do sinal. Isto corresponde a somente 6,25 segundos de gravação do equipamento, com taxa de 48 kHz.

Figura 33 - Evolução do erro quadrático médio da rede MLP.



Fonte: Elaborada pelo autor (2021).

Na análise inicial do Perceptron multiplas camadas, são apresentadas as variações do modelo em relação ao método de otimização e configuração dos parâmetros da rede neural. Esta análise é realizada na fase de treinamento a partir da variação no número de camadas e neurônios. A apresentação dos resultados na fase de treinamento é realizada em função do desempenho do modelo em relação ao método de otimização. Em seguida, é avaliada a função de ativação, e por fim, são apresentados os resultados do teste deste algoritmo.

6.3.1 BFGS quasi-Newton

Para análise do sinal bruto gravado durante as análises, o método BFGS quasi-Newton atingiu o tempo máximo de análise com apenas 50 iterações utilizando 600 pontos do sinal. Devido a maior necessidade de processamento do método, foi possível avaliar apenas 0,2% do sinal completo de 300.000 pontos.

Apesar do método ter um bom desempenho em relação a se aproximar do mínimo com poucas iterações, as iterações são consideravelmente lentas e com isto o tempo máximo para o treinamento é alcançado antes que ocorra a convergência, deixando o algoritmo com uma precisão insatisfatória. Utilizando mais que 1.000 pontos o algoritmo não ultrapassa algumas interações e tem uma precisão ainda inferior. Considerando 600 pontos, a precisão da rede calculada através de regressão foi satisfatória conforme pode ser visualizado na Tabela 5.

Utilizando o sinal completo não foi possível realizar muitas iterações e desta forma ficou claro que este método não é indicado para ser utilizado para este formato e volume de dados. Nesta análise, o tempo de validação é o tempo necessário para que haja a avaliação da regressão da rede.

Tabela 5 - Treinamento Utilizando o Método BFGS quasi-Newton.

Número de camadas ocultas	Neurônios nas camadas ocultas	Tempo treinamento (s)	Tempo validação (s)	Precisão (%)
1	5	551,330	0,034	97,590
	10	-	0,029	99,843
	15	-	0,038	39,526
2	5	-	0,056	97,590
	10	-	0,047	95,931
	15	-	0,053	44,434
3	5	460,310	0,052	99,999
	10	-	0,054	69,998
	15	-	0,052	74,401

Fonte: Elaborada pelo autor (2021).

6.3.2 Levenberg-Marquardt

Através do método LM foi possível analisar 2.400 pontos do sinal completo (0,8%). Utilizando mais que 2.500 pontos, a RNA atingiu o tempo máximo de treinamento sem que houvesse convergência, e desta forma, o resultado de precisão para o treinamento foi baixo. Mesmo utilizando um menor volume de dados (2.400 pontos), a rede não convergiu com 15 neurônios antes que o limite de tempo de treinamento fosse alcançado, conforme pode ser visto na Tabela 6.

Tabela 6 - Treinamento Utilizando o Método LM.

Número de camadas ocultas	Neurônios nas camadas ocultas	Tempo treinamento (s)	Tempo validação (s)	Precisão (%)
1	5	95,107	0,058	99,999
	10	-	0,558	97,006
	15	-	-	-
2	5	234,672	0,063	99,999
	10	-	0,294	89,012
	15	-	-	-
3	5	184,473	0,060	99,999
	10	-	0,316	60,958
	15	-	-	-

Fonte: Elaborada pelo autor (2021).

6.3.3 Gradiente descendente

A partir da utilização do método gradiente descendente foi possível analisar o sinal completo (300.000 pontos), tendo uma capacidade de processamento dos dados consideravelmente superior (maior que 10 vezes). Apesar deste método ter uma convergência inferior em cada iteração, é possível realizar uma quantidade maior de iterações dentro dos limites de tempo de treinamento e com isso o algoritmo converge e tem resultados superiores de precisão para o treinamento, conforme pode ser visto na Tabela 7.

Tabela 7 - Treinamento Utilizando o Método Gradiente Descendente.

Número de camadas ocultas	Neurônios nas camadas ocultas	Tempo treinamento (s)	Tempo validação (s)	Precisão (%)
1	5	191,777	0,163	57,957
	10	364,539	0,374	89,087
	15	502,392	0,384	99,975
2	5	138,314	0,158	94,281
	10	378,040	0,242	94,491
	15	478,361	0,284	99,999
3	5	184,118	0,161	97,590
	10	250,650	0,216	99,999
	15	519,291	0,310	99,999

Fonte: Elaborada pelo autor (2021).

A capacidade do método gradiente descendente pode ser ainda melhorada com a utilização do *momentum* e taxa adaptativa, conforme será visto na próxima subseção.

6.3.4 Gradiente descendente com *momentum* e taxa adaptativa

Utilizando o *momentum* e taxa adaptativa o algoritmo tem uma melhora em precisão e velocidade como pode ser visto na Tabela 8. Através destas técnicas o algoritmo apresenta esta melhora, pois reduz a taxa de aprendizado à medida que se aproxima do mínimo, que torna o algoritmo mais preciso e veloz.

Tabela 8 - Treinamento Utilizando o Método Gradiente com *Momentum* e Taxa Adaptativa.

Número de camadas ocultas	Neurônios nas camadas ocultas	Tempo treinamento (s)	Tempo validação (s)	Precisão (%)
1	5	53,614	0,132	88,976
	10	151,883	0,209	99,999
	15	541,436	0,283	99,999
2	5	194,055	0,255	99,403
	10	294,187	0,222	97,590
	15	109,410	0,284	99,999
3	5	195,398	0,163	99,364
	10	343,637	0,213	99,999
	15	193,728	0,452	99,999

Fonte: Elaborada pelo autor (2021).

6.3.5 Variações do método gradiente

Variações no método de retropropagação do gradiente conjugado serão apresentadas nesta subseção com o foco em avaliar a possibilidade de melhoria no treinamento utilizando técnicas mais sofisticadas de atualização para o método gradiente. Usando o método de retropropagação do gradiente conjugado com o método de reinício de Powell-Beale (STEFENON *et al.*, 2020b), houve uma melhora considerável no tempo de treinamento, como pode ser visto na Tabela 9.

Na análise da variação do método gradiente, a avaliação estatística dada pelo desvio padrão, variância e covariância é calculada em relação a precisão. Na configuração da rede o primeiro valor representa a quantidade de camadas ocultas e o segundo valor representa a quantidade de neurônios utilizada no modelo.

Tabela 9 - Treinamento com reinicializações de Powell-Beale.

Configuração da rede	Tempo treinamento (s)	Precisão (%)	Desvio padrão	Variância	Covariância
1_05	23,00	95,33	$4,3 \times 10^{-2}$	$1,9 \times 10^{-3}$	$3,5 \times 10^{-4}$
1_10	76,02	99,93	$2,1 \times 10^{-3}$	$4,5 \times 10^{-6}$	
1_15	52,64	99,99	$2,7 \times 10^{-8}$	$7,0 \times 10^{-16}$	
2_05	30,20	96,55	$4,5 \times 10^{-2}$	$2,1 \times 10^{-3}$	
2_10	48,88	99,93	$2,0 \times 10^{-3}$	$3,8 \times 10^{-6}$	
2_15	54,18	99,99	$1,8 \times 10^{-5}$	$3,0 \times 10^{-10}$	

Fonte: Elaborada pelo autor (2021).

Embora haja uma melhora considerável no tempo de treinamento, a configuração de uma camada oculta com cinco neurônios manteve um resultado de baixa precisão, bem como o método clássico de descida de gradiente. Neste caso, a análise estatística também estava dentro do esperado, mostrando que este método permaneceu estável.

Na retropropagação do gradiente conjugado com atualizações de Polak-Ribiére (STEFENON *et al.*, 2020b) houve uma melhora na precisão, porém houve um aumento do tempo de treinamento em relação ao método com a reinicialização de Powell-Beale, conforme pode ser observado na Tabela 10.

Tabela 10 - Treinamento com atualizações de Polak-Ribiére.

Configuração da rede	Tempo treinamento (s)	Precisão (%)	Desvio padrão	Variância	Covariância
1_05	40,75	97,35	$2,1 \times 10^{-2}$	$4,5 \times 10^{-4}$	$6,7 \times 10^{-5}$
1_10	186,66	99,55	$8,8 \times 10^{-3}$	$7,7 \times 10^{-5}$	
1_15	58,42	99,99	$4,8 \times 10^{-8}$	2×10^{-15}	
2_05	41,03	98,34	$3,6 \times 10^{-2}$	$1,3 \times 10^{-3}$	
2_10	97,74	99,86	$2,8 \times 10^{-3}$	$7,9 \times 10^{-6}$	
2_15	72,27	99,42	$1,7 \times 10^{-2}$	$3,0 \times 10^{-4}$	

Fonte: Elaborada pelo autor (2021).

Em termos de variabilidade estatística, houve um resultado satisfatório, bem como os métodos de descida de Gradiente apresentados até aqui. A retropropagação do gradiente conjugado com as atualizações de Fletcher-Reeves (STEFENON *et al.*, 2020b) resultou em um aumento no tempo de treinamento necessário para a convergência em comparação com a reinicialização de Powell-Beale. Além disso, a precisão usando 5 neurônios foi insatisfatória com 1 e 2 camadas ocultas, conforme pode ser visualizado na Tabela 11.

Tabela 11 - Treinamento com atualizações de Fletcher-Reeves.

Configuração da rede	Tempo treinamento (s)	Precisão (%)	Desvio padrão	Variância	Covariância
1_05	128,06	92,81	$3,9 \times 10^{-2}$	$1,5 \times 10^{-3}$	$7,5 \times 10^{-4}$
1_10	114,16	99,81	$2,6 \times 10^{-3}$	$6,6 \times 10^{-6}$	
1_15	100,13	99,99	$2,1 \times 10^{-5}$	$4,0 \times 10^{-10}$	
2_05	43,32	93,98	$5,8 \times 10^{-2}$	$3,4 \times 10^{-3}$	
2_10	117,69	97,44	$7,0 \times 10^{-2}$	$4,9 \times 10^{-3}$	
2_15	64,28	99,98	$7,1 \times 10^{-4}$	$5,0 \times 10^{-7}$	

Fonte: Elaborada pelo autor (2021).

A retropropagação do gradiente conjugado (NOCEDAL; WRIGHT, 2006) teve a melhor resposta na precisão comparando com todos os algoritmos baseados em gradiente, com uma baixa variabilidade como pode ser visto na Tabela 12. Além disso, o algoritmo tem um resultado satisfatório em relação ao tempo de treinamento.

Tabela 12 - Treinamento do gradiente conjugado em escala.

Configuração da rede	Tempo treinamento (s)	Precisão (%)	Desvio padrão	Variância	Covariância
1_05	283,68	99,99	$1,1 \times 10^{-4}$	$1,3 \times 10^{-8}$	$7,0 \times 10^{-9}$
1_10	113,55	99,99	$4,0 \times 10^{-16}$	$2,0 \times 10^{-31}$	
1_15	77,30	99,99	$3,0 \times 10^{-16}$	$1,0 \times 10^{-31}$	
2_05	300,39	99,92	$2,1 \times 10^{-3}$	$4,4 \times 10^{-6}$	
2_10	168,60	99,99	$3,0 \times 10^{-15}$	$7,0 \times 10^{-30}$	
2_15	151,85	99,99	$7,0 \times 10^{-16}$	$5,0 \times 10^{-31}$	

Fonte: Elaborada pelo autor (2021).

6.3.6 Retropropagação resiliente

Nesta avaliação, a retropropagação resiliente é mais rápida do que os algoritmos de descida mais íngreme tradicional (NOCEDAL; WRIGHT, 2006). Isso acontece porque esse algoritmo requer apenas um aumento modesto na necessidade de memória. Os valores de atualização para cada peso e tendência são armazenados, assim como no método gradiente. Apesar de ter o melhor resultado em velocidade de treinamento, a precisão do algoritmo foi relativamente baixa. A grande variabilidade em relação à precisão gera um resultado estatístico mais baixo, como mostrado na Tabela 13.

Tabela 13 - Treinamento de retropropagação resiliente.

Configuração da rede	Tempo treinamento (s)	Precisão (%)	Desvio padrão	Variância	Covariância
1_05	53,95	82,41	$1,1 \times 10^{-1}$	$1,3 \times 10^{-2}$	$6,8 \times 10^{-3}$
1_10	33,19	93,11	$1,3 \times 10^{-1}$	$1,6 \times 10^{-2}$	
1_15	41,09	96,03	$8,5 \times 10^{-2}$	$7,3 \times 10^{-3}$	
2_05	34,12	70,76	$2,2 \times 10^{-1}$	$5,0 \times 10^{-2}$	
2_10	37,43	92,02	$1,1 \times 10^{-1}$	$1,2 \times 10^{-2}$	
2_15	37,01	97,57	$5,5 \times 10^{-2}$	$3,0 \times 10^{-3}$	

Fonte: Elaborada pelo autor (2021).

6.3.7 Função de ativação

As análises apresentadas nas seções anteriores são referentes a utilização da função de ativação tangente hiperbólica. Para avaliar as diferenças entre as funções de ativação uma comparação é apresentada na Tabela 14, utilizando o método gradiente descendente com *momentum* e taxa adaptativa para 3 camadas ocultas.

Tabela 14 - Avaliação da função de ativação sobre o treinamento.

Função de ativação	Neurônios nas camadas ocultas	Tempo treinamento (s)	Tempo validação (s)	Precisão (%)
Tangente Hiperbólica	5	97,040	0,135	97,590
	10	102,605	0,223	99,999
	15	528,029	0,277	99,403
Sigmoide	5	194,872	0,180	98,055
	10	267,021	0,215	99,999
	15	534,560	0,304	99,999
Linear	5	195,523	0,164	99,999
	10	169,411	0,217	99,999
	15	314,901	0,318	99,999
Linear com Saturação	5	68,208	0,149	76,842
	10	363,512	0,240	83,808
	15	531,066	0,280	98,494

Fonte: Elaborada pelo autor (2021).

A função de ativação linear com saturação teve resultados inferiores de precisão na regressão da rede na grande maioria das análises. Com exceção desta as outras funções de ativação tiveram um bom desempenho, tanto em velocidade quanto em precisão.

