CENTRO UNIVERSITÁRIO SENAI CIMATEC

PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM COMPUTACIONAL E TECNOLOGIA INDUSTRIAL

Mestrado em Modelagem Computacional e Tecnologia Industrial

Dissertação de Mestrado

# Análise, avaliação e validação do uso de técnicas de aprendizado de máquina para detecção de falhas em turbinas eólicas do tipo PMSG - Gerador síncrono de ímãs permanentes

Apresentada por: Henrique Gomes Mergulhão
Orientador: Prof. Dr. André Telles da Cunha Lima
Coorientador: Prof. Dr. Oberdan Rocha Pinheiro

Maio de 2024

Henrique Gomes Mergulhão

# Análise, avaliação e validação do uso de técnicas de aprendizado de máquina para detecção de falhas em turbinas eólicas do tipo PMSG - Gerador síncrono de ímãs permanentes

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Modelagem Computacional e Tecnologia Industrial, Curso de Mestrado em Modelagem Computacional e Tecnologia Industrial do Centro Universitário SENAI CIMATEC, como requisito parcial para a obtenção do título de **Mestre em Modelagem Computacional e Tecnologia Industrial**.

Orientador: Prof. Dr. André Telles da Cunha Lima

Coorientador: Prof. Dr. Oberdan Rocha Pinheiro

Salvador
2024

Ficha catalográfica elaborada pela Biblioteca do Centro Universitário SENAI CIMATEC

M552a Mergulhão, Henrique Gomes

Análise, avaliação e validação do uso de técnicas de aprendizado de máquina para detecção de falhas em turbinas eólicas do tipo PMSG - Gerador síncrono de ímãs permanentes / Henrique Gomes Mergulhão. – Salvador, 2024.

87 f. : il. color.

Orientador: Prof. Dr. André Telles da Cunha Lima.
Coorientador: Prof. Dr. Oberdan Rocha Pinheiro.

Dissertação (Mestrado em Modelagem Computacional e Tecnologia Industrial) – Programa de Pós-Graduação, Centro Universitário SENAI CIMATEC, Salvador, 2024.
Inclui referências.

1. Aerogerador síncono de ímãs permanentes. 2. Multilayer perceptron. 3. Seleção de atributos. 4. Sistema de aquisição e monitoramento de dados. 5. Operação e Manutenção. I. Centro Universitário SENAI CIMATEC. II. Lima, André Telles da Cunha. III. Pinheiro, Oberdan Rocha. IV. Título.

CDD 621.816

NDI - 02

Sistema FIEB | SENAI CIMATEC
PELO FUTURO DA INOVAÇÃO

# CENTRO UNIVERSITÁRIO SENAI CIMATEC

**Mestrado Acadêmico em Modelagem Computacional e Tecnologia Industrial**

A Banca Examinadora, constituída pelos professores abaixo listados, aprova a Defesa de Mestrado, intitulada **"Análise, avaliação e validação do uso de técnicas de aprendizado de máquina para detecção de falhas em turbinas eólicas do tipo PMSG - Gerador síncrono de ímãs permanentes**" apresentada no dia 27 de maio 2024, como parte dos requisitos necessários para a obtenção do Título de Mestre em Modelagem Computacional e Tecnologia Industrial.

Assinado eletronicamente por:
André Telles da Cunha Lima
CPF: ***.510.438-**
Data: 18/06/2024 15:03:38 -03:00

Orientador: **Prof. Dr. André Telles da Cunha Lima**
SENAI CIMATEC

Assinado eletronicamente por:
Oberdan Rocha Pinheiro
CPF: ***.073.705-**
Data: 18/06/2024 07:57:18 -03:00

Coorientador: **Prof. Dr. Oberdan Rocha Pinheiro**
SENAI CIMATEC

Assinado eletronicamente por:
Alex Álisson Bandeira Santos
CPF: ***.191.765-**
Data: 17/06/2024 15:59:03 -03:00

Membro Interno: **Prof. Dr. Alex Álisson Bandeira Santos**
SENAI CIMATEC

Assinado eletronicamente por:
Helder Henrique Lima Diniz
CPF: ***.424.184-**
Data: 17/06/2024 17:52:42 -03:00

Membro Externo: **Prof. Dr. Helder Henrique Lima Diniz**
UPE

*Dedico este trabalho aos meus pais,*
*Antônio e Zuleide Mergulhão,*
*à minha irmã, Elyzângela,*
*à minha esposa Carol e filhos*
*Cassiano e Martim Campos Mergulhão.*

# Agradecimentos

Agradeço a Deus, sempre ao meu lado, dias, noites e madrugadas de estudo, dando-me saúde e disposição para buscar novos desafios e nunca desistir.

À minha família, em especial aos meus pais Antônio Lins Mergulhão e Zuleide Gomes Mergulhão e minha irmã Elyzângela, à minha amiga, companheira e esposa Carol, e filhos Cassiano e Martim Campos Mergulhão por sempre estarem do meu lado, me dando forças e incentivo além da compreensão em momentos em que eu não pude dar toda a atenção que eles merecem. À minha sogra Isaura e cunhada Tina por sempre darem suporte com meus filhos que tanto amo.

Agradeço ao Prof. Dr. André Telles da Cunha Lima pela orientação e incentivo na condução deste trabalho e direcionamento da pesquisa ao longo destes dois anos de trabalho conjunto. Imenso agradecimento ao Prof. Dr. Oberdan Rocha Pinheiro por sua orientação, sugestões e, como coordenador, pelo convite para integrar o seleto grupo, responsável pelo Subprojeto P4.3 da Planta Híbrida, o que me proporcionou grande conhecimento, participação em publicações e discussões que enriqueceram minha pesquisa. Aos amigos do P4.3 e ao amigo Rogério Santiago estendo meus agradecimentos, pelo apoio e discussões técnicas.

Ao professor e amigo, Prof. Dr. Márcio José das Chagas Moura, pela contribuição e incentivo ao ingresso no Mestrado, na fase de pré-projeto e pelo convite para participar de suas aulas na UFPE, o que me ajudaram bastante na consolidação das ideias e abordagens que mais tarde fariam parte do meu estudo.

Aos membros da banca pela disponibilidade de participar e contribuir, com avaliação criteriosa e construtiva, para o resultado da pesquisa.

Aos meus amigos do Mestrado Giorgio Onofre, Hugo Salvador, Elisabete Barreto, Carlos Eduardo Malaquias, Pedro Gomes e Tiago Richter, meu muito obrigado pela parceria, amizade e excelentes momentos de trabalho em equipe.

Agradeço ao Senai Cimatec, pelo suporte no mestrado e no desenvolvimento da pesquisa, à ANEEL (Agência Nacional de Energia Elétrica) e a Eletrobras CHESF pelo financiamento do projeto.

Salvador, Brasil — Henrique Gomes Mergulhão
27 de Maio de 2024

# Resumo

A implementação de fontes de energias, que atendam a requisitos de desenvolvimento sustentável, está em constante expansão, com destaque para a geração eólica, que desempenha um importante papel importante no cumprimento das metas de redução das emissões de carbono e diversidade do fornecimento de energia. As turbinas eólicas são sistemas bastante complexos, e os elevados custos relacionados à operação e manutenção tem estimulado à busca de procedimentos cada vez mais eficazes para sua redução. Neste estudo, propõe-se que desvios do comportamento esperados em nível de sistema em turbinas eólicas sejam detectados e classificados como falha usando uma abordagem baseada em dados operacionais, de diversos sensores instalados em aerogeradores de acionamento direto tipo PMSG - Gerador Síncrono de Imãs Permanentes, adquiridos pelo sistema supervisório de controle e aquisição de dados - SCADA, fornecidos pela Companhia ELETROBRAS, localizado no Brasil, mais especificamente no estado da Bahia, em Casa Nova. De modo a conseguir a detecção de falhas e capturar informações sobre o estado do aerogerador, utilizou-se uma metodologia baseada em aprendizado de máquina. Foram utilizados os métodos CFS (seleção de recursos baseada em correlação) e Random Forest para seleção de atributos mais relevantes do conjunto de dados. Em seguida, duas redes neurais MLP - Multilayer perceptron são configuradas, validadas e avaliadas, tendo como entrada os atributos mais relevantes selecionados pelos dois métodos anteriores para classificar se as turbinas eólicas apresentam comportamento de falha ou normalidade. Os resultados comprovam que a estratégia de seleção dos atributos mais relevantes para a rede neural Multilayer perceptron (MLP), para classificação de falhas resultou numa elevada taxa de reconhecimento de classificação de detecção de falhas em turbinas eólicas. A RNA CFS-MLP (rede neural com atributos selecionados pelo método CFS) demonstrou um excelente percentual de instâncias classificadas corretamente igual a 98,73% e precisão média de 98,9% para a ocorrência de falhas e de 98,5% para a classificação correta de funcionamento normal do aerogerador com a base de dados utilizada. Enquanto a RNA RF-MLP (rede neural com atributos selecionados pelo método Random Forest) demonstrou um percentual de instâncias classificadas corretamente igual a 97,88% e precisão média de 98,4% para a ocorrência de falhas e de 97,4% para a classificação correta de funcionamento normal do aerogerador. Todavia, somente a avaliação da performance não é suficiente, sendo importante se levar em consideração a expressiva redução da quantidade de sensores (atributos) selecionados.

**Palavras-chave**: Aerogerador síncono de ímãs permanentes, Multilayer perceptron, Seleção de atributos, sistema de aquisição e monitoramento de dados, Operação e manutenção.

# Abstract

The implementation of sustainable energy sources is constantly expanding, with particular emphasis on wind power, which plays a significant role in meeting carbon emission reduction targets and diversifying energy supply. Wind turbines are highly complex systems, and the high costs associated with operation and maintenance have spurred the search for increasingly effective procedures to reduce them. This study proposes that deviations from expected system behavior in wind turbines be detected and classified as faults using a data-driven approach, leveraging operational data from various sensors installed in direct-drive PMSG (Permanent Magnet Synchronous Generator) wind turbines. This data is acquired through the Supervisory Control and Data Acquisition (SCADA) system provided by ELETROBRAS, located in Brazil, specifically in the state of Bahia, Casa Nova. To achieve fault detection and capture information about the turbine's condition, a machine learning methodology is employed. The CFS (Correlation-based Feature Selection) and Random Forest methods are used to select the most relevant attributes from the dataset. Subsequently, two Multi-layer Perceptron (MLP) neural networks are configured, validated, and evaluated, using the most relevant attributes selected by the two aforementioned methods as input to classify whether wind turbines exhibit fault behavior or operate normally. The results demonstrate that the strategy of selecting the most relevant attributes for the MLP neural network for fault classification resulted in a high rate of fault detection in wind turbines. The CFS-MLP neural network showed an excellent percentage of instances correctly classified, at 98.73%, with an average precision of 98.9% for fault occurrences and 98.5% for correct classification of normal turbine operation, using the dataset. Meanwhile, the RF-MLP neural network exhibited a percentage of instances correctly classified at 97.88%, with an average precision of 98.4% for fault occurrences and 97.4% for correct classification of normal turbine operation. However, assessing performance alone is not sufficient; it is also important to consider the reduction of the number of selected sensors (attributes).

**Keywords**: Synchronous permanent magnet wind turbine, Multilayer perceptron, Attribute selection, Data acquisition and monitoring system, Operation and maintenance.

# Sumário

# Lista de Tabelas

# Lista de Figuras

# Lista de Siglas

| | |
|---|---|
| A.C ..... | Antes de Cristo |
| AM ..... | Aprendizado de máquina |
| ANEEL ..... | Agência Nacional de Energia Elétrica |
| ANN ..... | Artificial Neural Network |
| CBM ..... | Condition-Based Monitoring |
| CFS ..... | Algotimo de seleção de atributos baseado em correlação |
| CM ..... | Condition Monitoring |
| $CO_2$ ..... | Dióxido de carbono |
| CSV ..... | Comma-Separated-Values |
| ELETROBRÁS | Centrais Elétricas Brasileiras S.A. |
| FP ..... | False positive ou Falso positivo |
| GW ..... | Gigawatts |
| IA ..... | Inteligência Artificial |
| MAE ..... | Mean Absolute Error |
| MLP ..... | Multi-Layer Perceptron ou Perceptron de múltiplas camadas |
| MW ..... | Megawatts |
| O&M ..... | Operação e Manutenção |
| P&D ..... | Pesquisa e Desenvolvimento |
| PHM ..... | Prognostic and Health Management ou Prognóstico de Gestão de Saúde |
| PMSG ..... | Permanent Magnet Synchronous Generator , ou Gerador Síncrono de Imã Permanente |
| PPGMCTI .. | Programa de Pós Graduação de Modelagem Computacional e Tecnologia Industrial SENAI CIMATEC |
| PROP&D ... | Procedimento do Programa de Pesquisa e Desenvolvimento |
| REN ..... | Resolução Normativa |
| RNA ..... | Rede Neural Artificial |
| RF ..... | Random Forest |
| SCADA ..... | Supervisory Control and Data Acquisition ou Sistema de Supervisão e Aquisição de Dados |
| SGH ..... | Sistema de Geração Híbrido |
| ton ..... | tonelada |
| TP ..... | True Positive ou Verddeiro Positivo |
| WT ..... | Wind Turbine |
| WWW ..... | World Wide Web |

# Lista de Simbolos

| | |
|---|---|
| $\alpha$ ..... | termo de momento |
| $d_k(n)$ ..... | resposta desejada para o neurônio $k$ na iteração n |
| $e_k(n)$ ..... | sinal de erro na saída do neurônio k, para a iteração n. |
| $\xi_{media}$ ..... | Energia média do erro |
| $\xi(n)$ ..... | valor instantâneo da Energia do erro |
| $m$ ..... | quantidade neurônios de determinada camada |
| $n$ ..... | tempo discreto, quantidade de iterações ou épocas |
| $\eta$ ..... | taxa de aprendizagem |
| $t$ ..... | tempo contínuo |
| $v_j$ ..... | campo local induzido ou potencial de ativação do neurônio j |
| w* ..... | vetor de peso ótimo |
| $w_{ij}(n)$ ..... | peso sináptico conectando a saída do neurônio $i$ à entrada do neurônio $j$, na iteração $n$ |
| $x_k$ ..... | k-ésimo elemento do vetor de entrada $x$ |
| $y_i$ ..... | valor observado ou real para a amostra $i$; |
| $\hat{y}_i$ ..... | valor previsto pelo modelo para a amostra $i$. |
| $y_k$ ..... | o k-ésimo elemento do vetor de saída y |
| $\varphi_k(\cdot)$ ..... | função de ativação não linear do neurônio $k$ |

Capítulo Um

# Introdução

A energia é o vetor fundamental para o desenvolvimento da sociedade. A civilização, de forma global, tem recorrido a combustíveis fósseis como fonte histórica de energia, mas está em constante busca de soluções técnicas que alicercem o crescimento populacional, realizando transições energéticas para fontes alternativas de energia renováveis, que sejam economicamente eficientes e justifiquem seus investimentos, ao tempo que reduzam a emissão de poluentes no meio ambiente (GRIFFITH, 2022).

A energia eólica, por sua característica funcional, custo zero de insumo, não poluente e de disponibilidade infindável, sem deixar de considerar os benefícios sociais e ambientais atrelados, vem tendo papel de protagonismo no universo das fontes alternativas renováveis (VIAN et al., 2021), além do desenvolvimento recente, quando as turbinas eólicas comerciais passaram a ter capacidade de geração de kW para MW (kilowatts para Megawatts) (SI et al., 2017).

Para assegurar a viabilidade técnica, econômica e ambiental de um determinado empreendimento, devem ser considerados além dos custos de implantação, os de operação e manutenção (O&M). Sendo os últimos os mais representativos, quando o assunto é energia eólica, com 10-15% e 25-30% para parques eólicos onshore e offshore, respectivamente (ZHANG; ROBINSON; BASU, 2023). As turbinas eólicas, responsáveis pela geração eólica (TURNBULL; CARROLL; MCDONALD, 2021), normalmente funcionam 24 horas por dia, 7 dias por semana e como todos os equipamentos mecatrônicos, estão sujeitas a falhas durante seu ciclo de vida e o fator complicador deve-se ao fato de estarem em locais remotos, assim,interrupções inesperadas podem levar a enormes perdas financeiras. Estas condições e características combinadas tornam a operação e manutenção (O&M) de um parque eólico uma tarefa muito desafiadora, não só do ponto de vista técnico, mas também do ponto de vista econômico (LEITE; ARAÚJO; ROSAS, 2018).

Em busca de uma redução significativa de danos e manutenções da máquina, assim como aumento de vida úil, o monitoramento de condição eficaz e a detecção de certas falhas antes que elas atinjam níveis de gravidade catastróficos podem reduzir os custos de O&M juntamente com a otimização do intervalo de manutenção (ZHANG; ROBINSON; BASU, 2023) e tempo de inatividade (TURNBULL; CARROLL; MCDONALD, 2021).

Em geral, a maioria da literatura categoriza os algoritmos para previsão de falhas em abordagens baseadas em física e baseados em dados.(ATAMURADOV et al., 2017):(1) abordagem baseada em modelos numéricos, utilizando modelos físicos de uma turbina e

seus subcomponentes (caixa de engrenagens, pás, etc.) para identificar anomalias (PANDIT; INFIELD, 2018); (2) abordagem baseada em dados aquisitados de um sistema de controle e aquisição de supervisão (SCADA) para prever eventos futuros (SI et al., 2017) ou abordagem baseada em processamento de sinais para analisar estados, por exemplo, vibrações (QIAO; LU, 2015).

No caso de sistemas complexos, as abordagens baseadas na física dependem de uma modelagem precisa que são capazes de realizar uma detecção de falhas eficiente se a dinâmica do sistema puder ser bem descrita, mas têm dificuldades na prática, se as aplicações, como as turbinas eólicas industriais, não se adequarem suficientemente bem aos pressupostos. (SI et al., 2017)

Muitas publicações têm focado na abordagem baseada em dados, utilizando os dados de monitoramento da condição coletados dos sensores instalados junto com aprendizado de máquina O desempenho dessa abordagem, no entanto, depende muito do número e da qualidade dos dados, pois requer um grande número de tendências ou dados executados até a falha para a construção precisa do modelo (CHATTERJEE; DETHLEFS, 2021) e utilização de sensores apropriados (ZHANG; ROBINSON; BASU, 2023). Até o momento, vários tipos de algoritmos de detecção de falhas foram utilizados, como Kmeans (algoritmo de clusterização) (ELIJORDE; KIM; LEE, 2014), rede neural convolucional (CNN) (XIAO et al., 2021) (SANTOLAMAZZA; DADI; INTRONA, 2021) , KNN (K-ésimo vizinho mais próximo) (WANG; LIU, 2021), Random forests (SI et al., 2017), Suport vector machines (SVM) (TANG et al., 2014) (WENYI et al., 2013), Processos Gaussianos (PANDIT; INFIELD, 2018)

Mineração de dados e aprendizado de máquina andam em sintonia, dos quais vários insights podem ser derivados por meio de algoritmos de aprendizado adequados. Houve enorme progresso na mineração de dados e no aprendizado de máquina como resultado da evolução da tecnologia inteligente e da detecção automatizada de padrões significativos nos dados (KOTSIANTIS, 2007). O Sistema de Monitoramento de Condições pode facilitar a prevenção de falhas do sistema e a melhoria da disponibilidade do aerogerador por meio de detecções de falhas em estágio inicial (ZHANG; ROBINSON; BASU, 2023).

Existem etapas importantes, prévias (pré-processamento) e posteriores (pós-processamento) à mineração de dados para que o processo de descoberta de conhecimento da base de dados, seja realizado com sucesso. A representação e a qualidade dos dados do conjunto de amostras têm papel primoedial no aprendizado de máquina (GERON, 2019).

O monitoramento da condição é capaz de prever falhas com resolução de alta frequência, mas esta abordagem é mais cara em comparação ao SCADA (YANG et al., 2014). Assim, os sistemas SCADA tornam-se mais favoráveis para aplicação.(ZHANG; ROBINSON;

BASU, 2023). Entretanto, os sistemas SCADA, têm uma resolução de baixa frequência (TAUTZ-WEINERT; WATSON, 2017), uma vez que os dados SCADA são normalmente coletados sob uma taxa de amostragem de 10 minutos (média dos últimos 10 minutos).

Devido ao fato de que dados históricos de sensores (amostras de entradas) e informações sobre os momentos (saídas pretendidas que estão associadas a essas entradas) em que os problemas ocorreram poderem ser utilizados como preditores, o aprendizado de máquina supervisionado é um método eficaz para treinar algoritmos para identificação de falhas (KHAN; BYUN, 2023).

O método proposto nesta dissertação apresenta uma abordagem para o monitoramento de condição de turbinas eólicas do tipo acionamento direto e de imãs permanentes com dados extraídos do SCADA, sem instalação de sensores extras, contribuindo para a melhoria da confiabilidade e a redução dos custos de operação e manutenção de sistemas essenciais de geração de energia renovável.

Com a utilização de um algoritmo dotado de um agente classificador e um banco de dados, passando por etapas de pré-processamento e com atributos mais relevantes selecionados pelos métodos CFS (HALL, 1999) e Random Forest (BREIMAN, 2001), para permitir o treinamento de uma rede neural artificial Multilayer perceptron. Proporcionando que o algoritmo aprenda uma regra geral que mapeie entradas e saídas, tornando o processo de avaliação das condições dos aerogeradores eficiente, além de permitir reduções nos custos operacionais.

## *1.1 Definição do problema*

Para a tomada de decisão de qual fonte de energia é a mais adequada para implantação numa determinada localidade ou região, estudos são necessários e muitos custos são elencados e avaliados.

Esse trabalho faz parte do projeto de P&D ANEEL, Sistema Inteligente de Geração Híbrida com Armazenamento entre a ELETROBRAS e o Senai-Cimatec, com termos, definições e condições relacionados às fases na cadeia de inovação, observando diretrizes do PROP&D (Procedimento do Programa de Pesquisa e Desenvolvimento) de 2016. A Figura 1.1 apresenta o diagrama do Projeto de P&D.

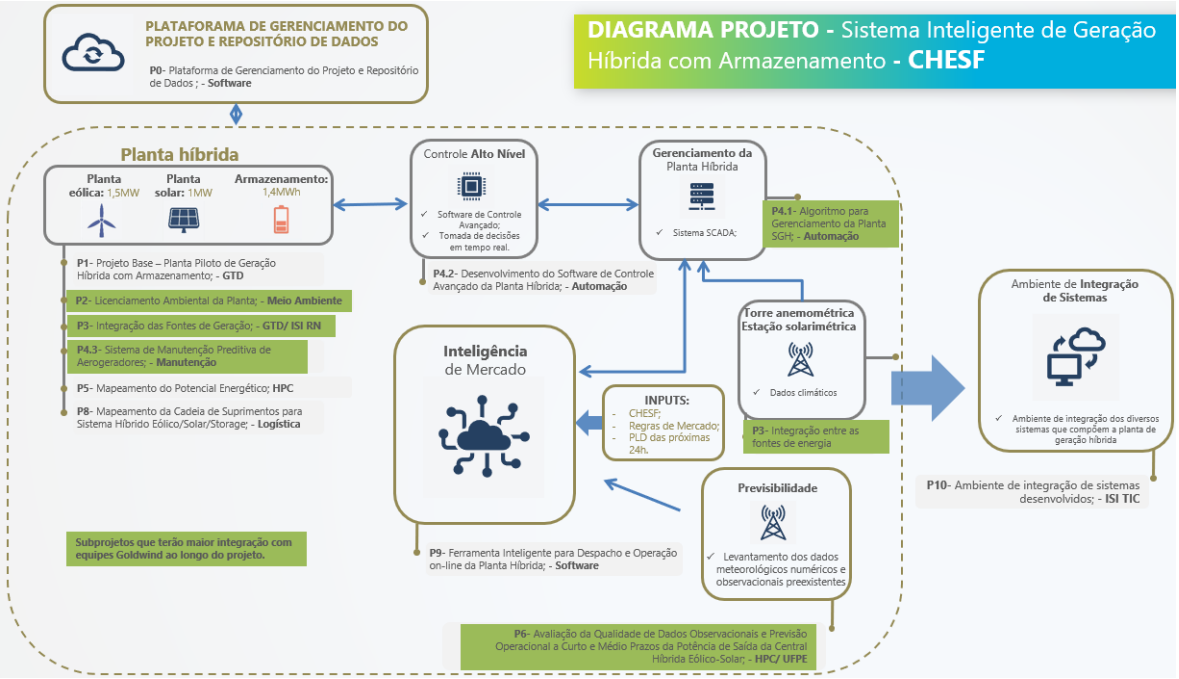


Figura 1.1: Diagrama do Projeto de P&D 00048/0217 ANEEL em Casa Nova/BA. Fonte: Elaborada pelo Projeto P&D.

A pesquisa busca contribuir com o desenvolvimento de uma aplicação de manutenção preditiva, que utilize aprendizado de máquina para processar os dados de monitoramento de um aerogerador de ímãs permanentes, visando detecção de falhas tendo como insumos os dados coletados pelo subprojeto P4.3.

Esta aplicação será integrada ao sistema híbrido eólico-solar-armazenamento proposto. Para alcançar com êxito o objetivo, etapas devem ser executads, consistindo na realização de análise de relatórios de manutenção, coleta de informações do sistema de aquisição de dados de monitoramento, pré-processamento, a redução da dimensionalidade dos dados com seleção dos atributos mais relevantes e modelagem utilizando arquiteturas de aprendizado de máquina buscando uma maior acurácia do modelo proposto.

## *1.2 Objetivo*

Neste trabalho, tem-se como objetivo geral o estudo e desenvolvimento de uma aplicação de manutenção preditiva para uma planta híbrida eólica-solar-armazenamento, que utilize aprendizado de máquina para processar os dados de monitoramento de um aerogerador de ímãs permanentes, seguindo pela construção de um modelo computacional capaz de selecionar os atributos mais relevantes visando detecção de falhas.

A hipótese explorada neste estudo é que avaliações por meio da estimativa da média do erro ao longo das épocas utilizando a melhor seleção de recursos para tarefas de classificação supervisionada para manutenção preditiva de aerogeradores pode ser realizada com base na seleção de atributos, e que tal processo de seleção pode ser benéfico para o algoritmo de aprendizado de máquina perceptron de múltiplas camadas - MLP, alcançando elevados níveis de classificação de padrões.

Os objetivos específicos são:

- Coletar e analisar dados de operação e relatórios de manutenção das turbinas eólicas de Gerador Síncrono de Imãs Permanentes, do inglês, (Permanent Magnet Synchronous Generator - PMSG);
- Desenvolver um agente inteligente para realizar a classificação de novos dados em falha e normalidade;
- Experimentalmente utilizar métodos para seleção de subconjuntos de atributos mais relevantes, que são úteis para construir um bom preditor;
- Comprovar a viabilidade técnica do modelo por meio de testes para monitorar elementos do aerogerador e fornecer subsídios para manutenção preditiva, utilizando ferramentas de aprendizado de máquina;
- A partir dos resultados obtidos, concluir se é possíver ter um retrato de quais sensores são indispensáveis de modo a possibilitar um mapeamento eficaz de falhas no aerogerador para a base de dados utilizada;
- Contribuir para o desenvolvimento de um sistema de manutenção preditiva de turbinas eólicas de uma planta híbrida.