A utilização de 15 neurônios na camada oculta teve resultados superiores de precisão, contudo esta configuração foi a mais lenta para alcançar a convergência. Há um compromisso para escolha do número de neurônios para este problema, com a utilização de mais neurônios a rede se torna mais lenta, mas em contrapartida tem uma precisão superior.

6.3.8 Resultados do teste da rede MLP usando a FFT e o sinal sem tratamento

Durante o teste da rede MLP, a partir do ganho do equipamento igual a 0,5 não houve convergência devido à similaridade entre as harmônicas para este tipo de rede. Desta forma, para

o treinamento desta rede foi utilizado ganho de 1,0. A análise baseada na característica FFT tem uma velocidade consideravelmente superior de treinamento do que a apresentada para os dados brutos, conforme pode ser visualizado na Tabela 15.

Tabela 15 - Resultados do treinamento em relação ao número de neurônios ocultos.

Configuração	Resultado para convergência	
# Neurônios ocultos	Tempo (s)	Iterações
1	-	Não convergiu
2	-	Não convergiu
3	1,143	2737
4	0,462	1685
5	0,366	1279
6	0,248	977
7	0,296	987
8	0,300	996
9	0,339	1145
10	0,308	1063
11	0,259	1067
12	0,377	1271
13	0,432	1729
14	0,485	2002
15	0,604	2567
16	1,956	9408
17	-	Não convergiu

Fonte: Elaborada pelo autor (2021).

Neste caso, a alteração do número de neurônios na camada oculta também tem influência sobre a convergência, sendo que a partir de 16 neurônios o treinamento não apresentou convergência. O melhor tempo de convergência versus flexibilidade teórica ocorreu com a utilização de 6 neurônios na camada oculta (em negrito), sendo este o valor definido para esta análise. O teste da rede foi realizado comparando 250 respostas em FFT e o resultado desta análise é apresentado na Tabela 16.

Tabela 16 - Resultados de precisão na classificação com FFT da rede MLP (10/250/1,0).

Condição	Precisão (%)
Isolador limpo e novo	84,40
Isolador contaminado	91,60
Isolador perfurado	97,60
Isolador quebrado	35,60

Fonte: Elaborada pelo autor (2021).

De forma geral o algoritmo clássico de MLP teve um resultado razoável, muito melhor que o SLP. Possivelmente, a utilização de mais camadas aplicando *deep learning* melhoraria este resultado, pois seria possível a rede estar mais adequada às não linearidades do sinal. Utilizando o sinal bruto em que cada ponto de medição é considerada uma entrada, com um período de amostragem de 0,1 segundos que gera o equivalente à 4.800 pontos, o resultado é ainda inferior mesmo aplicando uma rede MLP, conforme pode ser visto na Tabela 17.

Tabela 17 - Resultados de precisão na classificação com sinal bruto da rede MLP.

Condição	Precisão (%)
Isolador limpo e novo	13,51
Isolador contaminado	13,47
Isolador perfurado	89,16
Isolador quebrado	90,97

Fonte: Elaborada pelo autor (2021).

Além das dificuldades nesta análise, devido ao fato de que o sinal é parcialmente baseado em falhas esporádicas que geram não linearidades, os isoladores perfurados e quebrados foram retirados do sistema com presença de contaminação, o que torna mais difícil a diferenciação desta característica.

6.4 MÁQUINA DE APRENDIZAGEM EXTREMA OTIMIZADA

As simulações consistiram em 50 experimentos diferentes de 100 iterações cada. Os parâmetros escolhidos para o algoritmo PSO foram o tamanho de enxame definido em 20 partículas, e o peso de inércia, que tem o comportamento de decaimento linear, começando em 0,9, decaindo para 0,4 durante as primeiras 70 iterações e mantendo 0,4 durante as últimas 30 iterações.

6.4.1 Avaliação da configuração do conjunto

A Tabela 18 mostra a razão média de respostas corretas para as amostras testadas. A primeira e segunda coluna descrevem a configuração da RNA para cada conjunto, onde o primeiro número é o número de neurônios ocultos, seguido pelo número de neurônios lineares e não lineares para cada conjunto.

Os resultados para as diversas configurações apresentadas na Tabela 18, mostram que o PSO+EN-ELM é capaz de manter seu desempenho de classificação, ou mesmo aumentá-lo quando o método PCA para redução de dimensionalidade é aplicado.

Tabela 18 - Resultado da precisão na classificação para o teste das amostras.

Neurônios ocultos	Configuração do conjunto	Precisão (%)			
		Dados #1	Dados #2	Dados #3	Dados #4
5	(1:4/1:4)	97,17	97,00	97,66	97,35
	(4:1/4:1)	97,19	96,93	97,97	97,32
	(1:4/4:1)	97,29	96,97	97,81	97,34
10	(2:8/2:8)	97,29	96,96	97,73	97,33
	(8:2/8:2)	97,18	96,96	97,95	97,30
	(5:5/5:5)	97,34	97,06	98,00	97,20
	(2:8/8:2)	97,32	96,98	98,02	97,27
Média		97,25	96,98	97,88	97,30

Fonte: Elaborada pelo autor (2021).

Ao analisar os resultados médios entre os conjuntos de dados usando o mesmo número de atributos, os melhores desempenhos foram obtidos quando 60% das amostras foram utilizadas para o treinamento. No entanto, a diferença entre o conjunto de dados 1 e o conjunto de dados 3 é de 0,63% e entre o conjunto de dados 2 e o conjunto de dados 4 é de apenas 0,32%, mostrando que um conjunto de dados de treinamento menor não reduz significativamente a precisão do método.

A maior diferença entre empregar conjuntos de treinamento maiores ou menores pode ser observada quando o tempo de treinamento é analisado. A Tabela 19 mostra o consumo médio de tempo para o treinamento da RNA. Como pode ser visto, a configuração que requer menos parâmetros a serem ajustados pelo algoritmo PSO tem o menor consumo de tempo.

Tabela 19 - Tempo médio necessário para o treinamento da RNA.

Neurônios ocultos	Configuração do conjunto	Tempo médio de treinamento (s)			
		Dados #1	Dados #2	Dados #3	Dados #4
5	(1:4/1:4)	137	138	655	668
	(4:1/4:1)	157	160	679	695
	(1:4/4:1)	147	150	672	682
10	(2:8/2:8)	546	551	743	759
	(8:2/8:2)	563	573	766	791
	(5:5/5:5)	567	588	786	812
	(2:8/8:2)	565	572	765	790

Fonte: Elaborada pelo autor (2021).

Considerando a precisão obtida usando a configuração mais rápida, e a melhor precisão obtida em geral, a diferença de precisão entre eles é de cerca de 0,85%, enquanto a diferença de tempo é de cerca de 558%. Isso significa que a escolha de uma RNA maior não melhorará significativamente o desempenho.

6.4.2 Análise estatística

Apesar do tempo maior necessário para realizar o processo de treinamento para obter a melhor precisão, executar um teste utilizando o método é bastante rápido, como pode ser visto na Tabela 20. Neste caso, foram consideradas as configurações das RNAs que resultaram nas melhores taxas, usando 5 e 10 neurônios ocultos no conjunto de dados 3. As amostras de teste foram divididas em lotes com tamanhos diferentes para demonstrar possíveis situações práticas. Para alcançar resultados estatisticamente significativos, foram considerados 50 experimentos diferentes de 100 iterações cada.

Tabela 20 - Tempo médio necessário para o teste da RNA.

Neurônios ocultos	Amostras por lote	Análise estatística (s)		
		Média	Desvio padrão	Variância
5 (4:1/4:1)	1	$6,002 \times 10^{-4}$	$1,295 \times 10^{-4}$	$3,670 \times 10^{-14}$
	400	$9,719 \times 10^{-4}$	$4,006 \times 10^{-5}$	$3,874 \times 10^{-16}$
	800	$1,462 \times 10^{-3}$	$3,430 \times 10^{-4}$	$1,177 \times 10^{-7}$
10 (2:8/8:2)	1	$5,482 \times 10^{-4}$	$1,357 \times 10^{-4}$	$2,558 \times 10^{-13}$
	400	$1,155 \times 10^{-3}$	$1,090 \times 10^{-4}$	$6,053 \times 10^{-15}$
	800	$1,491 \times 10^{-3}$	$4,284 \times 10^{-4}$	$1,835 \times 10^{-7}$

Fonte: Elaborada pelo autor (2021).

A Tabela 21 apresenta uma análise estatística sobre simulações que obtiveram o máximo de respostas corretas possíveis, mais especificamente sobre a otimização do valor de adequação do PSO, sobre a razão de acertos obtidos durante os 50 experimentos realizados e sobre o peso da saída com valores obtidos a partir do método ELM.

Nesta análise, em menos de 2 iterações, a *fitness function* 2 alcança seu valor mínimo, enquanto a *fitness function* 1 permanece reduzindo progressivamente a aptidão por um período mais longo. Excluindo-se os pesos de saída da função (*fitness function* 2), foram obtidos pesos de saída consideravelmente menores e variações menores de amplitude de sinal durante sua

propagação ao longo da RNA. Esses resultados foram confirmados pelos valores médios do peso de saída absoluto e pela soma de todos os pesos de saída absolutos.

Tabela 21 - Detalhes da otimização do algoritmo (OEN-ELM) utilizando 2 *Fitness Functions*.

Parâmetros	Fitness Func. #1	Fitness Func. #2
Melhor resultado:	0,095582	0,092926
Média:	0,732508	0,094055
Variância:	1,094896	$7,316949 \times 10^{-7}$
Desvio padrão:	1,046373	$8,553917 \times 10^{-4}$
Melhor relação:	0,9825	0,9825
Média de relação:	0,9802	0,9814
Relação de desvio padrão:	$9,784020 \times 10^{-2}$	$7,840892 \times 10^{-2}$
Relação de variância:	$9,572704 \times 10^{-3}$	$6,147959 \times 10^{-3}$
Média de peso de saída absoluto:	$1,409148 \times 10^{-2}$	$2,76058 \times 10^{-4}$
Soma do peso absoluto de saída:	$2,818297 \times 10^{-1}$	$5,52115 \times 10^{-5}$

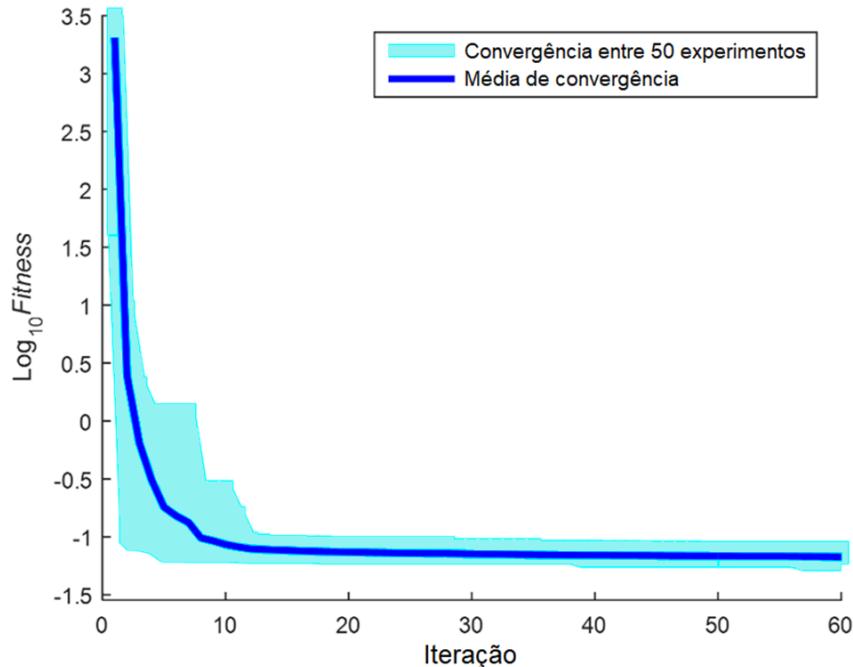
Fonte: Elaborada pelo autor (2021).

Com base nos indicadores estatísticos mencionados, pode-se verificar a robustez da PSO nesta aplicação, onde tanto a variância quanto o desvio padrão da razão de respostas corretas são pequenos, refletindo um valor médio próximo à melhor relação obtida em todos os experimentos. A *fitness function* 2 alcançou a melhor proporção média de respostas corretas, mas a qualidade da solução também deve se basear em seus valores de amplitude dos pesos.

Ao incluir o valor do peso de saída na função de adequação, o método de otimização tende a reduzir sua amplitude e reduz intrinsecamente os valores de amplitude dos pesos de entrada. A avaliação da convergência do algoritmo utilizando *fitness function* 2 é apresentada na Figura 34 para uma variação de 50 experimentos (em azul claro) e sua análise média (em azul escuro).

A partir da próxima subseção serão apresentadas as análises da previsão do sinal. Os resultados serão discutidos em relação a variação de parâmetros dos modelos avaliados. Será avaliada a utilização da transformada wavelet para redução do ruído do sinal.

Figura 34 - Avaliação da convergência do algoritmo em 50 experimentos.



Fonte: Elaborada pelo autor (2021).

6.5 ANÁLISE DE RESULTADOS ANFIS

Para configurar a estrutura de treinamento, alguns parâmetros foram definidos a fim de padronizar as análises. Definiu-se o número máximo de iterações igual a 1.000. Utilizou-se o algoritmo adaptativo com passo inicial igual a 0,01, taxa de diminuição igual a 0,9 e taxa de aumento igual a 1,1. Foi utilizado o método híbrido de otimização para o treinamento da rede, que é a combinação de *least-squares estimation* e *backpropagation*.

6.5.1 Avaliação da estrutura

Foram analisados três sistemas de estruturas de inferência fuzzy neste trabalho, sendo estes: estrutura de inferência fuzzy usando partição em grade, *cluster* subtrativo e FCM. Para realizar uma análise estatística o sinal foi dividido em cinco conjuntos de dados de 50.000 pontos. Na Tabela 22 são apresentados os valores com base na média das análises considerando inicialmente o sinal decomposto em pacotes wavelet até o terceiro nível com um nó para decomposição.

Na estrutura que utiliza o armazenamento em *cluster* subtrativo, especificou-se o alcance de influência de um centro de *cluster* em cada uma das dimensões de dados igual a 0,5. Este valor é aplicado a todas as dimensões de dados, ou seja, cada centro de *cluster* tem uma vizinhança esférica de influência com este raio.