## *1.3 Organização da Dissertação de Mestrado*

Este documento apresenta 5 capítulos e está estruturado da seguinte forma:

- **Capítulo 1 - Introdução**: Este capítulo apresenta a definição do problema, objetivos e justificativas da pesquisa e como esta dissertação de mestrado está estruturada;
- **Capítulo 2 - Revisão da Literatura**: Neste capítulo, é apresentada uma revisão da literatura. Estabelecendo um fluxo de informações conceituais sobre a importância da energia, energia eólica e perspectivas futuras. São abordadas as tecnologias existentes, forma de operação e conexão aos sistemas de potência, ou seja, isolados,

híbridos ou interligados à rede elétrica. Aborda também os tipos de aerogeradores citados na literatura. Um enfoque é dado à manutenção preditiva, apresentados termos e fornecida uma visão geral dos conceitos de aprendizado de máquina supervisionado com enfoque na manutenção. São discutidos aspectos relacionados à seleção de recursos de atributos mais relevantes para aprendizado de máquina utilizando os métodos CFS (Correlation-based Feature Selection) e Random Forest, além da rede neural artificial Multilayer perceptron, configuração de sua arquitetura e processos de aprendizagem;

- **Capítulo 3 - Materiais e Métodos**: Este capítulo descreve o conjunto de dados usados nos experimentos discutidos nos Capítulos 4 e 5, o modelo computacional que leva em consideração o método de seleção de atributos e redes neurais artificais analisando através de métricas de avaliação de modelos para melhor predição de falhas em aerogeradores do tipo PMSG;

- **Capítulo 4 - Resultados e Discussões**: Neste capítulo são apresentados os testes empíricos dos modelos CFS-MLP e RF-MLP com base no banco de dados, que foram aquisitados pelo sistema de controle supervisório e aquisição de dados - SCADA do aerogerador, fornecido pela Companhia ELETROBRAS;

- **Capítulo 5 - Considerações Finais**: Este capítulo apresenta as conclusões obtidas, contribuições e algumas sugestões de atividades de pesquisa a serem desenvolvidas no futuro.

Capítulo Dois

# Revisão da Literatura

## 2.1 A importância da Energia

A relação entre energia e sociedade é coevolutiva, com os insumos influenciando a estrutura e o tamanho da população de uma sociedade (CALVERT, 2015). Diferentes sistemas de energia, como os combustíveis fósseis, carvão e petróleo, tiveram um impacto significativo nas civilizações, levando ao crescimento das "sociedades de combustíveis fósseis" (URRY, 2014).

O uso de combustíveis fósseis levou ao acúmulo de dióxido de carbono na atmosfera, tornando urgente a redução das emissões de $CO_2$. A história das fontes de energia usadas pelas sociedades humanas mostra que as transições nas fontes de energia têm implicações fundamentais para o desenvolvimento social (FISCHER-KOWALSKI; SCHAFFARTZIK, 2015). Diversas tecnologias foram consolidadas para em benefício da humanidade, preservação do meio ambiente, geração de emprego e renda, e crescimento econômico (GRIFFITH, 2022).

## 2.2 Sistemas híbridos

As tecnologias de produção de energia elétrica têm diferentes pontos fortes e fracos quando se trata de apoiar o sistema de rede elétrica em massa para um funcionamento de baixo custo, seguro, estável e confiável, tanto a curto como a longo prazo (DYKES et al., 2020).

Entende-se como Sistema híbrido não apenas um sistema que que utiliza mais de uma fonte de energia, mas como aquele que se beneficia da complementaridade entre fontes e/ou tecnologias de armazenamento, assim, dependendo da disponibilidade dos recursos locais, possibilita,de forma natural ou controlada, que os pontos fracos de uma fonte sejam mitigados ou complementados pelos pontos fortes de outra, permitindo que o sistema seja projetado com a produção de energia maximizada e custos e riscos de interrupções de fornecimento mínimos (BARBOSA et al., 2016).

O comportamento da radiação solar ao longo do dia segue um padrão razoavelmente previsível, iniciando no início da manhã com valores discretos, atingindo um máximo próximo ao meio-dia, e decrescendo até o entardecer. Em contrapartida, o vento pode ser originado por contrastes térmicos em grande escala ou escala local, sendo assim menos

previsível ao longo do dia, mas que tende a ser maior em período de estiagem (MELO; ARAGÃO; CORREIA, 2014).

Como a potência gerada pelas turbinas eólicas é diretamente proporcional à velocidade dos ventos (BURTON et al., 2011),é possível que a geração proveniente dos parques eólicos operem em complementação a outros tipos de geração. A exploração da complementaridade da produção de energias renováveis, tais como, eólica e solar fotovoltaica, podem representar uma oportunidade técnica e economicamente sustentável para suprimento e diversificação energética principalmente em áreas remotas além de diversificar a matriz energética, proporcionando uma maior confiabilidade ao sistema (COUTO; FERREIRA; ESTANQUEIRO, 2020) (PINTO, 2013).

Os sistemas híbridos se beneficiam de forma otimizada das infraestruturas existentes (por exemplo, subestação), e com custos mínimos, para alimentação a uma determinada carga ou conexão a uma rede elétrica, isolada ou conectada a outras redes (COUTO; FERREIRA; ESTANQUEIRO, 2020). Neste sentido, o sistema de armazenamento de energia guarda o excesso de energia não consumida, e despacha em momentos de alta demanda da rede, além de ser utilizado como um meio para garantir a estabilidade do sistema em situações de variações bruscas de geração e carga, contribuindo para a diversificação da matriz energética, evitando assim o uso de gerações não eficientes. (BARBOSA et al., 2016) (LEE; ZHAO, 2022).

A Empresa de Pesquisa Energética demonstrou estudo sobre usinas híbridas em outros países que, apesar de haver potenciais benefícios, ainda existem grandes dificuldades, dais quais destacam-se comerciais e regulatórias para sua implementação (EPE, 2019).

No Brasil, o P&D 00048/0217 ANEEL foi aprovado abordando o tema Planta Híbrida inteligente - Sistema Inteligente com Aerogerador Integrado às Fontes de Energia Eólica, Solar e Storage (banco de baterias) como plataforma de desenvolvimento visando melhorias contínuas no processo de geração de energia elétrica. Na figura 2.1 vemos um esquemático de como a mesma será implantada.

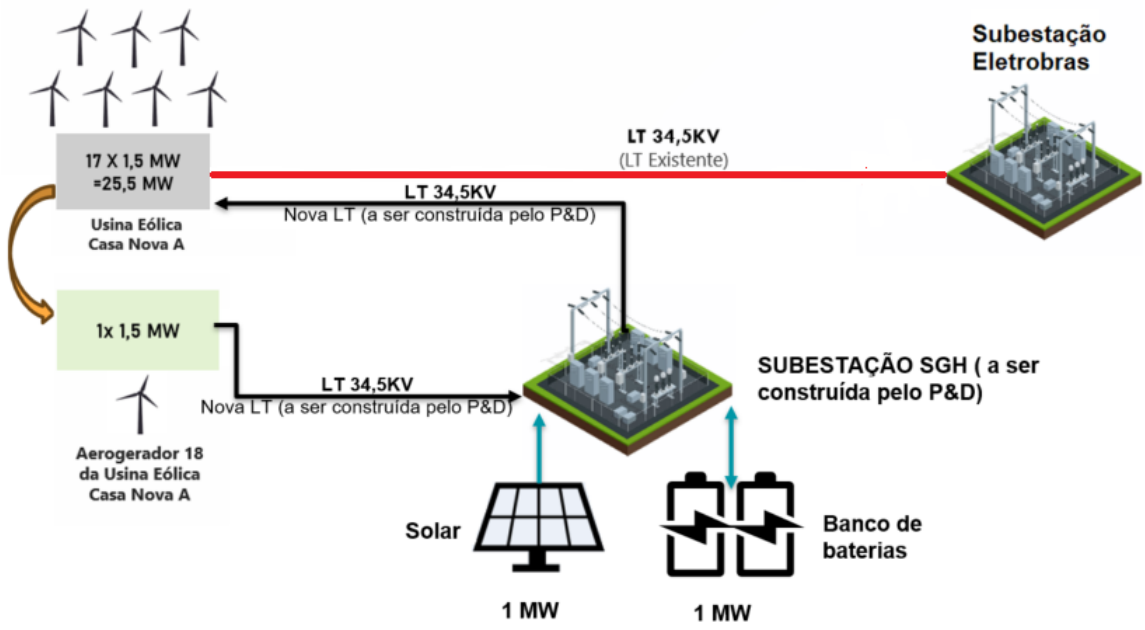


Figura 2.1: Esquemático da Planta Híbrida a ser instalada conforme P&D 00048/0217 ANEEL em Casa Nova/BA. Fonte: Adaptada do Projeto P&D.

## *2.3 Energia eólica*

Energia eólica pode ser entendida como a energia cinética contida nas massas de ar em movimento, independente de sua aplicação (VIAN et al., 2021).

O perfil do crescimento da energia eólica mundial nos últimos anos indica perspectivas promissoras para as próximas décadas, apontando para um crescimento sustentável. O total global de potência instalada de sistemas eólicos interligados à rede somam aproximadamente 934 GW. Seu crescimento ao longo dos últimos anos pode ser visualizado no gráfico 2.2 (WWEA, 2022).

A geração de energia eólica no Brasil não seguiu a tendência mundial de perto (na Europa a produção em grande escala iniciou-se ainda no século XX). Mas após a realização do PROINFA (Programa de Incentivo às Fontes Alternativas) em 2002, considerado um sucesso por ter viabilizado a entrada da tecnologia no país, desenvolvendo fornecedores, fabricantes, instaladores, entre outros e o segundo leilão de energia de reserva realizado em 2009 (o primeiro voltado exclusivamente à fonte eólica), o desenvolvimento do setor no Brasil finalmente engrenou com a contratação de 1,8 GW de capacidade (VIAN et al., 2021).

O Brasil, possui um potencial eólico de 1.500 GW onshore e offshore. Hoje, com 1.039

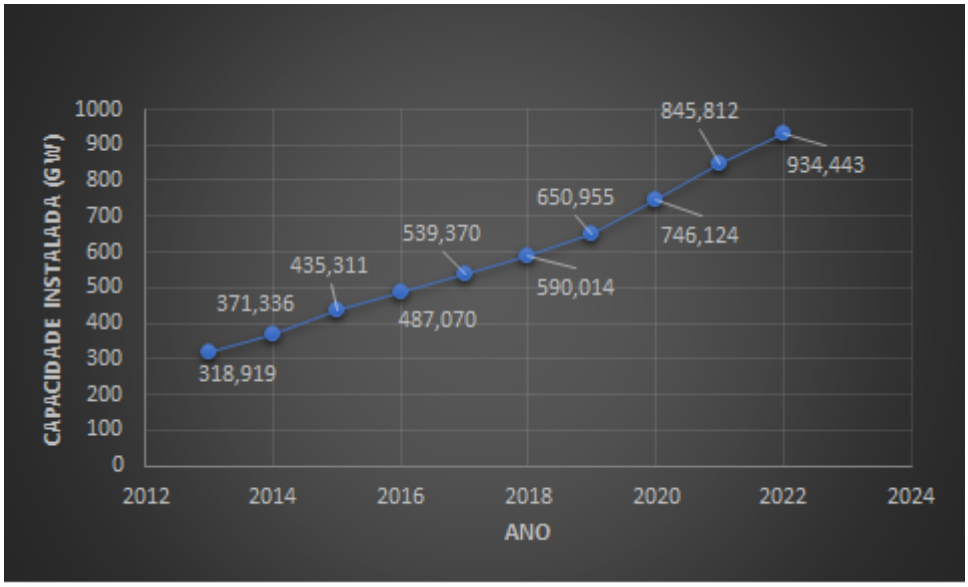


Figura 2.2: Evolução anual da potência eólica instalada mundial. Fonte: (WWEA, 2022).

parques eólicos e mais de 11.000 aerogeradores em operação, ocupa o sexto lugar global onshore de capacidade instalada. Em 2022, 81,45TWh de energia eólica foram gerados, o que representa 12,4% de toda a geração injetada no Sistema Interligado Nacional e um crescimento de 18,85% com relação à geração em 2021. A figura 2.3 mostra a representa a composição da matriz elétrica brasileira atual e a evolução da capacidade instalada de 2005 até 2023.

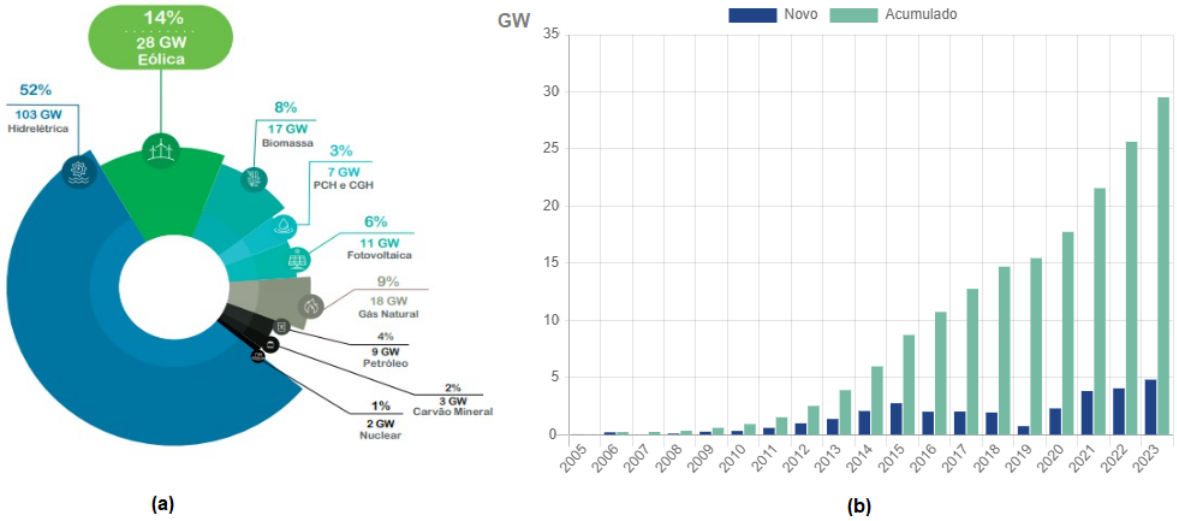


Figura 2.3: (a) Matriz elétrica brasileira (GW) e (b) Evolução da capacidade instalada (GW). Fonte: (ABEEóLICA., 2024).

## 2.4 Aerogeradores

O processo de conversão de energia eólica em energia elétrica ocorre por intermédio de um aerogerador. A rotação produzida (captação da energia cinética), nas pás do rotor, pelo vento é transferida através de um eixo para o gerador (a depender do tipo pode necessitar ou não de um sistema de engrenagens para converter a baixa rotação para uma adequada ao gerador) que por sua vez, através do movimento rotacional converte a energia mecânica em energia elétrica (ROCHA et al., 2023).

Ao longo dos anos de desenvolvimento, variados tipos e configurações de turbinas eólicas, foram propostas: turbinas com eixos vertical e horizontal; com uma, duas ou três pás; com ou sem caixa de engrenagens; com ou sem rotação das pás para controle de potência; e operação com velocidade fixa ou variável (VIAN et al., 2021).

Os tipos de classificação para aerogeradores mais comumente utilizada é com relação ao posicionamento do eixo ao redor do qual as pás giram, se o rotor é de eixo vertical ou horizontal, estes últimos mais comuns, e grande parte da experiência mundial está voltada para sua utilização (ROCHA et al., 2023).

As principais configurações de um aerogerador de eixo horizontal podem ser vistas na figura 2.4.

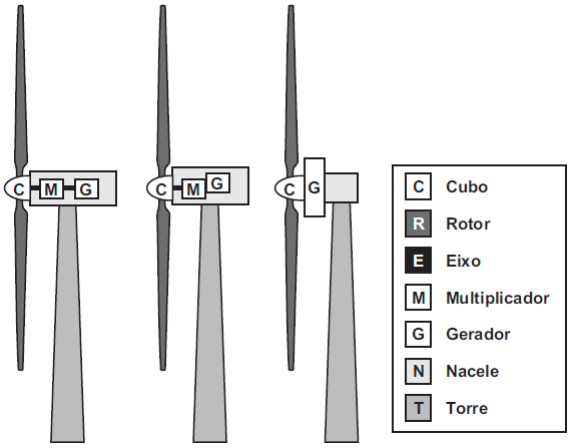


Figura 2.4: Componentes de um aerogerador de eixo horizontal. Fonte: (CRESESB, 2008).

Quanto ao tipo de componente gerador elétrico, diferentes tipos de arranjo são comercializados. Podendo ser comumente divididos em dois grupos: o de velocidade fixa e o de velocidade variável. O sistema de velocidade variável podem utilizar geradores síncronos, ou ainda, dependendo do arranjo, máquinas de indução. Já os sistemas de velocidade fixa tem os geradores de indução como opção mais utilizada. Os aerogeradores de velo-

cidade fixa, usualmente utilizam um sistema de caixa de engrenagens (de transmissão, multiplicadora ou gearbox) responsável por converter a rotação baixa e variável das pás eólicas em uma rotação fixa e elevada, a qual é essencial para o tipo de gerador utilizados nessa categoria de turbina. Nas Figuras 2.5 e 2.6 estão ilustrados os dois tipos de sistema, velocidade fixa e variável (direct-drive) (PINTO, 2013). Dentre as possibilidades de aerogeradores direct-drive está o gerador síncrono de imãs permanentes tipo PMSGs - do inglês, Permanent Magnet Synchronous Generators.

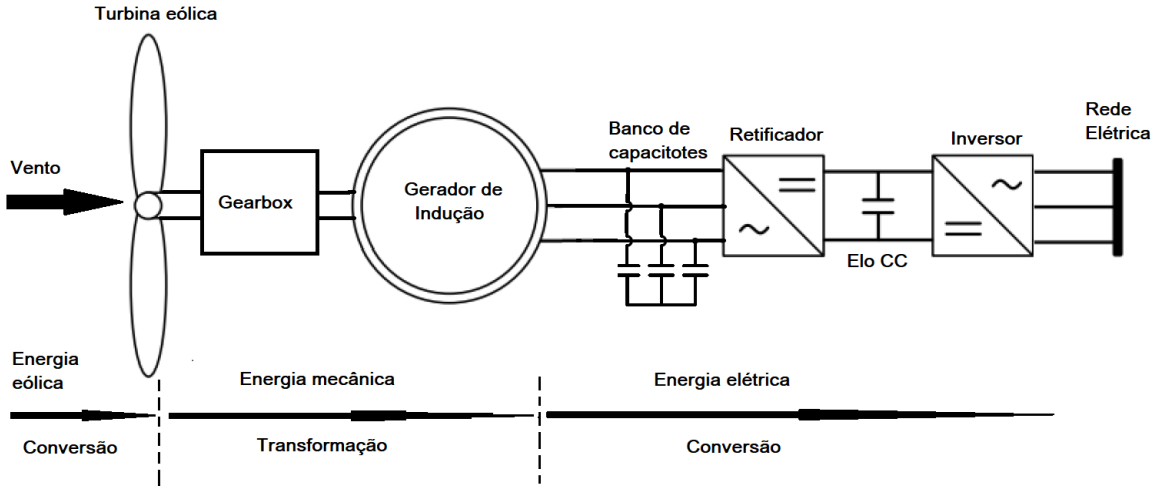


Figura 2.5: Conversão da energia do vento em energia elétrica utilizando aerogerador de indução e Gearbox. Fonte: Adaptação de (PINTO, 2013) e (BONELLI, 2010).

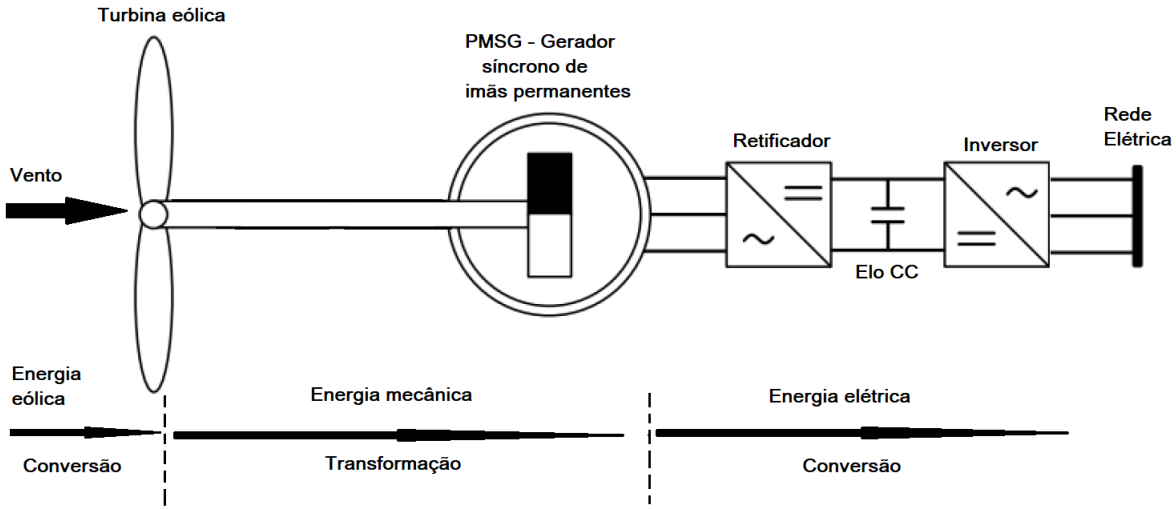


Figura 2.6: Conversão da energia do vento em energia elétrica utilizando Gerandor síncrono de ímãs permanentes. Fonte: Adaptação de (PINTO, 2013) e (BONELLI, 2010).

Cada solução apresenta vantagens e desvantagens que devem ser analisadas na sua incor-

poração ao sistema de conversão de energia eólica. A tecnologia líder de mercado, é a do DFIG, que permite o uso de geradores de massa menor, evita o uso de equipamentos complexos de geração de energia com grande número de pólos e requer apenas um conversor. Uma das desvantagens é que na topologia para emprego do gerador tipo DFIG é necessária uma caixa multiplicadora de velocidades, implicando em manutenção extra e diminuição da confiabilidade, sendo assim um componente crítico e responsável por parcela significativa das falhas do sistema de conversão eólico. A mesma tem custo que pode chegar a até 25% do total do custo de uma turbina, além de apresentar maior susceptibilidade a distúrbios na rede externa ao parque eólico (OLIVEIRA, 2018) (PINTO, 2013).

Um dos benefícios dos geradores síncronos de ímãs permanentes é que, neles, a potência pode ser gerada a qualquer velocidade, pelo fato de ele usar um conversor eletrônico de potência ao ajustar as condições das correntes. Outro benefício importante é o fato de ele ser projetado com pólos salientes e com elevado número de pólos, permitindo, assim, sua operação em baixa velocidade, eliminando, com isso, a necessidade do uso de caixa de engrenagens (ROCHA et al., 2023). Esta redução de partes móveis representa um ganho em confiabilidade e consequentemente em redução de manutenções (OLIVEIRA, 2018).

A construção dos enrolamentos do estator do gerador síncrono e do gerador de indução é idêntica. No entanto, os seus rotores diferem entre si, tanto na forma geométrica quanto no tipo de enrolamento. Um gerador síncrono com ímãs permanentes apresenta uma eficiência maior, pois a excitação do rotor não é fornecida por uma fonte ou por indução. No entanto, há algumas desvantagens. Sua aplicação exige o uso de um conversor eletrônico de potência para ajustar a tensão e a frequência do gerador para a tensão da rede elétrica, exigindo um gasto adicional. As máquinas com ímãs permanentse são mais caras e seus matérias magnéticos são sensíveis à temperatura. Assim, a temperatura tem que ser supervisionada (ROCHA et al., 2023).

Nos últimos anos, avanços tecnológicos na fabricação de ímãs permanentes, fazem com que os aerogeradores direct-drive do tipo PMSG tenham aumento de sua utilização nos processos comerciais de conversão de energia (BARROS; BARROS, 2017) (KUCHENBECKER; TEIXEIRA, 2015) (LAWSON, 2012).

## *2.5 Manutenção*

Para o adequado funcionamento de um sistema de geração de energia, deve-se haver um plano de manutenção e operação adequado, fundamental para as operações dos parques eólicos, com implicações de longo alcance nos custos dos equipamentos, nas receitas do mercado e na logística da equipe de manutenção. Em aplicações típicas, os custos de manutenção constituem de 10-15% e 25-30% dos custos operacionais dos parques eólicos

onshore e offshore, respectivamente (ZHANG; ROBINSON; BASU, 2023), o que certamente encoraja os operadores a empregar a metodologia de O&M mais avançada e adaptada e a focar nos componentes mais críticos (aqueles cujas falhas inesperadas causam paradas não programadas que terão grande impacto na disponibilidade e produtividade) para reduzir a taxa de falhas, o tempo de reparo e maximizar o desempenho do parque eólico (LEITE; ARAÚJO; ROSAS, 2018).

Em gestão da manutenção, o planejamento pode ser realizado de maneiras diferentes, sendo geralmente divididas em não planejada e planejada. Apesar da sua importância, as políticas de O&M de parques eólicos normalmente dependem de decisões *ad hoc*[1] na prática (RUIZ et al., 2020) (LEITE; ARAÚJO; ROSAS, 2018). Atualmente, a forma típica de políticas de manutenção para parques eólicos é uma combinação de políticas estritamente corretivas que iniciam ações de manutenção após a observação de falhas e políticas baseadas em tempo que realizam manutenção em intervalos de tempo fixos (BAKIR; YILDIRIM; URSAVAS, 2021).

Enquanto a manutenção corretiva implica na restauração de um equipamento para as condições normais de funcionamento após a ocorrência de alguma falha, a manutenção emergencial abrange ações tomadas quando surge uma emergência referente à segurança ou desempenho do sistema (VATHOOPAN et al., 2018). Nenhum monitoramento da condição é realizado. Falhas incipientes podem levar a falhas catastróficas nos aerogeradores, mesmo que sejam encontradas em componentes não críticos (LEITE; ARAÚJO; ROSAS, 2018).

Já a manutenção preventiva, está relacionada ao agendamento de atividades voltadas para a prevenção de falhas ou quebras futuras. Desta forma, algumas atividades de manutenção e inspeção são implementadas seguindo uma rotina para favorecer a confiabilidade do equipamento (POHAN; SAPUTRA; TUA, 2023), sem o aporte de indicadores da saúde dos ativos industriais. Dessa forma, à medida que a demanda por confiabilidade aumenta, a frequência das atividades preventivas aumenta drasticamente, tornando o custo dessa política de manutenção elevada (AHMAD; KAMARUDDIN, 2012).

Ao longo dos últimos anos, temos observado avanços sem precedentes em tecnologias de sensoriamento e instrumentação industrial, visando a coleta de indicadores de saúde de parques eólicos, tais como: vibração, temperatura, potência, corrente e etc, que interferem na forma como as organizações entendem as políticas de manutenção, fazendo com que elas tomem decisões baseadas em torno do monitoramento online dos equipamentos (BAKIR; YILDIRIM; URSAVAS, 2021). Assim, a máquina só é parada quando um nível de alarme pré-estabelecido é atingido (LEITE; ARAÚJO; ROSAS, 2018). Consequentemente, essas

[1] termo frequentemente utilizado para descrever algo que é feito de forma improvisada, temporária ou específica para uma determinada situação ou necessidade.

ações tornam possível a redução de tarefas dispendiosas e desnecessárias nos sistemas. Habitualmente, esse conjunto de ações de manutenção é denominada Manutenção Baseada nas Condições (CBM – do inglês – Condition Based Maintenance)(BAKIR; YILDIRIM; URSAVAS, 2021).

Para a aplicação da Manutenção baseada nas condições, é necessário implementar uma estrutura de modelos composta por sistemas, equipamento e pessoas capazes de realizar: coleta e armazenamento de dados; detecção de falhas de equipamentos; diagnóstico e prognóstico de falhas; e, por último, garantir que essas informações sejam adequadamente utilizadas para sustentar tomadas de decisões. Tal estrutura é denominada Prognóstico e Gestão da Saúde de Máquinas (PHM – do inglês Prognostic and Health Management)(PECHT; KANG, 2018). A Figura 2.7 apresenta um modelo para a implementação e execução do PHM.