Tabela 22 - Resultado usando diferentes estruturas ANFIS.

Estruturas	Tempo (s)	Média de Precisão (%)			RMSE	MAE	MAPE	Desvio Padrão
		Treino	Teste	Valid.				
Partição em grade	115,03	95,88	96,02	95,92	0,0155	$8,5 \times 10^{-3}$	0,9309	$2,48 \times 10^{-3}$
FCM clustering	158,52	96,35	96,43	96,37	0,0142	$6,3 \times 10^{-3}$	0,8580	$2,55 \times 10^{-3}$
Cluster subtrativo	956,59	96,78	96,85	96,86	0,0141	$9,6 \times 10^{-4}$	0,1737	$2,22 \times 10^{-3}$

Fonte: Elaborada pelo autor (2021).

Analizando o desvio padrão do RMSE de todos os métodos, foi obtido um pequeno valor. O MAE também forneceu um baixo valor de desvio padrão entre os métodos analisados. Finalmente, os valores MAPE seguem a tendência do RMSE. Considerando essas informações, a análise de desempenho apresentada na sequência considerou o coeficiente de determinação como o principal fator.

A estrutura que foi mais rápida foi a com participação em grade, contudo, teve a menor precisão em relação a validação dos dados. A estrutura utilizando *cluster* subtrativo teve um resultado superior, no entanto, foi consideravelmente mais demorada que as outras estruturas. Em todas as análises o desvio padrão se manteve baixo o que mostrou que os algoritmos utilizados são estáveis, mesmo utilizando dados em diferentes janelas de tempo. Neste trabalho a estrutura FCM *clustering* teve um resultado satisfatório em termos de precisão e em relação ao tempo de análise. Desta forma, esta estrutura é estratificada e amplamente analisada nesta subseção.

6.5.2 Número de clusters

Utilizando o FCM na Tabela 23 se apresenta uma avaliação em relação ao número de *clusters*. A avaliação leva em consideração o tempo necessário para o treinamento do modelo, a precisão no treino, teste e validação do modelo; além do erro para a previsão.

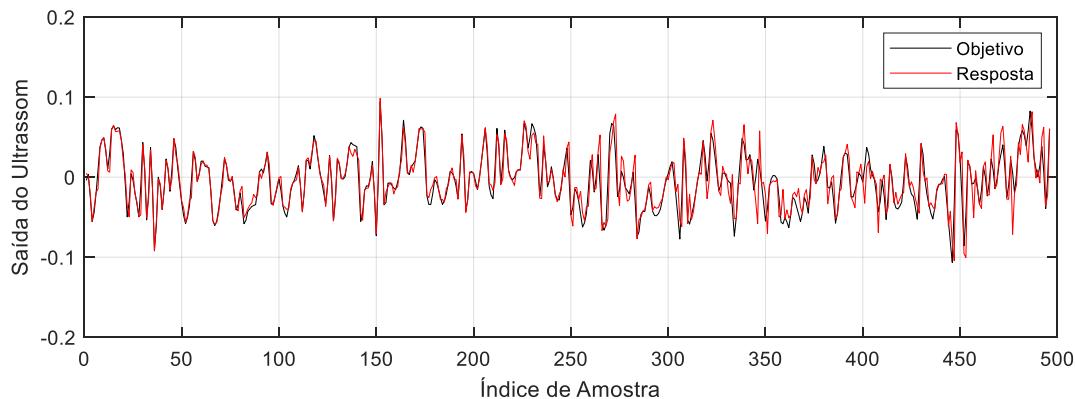
Tabela 23 - Resultado usando diferentes números de clusters.

Número de Clusters	Tempo Médio (s)	Média de Precisão (%)			RMSE	MAE	MAPE	Desvio Padrão
		Treino	Teste	Valid.				
5	78,91	95,88	96,04	95,92	0,0146	$6,5 \times 10^{-5}$	0,8576	$2,24 \times 10^{-3}$
10	165,77	96,20	96,33	96,23	0,0140	$1,2 \times 10^{-4}$	0,7639	
15	279,26	96,28	96,41	96,32	0,0136	$1,6 \times 10^{-4}$	0,7574	
20	430,61	96,38	96,52	96,41	0,0135	$8,1 \times 10^{-5}$	0,6211	
25	620,44	96,47	96,59	96,50	0,0132	$3,7 \times 10^{-5}$	0,5990	
30	826,00	96,51	96,61	96,53	0,0132	$5,1 \times 10^{-5}$	0,5388	

Fonte: Elaborada pelo autor (2021).

Em termos de precisão a maior variação se deu na mudança de 5 para 10 clusters, e desta forma se definiu o valor de 10 *cluster* para uma avaliação comparativa geral deste algoritmo em relação as configurações do WPT. Utilizando esta configuração, a Figura 35 apresenta uma comparação do sinal original (objetivo) e da previsão de sinal (resposta). Em termos de tempo total da análise houve um grande aumento progressivo juntamente com o aumento do número de *clusters*.

Figura 35 - Comparaçao do sinal original (objetivo) e previsão de sinal (resposta).



Fonte: Elaborada pelo autor (2021).

6.5.3 Utilização da transformada wavelet

Para ampliação da abrangência de avaliação, nesta subseção serão apresentados os dados dos 5 conjuntos de 50.000 pontos. Serão apresentados resultados em relação a decomposição de pacotes wavelet do terceiro, quarto e quinto nível com dois e três nós. A Tabela 24 apresenta os

resultados em relação a avaliação do treinamento, em que o primeiro número da primeira coluna representa o nível de decomposição e o segundo número representa a quantidade de nós.

Tabela 24 - Resultado para treinamento do FCM *clustering*.

Níveis_Número de Nós	Média de Precisão (%)				
	Dados #1	Dados #2	Dados #3	Dados #4	Dados #5
3_1	96,18	96,16	96,77	96,12	96,49
3_2	96,65	96,67	96,66	96,82	96,82
4_1	96,18	96,16	96,83	96,11	96,50
4_2	96,29	96,53	96,99	96,23	96,94
5_1	96,18	96,17	96,84	96,04	96,53
5_2	96,72	96,86	96,63	96,53	96,32

Fonte: Elaborada pelo autor (2021).

Nesta análise preliminar o algoritmo apresentou um bom resultado para 4 níveis de decomposição com 2 nós. A avaliação do teste destas configurações é apresentada na Tabela 25.

Tabela 25 - Resultado para teste do FCM *clustering*.

Níveis_Número de Nós	Média de Precisão (%)				
	Dados #1	Dados #2	Dados #3	Dados #4	Dados #5
3_1	96,32	96,22	96,90	96,60	96,09
3_2	96,69	96,69	96,72	96,63	96,65
4_1	96,32	96,22	96,98	96,60	96,09
4_2	96,29	96,56	97,05	96,15	96,59
5_1	96,31	96,23	96,98	96,55	96,13
5_2	96,81	96,92	96,63	96,54	96,30

Fonte: Elaborada pelo autor (2021).

Os resultados do teste do algoritmo, foram equivalentes aos resultados do treinamento. Para esta avaliação 75% do conjunto de dados considerado foi utilizado para treinamento e 25% foi utilizado para o teste do algoritmo. A Tabela 26 apresenta os resultados da validação da rede em relação aos dados utilizados para o teste.

No teste e na validação da análise o resultado manteve-se superior em relação a precisão na configuração de 4 níveis de decomposição com 2 nós. A comparação entre os diferentes conjuntos de dados na validação, mostram que o algoritmo é estável pois não apresenta resultados de variação superiores a 0,79%. A análise estatística completa da validação é apresentada na Tabela 27. Na análise estatística a covariância é calculada em relação a variação do número de nós para cada nível de decomposição.

Tabela 26 - Resultado para validação do FCM *clustering*.

Níveis_Número de Nós	Média de Precisão (%)				
	Dados #1	Dados #2	Dados #3	Dados #4	Dados #5
3_1	96,22	96,18	96,80	96,26	96,39
3_2	96,66	96,68	96,67	96,77	96,78
4_1	96,22	96,17	96,87	96,25	96,40
4_2	96,29	96,54	97,00	96,21	96,85
5_1	96,22	96,19	96,88	96,18	96,43
5_2	96,74	96,87	96,63	96,54	96,30

Fonte: Elaborada pelo autor (2021).

Tabela 27 - Resultado estatístico para validação do FCM.

Níveis_Número de Nós	Média (%)	RMSE	MAE	MAPE	Desvio Padrão	Variância	Covariância
3_1	96,37	0,0140	$1,2 \times 10^{-4}$	0,7714	$2,55 \times 10^{-3}$	$6,49 \times 10^{-6}$	$1,46 \times 10^{-7}$
3_2	96,71	0,0064	$2,9 \times 10^{-6}$	0,2325	$5,63 \times 10^{-4}$	$3,17 \times 10^{-7}$	
4_1	96,38	0,0139	$1,1 \times 10^{-4}$	0,7431	$2,88 \times 10^{-3}$	$8,27 \times 10^{-6}$	$6,42 \times 10^{-6}$
4_2	96,58	0,0061	$1,3 \times 10^{-4}$	0,0368	$3,45 \times 10^{-3}$	$1,19 \times 10^{-5}$	
5_1	96,38	0,0140	$1,2 \times 10^{-4}$	0,7728	$2,97 \times 10^{-3}$	$8,80 \times 10^{-6}$	$1,30 \times 10^{-6}$
5_2	96,62	0,0063	$6,5 \times 10^{-5}$	0,2540	$2,18 \times 10^{-3}$	$4,74 \times 10^{-6}$	

Fonte: Elaborada pelo autor (2021).

A importância de avaliar mais medidas de desempenho pode ser destacada neste momento, pois para o RMSE, três configurações distintas forneceram resultados semelhantes, usando dois nós. O fato de adicionar a métrica R^2 contribui para a seleção do melhor modelo. Os valores do MAE obtidos neste caso ajudaram a confirmar que mais nós fornecem a melhor configuração de modelo.

A análise estatística deixa claro a pouca variação de resultados do algoritmo, que mostra que a aplicação do WPT para normalização reduz ruído e deixa o algoritmo mais estável, sendo que a maior diferença entre a média dos conjuntos de dados é de apenas 0,34%.

6.6 AVALIAÇÃO DOS MÉTODOS AUTOREGRESSIVOS

Para uma avaliação comparativa completa dos métodos autoregressivos, será apresentada uma NARX e NAR. Para esses métodos de previsão, são aplicadas as formas de otimização LM, *bayesian regularization* (BR) e *scaled conjugate gradient* (SCG). A rede é configurada com diferentes números de neurônios ocultos e números de atrasos, conforme Tabela 28.

Tabela 28 - Precisão do método autorregressivo não linear.

Neurônios Ocultos	Nº de Atrasos	NAR - Precisão (%)			NARX - Precisão (%)		
		LM	BR	SCG	LM	BR	SCG
05	05	74,414	74,405	74,252	74,091	74,120	73,647
	10	78,765	78,792	78,687	78,642	78,693	78,563
	15	81,890	81,930	81,794	81,726	81,806	81,713
10	05	74,397	74,404	74,180	74,113	74,135	73,670
	10	78,738	78,796	78,563	78,678	78,760	77,872
	15	81,909	81,976	81,755	81,762	81,912	81,615
15	05	74,348	74,404	73,871	74,098	74,126	73,920
	10	78,791	78,794	78,453	78,666	78,826	76,664
	15	81,898	82,010	81,448	81,754	81,994	81,413

Fonte: Elaborada pelo autor (2021).

Nesta análise, os métodos NAR e NARX tiveram um resultado baixo em comparação ao método *wavelet neuro-fuzzy* aplicado neste trabalho para o mesmo objetivo, tanto em termos de acurácia, quanto em termos do RMSE conforme apresentado na Tabela 29. A variação do método de otimização dos algoritmos e do número de neurônios ocultos não gerou grandes alterações em termos de precisão.

Percebe-se uma melhoria com a utilização de mais atrasos na rede autoregressiva, contudo este aumento exige maior esforço computacional, sendo que a partir de 15 atrasos o algoritmo atinge o limite especificado de 1.000 iterações. Os valores de RMSE, MAE e MAPE para a fase de teste foram menores usando mais *clusters*, no entanto, o tempo necessário para a convergência foi maior. Novamente, foram obtidas pequenas variações em termos do número de *clusters* para RMSE e MAE.

Tabela 29 - Erro do método autorregressivo não linear.

Neurônios Ocultos	Nº de Atrasos	NAR - RMSE			NARX - RMSE		
		LM	BR	SCG	LM	BR	SCG
05	05	0,0374	0,0377	0,0373	0,0374	0,0375	0,0372
	10	0,0350	0,0351	0,0345	0,0345	0,0348	0,0352
	15	0,0323	0,0318	0,0350	0,0323	0,0320	0,0330
10	05	0,0378	0,0379	0,0378	0,0374	0,0379	0,0376
	10	0,0352	0,0345	0,0344	0,0349	0,0347	0,0343
	15	0,0325	0,0327	0,0328	0,0323	0,0327	0,0325
15	05	0,0377	0,0373	0,0380	0,0389	0,0384	0,0375
	10	0,0348	0,0346	0,0350	0,0347	0,0334	0,0347
	15	0,0324	0,0325	0,0325	0,0327	0,0323	0,0328

Fonte: Elaborada pelo autor (2021).

6.7 ANÁLISE DO MÉTODO DE GRUPO DE TRATAMENTO DE DADOS

Inicialmente são avaliados os parâmetros de configuração do algoritmo. Neste algoritmo é utilizada a validação cruzada, na qual os dados utilizados para o treinamento não são utilizados para o teste da rede. A Tabela 30 apresenta os resultados do teste da rede em relação ao percentual de dados usados para treinamento e teste (*treino* _ *teste*). Para esta avaliação inicial foi considerado o número máximo de 5 camadas com 25 neurônios por camada.

Tabela 30 - Resultado para teste da rede usando o GMDH.