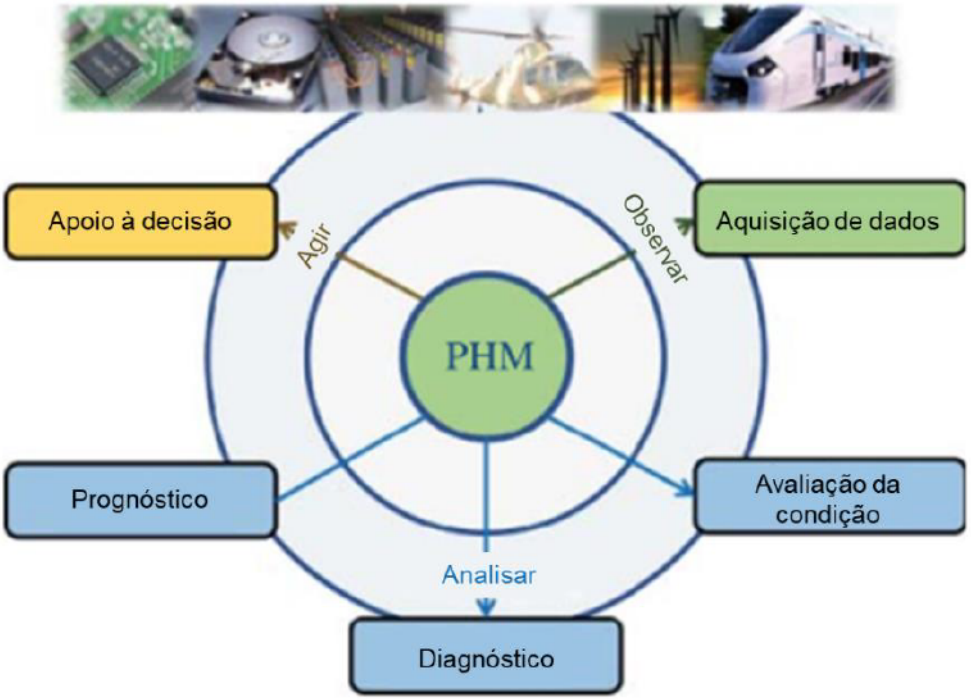


Figura 2.7: Modelo de PHM. Fonte: (PECHT; KANG, 2018).

Para a aplicação dessa estrutura de modelos, são necessárias aquisições de dados de saúde dos equipamentos para pré-processamento e análise qualitativa e quantitativa dos dados de monitoramento visando otimização da qualidade. Modelos baseados em grande quantidade de dados são amplamente utilizados. Nesses casos, os dados são coletados e armazenados, e modelos de estatística ou aprendizagem de máquina reconhecem e reproduzem os padrões entre os fenômenos de falha do equipamento e seus indicadores de saúde (MARTI-PUIG et al., 2018) (SAIDI et al., 2018).

Quando nos referimos a aerogeradores, as falhas são geralmente classificadas em 3 tipos,

as falhas elétricas (problemas de curto-circuito) que podem levar à necessidade de desligamento e reduzem rendimento dos aerogeradores, de controle (falhas no sistema pitch, yaw, hidráulico, sensores e etc.) e mecânicas (geralmente associado à caixa de engrenagens e pás) Entretanto é imprescindível não generalizar e sim analisar cada caso em particular (PINTO, 2013).

O aerogerador PMSG por exemplo tem ocorrência maior de falhas elétricas e eletrônicas, uma vez que não possuem caixa de engrenagens. Por outro lado, segundo (KUCHENBECKER; ROSA; TEIXEIRA, 2018), a detecção de falhas em geradores do tipo PMSG é uma tarefa um tanto complexa de ser realizada, que devido a utilização de ímãs permanentes e sem caixa de velocidades este tipo de gerador continua girando normalmente durante ocorrências de falhas internas da máquina, o que pode mascarar essas falhas, principalmente nos estágios iniciais (SÁ et al., 2019). Sem caixa de velocidades a análise de vibração pode ser menos eficaz para detectar falhas em baixas velocidades, pois podem não sensibilizar os sensores vibracionais (KHAN; BYUN, 2023). Assim, a análise de sinais elétricos e de temperatura do gerador pode ser mais eficiente (YANG; TAVNER; WILKINSON, 2009) para detectar variações no campo magnético, o que pode indicar desalinhamento do eixo ou outros problemas. No geral, uma combinação de diferentes métodos de detecção e análise pode ser necessária para diagnosticar efetivamente falhas em turbinas eólicas baseadas em PMSG (KHAN; BYUN, 2023).

A figura 2.8 mostra uma análise feita por Tavner, Bussel e Spinato (2006), que em seu estudo de pesquisa comparativo sobre falhas em turbinas eólicas de acionamento direto e indireto, concluíram que no conceito acionamento direto, a taxa de falha da eletrica é muito significativa. Uma leitura do que foi apresentado, mostra que, no quadrante que estratificou as falhas eléticas, as falhas das turbinas de acionamento direto foram mais significativas ao tempo que no quadrante trem de potência, as mais representativas foram para os aerogeradores de acionamento indireto.

Os componentes do trem de potência estão sujeitos a cargas altamente dinâmicas devido à interação de forças inerciais, aerodinâmicas, estruturais e vibração mecânica, fazendo com que a fadiga experimentada por seus componentes possa ser ordens de grandeza maior do que a de outros sistemas ou máquinas, o que torna a O&M dos parques eólicos ainda mais distintos (LEITE; ARAÚJO; ROSAS, 2018).

Comparação entre as taxas de falha da caixa de engrenagens no conceito de Velocidade Fixa com a taxa de falha do inversor no Conceito Direct Drive mostra que a taxa de falhas do inversor é consistentemente muito menor do que a da gearbox. A vantagem torna-se avassaladora quando se considera que o inversor é mais simples e rápido de se manter, enquanto reparos ou substituições imprevistas de rolamentos, eixos e rodas dentadas podem ser muito caros, exigindo às vezes a desmontagem de todo o rotor e do

trem de potência do aerogerador (LEITE; ARAÚJO; ROSAS, 2018) (TAVNER; BUSSEL; SPINATO, 2006).

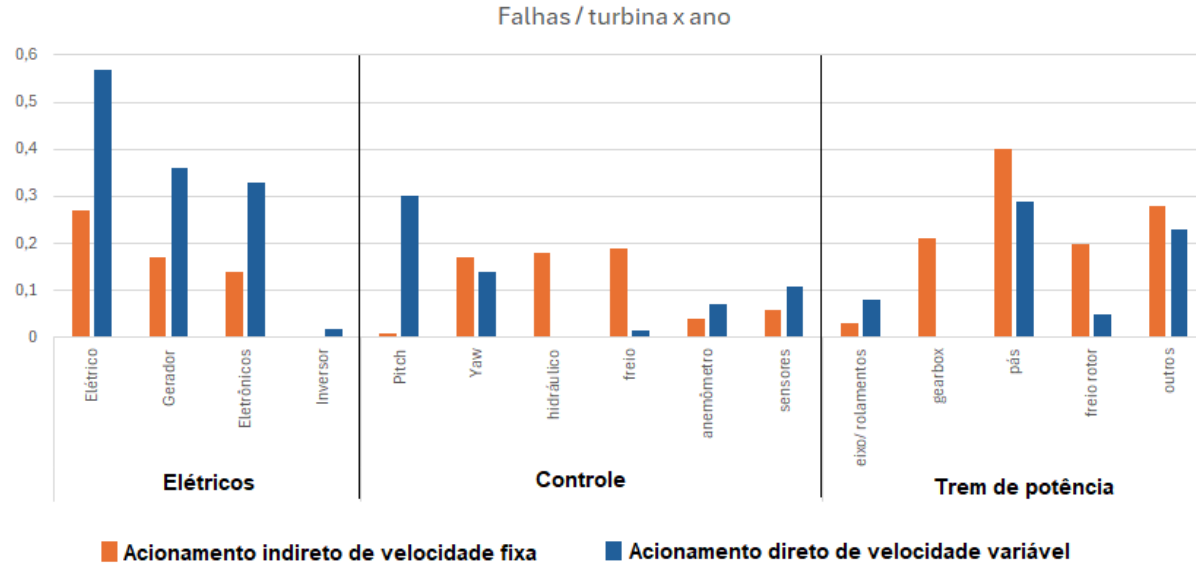


Figura 2.8: Comparação entre taxas de falha anuais entre turbinas de acionamento direto e indireto. Fonte: Adaptado de(TAVNER; BUSSEL; SPINATO, 2006).

Em geral, a maioria da literatura categoriza os algoritmos para detecção de falhas em abordagens baseadas em física e abordagens baseados em dados para utilização do CBM. as abordagens baseadas na física dependem de uma modelagem precisa que são capazes de realizar uma detecção de falhas eficiente se a dinâmica do sistema puder ser bem descrita, uma vez que utiliza resíduos entre as medições do sistema real e a saída esperada do modelo físico, podem ser usados para monitorar o seu estado operacional. Mas têm dificuldades na prática se as aplicações, como as turbinas eólicas industriais, não se adequarem suficientemente bem aos pressupostos (SI et al., 2017).

O outro tipo de abordagem é a orientada a dados, que não requer o conhecimento profundo do sistema que será estudado, no entanto, mas há uma necessidade de utilização dos dados de monitoramento da condição coletados dos sensores instalados com aprendizado de máquina O desempenho dessa abordagem, no entanto, depende muito do número e da qualidade dos dados, pois requer um grande número de tendências ou dados executados até a falha para a construção precisa do modelo (CHATTERJEE; DETHLEFS, 2021) e utilização de sensores apropriados (SI et al., 2017).

A detecção de falhas refere-se ao processo de reconhecer a condição ou declínio de desempenho de um sistema ou estar ciente da condição ou diminuição de desempenho de um sistema (WANG; WANG; JI, 2022). Esta abordagem permite a aplicação operar online

ou em tempo real usando streaming de dados de sistemas de monitoramento de condições, possibilitando o monitoramento contínuo das turbinas. Além do aproveitar volumes crescentes de dados de monitoramento de parques eólicos de mesma tecnologia, quando mais estudos baseados em dados podem ser realizados para otimizar algoritmos e obter insights mais precoces (MUNGUBA et al., 2024)

## 2.6 Aprendizado de máquina

Nos últimos anos, processamento de dados através do uso de Aprendizado de máquina - AM vem sendo explorado de forma intensa, apresentando desempenho significativo para o contexto do monitoramento e manutenção de máquinas frente aos modelos matemáticos tradicionais. As técnicas de AM vinculam seu desempenho ao volume de dados disponíveis para a etapa de aprendizado. a mineração de dados tem papel significativo. Através da mineração de dados, podemos nos aprofundar em grandes quantidade de dados e descobrir padrões que não eram aparentes (GERON, 2019) (GOODFELLOW; BENGIO; COURVILLE, 2016).

Assim, quanto maior a representatividade dos dados disponibilizados melhor será a qualidade do treinamento do modelo quanto ao fenômeno estudado (GERON, 2019)(MENDONçA et al., 2024).

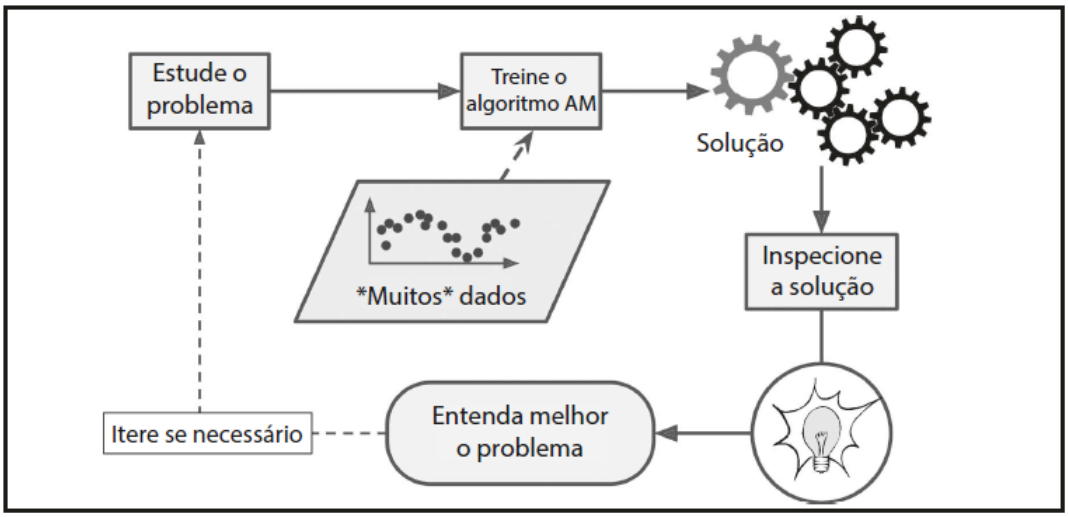


Figura 2.9: Fluxo simplificado do aprendizado de máquina. Fonte: (GERON, 2019).

Um modelo é uma especificação de uma relação matemática (ou probabilística) entre diferentes variáveis (GRUS, 2021). Ele recebe entradas e produz saídas. Para saber se o modelo está funcionando bem, deve-se medir o quão bom seu modelo é, ou seja, as

previsões e os exemplos de treinamento devem corresponder o mais próximo possível. (GERON, 2019)

Sintetizando, para desenvolvimento de um projeto de aprendizagem de máquina deve-se estudar os dados, selecionar um modelo, treinar o modelo com os dados de treinamento (o algoritmo de aprendizado procura os valores dos parâmetros do modelo que minimizam o erro) e por último, validar (aplicar o modelo para fazer previsões em novos casos, na expectativa de que esse modelo generalize bem). Na figura 2.9 vemos um fluxo simplificado do aprendizado de máquina (GERON, 2019).

### *2.6.1 Processos de aprendizagem*

Segundo Goodfellow, Bengio e Courville (2016) os algoritmos de aprendizado de máquina são organizados em taxonomia, com base no resultado desejado do algoritmo para as saídas. Os tipos de algoritmos comuns incluem:

- Aprendizado supervisionado: é a busca por algoritmos que raciocinam a partir de instâncias fornecidas externamente para produzir hipóteses gerais, que então fazem previsões sobre instâncias futuras; (OSISANWO et al., 2017)

- Aprendizado não-supervisionado: Diferentemente, algoritmos não-supervisionados não requerem a presença do rótulo. Eles aprendem forma implícita ou explícita a densidade de probabilidade que gerou tais dados, as regras de associação e os grupos de padrões existentes, o que levam a um processo mais complexo de análise que os supervisionados.

- Aprendizado semi-supervisionado: Quando o algoritmo combina exemplos rotulados e não rotulados. O algoritmo é treinado apenas com dados que apresentam um comportamento específico. E quando da submissão ao conjunto de teste, os pontos que diferenciam da distribuição original de treinamento são identificados (ALLA; ADARI, 2019)

- Aprendizado por reforço: Quando o algoritmo aprende a como agir, dada uma obervação do ambiente;

O tipo de aprendizagem supervisionada depende do tipo da variável dependente. Cada amostra é descrita por um número fixo de atributos associadas a uma classe, que caso

pertença ao conjunto dos números reais, denominada de regressão ou que pode ser discreta, para os casos de variáveis categóricas, sendo nesse caso denominada de classificação (GERON, 2019).

A classificação de padrões supervisionada é bastante comum. O objetivo geralmente é fazer com que o computador aprenda um sistema de classificação previamente modelado. Também pode ser definida como o processo de atribuir um ou alguns dentre os categorias predefinidas para cada item (JO, 2023).

O formato matricial do atributo, usado para representar amostras ou instâncias, é apresentado na tabela 2.1.

| Amostras | Atributos | | | | |
|---|---|---|---|---|---|
| | $X_1$ | $X_2$ | ... | $X_N$ | Classe (Y) |
| $A_1$ | $X_{11}$ | $X_{12}$ | ... | $X_{1M}$ | $y_1$ |
| $A_2$ | $X_{21}$ | $X_{22}$ | ... | $X_{2M}$ | $y_2$ |
| $A_3$ | $X_{31}$ | $X_{32}$ | ... | $X_{3M}$ | $y_3$ |
| ... | ... | ... | ... | ... | ... |
| $A_N$ | $X_{N1}$ | $X_{N2}$ | ... | $X_{NM}$ | $y_N$ |

Tabela 2.1: Formato padrão do conjunto de amostras representando atributos e classes.

### *2.6.2 Seleção de atributos*

Muitas situações em aprendizado de máquina - AM, o banco de dados é muito extenso e envolvem milhares ou até milhões de recursos para cada instância de treinamento (GERON, 2019). Caso a opção seja medir tudo que esteja disponível, na esperança que as características corretas ou mais informativas isoladas, cria-se um processo extenuante, apelidado por Ayodele (2010) como "força-bruta".

Grandes volumes de dados observados tendem a ser acompanhados de ruído (outliers) e valores ausentes, e os mecanismos exatos que os geram são geralmente desconhecidos (MORALES; MÉNDEZ, 2008). Todos esses recursos exigem um grande poder computacional para processamento e treinamento do modelo computacional e sem um pré-processamento significativo não são adequados para indução, pois podem tornar muito mais difícil encontrar um boa solução. O fato de muitos recursos dependerem uns dos outros muitas vezes influencia indevidamente a precisão dos modelos de classificação de AM supervisionados. Este problema é muitas vezes referido como maldição da dimensionalidade (GERON, 2019).

A seleção de recursos faz parte da engenharia de recursos e visa reduzir a dimensionalidade, eliminando recursos de menor importância e melhorando a eficiência computacional das redes neurais de aprendizagem (ZHANG; ROBINSON; BASU, 2023). Consiste em escolher um subconjunto ótimo de A' atributos relevantes, de acordo com um determinado critério, a partir do conjunto original com A atributos, de maneira que A' $\leq$ A e de processamento mais rápido (LEE, 2005).

O método de seleção de atributos para aprendizado supervisionado geralmente consiste em três etapas descritas abaixo (KAREGOWDA; JAYARAM; MANJUNATH, 2011):

(a) Gerar subconjunto candidato: procedimento de busca que produz subconjuntos de recursos candidatos para avaliação com base em um determinada estratégia de pesquisa.

(b) Avaliação do subconjunto candidato.

(c) Critério de parada: Mesmo após a seleção, o número de subconjuntos pode ser elevado, sendo assim algum tipo de critério de parada, baseado em um número pré-definido de recursos, se foi atingido um número predefinido de iterações, por exemplo, ou ainda se a adição (ou exclusão) de qualquer recurso não produzir um subconjunto melhor.

#### *2.6.2.1 Seleção de recursos baseados em correlação - CFS subset evaluator*

O algoritmo CFS subset evaluator é um algoritmo utilizado na abordagem filtro para a seleção de atributos, que consiste em selecionar recursos para aprendizado de máquina por meio de uma abordagem baseada em relação entre atributos e correlação entre atributos e classes. Avaliando o valor de um subconjunto de atributos considerando a capacidade preditiva individual de cada atributo junto com o grau de redundância entre eles, independente da quantidade de dados disponíveis (PALMA-MENDOZA et al., 2019)

> Um bom subconjunto de características é aquele que contém características altamente correlacionadas com (preditivos da) classe, mas não correlacionados (não preditivos) entre si. (HALL, 1999)

O seletor de recursos CFS por ser um filtro, é simples e não incorre em alto custo computacional associado ao invocar repetidamanete um algoritmo de aprendizagem (VELMURUGAN; ANURADHA, 2016). Este seletor classifica subconjuntos de atributos de acordo com uma função de avaliação heurística baseada em correlação (SAARI; EEROLA; LARTILLOT, 2011). Atributos irrelevantes devem ser ignorados, uma vez que terão baixa correlação com a classe, assim como os redundantes devem ser excluídos, pois estarão altamente correlacionados com um ou mais dos atributos restantes. A aceitação de um

atributo dependerá da extensão em que ele prevê classes em áreas do espaço de instâncias ainda não previstas por outros (VELMURUGAN; ANURADHA, 2016).

Considerando um conjunto de dados extenso e completo, existirá multicolinearidade entre seus atributos, assim, quando um determinado atributo puder ser expresso como uma combinação linear de outros, o cálculo do diagnóstico de falhas aumentará dramaticamente, mas a precisão não será significativamente melhorada se os atributos mais relevantes não forem selecionados. De outra perspectiva, há também uma correlação entre atributoss e falhas (HAN; YANG, 2023).

As métricas de seleção podem ser avaliadas pela fórmula 2.1. O numerador pode ser interpretado como um indicador de quão preditivo o conjunto de recursos é, por outro lado, o denominador indica quão redundantes os recursos são (PALMA-MENDOZA et al., 2019).

$$Merit_m = \frac{m \cdot \bar{r_{ca}}}{\sqrt{m + m(m-1) \cdot \bar{r_{a_i a_j}}}} \tag{2.1}$$

Onde:

- $Merit_m$ é o resultado da avaliação do subconjunto $m$ de atributos;
- $m$ é o número total de atributos no subconjunto;
- $\bar{r}_{ca}$ é a média das correlações entre cada atributo $a$ e a classe $c$, representada pela equação 2.2;
- $\bar{r}_{a_i a_j}$ é a média das corelações entre as combinações de subconjuntos de atributos, representada pela equação 2.3.

$$\bar{r}_{ca} = \left( \frac{r_{ca_1} + r_{ca_2} + ... + r_{ca_m}}{m} \right) \tag{2.2}$$

$$\bar{r}_{a_i a_j} = \left( \frac{r_{a_1 a_2} + r_{a_1 a_3} + ... + r_{a_1 a_m}}{\frac{m(m-1)}{2}} \right) \tag{2.3}$$

Este algoritmo permite realizar a busca das seguintes maneiras: forward selection, backward selection e best first. Na busca busca Best-first (melhor primeiro), o algoritmo começa com um conjunto vazio de recursos e em cada etapa da pesquisa todas as expansões possíveis de recursos únicos são geradas. Os novos subconjuntos são avaliados usando a equação 2.1

e então adicionados a uma fila de prioridade de acordo com a classificação do mérito. Na iteração subsequente, o melhor subconjunto da fila é selecionado para expansão da mesma forma que foi feito para o primeiro subconjunto vazio. Se a expansão do melhor subconjunto não produzir uma melhoria no mérito geral, isso conta como uma falha e o próximo melhor subconjunto da fila é selecionado até atingir o limote de cinco falhas consecutivas como critério de parada e como limite no comprimento da fila (PALMA-MENDOZA et al., 2019).

A figura 2.10 mostra uma visualização das etapas do CFS e a tabela 2.2 mostra um pseudocódigo de implementação do algoritmo CFS utilizando a busca best-first.

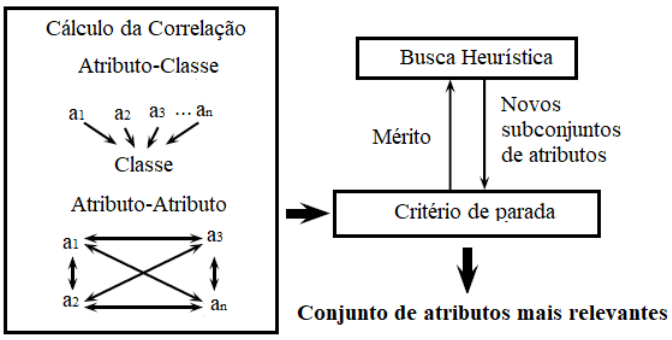


Figura 2.10: Etapas do método CFS. Fonte: Adaptado de (CHAMBY-DIAZ; RECAMONDE-MENDOZA; BAZZAN, 2019).

| **Algoritmo** CFS. |
|---|
| 1. Para cada par de atributos do banco de dados (i, j) em X: |
| Calcule a correlação entre os atributos $i$ e $j$; |
| 2. Retorne a matriz de redundância; |
| 3. Calcule a correlação entre cada atributo e a variável de classe Y; |
| 4. Execute a busca best-first, iniciando com um conjunto vazio; |
| 5. Para cada atributo em X: Calcule o mérito usando a fórmula do CFS 2.1; |
| 6. Adicione à fila de prioridade; |
| 7. Ordene os atributos em ordem decrescente de mérito; |
| 8. Se a busca não produzir melhoria após 5 interações consecutivas, pare; |
| 9. Retorne os atributos selecionados com base no mérito calculado. (Melhor subconjunto) |

Tabela 2.2: Pseudocódigo do Algoritmo CFS.

Hosseinpour-Zarnaq, Omid e Biabani-Aghdam (2022) usaram o método CFS para encontrar os melhores recursos e Random Forests e perceptron multicamadas (MLP) foram empregadas para classificar os dados de uma gearbox. A precisão geral do classificador RF sem seleção de recursos foram 86,25% e a precisão com os 6 atributos ideais dos

56 disponíveis (90% de redução na dimensão de atributos, reduzindo a complexidade do modelo) usando CFS foi de 92,5%.

#### *2.6.2.2 Random Forests*

Random Forests (RF) ou Florestas aleatórias é um método de aprendizado de máquina supervisionado muito poderoso, usado frequentemente em análise de dados para resolver vários problemas em todos os setores (DANGETI, 2017). RF consiste num conjunto de árvores de decisão, algoritmos versáteis de aprendizado de máquina que podem executar tanto tarefas de classificação quanto de regressão e até tarefas de múltiplas saídas, capazes de ajustar conjuntos de dados complexos e por sua capacidade de lidar com overfitting devido à diversidade das árvores no ensemble (GERON, 2019).

O processo é iniciado particionando conjunto de treinamento utilizando uma amostragem aleatória com reposição, o que ajuda a aumentar a diversidade entre as árvores e a reduzir a correlação entre elas. Posteriormente, cada novo conjunto menor que o inicial será utilizado para construir uma nova árvore de decisão. As saídas das várias árvores de decisão, são combinadas para alcançar um único resultado, daí provém o nome floresta.(GOODFELLOW; BENGIO; COURVILLE, 2016). A distinção entre florestas aleatórias para classificação e para regressão é que quando usada para classificação, uma floresta aleatória obtém um voto de classe de cada árvore e, em seguida, classifica usando o voto da maioria. Já, quando usadas para regressão, o as previsões de cada árvore em um ponto alvo $x$ são simplesmente calculadas sua média. Sua efetividade e não ocorrência de overfitting está relacionada com a Lei dos Grandes Números (LGN), em que a frequência de um evento se estabiliza se o experimento for repetido muitas vezes (RASCHKA; MIRJALILI, 2019).

O Bagging [2] participa do processo reduzindo a variância e influência de sobreajustes das árvores de decisão. Seja um conjunto de treinamento, dado por x = $\{x_1, x_2, ..., x_n\}$, e com resposta y = $\{y_1, y_2, ..., y_n\}$. O bagging irá repetir B vezes para selecionar uma amostra aleatória com substituição do conjunto de treinamento e ajusta as árvores a essas amostras. Uma árvore $T_b$, (b=1,2,...B) será treinada por vez. Após o treinamento, o modelo de previsão final será determinado obtendo o voto majoritário das árvores de decisão B. (SI et al., 2017)

A tabela 2.3 mostra um pseudocódigo de implementação do algoritmo Random Forest para classificação.

[2] abreviação para agregação de bootstrap, que é um algoritmo de conjunto projetado para melhorar a estabilidade e a precisão de modelos preditivos individuais, como árvores (RASCHKA; MIRJALILI, 2019)

Este algoritmo é muito usado na aplicação da manutenção preditiva como em alguns estudos para melhorar a proteção e manutenção de equipamentos (INGOLE et al., 2022) (VINH; HUY, 2022). também usado para estimar a vida útil remanescente (RUL) de máquinas industriais, permitindo a previsão de manutenção e o planejamento apropriado (BAKDI; KRISTENSEN; STAKKELAND, 2022).