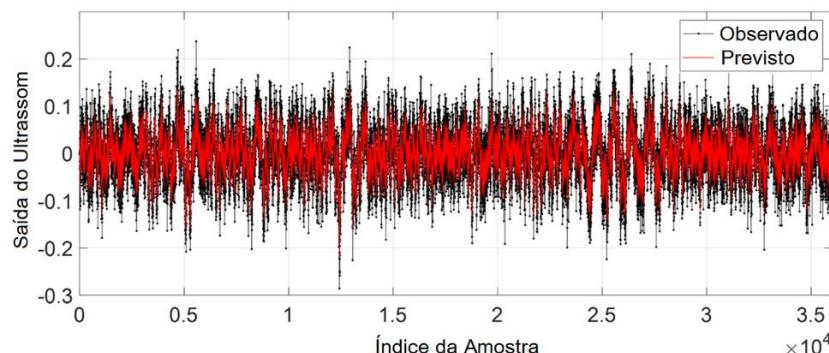
Treino	Teste (%)	Tempo (s)	RMSE	MAE	MAPE	R^2
10	90	10,1	0,0381	$3,4 \times 10^{-5}$	0,1966	0,7381
25	75	12,6	0,0379	$4,6 \times 10^{-5}$	0,7324	0,7381
50	50	20,5	0,0377	$9,1 \times 10^{-5}$	0,7595	0,7452
75	25	27,6	0,0376	$3,0 \times 10^{-5}$	1,1205	0,7430
90	10	30,8	0,0373	$1,7 \times 10^{-4}$	1,1071	0,6840

Fonte: Elaborada pelo autor (2021).

A utilização de um conjunto de dados maior para treinamento faz com que a rede leve mais tempo para convergir. Considerando que a análise é realizada *offline*, o tempo para treinamento não é critério para avaliar o melhor desempenho da rede. O menor valor de RMSE foi obtido utilizando 90% dos dados para treinamento, porém houve um menor R^2 nesta configuração.

A utilização de 75% dos dados para o treinamento apresentou RMSE baixo e R^2 superior, sendo uma configuração adequada para esta análise. A partir desse resultado, todas as análises para este algoritmo são realizadas com 75% dos dados para treinamento da rede. A avaliação da rede em fase de treinamento pode ser vista na Figura 36.

Figura 36 - Regressão de rede para treinamento usando o GMDH.



Fonte: Elaborada pelo autor (2021).

6.7.1 Número de neurônios e camadas

Os valores previstos seguem a tendência dos valores observados, com uma pequena redução da amplitude e desconsiderando as altas frequências. A Figura 37 mostra a avaliação da rede na fase de teste. A Tabela 31 apresenta uma avaliação da rede para a variação do número de neurônios, usando 5 camadas.

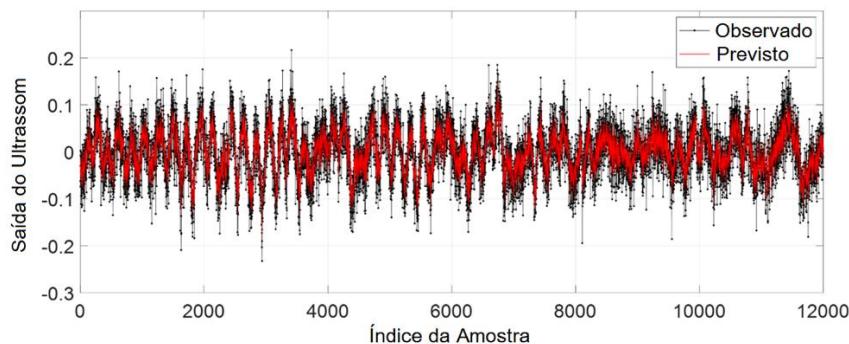
Tabela 31 - Resultados da variação no número de neurônios.

Neurônios	Tempo (s)	RMSE	MAE	MAPE	R^2
5	7,3	0,0386	$1,5 \times 10^{-4}$	0,5736	0,7268
10	10,0	0,0380	$1,4 \times 10^{-4}$	0,3790	0,7370
15	13,5	0,0379	$1,2 \times 10^{-4}$	1,3140	0,7387
20	20,5	0,0377	$9,4 \times 10^{-4}$	0,0306	0,7417
25	25,9	0,0377	$5,3 \times 10^{-4}$	0,0032	0,7431
30	39,6	0,0377	$1,8 \times 10^{-4}$	0,0631	0,7428
35	48,5	0,0377	$1,2 \times 10^{-4}$	0,7056	0,7424
40	63,0	0,0377	$1,1 \times 10^{-4}$	0,3301	0,7423
45	74,4	0,0377	$1,4 \times 10^{-4}$	0,5120	0,7428
50	87,7	0,0377	$1,9 \times 10^{-4}$	0,0195	0,7428

Fonte: Elaborada pelo autor (2021).

A utilização de mais de 20 neurônios não resultou necessariamente em uma melhora nos valores de R^2 e RMSE. O tempo necessário para a convergência aumentou conforme o número de neurônios foi aumentado. Como não há melhora significativa nos resultados a partir de 20 neurônios, esse valor foi definido para as análises posteriores. A Tabela 32 apresenta os resultados usando 20 neurônios para a variação no número de camadas.

Figura 37 - Regressão de rede para teste para o GMDH.



Fonte: Elaborada pelo autor (2021).

Tabela 32 - Resultados para variação no número de camadas.

Camadas	Tempo (s)	RMSE	MAE	MAPE	R^2
2	10,6	0,0392	$4,6 \times 10^{-5}$	0,3451	0,7168
3	12,0	0,0382	$3,3 \times 10^{-5}$	0,5310	0,7340
4	19,8	0,0378	$6,0 \times 10^{-5}$	1,1330	0,7407
5	26,8	0,0377	$8,8 \times 10^{-5}$	0,5861	0,7429
6	32,1	0,0377	$3,5 \times 10^{-4}$	0,1240	0,7419
7	33,3	0,0377	$2,0 \times 10^{-4}$	0,5447	0,7425
8	32,7	0,0377	$1,1 \times 10^{-4}$	0,5032	0,7415
9	44,6	0,0377	$1,8 \times 10^{-4}$	0,0056	0,7423
10	87,5	0,0377	$1,8 \times 10^{-4}$	1,0221	0,7419

Fonte: Elaborada pelo autor (2021).

A partir de mais de 5 camadas, a rede resultou em valores equivalentes para R^2 e idênticos para RMSE, com maior tempo para convergência da rede. Com base neste resultado, a melhor configuração encontrada do GMDH para este problema, foi utilizar 75% dos dados para treinamento, com 20 neurônios por camada oculta, utilizando 5 camadas ocultas.

6.7.2 Utilização da transformada wavelet

Usando a melhor configuração encontrada para este problema, a Tabela 33 apresenta uma avaliação do uso do WEC para extração de características. O uso do WEC mostrou-se promissor em todas as configurações utilizadas, tanto em relação ao RMSE quanto ao R^2 , deixando claro que a combinação entre o GMDH e WEC é adequada para o problema em avaliação.

Os melhores resultados foram obtidos usando 3 nós da árvore de coeficientes. A variação na profundidade da árvore de coeficientes não resultou em grandes variações nos resultados, e o melhor resultado foi obtido usando uma profundidade de 4 níveis com 3 nós, portanto, essa configuração foi usada nas análises estatísticas.

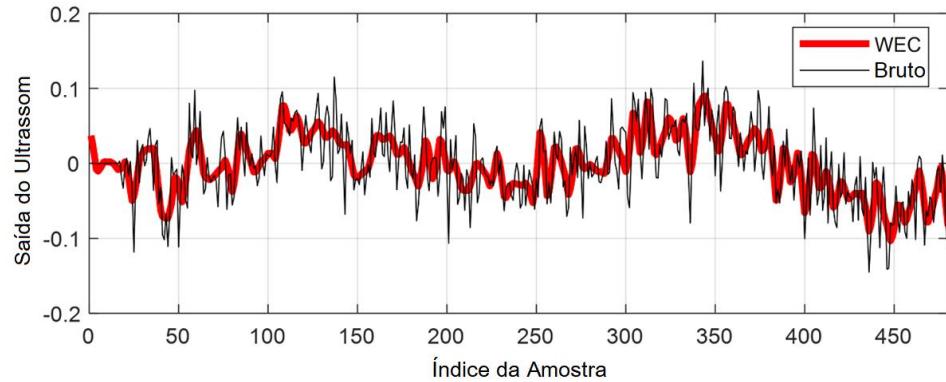
A partir da árvore de pacotes binários ótima, um sinal com menos ruído é obtido, como pode ser visto na Figura 38 para uma janela de 480 pontos, correspondendo a 1% do sinal total analisado. O WEC melhora a previsibilidade do algoritmo e mantém a característica do sinal ultrassônico processado de um isolador contaminado.

Tabela 33 - Resultados para aplicação do WEC.

Profundidade Nós	Tempo (s)	RMSE	MAE	MAPE	R^2
2_1	8,1	0,0190	$1,0 \times 10^{-4}$	0,1485	0,9266
2_2	8,2	0,0121	$4,8 \times 10^{-5}$	1,6585	0,8692
2_3	8,3	0,0058	$2,5 \times 10^{-5}$	0,0554	0,9922
2_4	8,5	0,0092	$3,1 \times 10^{-5}$	0,3370	0,8909
3_1	8,3	0,0191	$4,6 \times 10^{-5}$	0,2491	0,9265
3_2	8,1	0,0121	$6,5 \times 10^{-5}$	0,4439	0,8691
3_3	8,7	0,0058	$2,1 \times 10^{-5}$	0,0988	0,9920
3_4	8,5	0,0093	$5,6 \times 10^{-5}$	0,4748	0,8897
4_1	8,3	0,0192	$1,0 \times 10^{-4}$	1,0843	0,9252
4_2	8,0	0,0121	$6,9 \times 10^{-6}$	0,2993	0,8704
4_3	8,2	0,0057	$5,2 \times 10^{-6}$	0,2307	0,9923
4_4	9,1	0,0093	$2,4 \times 10^{-5}$	0,5997	0,8886

Fonte: Elaborada pelo autor (2021).

Figura 38 - Sinal com aplicação de WEC em janela de 480 pontos.

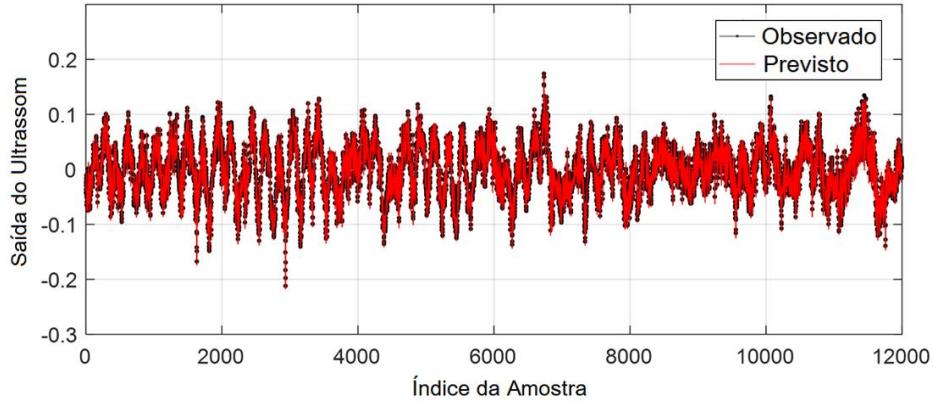


Fonte: Elaborada pelo autor (2021).

Para melhor visualização, a Figura 39 apresenta os sinais previstos e observados, mostrando que o algoritmo combinado ao WEC tem boa assertividade e é indicado para o problema discutido neste trabalho.

Para uma avaliação final da robustez do algoritmo, a Tabela 34 apresenta uma análise estatística completa, a partir de 50 simulações com a mesma configuração. A análise estatística mostra que a variação dos valores é baixa para o cálculo do RMSE, MAE, MAPE e R^2 .

Figura 39 - Resultado do teste final utilizando o WEC com o GMDH.



Fonte: Elaborada pelo autor (2021).

Tabela 34 - Análise estatística utilizando o WEC com o GMDH.

Indicator	RMSE	MAE	MAPE	R^2
Média	0,0058	$1,9 \times 10^{-5}$	0,0688	0,9921
Desvio Padrão	$4,2 \times 10^{-5}$	$1,5 \times 10^{-5}$	$2,8 \times 10^{-13}$	$1,2 \times 10^{-4}$
Variância	$1,8 \times 10^{-9}$	$2,1 \times 10^{-10}$	$8,0 \times 10^{-26}$	$1,3 \times 10^{-8}$
Covariância	$3,3 \times 10^{-10}$	$3,9 \times 10^{-11}$	$2,9 \times 10^{-25}$	$2,4 \times 10^{-9}$

Fonte: Elaborada pelo autor (2021).

6.8 CONSIDERAÇÕES FINAIS DO CAPÍTULO

Este capítulo ressaltou que o OEN-ELM proposto nesta tese é uma técnica superior aos modelos clássicos de classificação. Além de resultar em valores altos de acurácia, o OEN-ELM se mostrou um algoritmo rápido, que exige baixo esforço computacional. No próximo capítulo é apresentada a conclusão desta tese.

7 CONCLUSÕES

Neste trabalho comparou-se a utilização de modelos de inteligência artificial para realizar a classificação e previsão de um sinal ultrassônico, capturado em laboratório, durante análises de isoladores do tipo pino de média tensão. A partir dos resultados obtivos, várias técnicas foram utilizadas para propor um novo modelo de rede neural artificial. O uso destas técnicas para classificar isoladores provou ser promissor para identificação de falhas nestes componentes.

Percebeu-se em todas as análises que a identificação do isolador contaminado é difícil. Esse resultado era esperado porque os isoladores perfurados e quebrados que foram removidos do sistema de distribuição, também tinham contaminação incrustada em sua superfície. Outro ponto importante na análise do sinal é que o ganho de equipamento altera as assinaturas de sinal. Com o ganho baixo, a assinatura do sinal fica semelhante em diversas condições, o que torna difícil a multiclassificação das condições do isolador. Em contrapartida, utilizando um ganho alto, o sinal resultante é mais ruidoso e pode saturar a medição.

Para classificação das condições dos isoladores, a rede MLP baseada na FFT teve um bom resultado, pois foi possível ter uma validação satisfatória para multiclassificação. A utilização da rede SLP se mostrou pouco efetiva, e este resultado era esperado para este modelo devido a característica de não linearidade no sinal. Na análise baseada nos dados brutos, várias tentativas de combinações da quantidade de nós de entrada e nós ocultos foram avaliadas. Entretanto, devido a esta técnica não extrair as características predominantes no sinal a precisão na análise foi baixa.