Random Forests alcança resultados promissores em termos de precisão. São capazes de lidar tanto com problemas de regressão como classificação. Podem ser utilizadas com problemas de alta dimensionalidade e são rápidas para treinamento e implementação(CUTLER; CUTLER; STEVENS, 2011). Este método fornece pontuações de importância para cada atributo, que podem ser usadas para selecionar as características mais relevantes (GERON, 2019), como a diminuição calculada da impureza média a partir de todas as árvores de decisão na floresta, sem fazer qualquer suposições sobre se nossos dados são linearmente separáveis ou não (RASCHKA; MIRJALILI, 2019).

Em seu estudo, (ZHANG; ROBINSON; BASU, 2023), comparou vários classificadores para classificação de dados em turbina eólica. Random Forests alcançou a melhor acurácia. O classificador conduziu a seleção de atributos usando uma validação cruzada de 10 partições para o conjunto de dados. Com base no processo de seleção de atributos, na Figura 2.11, a melhor precisão (0,9888) foi observada selecionando apenas 33 recursos mais relevantes do total de 60 disponíveis para a tarefa de diagnóstico de falhas na classificação. Com qualquer quantidade ou seleção de recursos diferente da escolhida, a precisão diminui.

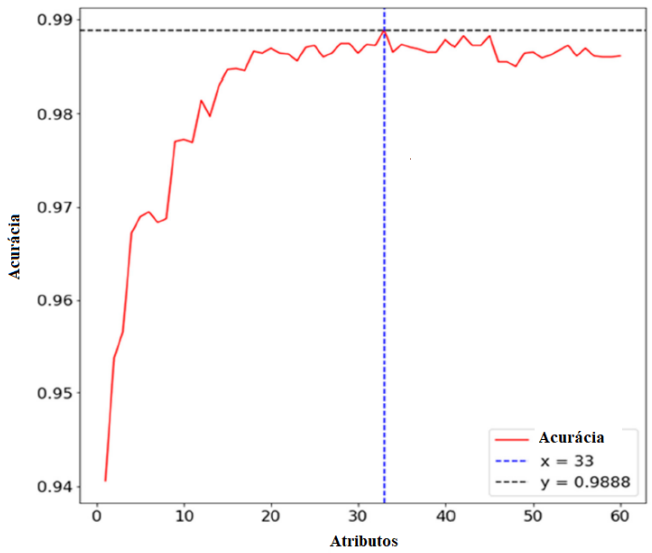


Figura 2.11: Pontuações da validação cruzada plotadas em relação ao número de recursos selecionados pelo Random Forests. Fonte: (ZHANG; ROBINSON; BASU, 2023).

| **Algoritmo** Random Forests. |
|---|
| 1. Para b=1 até B: (onde B é o número de árvores escolhidas para a implementação) |
| (a) Tome uma amostra **A** de tamanho N dos dados de treinamento. (A amostra pode ser o próprio conjun |
| (b) Construa uma árvore de floresta aleatória $T_b$ para os dados inicializados, repetindo recursivamente as seguintes etapas para cada nó terminal de árvore, até que o tamanho mínimo do nó $n_{min} = 1$ seja atingido. (A idéia é realizar uma escalada de melhoria até que o nó seja puro, nem que para isso o nó terminal só contenha 1 elemento). |
| i. Selecione aleatoriamente a quantidade $m$ de atributos dos $p$ atributos possíveis: Pode-se considerar o valor para $m$ sendo o primeiro inteiro menor que $\log_2 p + 1$, $\sqrt{p}$ ou variações. |
| ii. Escolha se deseja amostrar tanto instâncias quanto atributos ou utilizar todas as instâncias e apenas amostrar os atributos; |
| iii. Escolha o melhor atributo/ponto de divisão entre os atributos de m. (São comumente utilizados Coeficiente de Gini ou Entropia) |
| iv. Divida o nó em dois nós filhos. |
| 2. Permita que as amostragens sejam realizadas com reposição (São geradas amostras diferentes mesmo se a amostra for o conjunto de dados completo). |
| 3. Produza o conjunto de árvores $\{T_b\}_1^B$, com amostras independentes. |
| 4.Para fazer uma previsão em uma nova instância de $x$: Seja $C_b(x)$ a previsão de classe da $b$-ésima árvore da floresta aleatória. Então $C_{rf}^B(x) =$ voto majoritário $\{C_b(x)\}_1^B$. |
| 5. Cálculo do erro OOB (out of bag) Média do erro de classificação para todas as instâncias que estavam fora do saco (bag) em cada árvore. |
| 6. Cálculo da importância das variáveis: |
| (a) Para cada atributo $a$: |
| i. Para cada árvore $t$, pegue as amostras de instâncias out of bag desta árvore. |
| ii. Permute os valores de $a$ nas amostras de instâncias *out of bag* de $t$. |
| iii. Classifique as instâncias usando a árvore $t$. |
| iv. Com base nessas novas classificações, calcule $OOB_p$ |
| v. Importância de $a = (OOB_p - OOB)/OOB$ |
| 7. Classifique as importâncias dos atributos. |

Tabela 2.3: Pseudocódigo do Algoritmo Random Forests para seleção de atributos mais relevantes para classificação de padrões.

### *2.6.3 Algoritmos classificadores ou de indução*

Algoritmos classificadores são aqueles que possuem como objetivo prever a classe de um novo dado baseado no aprendizado supervisionado sobre dados semelhantes em observações passadas (RASCHKA; MIRJALILI, 2017). Numa classificação binária, as categorias

pré-definidas são a classe positiva e a classe negativa. Amostras de cada um dos dados rotulados com uma das categorias são coletados como exemplos de treinamento, e a aprendizagem é o processo de minimização da função do erro de classificação nestes exemplos (HAYKIN, 2009).

Um dos problemas mais antigos da ciência experimental é encontrar funções que se ajustem ou expliquem os dados observados dos fenômenos naturais (MORALES; MÉNDEZ, 2008). Dado um conjunto de treinamento de N exemplos de pares de entrada-saída, que não precisam ser números, $(x_1, y_1), (x_2, y_2), ...(x_N, y_N)$, onde cada $y_i$ foi gerado por uma função desconhecida $y = f(x)$, a tarefa da aprendizagem supervisionada é descubrir uma função **hipótese** $h$ que se aproxima da verdadeira função $f$ mesmo em novos exemplos além do conjunto de treinamento, conforme equação 2.4 (GOODFELLOW; BENGIO; COURVILLE, 2016) (HAYKIN, 2009).

$$\| h(x) - f(x) \| < \varepsilon \tag{2.4}$$

para todo **x**. Onde $\varepsilon$ é suficientemente pequeno.

Dizemos uma hipótese generalizou bem, quando ela tem capacidade de predizer a correta saída quando um conjunto de teste com dados distintos do conjunto de treinamento é utilizado Em geral, há uma compensação entre hipóteses complexas que se ajustam aos dados de treinamento bem e hipóteses mais simples que podem generalizar melhor (GERON, 2019).

### *2.6.4 Redes neurais artificiais - RNA*

As redes neurais se inspiram no cérebro, utilizando neurônios artificiais interconectados que funcionam em uníssono através de um processo de aprendizagem para realizar cálculos como reconhecimento de padrões e percepção por meio de um processo de aprendizagem (MORALES; MÉNDEZ, 2008)(RASCHKA; MIRJALILI, 2019).

A figura 2.12 mostra o grafo estrutural de um perceptron de múltiplas camadas-MLP, com uma camada de entrada, isto é, cada uma das covariáveis do banco de dados. Duas camadas ocultas da rede e uma camada de saída. Cada flecha representa um peso e cada nó, numa determinada camada, representa uma transformação na camada anterior (IZBICKI; SANTOS, 2020).

Cada neurônio oculto é um detector de recursos e junto com os neurônios de saída re-

presentam uma transformação não linear do sinal recebido nos neutônios predecessores e pesos sinápticos, denominado cálculo funcional, da esquerda para a direita. Existe também em sentido contráio, o cálculo do sinal de erro, que envolve uma computação de comparação entre o sinal obtido e o desejado, necessário para a retropropagação. Na figura 2.12, por questões de uma melhor visualização, estão representados apenas os fluxos de sinais funcionais, que seguem da esquerda para a direita, de camada em camada (RASCHKA; MIRJALILI, 2017).

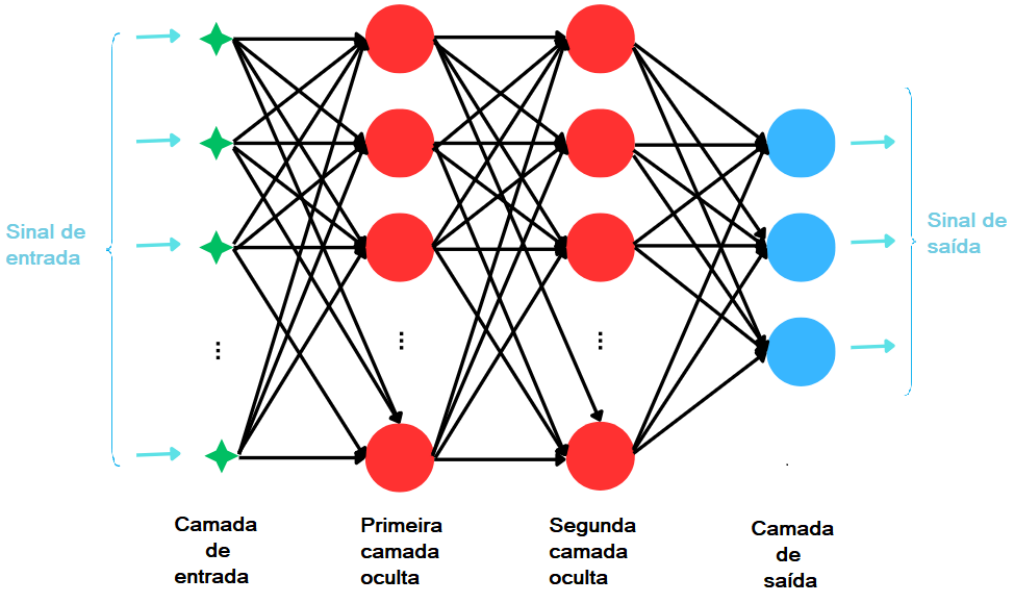


Figura 2.12: Grafo de estrutura de MLP com duas camadas ocultas.

### *2.6.4.1 Algoritmo de retropropagação*

Na figura 2.13, vemos o grafo de fluxo de sinal de uma rede neural MLP com camadas ocultas. O Multilayer perceptron pode ser compreendido por um nó aditivo que calcula uma combinação linear das entradas, que segue para uma função de ativação $\phi$, que realiza a função sinal. O neurônio $j$ representa uma camada oculta enquanto o neurônio $k$ representa a saída. No grafo, para efeito de representação, são dois perceptrons de camada única conectados com a saída do neurônio $j$ servindo de entrada para o neurônio $k$ (HAYKIN, 2009)

Chamamos de algoritmo de aprendizagem, o ajuste de pesos sinápticos, representados por $w_{ij}$, para de classificar corretamente os estímulos aplicados na entrada ou no neurônio anterior, $x_i$ ou $y_i$ da rede de forma ordenada para atingir um objetivo performance desejado ou classificar classes corretamente (RASCHKA; MIRJALILI, 2019).

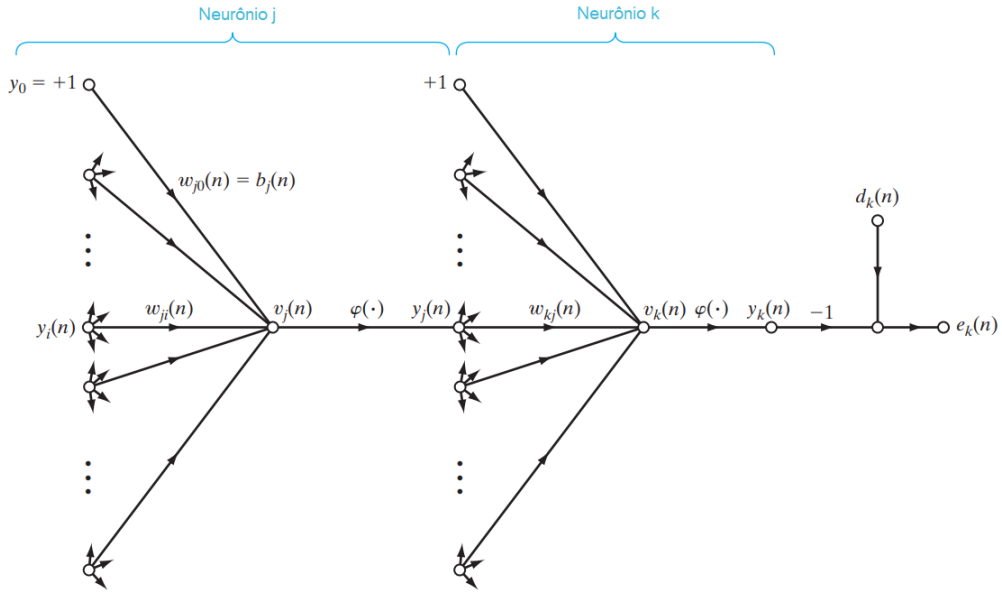


Figura 2.13: Grafo de fluxo de sinal de uma MPL com duas camadas ocultas. Fonte: (HAYKIN, 2009).