O algoritmo desenvolvido para classificação intitulado OEN-ELM, mostrou ter um índice de acurácia superior a 98% para classificação de isoladores multiclasse considerando a média de análise de 50 experimentos. Aplicando o PSO, foi possível obter respostas 588% mais rápidas em comparação ao EN-ELM tradicional na fase de treinamento. Isso corresponde a um modelo que convergiu em 137 segundos usando a melhor configuração encontrada.

A otimização do EN-ELM usando PSO e PCA para otimização e redução de dimensionalidade provou ser uma abordagem rápida, robusta e precisa para resolver o problema de classificação multiclasse. Baseado nestes resultados, pode-se verificar que é possível classificar automaticamente as condições dos isoladores, melhorando a confiabilidade da análise para aplicação em inspeções do sistema elétrico de potência.

Considerando a avaliação de previsão de séries temporais, a aplicação do ANFIS se mostrou promissora, tanto em termos de velocidade quanto em termos de precisão. Utilizar um número muito maior de *clusters* resultou em um aumento considerável em tempo para convergência e não necessariamente trouxe uma grande melhora em termos de precisão. Em uma avaliação específica da configuração do algoritmo, o FCM *clustering* gerou um resultado equilibrado em relação a velocidade de convergência e precisão. A análise estatística para FCM *Clustering* mostrou que os resultados têm pouca variabilidade, mesmo utilizando diferentes conjuntos de dados, o que torna claro que o algoritmo é estável para esta aplicação. O resultado desta estabilidade é devido a aplicação de wavelet para redução de ruído e extração de características, tornando o wavelet-ANFIS uma ótima alternativa para a previsão de séries temporais.

O WEC combinado ao GMDH teve um resultado promissor para a previsão do sinal do isolador contaminado de média tensão. O uso do WEC para extração de características provou ser muito adequado para filtrar o sinal e obter maior precisão. Esta afirmação é apoiada pelo modelo sendo mostrado muito mais rápido do que algoritmos bem estabelecidos, como ANFIS e LSTM, que estão sendo amplamente estudados.

7.1 TRABALHOS FUTUROS

Para o problema proposto, algoritmos híbridos apresentam um resultado superior aos algoritmos sem a aplicação de técnicas de filtragem de sinais. Outros métodos podem ser aplicados com a mesma finalidade, reduzindo o ruído gerado pelo equipamento de ultrassom e mantendo sua característica necessária para identificar a possível falha no isolador de média ou alta tensão. Dessa forma, será possível aplicar outras técnicas combinadas para melhorar ainda mais os resultados dessas análises.

Especificamente se mostrou interessante a combinação entre técnicas de classificação baseadas no sinal ultrassônico e análise por visão computacional. Futuramente, um modelo que tenha a capacidade de realizar a classificação das condições dos isoladores a partir de múltiplas entradas, como sinal ultrassônico e imagens do componente, poderá ser uma alternativa para melhorar a identificação de falhas.

A solução apresentada nesta tese poderá futuramente ser embarcada para auxiliar as inspeções das redes de distribuição de energia elétrica. A partir de um modelo otimizado, que tem

maior eficiência para classificação, se torna promissor desenvolver um projeto comercial a ser utilizado pelas concessionárias de energia elétrica.

A utilização do sinal pré-processado teve sucesso na avaliação apresentada nesta tese. Utilizar o sinal originalmente registrado pelo ultrassom pode melhorar ainda mais a capacidade de se desenvolver um equipamento eficiente para localização de condições adversas em isoladores. A previsão da série temporal a partir de um passo a frente resultou em baixo erro e alta acurácia em relação ao sinal observado. Uma avaliação de um número maior de passos pode representar uma previsão mais longa, sendo mais interessante para prever falhas no sistema elétrico.

REFERÊNCIAS

ABBO, K. K.; KHUDHUR, H. M. A Hybrid Conjugate Gradient Fletcher-Reeves and Polak-Ribiere Algorithm for Unconstrained Optimization. **Tikrit Journal of Pure Science**, v. 21, n. 1, p. 124-129, 2018.

ABUASSBA, A. O. M. *et al.* A Heterogeneous Ensemble of Extreme Learning Machines with Correntropy and Negative Correlation. **Tsinghua Science and Technology**, v. 22, n. 6, p. 691-701, dez. 2017.

AHMAD, A. S. S.; MATTI, M. S.; AL-HABIB, O. A.; SHAIKHOW, S. ECG Signal Classification Using Scaled Conjugate Gradient Learner Algorithm. **International Journal of Medical Research & Health Sciences**, v. 7, n. 5, p. 11-17, 2018.

AÏT-SAHALIA, Yacine; XIU, Dacheng. Principal Component Analysis of High-Frequency Data. **Journal of the American Statistical Association**, p. 1-17, 2018.

ALBANO, Maurizio *et al.* Infrared Analysis of Dry-band Flashover of Silicone Rubber Insulators. **IEEE Transactions on Dielectrics and Electrical Insulation**, v. 23, n. 1, p. 304-310, 2016.

ALBARRACIN, R. *et al.* Separation of Sources in Radiofrequency Measurements of Partial Discharges Using Time Power Ratio Maps. **ISA Transactions**, v. 58, p. 389-397, set. 2015.

ALHASSAN, Ahmad Bala *et al.* Power Transmission Line Inspection Robots: A Review, Trends and Challenges for Future Research. **International Journal of Electrical Power & Energy Systems**, v. 118, p. 105862, 2020.

ALKHRAIJAH, M. M.; ABIDO, M. A. Power Quality Classification Using Neuro Fuzzy Logic Inference System. In: IEEE-GCC CONFERENCE AND EXHIBITION, 9, 2018, Manama. *IEEE*, Manama: GCCCE, 2017. p. 1-4.

ALI, H. Leakage Current Prediction for High Voltage Insulators Flashover Based on Extreme Value Theory. In: INTERNATIONAL SYMPOSIUM ON COMPUTER, CONSUMER AND CONTROL, 2016, Xi'an. *IEEE*, Xi'an: IS3C, 2016, p. 870-873.

AMINI, Mohamad Ali; SEDIGHI, Ali Reza. A New Procedure for Determination of Insulators Contamination in Electrical Distribution Networks. **International Journal of Electrical Power & Energy Systems**, v. 61, p. 380-385, 2014.

BABAYOMI, Oluleke *et al.* Fault Diagnosis in an Extra High Voltage Power Line. In: IEEE PES POWER AFRICA, 2017, Accra. *IEEE*, Accra, 2017. p. 311-316.

BAKAR, Ab Halim Abu *et al.* High Impedance Fault Location in 11 kV Underground Distribution Systems Using Wavelet Transforms. **International Journal of Electrical Power & Energy Systems**, v. 55, p. 723-730, 2014.

BALLABIO, Davide; GRISONI, Francesca; TODESCHINI, Roberto. Multivariate Comparison of Classification Performance Measures. **Chemometrics and Intelligent Laboratory Systems**, v. 174, p. 33-44, mar. 2018.

BASANTAK, Panigrahi *et al.* Fault Detection and Classification Using Wavelet Transform and Neuro Fuzzy System. In: INTERNATIONAL CONFERENCE ON CURRENT TRENDS TOWARDS CONVERGING TECHNOLOGIES, 2018, Coimbatore. *IEEE*, Coimbatore: ICCTCT, 2018. p. 1-5.

BEHERA, L.; KUMAR, S.; PATNAIK, A. On Adaptive Learning Rate that Guarantees Convergence in Feedforward Networks. **IEEE Transactions on Neural Networks**, v. 17, n. 5, p. 1116-1125, set. 2006.

BELHOUCHET, K.; BAYADI, A.; BENDIB, M. Elhadi. Artificial Neural Networks (ANN) and Genetic Algorithm Modeling and Identification of Arc Parameter in Insulators Flashover Voltage and Leakage Current. In: INTERNATIONAL CONFERENCE ON ELECTRICAL ENGINEERING, 4, 2015, Boumerdes. *ICEE*, Boumerdes: IEEE, 2016. p. 1-6.

BHATTACHARYA, B.; SINHA, A. Intelligent Fault Analysis in Electrical Power Grids. In: IEEE INTERNATIONAL CONFERENCE ON TOOLS WITH ARTIFICIAL INTELLIGENCE, 29, 2017, Boston. *IEEE*, Boston: ICTAI, 2017, p. 985-990.

BISGIN, Halil *et al.* Diagnosis of Long QT Syndrome via Support Vector Machines Classification. **Journal of Biomedical Science and Engineering**, v. 4, n. 04, p. 264, 2011.

BRO, Rasmus; SMILDE, Age K. Principal Component Analysis. **Analytical Methods**, v. 6, n. 9, p. 2812-2831, 2014.

BISHOP, Christopher M. **Neural Networks for Pattern Recognition**, Cambridge: Oxford University, 1995.

BRITO, J. M. E. **Deteção de Efeito Coroa em Isoladores Compósitos**. 2017. 81 f. Dissertação de Mestrado em Engenharia Electrotécnica e de Computadores – Técnico Lisboa, Portugal, 2017.

CAO, Weipeng *et al.* Fuzziness Based Online Sequential Extreme Learning Machine for Classification Problems. **Soft Computing**, v. 22, n. 11, p. 3487-3494, fev. 2018a.

CAO, Zhen *et al.* An Image Steganalysis Algorithm Based on Rotation Forest Transformation and Multiple Classifiers Ensemble. In: International Conference on Emerging Internetworking, Data & Web Technologies, 6, Tirana, **Advances in Internetworking, Data & Web Technologies**, Albania, Springer, 2018b, p. 1-12.

CHAI, Songjian; XU, Zhao; JIA, Youwei. Conditional Density Forecast of Electricity Price based on Ensemble ELM and Logistic EMOS. **IEEE Transactions on Smart Grid**, abr. 2018.

CHENG, L. *et al.* Research on aging evaluation and remaining lifespan prediction of composite insulators in high temperature and humidity regions. **IEEE Transactions on Dielectrics and Electrical Insulation**, v. 23, n. 5, p. 2850-2857, 2016.

CHEN, Shaowei; SHANG, Yue; WU, Minhua. Application of PSO-ELM in Electronic System Fault Diagnosis. In: INTERNATIONAL CONFERENCE ON PROGNOSTICS AND HEALTH MANAGEMENT, 2016, Ottawa. *IEEE*, Ottawa: ICPHM, 2016. p. 1-5.

COETZER, C. J.; LEUSCHNER, F. W. The Influence of a Camera's Spectral Transfer Function Used for Observing High Voltage Corona on Insulators. **IEEE Transactions on Dielectrics and Electrical Insulation**, v. 23, n. 3, p. 1753-1759, 2016.

CORSO, Marcelo Picolotto *et al.* Evaluation of Methods for Electric Field Calculation in Transmission Lines. **IEEE Latin America Transactions**, v. 16, n. 12, p. 2970-2976, 2018.

COSTA, E. G.; FERREIRA, T. V.; NERI, M. G. G. Trilhamento e Erosão em Núcleo de Isoladores Poliméricos. In: SEMINÁRIO NACIONAL DE RODUÇÃO E TRANSMISSÃO DE ENERGIA ELÉTRICA, 19, 2007, Rio de Janeiro. Rio de Janeiro: SNPTEE, 2017. p. 1-8.

CUI, Hang; TU, Niannian. Generalized Regression Neural Networks Based HVDC Transmission Line Fault Localization. In: INTERNATIONAL CONFERENCE ON INTELLIGENT HUMAN-MACHINE SYSTEMS AND CYBERNETICS, 7, 2015, Hangzhou. *IEEE*, Hangzhou: IHMSC, 2015. p. 25-29.

CYBENKO, George. Approximation by Superpositions of a Sigmoidal Function. **Mathematics of Control, Signals and Systems**, v. 2, n. 4, p. 303-314, dez. 1989.

DAG, Osman; YOZGATLIGIL, Ceylan. GMDH: An R Package for Short Term Forecasting via GMDH-Type Neural Network Algorithms. **The R Journal**, v. 8, n. 1, p. 379, 2016.

DAMAYANTI, Auli. Fuzzy Learning Vector Quantization, Neural Network and Fuzzy Systems for Classification Fundus Eye Images with Wavelet Transformation. In: INTERNATIONAL CONFERENCES ON INFORMATION TECHNOLOGY, INFORMATION SYSTEMS AND ELECTRICAL ENGINEERING, 2, 2017, Yogyakarta. *IEEE*, Yogyakarta: ICITISEE, 2017, p. 331-336.

DE MELLO CANALLI, Ygor. **Funções de Ativação Hiperbólicas em Redes Neurais**. 2017. 92 f. Dissertação. Mestrado em Engenharia de Sistemas e Computação - Universidade Federal do Rio de Janeiro. Programa de Pós-graduação em Engenharia de Sistemas e Computação, Rio de Janeiro, 2017.

DEL VALLE, Yamille *et al.* Particle Swarm Optimization: Basic Concepts, Variants and Applications in Power Systems. **IEEE Transactions on Evolutionary Computation**, v. 12, n. 2, p. 171-195, 2008.

DONG, M. *et al.* Joint Visualization Diagnosis of Outdoor Insulation Status with Optical and Acoustical Detections. **IEEE Transactions on Power Delivery**, set. 2018.

EBTEHAJ, Isa *et al.* GMDH-type Neural Network Approach for Modeling the Discharge Coefficient of Rectangular Sharp-crested Side Weirs. **Engineering Science and Technology, an International Journal**, v. 18, n. 4, p. 746-757, 2015.

EIBAR, Carlos *et al.* Design and Simulation of a Mobile Robot, for the Inspection of Electrical Lines. In: Simposio Iberoamericano en Programación Informática, 2017, Riobamba, **KnE Engineering**, Ecuador, Escuela Superior Politécnica de Chimborazo, 2018, p. 141-152.

EOM, Yong Hwan *et al.* Refrigerant Charge Fault Detection Method of Air Source Heat Pump System Using Convolutional Neural Network for Energy Saving. **Energy**, v. 187, p. 115877, 2019.

FEBRERO-BANDE, Manuel; GALEANO, Pedro; GONZÁLEZ-MANTEIGA, Wenceslao. Functional Principal Component Regression and Functional Partial Least-Squares Regression: An Overview and a Comparative Study. **International Statistical Review**, v. 85, n. 1, p. 61-83, 2017.

FERRARINI, G. *et al.* Thermal Response Measurement of Building Insulating Materials by Infrared Thermography. **Energy and Buildings**, v. 133, p. 559-564, 2016.

FEZAI, R. *et al.* Online Reduced Kernel GLRT Technique for Improved Fault Detection in Photovoltaic Systems. **Energy**, v. 179, p. 1133-1154, 2019.

FINARDI, Erlon Cristian. **Métodos de Busca Linear para Problemas Irrestritos**. Florianópolis: Laboratório de Planejamento de Sistemas de Energia Elétrica, 2017. 56 slides, color.

FREER, Benjamin Avery; MANAHAN, Joseph Michael. **Electrical Network Inspection Devices**, US Patent Application 15/645.459, 2018, 18p.

FU, Si-Yao *et al.* Motion Based Image Deblur Using Recurrent Neural Network for Power Transmission Line Inspection Robot. In: INTERNATIONAL JOINT CONFERENCE ON NEURAL NETWORKS, 2006, Vancouver. *IEEE*, Vancouver, 2006. p. 3854-3859.

FU, Zhao *et al.* Prediction of Industrial Wastewater Quality Parameters Based on Wavelet Denoised ANFIS Model. In: ANNUAL COMPUTING AND COMMUNICATION WORKSHOP AND CONFERENCE, 8, 2018, Las Vegas. *IEEE*, Las Vegas: CCWC, 2018. p. 301-306.

GAJJAR, Shriram; KULAHCI, Murat; PALAZOGLU, Ahmet. Real Time Fault Detection and Diagnosis Using Sparse Principal Component Analysis. **Journal of Process Control**, v. 67, p. 112-128, 2018.

GAN, Meng *et al.* Construction of Hierarchical Diagnosis Network Based on Deep Learning and its Application in the Fault Pattern Recognition of Rolling Element Bearings. **Mechanical Systems and Signal Processing**, v. 72, p. 92-104, 2016.

GARSON, G. David. **Neural networks: An Introductory Guide for Social Scientists**. London: Sage, 1998.

GHOSH, Riddhi; CHATTERJEE, Biswendu; CHAKRAVORTI, Sivaji. A Novel Leakage Current Index for the Field Monitoring of Overhead Insulators Under Harmonic Voltage. **IEEE Transactions on Industrial Electronics**, v. 65, n. 2, p. 1568-1576, jul. 2018.

GIL, Paulo; OLIVEIRA, Tiago; PALMA, Luis. Adaptive Neuro-fuzzy Control for Discrete-time Non-affine Nonlinear Systems. **IEEE Transactions on Fuzzy Systems**, 2018.

GRAPS, Amara. An Introduction to Wavelets. **IEEE computational science and engineering**, v. 2, n. 2, p. 50-61, 1995.

GREFF, Klaus *et al.* LSTM: A Search Space Odyssey. **IEEE transactions on neural networks and learning systems**, v. 28, n. 10, p. 2222-2232, 2016.

GRIMM, Kevin J.; MAZZA, Gina L.; DAVOUDZADEH, Pega. Model Selection in Finite Mixture Models: A k-fold Cross-validation Approach. **Structural Equation Modeling: A Multidisciplinary Journal**, v. 24, n. 2, p. 246-256, 2017.

GÓMEZ, F. Á. *et al.* Diagnosis of Insulation Condition of MV Switchgears by Application of Different Partial Discharge Measuring Methods and Sensors. **Sensors**, v. 18, n. 3, p. 720, fev. 2018.

HANSEN, Lars Kai; SALAMON, Peter. Neural Network Ensembles. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 12, n. 10, p. 993-1001, out. 1990.

HAO, Qiushi *et al.* A Novel Rail Defect Detection Method Based on Undecimated Lifting Wavelet Packet Transform and Shannon Entropy Improved Adaptive Line Enhancer. **Journal of Sound and Vibration**, v. 425, p. 208-220, 2018.

HAVENS, Timothy C. *et al.* Fuzzy C-means Algorithms for Very Large Data. **IEEE Transactions on Fuzzy Systems**, v. 20, n. 6, p. 1130-1146, 2012.

HEBB, Donald Olding. **The Organizations of Behavior: a Neuropsychological Theory**. London: Chapman and Hall, 1957.

HOCHREITER, Sepp; SCHMIDHUBER, Jürgen. Long Short-Term Memory. **Neural Computation**, v. 9, n. 8, p. 1735-1780, 1997.

HORNIK, Kurt. Approximation Capabilities of Multilayer Feedforward Networks. **Neural Networks**, v. 4, n. 2, p. 251-257, 1991.

HOSSEINPOUR, Soleiman; AGHBASHLO, Mortaza; TABATABAEI, Meisam. Biomass Higher Heating Value (HHV) Modeling on the Basis of Proximate Analysis Using Iterative Network Based Fuzzy Partial Least Squares Coupled with Principle Component Analysis (PCA-INFPLS). **Fuel**, v. 222, p. 1-10, 2018.

HUANG, Guang-Bin; BABRI, Haroon A. Upper Bounds on the Number of Hidden Neurons in Feedforward Networks with Arbitrary Bounded Nonlinear Activation Functions. **IEEE Transactions on Neural Networks**, v. 9, n. 1, p. 224-229, jan. 1998.

HUANG, Guang-Bin; CHEN, Lei. Convex Incremental Extreme Learning Machine. **Neurocomputing**, v. 70, n. 16-18, p. 3056-3062, out. 2007.

HUANG, Guang-Bin; CHEN, Lei. Enhanced Random Search Based Incremental Extreme Learning Machine. **Neurocomputing**, v. 71, n. 16-18, p. 3460-3468, out. 2008.

HUANG, Guang-Bin; DING, Xiaojian; ZHOU, Hongming. Optimization Method Based Extreme Learning Machine for Classification. **Neurocomputing**, v. 74, n. 1-3, p. 155-163, dez. 2010.

HUANG, Guang-Bin *et al.* Extreme Learning Machine for Regression and Multiclass Classification. **IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)**, v. 42, n. 2, p. 513-529, abr. 2012.

HUANG, Guang-Bin; WANG, D. H.; LAN, Y. Extreme Learning Machines: A Survey. **International Journal of Machine Learning and Cybernetics**, v. 2, n. 2, p. 107-122, 2011.

HUANG, Guang-Bin; ZHU, Qin-Yu; SIEW, Chee-Kheong. Extreme Learning Machine: A New Learning Scheme of Feedforward Neural Networks. In: International Joint Conference on Neural Networks, 2004, Budapest. Budapest: IEEE, 2005. p. 985-990.

HUANG, Guang-Bin; ZHU, Qin-Yu; SIEW, Chee-Kheong. Extreme Learning Machine: Theory and Applications. **Neurocomputing**, v. 70, n. 1-3, p. 489-501, 2006.

HUANG, Wei; OH, Sung-Kwun; PEDRYCZ, Witold. Fuzzy Wavelet Polynomial Neural Networks: Analysis and Design. **IEEE Transactions on Fuzzy Systems**, v. 25, n. 5, p. 1329-1341, 2016.

HUANG, Wei; OH, Sung-Kwun; PEDRYCZ, Witold. Hybrid Fuzzy Wavelet Neural Networks Architecture Based on Polynomial Neural Networks and Fuzzy Set/relation Inference-based Wavelet Neurons. **IEEE Transactions on Neural Networks and Learning Systems**, v. 29, n. 8, p. 3452-3462, 2017.

HUI, Xiaolong *et al.* Vision-Based Autonomous Navigation Approach for Unmanned Aerial Vehicle Transmission-line Inspection. **International Journal of Advanced Robotic Systems**, v. 15, n. 1, p. 1-15, 2018.

IIGUNI, Youji; SAKAI, Hideaki; TOKUMARU, Hidekatsu. A Real-time Learning Algorithm for a Multilayered Neural Network Based on the Extended Kalman Filter. **IEEE Transactions on Signal Processing**, v. 40, n. 4, p. 959-966, 1992.

JIA, Feng *et al.* Deep Neural Networks: A Promising Tool for Fault Characteristic Mining and Intelligent Diagnosis of Rotating Machinery with Massive Data. **Mechanical Systems and Signal Processing**, v. 72, p. 303-315, 2016.

JUFRI, Fauzan Hanif; OH, Seongmun; JUNG, Jaesung. Development of Photovoltaic Abnormal Condition Detection System Using Combined Regression and Support Vector Machine. **Energy**, v. 176, p. 457-467, 2019.

JUNIOR, Elias Teodoro Silva *et al.* Corona Effect Detection in Energized Polymeric Insulators Using Machine Learning and Ultrasonic Emissions. **IEEE Latin America Transactions**, v. 16, n. 6, p. 1587-1594, 2018.

KALLA, Ujjwal Kumar *et al.* An Experimental Study of Power Quality Issues in Ceramic Insulator Industry. In: INTERNATIONAL CONFERENCE ON POWER SYSTEMS, 6, 2016, New Delhi. *IEEE*, New Delhi: ICPS, 2016. p. 1-6.

KANELLOPOULOS, I.; WILKINSON, G. G. Strategies and Best Practice for Neural Network Image Classification. **International Journal of Remote Sensing**, v. 18, n. 4, p. 711-725, 1997.

KARRI, Chandram; DURGAM, Rajababu; RAGHURAM, K. Electricity Price Forecasting in Deregulated Power Markets using Wavelet-ANFIS-KHA. In: INTERNATIONAL CONFERENCE ON COMPUTING, POWER AND COMMUNICATION TECHNOLOGIES, 2018, Greater Noida. *IEEE*, Greater Noida: GUCON, 2018. p. 982-987.

KASBURG, Christopher; STEFENON, Stéfano Frizzo. Deep Learning for Photovoltaic Generation Forecast in Active Solar Trackers. **IEEE Latin America Transactions**, v. 17, n. 12, p. 2013-2019, 2019.

KENNEDY, J.; EBERHART, R. Particle swarm optimization. In: PROCEEDINGS IEEE INTERNATIONAL CONFERENCE ON NEURAL NETWORKS, 1995, Perth, WA, Australia. *IEEE*, Perth: ICNN, 1995. P. 1942–1948.

KHAMIS, A.; XU, Y.; DONG, Z. Y.; ZHANG, R. Faster Detection of Microgrid Islanding Events Using an Adaptive Ensemble Classifier. **IEEE Transactions on Smart Grid**, v. 9, n. 3, p. 1889-1899, May 2018.

KHAN, Hidayatullah; AMIN, Muhammad; AHMAD, Ayaz. Multistress Accelerated Aging and Tracking/erosion-resistance Investigation of High Voltage Polymeric Insulators. **Arabian Journal for Science and Engineering**, v. 42, n. 12, p. 5101-5120, 2017.

KHAN, Muhammad Mohsin; MOKHTAR, Ainul Akmar; HUSSIN, Hilmi. A Neural Based Fuzzy Logic Model to Determine Corrosion Rate for Carbon Steel Subject to Corrosion Under Insulation. **Applied Mechanics and Materials**, v. 789-790, p. 526-530, set. 2015.

KHAN, U. F. *et al.* An Efficient Algorithm for Partial Discharge Localization in High-Voltage Systems Using Received Signal Strength. **Sensors**, v. 18, n. 11, 2018.

KINGMA, Diederik P.; BA, Jimmy. ADAM: A Method for Stochastic Optimization. In: INTERNATIONAL CONFERENCE FOR LEARNING REPRESENTATIONS, 2015, San Diego. *ICLR*, San Diego: Cornell University, 2015. P. 1-15.

KRÖSE, Ben; VAN DER SMAGT, Patrick. **An Introduction to Neural Networks**. Amsterdam: University of Amsterdam, 1996.

KUFFEL, E.; ZAENGL, W.S.; KUFFEL, J. **High Voltage Engineering Fundamentals**. 2 ed. 2000, p. 281-365.

KUMAR, B. Pradeep *et al.* Online Fault Detection and Diagnosis in Photovoltaic Systems Using Wavelet Packets. **IEEE Journal of Photovoltaics**, v. 8, n. 1, p. 257-265, 2018.

KUMAR, N. Krishna; SAVITHA, R.; AL MAMUN, Abdullah. Ocean Wave Height Prediction Using Ensemble of Extreme Learning Machine. **Neurocomputing**, v. 277, p. 12-20, 2018.

LAINE, Andrew; FAN, Jian. Texture Classification by Wavelet Packet Signatures. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 15, n. 11, p. 1186-1191, 1993.

LAN, Yuan; SOH, Yeng Chai; HUANG, Guang-Bin. Ensemble of Online Sequential Extreme Learning Machine. **Neurocomputing**, v. 72, n. 13-15, p. 3391-3395, 2009.

LASALVIA, L. A. B. *et al.* Intelligent Acoustic Detection of Defective Porcelain Station Post Insulators. In: ELECTRICAL INSULATION CONFERENCE, 2015, Seattle. *IEEE*, Seattle: EIC, 2015. p. 118-122.

LEDOIT, Olivier; WOLF, Michael. Spectrum Estimation: A Unified Framework for Covariance Matrix Estimation and PCA in Large Dimensions. **Journal of Multivariate Analysis**, v. 139, p. 360-384, 2015.

LESHNO, Moshe *et al.* Multilayer Feedforward Networks with a Nonpolynomial Activation Function can Approximate any Function. **Neural Networks**, v. 6, n. 6, p. 861-867, 1993.

LEUNG, Henry; HAYKIN, Simon. The Complex Backpropagation Algorithm. **IEEE Transactions on Signal Processing**, v. 39, n. 9, p. 2101-2104, 1991.

LEVENBERG, Kenneth. A Method for the Solution of Certain Non-linear Problems in Least Squares. **Quarterly of Applied Mathematics**, v. 2, n. 2, p. 164-168, 1944.

LIANG, Nan-Ying *et al.* A Fast and Accurate Online Sequential Learning Algorithm for Feedforward Networks. **IEEE Transactions on Neural Networks**, v. 17, n. 6, p. 1411-1423, 2006.

LI, B.; LI, Y.; RONG, X. The Extreme Learning Machine Learning Algorithm with Tunable Activation Function. **Neural Computing and Applications**, v. 22, n. 3-4, p. 531-539, 2013.

LI, Botong *et al.* A Permanent Fault Identification Method for Single-pole Grounding Fault of Overhead Transmission Lines in VSC-HVDC Grid Based on Fault Line Voltage. **International Journal of Electrical Power & Energy Systems**, v. 117, p. 105603, 2020.

LI, Rita Yi Man; FONG, Simon; CHONG, Kyle Weng Sang. Forecasting the REITs and stock indices: group method of data handling neural network approach. **Pacific Rim Property Research Journal**, v. 23, n. 2, p. 123-160, 2017.

LIU, Bin *et al.* Estimating B-coefficients of Power Loss Formula Considering Volatile Power Injections: An Enhanced Least Square Approach. **IET Generation, Transmission & Distribution**, v. 12, n. 12, p. 2854-2860, 2018a.

LIU, Lishuai *et al.* An Active Infrared Thermography Method for Nondestructive Testing of Porcelain Post Insulators. **Power Research**, v. 13, n. 1, p. 7-12, 2017a.

LIU, Nan; WANG, Han. Ensemble Based Extreme Learning Machine. **IEEE Signal Processing Letters**, v. 17, n. 8, p. 754-757, 2010.

LIU, Yunpeng *et al.* The Discrimination Method as Applied to a Deteriorated Porcelain Insulator Used in Transmission Lines on the Basis of a Convolution Neural Network. **IEEE Transactions on Dielectrics and Electrical Insulation**, v. 24, n. 6, p. 3559-3566, 2017b.

LIU, Wei *et al.* Short-term Load Forecasting Based on Elastic Net Improved GMDH and Difference Degree Weighting Optimization. **Applied Sciences**, v. 8, n. 9, p. 1603, 2018b.

LI, Xuewen *et al.* Surface Discharge Detection Method of Contaminated Insulators Based on Ultraviolet Images Parameters. In: INTERNATIONAL CONFERENCE ON ELECTRICAL MATERIALS AND POWER EQUIPMENT, 2017, Xi'an. *IEEE*, Xi'an: ICEMPE, 2017. p. 155-158.

LU, W. *et al.* Early Fault Detection Approach with Deep Architectures. **IEEE Transactions on Instrumentation and Measurement**, v. 67, n. 7, p. 1679-1689, 2018.

LV, F.; HAN, M.; QIU, T. Remote Sensing Image Classification Based on Ensemble Extreme Learning Machine with Stacked Autoencoder. **IEEE Access**, v. 5, p. 9021-9031, 2017.

LV, Li-Jun *et al.* An Approach for Fault Monitoring of Insulators Based on Image Tracking. In: INTERNATIONAL CONFERENCE ON MECHATRONICS AND MACHINE VISION IN PRACTICE, 24, 2017, Auckland. *IEEE*, Auckland: M2VIP, 2017. p. 1-6.

MAHMOUDI, Javad; MOUSSAVI, Seyed Zeinolabedin; SHARIFIAN, Hamid. Partial Discharge Diagnosis of Ceramic Pin Insulators Case Study in a Medium Voltage Feeder. In: CONFERENCE ON ELECTRICAL POWER DISTRIBUTION NETWORKS CONFERENCE, 2017, Semnan. *IEEE*, Semnan: EPDC, 2017. p. 68-73.

MALANGE, Fernando Cezar Vieira. **Rede Neuro-Fuzzy-Wavelet para Detecção e Classificação de Anomalias de Tensão em Sistemas Elétricos de Potência**. 2010. 128 f. Tese (Doutorado) - Programa de Pós-Graduação em Engenharia Elétrica, Faculdade de Engenharia de Ilha Solteira, Universidade Estadual Paulista, Ilha Solteira, 2010.

MANDT, Stephan; HOFFMAN, Matthew D.; BLEI, David M. Stochastic Gradient Descent as Approximate Bayesian Inference. **The Journal of Machine Learning Research**, v. 18, n. 1, p. 4873-4907, 2017.

MANSOURIAN, A.; GHANIZADEH, A. R.; GOLCHIN, B. Modeling of Resilient Modulus of Asphalt Concrete Containing Reclaimed Asphalt Pavement using Feed-Forward and Generalized Regression Neural Networks. **Journal of Rehabilitation in Civil Engineering**, v. 7, n. 1, p. 74-92, 2019.

MARQUARDT, Donald W. An Algorithm for Least-squares Estimation of Nonlinear Parameters. **Journal of the Society for Industrial and Applied Mathematics**, v. 11, n. 2, p. 431-441, 1963.

MARAABA, L.; AL-HAMOUZ, Z.; AL-DUWAISH, H. Estimation of High Voltage Insulator Contamination Using a Combined Image Processing and Artificial Neural Networks. In: INTERNATIONAL POWER ENGINEERING AND OPTIMIZATION CONFERENCE, 8, 2014, Langkawi. *IEEE*, Langkawi: PEOCO, 2014. p. 214-219.

MARTINEZ, José Mario; SANTOS, Sandra Augusta. **Métodos Computacionais de Otimização**. Departamento de Matemática Aplicada IMECC-UNICAMP, 1995.

MAS'UD, A. A. *et al.* An Ensemble-Boosting Algorithm for Classifying Partial Discharge Defects in Electrical Assets. **Machines**, v. 5, n. 3, p. 18, 2017a.

MAS'UD, A. A. *et al.* Artificial Neural Network Application for Partial Discharge Recognition: Survey and Future Directions. **Energies**, v. 9, n. 8, p. 574:1-18, 2016.

MAS' UD, A. A. *et al.* Comparison of Artificial Neural Network and Multiple Regression for Partial Discharge Sources Recognition. In: CONFERENCE AND EXHIBITION, 9, 2017, Manama. *IEEE*, Manama: GCCCE, 2017b. p. 1-9.

MAS'UD, A. A. *et al.* Comparison of the Performance of Artificial Neural Networks and Fuzzy Logic for Recognizing Different Partial Discharge Sources. **Energies**, v. 10, n. 7, 2017c.

MAVRIKAKIS, N. C. *et al.* Insulators Pollution Problem: Experience from the Coastal Transmission System of Crete. In: INTERNATIONAL UNIVERSITIES POWER ENGINEERING CONFERENCE, 52, 2017, Heraklion. *IEEE*, Heraklion: UPEC, 2017. p. 1-6.

MCCULLOCH, Warren S.; PITTS, Walter. A Logical Calculus of the Ideas Immanent in Nervous Activity. **The Bulletin of Mathematical Biophysics**, v. 5, n. 4, p. 115-133, 1943.

MENDEL, J. M. **Uncertain Rule-based Fuzzy Systems**. Los Angeles: Springer International Publishing, 2017. p. 684.

MINSKY, Marvin L. Some Methods of Artificial Intelligence and Heuristic Programming. In: SYMPOSIUM ON THE MECHANIZATION OF THOUGHT PROCESSES, 1958, Teddington. Teddington, 1958. p. 1-36.

MINSKY, Marvin L.; PAPERT, Seymour A. **Perceptrons: An Introduction to Computational Geometry**. Boston: Massachusetts Institute of Technology, 2017.

MIRANDA, Vladimiro; FONSECA, Nuno. EPSO-evolutionary Particle Swarm Optimization, a New Algorithm with Applications in Power Systems. In: TRANSMISSION AND DISTRIBUTION CONFERENCE AND EXHIBITION, 2002, Yokohama. *IEEE*, Yokohama: IEEE/PES, 2002. p. 745-750.

MIŠÁK, Stanislav *et al.* A novel approach of partial discharges detection in a real environment. In: INTERNATIONAL CONFERENCE ON ENVIRONMENT AND ELECTRICAL ENGINEERING, 16, 2016, Florence. *IEEE*, Florence: EEEIC, 2016. p. 1-5. 2016.

MO, Lili *et al.* GMDH-based Hybrid Model for Container Throughput Forecasting: Selective Combination Forecasting in Nonlinear Subseries. **Applied Soft Computing**, v. 62, p. 478-490, 2018.

MORÉ, Jorge J. **The Levenberg-Marquardt algorithm: Implementation and Theory**. Berlin: Springer Numerical Analysis, 2006.

MOREIRA, Michel Castro *et al.* Desenvolvimento e Análise de uma Rede Neural Artificial para Estimativa da Erosividade da Chuva para o Estado de São Paulo. **Revista Brasileira de Ciência do Solo**, v. 30, n. 6, p. 1069-1074, 2006.

MURPHY, K. P. **Machine Learning: A Probabilistic Perspective**. Cambridge: The MIT Press, 2012.

NASCIMENTO, Wallas Santos. **Sobre Algumas Características da Entropia de Shannon para Sistemas Atômicos Confinados**. 2013. 133 f. Dissertação. Mestrado em Física - Universidade Federal da Bahia, Instituto de Física. Programa de Pós-graduação em Física, Salvador, 2013.

NASLE, Adib. **Real-time Predictive Systems for Intelligent Energy Monitoring and Management of Electrical Power Networks**. US Patent 9.557.723, 2017, 1-62.

NEVES, José Luis. Pesquisa qualitativa: características, usos e possibilidades. **Caderno de Pesquisas em Administração, São Paulo**, v. 1, n. 3, p. 1-5, 1996.

NGUYEN, Van Nhan; JENSSEN, Robert; ROVERSO, Davide. Automatic Autonomous Vision-based Power Line Inspection: A Review of Current Status and the Potential Role of Deep Learning. **International Journal of Electrical Power & Energy Systems**, v. 99, p. 107-120, 2018.

NIED, Ademir. **Treinamento de Redes Neurais Artificiais Baseado em Sistemas de Estrutura Variável com Taxa de Aprendizado Adaptativa**. 2007. 108 f. Tese. Doutorado em Engenharia Elétrica - Universidade Federal de Minas Gerais. Programa de Pós-graduação em Engenharia Elétrica, Belo Horizonte, 2007.

NOCEDAL, Jorge; WRIGHT, Stephen J. **Numerical Optimization**. New York: Springer, 2006.

PAIVA, Sâmara Cavalcante *et al.* A Wavelet-based Hybrid Islanding Detection System Applied for Distributed Generators Interconnected to AC Microgrids. **International Journal of Electrical Power & Energy Systems**, v. 121, p. 106032, 2020.

PAN, Jun *et al.* LiftingNet: A Novel Deep Learning Network With Layerwise Feature Learning From Noisy Mechanical Data for Fault Classification. **IEEE Transactions on Industrial Electronics**, v. 65, n. 6, p. 4973-4982, 2018.

PARISI, Raffaele *et al.* A Generalized Learning Paradigm Exploiting the Structure of Feedforward Neural Networks. **IEEE Transactions on Neural Networks**, v. 7, n. 6, p. 1450-1460, 1996.

PARK, Kyu-Chil; MOTAI, Yuichi; YOON, Jong Rak. Acoustic Fault Detection Technique for High-Power Insulators. **IEEE Transactions on Industrial Electronics**, v. 64, n. 12, p. 9699-9708, 2017.

PARSAPOOR, Mahboobeh; BILSTRUP, Urban; SVENSSON, Bertil. Forecasting Solar Activity with Computational Intelligence Models. **IEEE Access**, v. 6, p. 70902-70909, 2018.

PASCANU, Razvan; MIKOLOV, Tomas; BENGIO, Yoshua. On the Difficulty of Training Recurrent Neural Networks. In: INTERNATIONAL CONFERENCE ON MACHINE LEARNING, 2013, Atlanta. JMLR, Atlanta: W&CP, v. 28, 2013. p. 1310-1318.

POLISETTY, Satish Kumar; JAYARAM, Shesha; EL-HAG, Ayman. Partial Discharge Classification Using Acoustic Signals and Artificial Neural Networks. In: ELECTROSTATICS JOINT CONFERENCE, 2018, Boston. Boston: Boston University, 2018. p. 1-8.

REZAEIAN, Mohammad Hossein; ESMAEILI, Saeid; FADAEINEDJAD, Roohollah. Generator Coherency and Network Partitioning for Dynamic Equivalencing Using Subtractive Clustering Algorithm. **IEEE Systems Journal**, v. 12, n. 4, p. 3085-3095, 2018.

RIBEIRO, Matheus Henrique Dal Molin; COELHO, Leandro dos Santos. Ensemble Approach Based on Bagging, Boosting and Stacking for Short-term Prediction in Agribusiness Time Series. **Applied Soft Computing**, v. 86, p. 105837, 2020.

ROBERTS, David R. *et al.* Cross-validation Strategies for Data with Temporal, Spatial, Hierarchical, or Phylogenetic Structure. **Ecography**, v. 40, n. 8, p. 913-929, 2017.

ROCHA, P. H. V. *et al.* Inspection in Overhead Insulators Through the Analysis of the Irradiated RF Spectrum. **International Journal of Electrical Power & Energy Systems**, v. 113, p. 355-361, 2019.

ROSENBLATT, Frank. The Perceptron: a Probabilistic Model for Information Storage and Organization in the Brain. **Psychological Review**, v. 65, n. 6, p. 386, 1958.

RUMELHART, David E.; HINTON, Geoffrey E.; WILLIAMS, Ronald J. **Learning Internal Representations by Error Propagation**. California: Univ. San Diego La Jolla Inst. for Cognitive Science, 1985.

RUMELHART, David E.; HINTON, Geoffrey E.; WILLIAMS, Ronald J. Learning Representations by Back-propagating Errors. **Nature**, v. 323, n. 6088, p. 533, 1986.

SAINI, Manish Kumar; BENIWAL, Rajender Kumar. Detection and Classification of Power Quality Disturbances in Wind-grid Integrated System Using Fast Time-time Transform and Small Residual Extreme Learning Machine. **International Transactions on Electrical Energy Systems**, p. 2519, 2018.

SHANNON, C. A Mathematica Theory of Communication. **The Bell System Technical Journal**, v. 27, p. 379-423, 1948.

SHEN, Chuanbo *et al.* Group Method of Data Handling (GMDH) Lithology Identification Based on Wavelet Analysis and Dimensionality Reduction as Well Log Data Pre-processing Techniques. **Energies**, v. 12, n. 8, p. 1509, 2019.

SHEN-PEI, Zhou *et al.* Research on Insulator Fault Diagnosis and Remote Monitoring System Based on Infrared Images. **Procedia Computer Science**, v. 109, p. 1194-1199, 2017.

SHI, Yuhui; EBERHART, Russell C. Empirical Study of Particle Swarm Optimization. In: **PROCEEDINGS OF THE 1999 CONGRESS ON EVOLUTIONARY COMPUTATION**, 1999, Washington. *IEEE*, Washington: CEC99, 1999. p. 1945-1950.

SILVA, Fernando M.; ALMEIDA, Luis B. Speeding up Backpropagation. **Advanced neural computers**, p. 151-158, 1990.

SINGH, Manmohan; DHILLON, J. S. Multiobjective Thermal Power Dispatch Using Opposition-based Greedy Heuristic Search. **International Journal of Electrical Power & Energy Systems**, v. 82, p. 339-353, 2016.

SOKOLOVA, Marina; LAPALME, Guy. A systematic analysis of performance measures for classification tasks. **Information processing & management**, v. 45, n. 4, p. 427-437, 2009.

SONG, Yan *et al.* Gaussian Derivative Models and Ensemble Extreme Learning Machine for Texture Image Classification. **Neurocomputing**, v. 277, p. 53-64, 2018.

STEFENON, S. F. *et al.* Analysis of Training Techniques of ANN for Classification of Insulators in Electrical Power Systems. **IET Generation Transmission & Distribution** (Wiley), v. 14, n. 8, p. 1591-1597, 2020a.

STEFENON, S. F. *et al.* Analysis of the electric field in porcelain pin-type insulators via finite elements software. **IEEE Latin America Transactions**, v. 16, n. 10, p. 2505-2512, 2018.

STEFENON, S. F. *et al.* Diagnostic of Insulators of Conventional Grid Through LabVIEW Analysis of FFT Signal Generated from Ultrasound Detector. **IEEE Latin America Transactions**, v. 15, n. 5, p. 884-889, 2017.

STEFENON, S. F. *et al.* Electrical Insulator Fault Forecasting Based on a Wavelet Neuro-Fuzzy System. **Energies**, v. 13, n. 2, p. 484, 2020b.

STEFENON, S. F.; MEYER, L. H. **Inspection of Electrical Distribution Network**. Saarbrücken: Lambert Academic Publishing, 2015.

STEFENON, S. F.; MEYER, L. H.; MOLINA, F. H. Analysis of the Ultrasound Emitted from Defective Insulators. In: **INTERNATIONAL CONFERENCE ON CONDITION MONITORING AND DIAGNOSIS**, 2014, Jeju Island. *CMD*, Jeju Island, 2014. p. 1-4.

STEFENON, S. F.; MEYER, L. H.; MOLINA, F. H. Real Time Automated Diagnosis of Insulating System Employing Ultrasound Inspection. In: INTERNATIONAL CONFERENCE ON ELECTRICITY DISTRIBUTION, 23, 2015, Lion. *CIRED*, Lion, 2015. p. 1-4.

STEFENON, S. F.; NIED, A. FEM applied to evaluation of the influence of electric field on design of the stator slots in PMSM. **IEEE Latin America Transactions**, v. 17, n. 04, p. 590-596, 2019.

STRACUZZI, David J. **Scalable Knowledge Acquisition Through Cumulative Learning and Memory Organization**. 2006. 147 f. Tese. Doutorado em Computer Science - University of Massachusetts Amherst. Graduate School, Amherst, 2006.

SUN, Zhan-Li *et al.* Sales Forecasting Using Extreme Learning Machine With Applications in Fashion Retailing. **Decision Support Systems**, v. 46, n. 1, p. 411-419, 2008.

TARASSENKO, Lionel. **Guide to Neural Computing Applications**. Burlington: Elsevier, 1998.

TIELEMAN, Tijmen; HINTON, Geoffrey. Lecture 6.5-RMSPROP: Divide the Gradient by a Running Average of its Recent Magnitude. **COURSEERA: Neural Networks for Machine Learning**, v. 4, n. 2, p. 26-31, 2012.

TOLLENAERE, Tom. SuperSAB: Fast Adaptive Back Propagation with Good Scaling Properties. **Neural Networks**, v. 3, n. 5, p. 561-573, 1990.

ÜNAL, Fatih; EKICI, Sami. A Fault Location Technique for HVDC Transmission Lines Using Extreme Learning Machines. In: INTERNATIONAL ISTANBUL SMART GRID AND CITIES CONGRESS AND FAIR, 5, 2017, Istanbul. *IEEE*, Istanbul, 2017, p. 125-129.

VALPOLA, Harri. From Neural PCA to Deep Unsupervised Learning. In: BINGHAM, Ella *et al.* **Advances in Independent Component Analysis and Learning Machines**. London: Academic Press Elsevier, 2015. p. 143-171.

VAN HEESWIJK, M. *et al.* Adaptive Ensemble Models of Extreme Learning Machines for Time Series Prediction. In: International Conference on Artificial Neural Networks, 2009, Limassol, **Artificial Neural Networks**, Cyprus, Springer, 2009, p. 305-314.

VAN HEESWIJK, M. *et al.*, GPU-accelerated and Parallelized ELM Ensembles for Large-scale Regression. **Neurocomputing**, v. 74, n. 16, p. 2430-2437, 2011.

VINAY, A. *et al.* Face Recognition Using Interest Points and Ensemble of Classifiers. In: INTERNATIONAL CONFERENCE ON RECENT ADVANCES IN INFORMATION TECHNOLOGY, 4, 2018, Dhanbad. *IEEE*, Dhanbad, 2018. p. 1-8.

VELASCO, Jose Angel; AMARIS, Hortensia; ALONSO, Monica. Deep Learning Loss Model for Large-scale Low Voltage Smart Grids. **International Journal of Electrical Power & Energy Systems**, v. 121, p. 106054, 2020.

VELÁSQUEZ, R. M. Arias; LARA, J. V. Mejía. Robot Unit for Cost and Time Balance Using Automatic Inspection on Overhead Lines. In: ANDESCON, 2016, Arequipa. *IEEE*, Arequipa, 2017, p. 1-4.

VERSTRAETE, Jörg. The Spatial Disaggregation Problem: Simulating Reasoning Using a Fuzzy Inference System. **IEEE Transactions on Fuzzy Systems**, v. 25, n. 3, p. 627-641, 2017.

VLACHOGIANNIS, John G.; LEE, Kwang Y. Economic Load Dispatch: A Comparative Study on Heuristic Optimization Techniques with an Improved Coordinated Aggregation-based PSO. **IEEE Transactions on Power Systems**, v. 24, n. 2, p. 991-1001, 2009.

VON NEUMANN, John. The General and Logical Theory of Automata. **Cerebral Mechanisms in Behavior**, v. 1, n. 41, p. 1-2, 1951.

WANG, G. T.; LI, P.; CAO, J. T. Variable Activation Function Extreme Learning Machine Based on Residual Prediction Compensation. **Soft Computing**, v. 16, n. 9, p. 1477-1484, 2012.

WANG, H.; JIN, L.; LU, P.; GAO, K. Physical Stages Dividing and Quantitative Study of Insulator Surface Discharges. In: INTERNATIONAL SYMPOSIUM ON DISCHARGES AND ELECTRICAL INSULATION IN VACUUM, 28, 2018, Greifswald. *IEEE*, Greifswal: ISDEIV, 2018, p. 131-134.

WANG, Yueying *et al.* Dissipativity-based Fuzzy Integral Sliding Mode Control of Continuous-Time TS Fuzzy Systems. **IEEE Transactions on Fuzzy Systems**, v. 26, n. 3, p. 1164-1176, 2017.

WANG, Zhichao *et al.* High Voltage Transmission Lines Remote Distance Inspection System. In: INTERNATIONAL CONFERENCE ON MECHATRONICS AND AUTOMATION, 2015, Beijing. *IEEE*, Beijing: ICMA, 2015. p. 1676-1680.

WANG, Haizheng *et al.* Probability Modeling for PV Array Output Interval and its Application in Fault Diagnosis. **Energy**, v. 189, p. 116248, 2019a.

WANG, Shouxiang *et al.* Bi-directional Long Short-term Memory Method Based on Attention Mechanism and Rolling Update for Short-term Load Forecasting. **International Journal of Electrical Power & Energy Systems**, v. 109, p. 470-479, 2019b.

WIDROW, Bernard; LEHR, Michael A. 30 Years of Adaptive Neural Networks: Perceptron, Madaline, and Backpropagation. **Proceedings of the IEEE**, v. 78, n. 9, p. 1415-1442, 1990.

XIA, Y.; LU, J.; LI, H.; XU, H. A Deep Learning Based Image Recognition and Processing Model for Electric Equipment Inspection. In: CONFERENCE ON ENERGY INTERNET AND ENERGY SYSTEM INTEGRATION, 2, 2018, Beijing. *IEEE*, Beijing: EI2, 2018, p. 1-6.

XU, Fang *et al.* Fast Nonlinear Model Predictive Control on FPGA Using Particle Swarm Optimization. **IEEE Transactions on Industrial Electronics**, v. 63, n. 1, p. 310-321, 2015.

XU, Lei. Least Mean Square Error Reconstruction Principle for Self-organizing Neural-nets. **Neural Networks**, v. 6, n. 5, p. 627-648, 1993.

YAMASHITA, Takahiko *et al.* Critical Current for Phase Shift of Dry-band Discharge on Wet Polluted Insulators. **Journal of Electrostatics**, v. 97, p. 51-57, 2019.

YAM, J. Y. F; CHOW, T. W. S. Extended Least Squares Based Algorithm for Training Feedforward Networks. **IEEE Transactions on Neural Networks**, v. 8, n. 3, p. 806-810, 1997.

YANG, Hui; FU, Yating; WANG, Dianhui. Multi-ANFIS Model Based Synchronous Tracking Control of High-speed Electric Multiple Unit. **IEEE Transactions on Fuzzy Systems**, v. 26, n. 3, p. 1472-1484, 2017.

YANG, Longzhi; CHAO, Fei; SHEN, Qiang. Generalized Adaptive Fuzzy Rule Interpolation. **IEEE Transactions on Fuzzy Systems**, v. 25, n. 4, p. 839-853, 2016.

YANG, Miin-Shen; NATALIANI, Yessica. A Feature-reduction Fuzzy Clustering Algorithm Based on Feature-weighted Entropy. **IEEE Transactions on Fuzzy Systems**, v. 26, n. 2, p. 817-835, 2018.

YANG, Shanchao; LIU, Jing. Time-series Forecasting Based on High-order Fuzzy Cognitive Maps and Wavelet Transform. **IEEE Transactions on Fuzzy Systems**, v. 26, n. 6, p. 3391-3402, 2018.

YANG, Xin-She; DEB, Suash; FONG, Simon. Accelerated Particle Swarm Optimization and Support Vector Machine for Business Optimization and Applications. In: INTERNATIONAL CONFERENCE ON NETWORKED DIGITAL TECHNOLOGIES, 2011, Berlin. Berlin: Springer, 2011. p. 53-66.

YIN, Peng-Yeng *et al.* A Complementary Cyber Swarm Algorithm. **International Journal of Swarm Intelligence Research (IJSIR)**, v. 2, n. 2, p. 22-41, 2011.

YU, X.; EFE, M. O.; KAYNAK, O. A General Backpropagation Algorithm for Feedforward Neural Networks Learning. **IEEE Transactions on Neural Networks**, v. 13, n. 1, p. 251-254, 2002.

ZHAI, J.; ZHANG, S.; ZHANG, M.; LIU, X. Fuzzy Integral-based ELM Ensemble for Imbalanced Big Data Classification. **Soft Computing**, v. 22, n. 11, p. 3519-3531, 2018.

ZHAI, J.; ZHANG, S.; WANG, C. The Classification of Imbalanced Large Data Sets Based on MapReduce and Ensemble of ELM Classifiers. **International Journal of Machine Learning and Cybernetics**, v. 8, n. 3, p. 1009-1017, 2017.

ZHAN, Zhi-Hui *et al.* Adaptive Particle Swarm Optimization. **IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)**, v. 39, n. 6, p. 1362-1381, 2009.

ZHANG, Boyang *et al.* Ensemble Based Reactivated Regularization Extreme Learning Machine for Classification. **Neurocomputing**, v. 275, p. 255-266, 2018.

ZHANG, Da; MENG, Fancui. Research on the Interrelation between Temperature Distribution and Dry Band on Wet Contaminated Insulators. **Energies**, v. 12, n. 22, p. 4289, 2019.

ZHANG, S.; WANG, Y.; LIU, M.; BAO, Z. Data-Based Line Trip Fault Prediction in Power Systems Using LSTM Networks and SVM. **IEEE Access**, v. 6, p 7675-7686, 2018.

ZHAO, Jing; LIN, Chih-Min. Wavelet-TSK-type Fuzzy Cerebellar Model Neural Network for Uncertain Nonlinear systems. **IEEE Transactions on Fuzzy Systems**, v. 27, n. 3, p. 549-558, 2018.

ZHAO, Minghang *et al.* Multiple Wavelet Coefficients Fusion in Deep Residual Networks for Fault Diagnosis. **IEEE Transactions on Industrial Electronics**, v. 66, n. 6, p. 4696-4706, 2019.

ZHAO, Y. On-line Neural Network Learning Algorithm with Exponential Convergence Rate. **Electronics Letters**, v. 32, n. 15, p. 1381-1382, 1996.

ZHAO, Zheng *et al.* LSTM Network: A Deep Learning Approach for Short-term Traffic Forecast. **IET Intelligent Transport Systems**, v. 11, n. 2, p. 68-75, 2017.

ZHOU, Guian; SI, Jennie. Advanced Neural-network Training Algorithm with Reduced Complexity Based on Jacobian Deficiency. **IEEE Transactions on Neural Networks**, v. 9, n. 3, p. 448-453, 1998.

ZHOU, Z. *et al.* RFSEN-ELM: Selective Ensemble of Extreme Learning Machines Using Rotation Forest for Image Classification. **Neural Network World**, v. 27, n. 5, p. 499-517, 2017.

ZOELLER, Clemens *et al.* Evaluation and Current-response-based Identification of Insulation Degradation for High Utilized Electrical Machines in Railway Application. **IEEE Transactions on Industry Applications**, v. 53, n. 3, p. 2679-2689, 2017.