As notações a seguir foram utilizadas para representar as informações repassadas e obtidas pelo algoritmo da rede neural (HAYKIN, 2009):

- Índices $i, j$ e $k$ são neurônios diferentes da rede. Sinais funcionais se propagam da esquerda para a direita na sequência, $i$, $j$ e $k$;
- $x_i$ e $y_i$ representam o i-ésimo elemento dos vetor de entrada $x$ e saída $y$, respectivamente;
- $y_i(n)$ e $y_j(n)$ referem-se aos sinais funcionais nas saídas dos neurônios $i$ e $j$, na iteração $n$;
- Época $(n)$ representa a quantidade de iterações a que o algoritmo foi submetido;
- Número de nós $(m)$ representa a quantidade de nós de uma determinada camada;
- Símbolo $d_k(n)$ se refere à resposta desejada para o neurônio $k$;
- Símbolo $w_{ij}$ representa o peso sináptico conectando a saída do neurônio $i$ à entrada do neurônio $j$, na iteração $n$;
- Símblo $\Delta w_{ij}$ representa a correção do peso sináptico conectando a saída do neurônio $i$ à entrada do neurônio $j$, na iteração $n$. Assim,

$$\Delta w_{kj} = w_{kj}(n+1) - w_{kj}(n) \tag{2.5}$$

- Os símbolos $v_j(n)$ e $v_k(n)$ representam a soma ponderada de todas as entradas, acrescidas do viés nos neurônios $j$ e $k$ respectivamente, na iteração n. É o sinal funcional aplicado à função de ativação associada ao respectivo neurônio. neurônio, representado por:

$$v_j(n) = \sum_{j=0}^{m} w_{kj}(n) y_j(n) \tag{2.6}$$

  Onde $m$ é o número total de entradas no neuônio;

- Função de ativação $\phi_j$ ou $\phi_k$ descrevem a relação funcional de entrada e saída associadas aos neurônios $j$ e $k$, respectivamente, em função do campo local induzido $v_j(n)$ e $v_k(n)$.

- Símbolo $e_k(n)$ é o sinal de erro na saída do neurônio $k$, para a iteração n. Pode ser expresso através da diferença entre a saída desejada $d_k(n)$ e o resultado realmente obtido $y_k(n)$, conforme:

$$e_k(n) = d_k(n) - y_k(n) = d_k(n) - v_k(n) \tag{2.7}$$

  e o valor instantâneo da energia do erro para um determinado neurônio:

$$\xi = \frac{1}{2} {e_k}^2(n) \tag{2.8}$$

- Símbolo $\xi(n)$ representa a soma dos erros quadráticos na época n.

$$\xi(n) = \frac{1}{2} \sum_{j \in C}^{m} {e_k}^2(n) \tag{2.9}$$

  Onde C, inclui todos os neurônios da camada de saída da rede.

Na saída de cada neurônio, o sinal gerado passa por uma função de ativação, responsável por ponderar o efeito de cada saída na camada subsequente. O aprendizado do Perceptron é atualização gradativa de seus vetores de peso, minimizando os erros de classificação nas amostras de treinamento (JO, 2023).

Existem diversos tipos de funções de ativação, tais quais:

- Função de limiar: Também conhecida como função de *Heaviside*, Apresenta um comportamento binário, resultando em uma saída em 1 ou 0, equação 2.10, dependendo do valor obtido pelo neurônio. Exemplo na fig 2.14.

$$\begin{cases} 1, & \text{se } v_k \geq 0 \\ 0, & \text{se } v_k < 0 \end{cases} \tag{2.10}$$

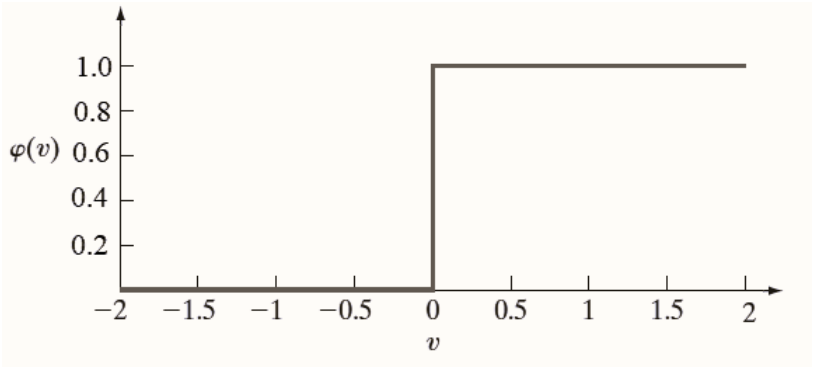


Figura 2.14: Gráfico exemplo de função de limiar. Fonte: Adaptado de (HAYKIN, 2009).

- Função linear por partes : Apresenta um comportamento que se aproxima da função de limiar, se o fator de amplificação da região linear é relativamente grande. Enquanto, na região intermediária, funciona como um combinador linear.

- Função sigmóide : Apresenta um comportamento abrupto da funçãoo limiar, matematicamente representado através da equação a seguir. Dessa forma, a função sigmóide fornece níveis intermediários entre os valores 0 e 1, podendo ter o aclive ajustado conforme a necessidade. A função sigmóide, cujo gráfico tem formato de "S", é definida como uma função estritamente crescente que exibe um equilíbrio entre comportamento linear e não linear. Um exemplo de função sigmóide é a logística com exemplo na fig 2.15 e com função definida por:

$$\varphi(v) = \frac{1}{1 + exp(-av)} \tag{2.11}$$

onde "$a$ "é o parâmetro de inclinação da função sigmóide. Variando-se o parâmetro a, temos obtemos funções sigmóides de diferentes inclinações, conforme ilustrado na Fig. Na verdade, a inclinação na origem é igual a "$a/4$". No limite, à medida que o parâmetro de inclinação se aproxima do infinito, A função sigmóide torna-se simplesmente uma função de limiar. Enquanto uma função de limite assume o valor de 0 ou 1, uma função sigmóide assume um intervalo contínuo de valores de 0 a 1. A função sigmóide é diferenciável, enquanto a função de limiar não é. (HAYKIN, 2009)

- Função tangente hiperbólica:

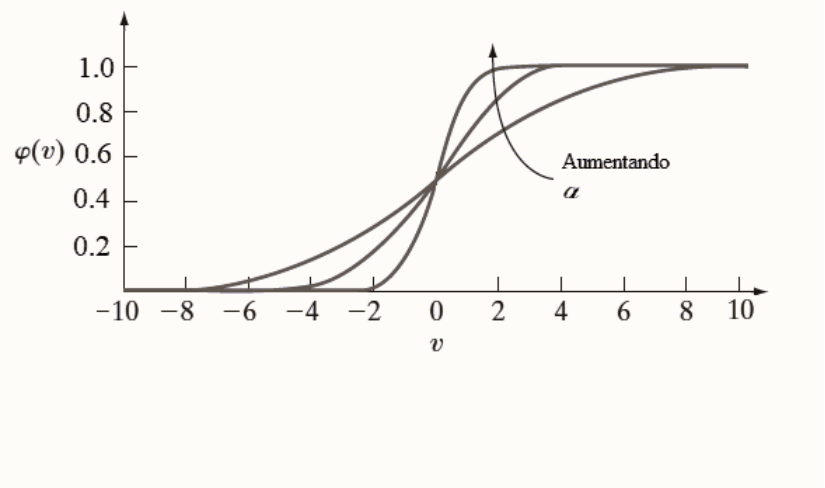


Figura 2.15: Gráfico exemplo de função Sigmóide. Fonte: Adaptado de (HAYKIN, 2009).

A função tangente hiperbólica tem um comportamento similar à sigmóide mas pode assumir valores negativos, definida pela equação 2.12:

$$\varphi(v) = tanh(av) \tag{2.12}$$

Uma MLP com cada saída correspondente a uma classe binária é frequentemente utilizado para classificação. Quando as classes são mutuamente exclusivas, a função de ativação softmax geralmente é uma boa opção para as tarefas de classificação na camada de saída. Quando não são (ou quando existem apenas duas classes), utilize a função sigmóid (GERON, 2019).

Para um banco de dados de treinamento, a soma das contribuições de todos os neurônios na camada de saída é a medida de desempenho da aprendizagem, que é função de todos os hiperparâmetros. É obtida somando-se os $\xi(n)$ para todos as épocas $n$ e normalizando para o conjunto de N amostras, o erro quadrático médio, é:

$$\xi_{medio}(N) = \frac{1}{N}\sum_{n=1}^{N}\xi(n) = \frac{1}{2N}\sum_{n=1}^{N}\sum_{k\in C}{e_k}^2(n) \tag{2.13}$$

Deste modo, o objetivo é o ajuste dos hiperparâmetros de acordo com os respectivos erros calculados apresentado à rede até formar uma época de modo a minimizar o erro médio $\xi_{medio}$ sobre o conjunto de treinamento inteiro ou estejam o mais próximo possível. A soma dos erros quadrados na amostra de treinamentos pode ser definida como função dos parâmetros livres do sistema. A função que melhor ajusta o erro pode ser visualizada

como uma superfície multidimensional de desempenho do erro, assim qualquer operação neste sistema supervisionado, funciona como um ponto sobre tal superfície. (HAYKIN, 2009)(MORALES; MÉNDEZ, 2008)

O vetor $w$ que minimiza o $\xi_{medio}(N)$ não possui uma solução analítica. No contexto de redes neurais, o método da retropropagação aplica uma correção $\Delta w_{ij}(n)$ ao peso sináptico $w_{ij}(n)$ determinando a direção de busca no espaço de pesos), que é proporcional à derivada parcial $\delta\xi(n)/\delta w_{ij}(n)$. Este método consiste em calcular o gradiente descendente e propagando esse erro para as camadas anteriores, iniciando da última camada até a primeira (IZBICKI; SANTOS, 2020).

O gradiente numa superficie de erro é o vetor que aponta na direção da descida mais íngreme. Pela regra da cadeia, para derivada de funções de multiplas variáveis, podemos expressar este gradiente como:

$$\frac{\delta\xi(n)}{\delta w_{ji}(n)} = \frac{\delta\xi(n)}{\delta e_j(n)}\frac{\delta e_j(n)}{\delta y_j(n)}\frac{\delta y_j(n)}{\delta v_j(n)}\frac{\delta v_j(n)}{\delta w_{ij}(n)} \tag{2.14}$$

Onde a correção $\Delta w_{ji}$ pode ser definida por:

$$\Delta w_{ij}(n) = -\eta\frac{\delta\xi(n)}{\delta w_{ji}(n)} \tag{2.15}$$

onde $\eta$ é o número adimensional denominado parâmetro da taxa de aprendizagem do algoritmo de retropropagação, e o sinal negativo indica a descida do gradiente no espaço de pesos, quando avançamos passo a passo no processo de aprendizagem, sinalizando a busca sucessiva de uma direção que reduza o valor de $\xi(n)$. (HAYKIN, 2009)

### *2.6.4.2 Fronteira de decisão*

Para o caso de uma MLP de apenas uma camada oculta, o campo local induzido é dado por:

$$v = \sum_{i=1}^{m} w_i x_i + b \tag{2.16}$$

Da equação com $m$ variáveis de entrada $x_1, x_2, x_3, ..., x_m$ num espaço $m$-dimensional, pode-

se traçar um mapa de $m$ regiões. Onde, $x_i$ é o vetor de entrada que representa as amostras e $w_i$ é o vetor perpendicular a um dado hiperplano, e representa o peso, que é otimizado através do aprendizado (HAYKIN, 2009). Portanto, o hiperplano (é usado como limite de classificação) no espaço em duas dimensões é visto como uma linha, já no espaço tridimensional é visto como um plano. O hiperplano no espaço $m$-dimensional é dado como um espaço de dimensão $m$-1, e os valores das variáveis no hiperplano indicam se a equação que descreve o hiperplano é satisfeita ou não (JO, 2023).

Assim, para o caso de duas variáveis, $x_1, x_2$, a separabilidade linear que caracteriza a classificação pode ser visualizada no espaço bidimensional conforme figura 2.16 (JO, 2023). Caso um ponto$(x_1, x_2)$ se encontre acima da linha de fronteira, é atribuído à classe $C_1$, caso contrário à classe $C_2$. O efeito do viés $b$ é deslocar a fronteira de decisão com relação à origem. (HAYKIN, 2009)

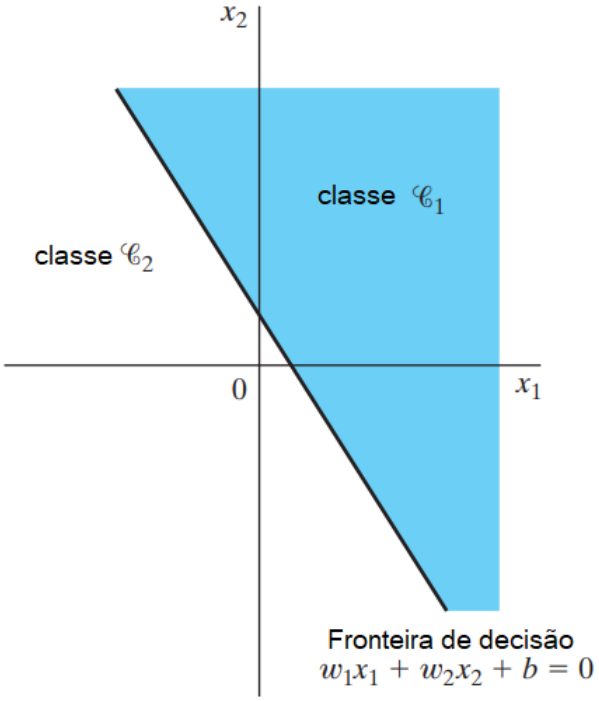


Figura 2.16: Hiperplano como fronteira de decisão para um problema de classificação de padrões entre duas classes. Fonte: (HAYKIN, 2009).

Mesmo que as amostras de treinamento sejam classificadas perfeitamente, não há garantia de que amostras de teste sejam classificados corretamente. Em algumas situações não é possível traçar um hiperplano, que defina uma fronteira entre duas classes, assim, ao aplicá-lo para a classificação de dados, o hiperplano deve ser otimizado para minimizar a classificação incorreta no treinamento exemplos, em vez de eliminá-los completamente (JO, 2023).

### *2.6.4.3 Teorema da convergência do Perceptron*

A partir das definições e equações mostradas anteriormente. Temos o vetor de entrada $\mathbf{x}$(n) e o vetor de pesos $\mathbf{w}$(n) de $m+1$ linhas e 1 coluna, com $n$ representando o passo de iteração ao longo das épocas do algoritmo respectivamente como:

$$\mathbf{x}(n) = [x_1(n), x_2(n), ..., x_m(n)]^T \tag{2.17}$$

e

$$\mathbf{w}(n) = [w_1(n), w_2(n), ..., w_m(n)]^T \tag{2.18}$$

Assim, reescrevendo a saída do combinador linear:

$$v(n) = \sum_{i=0}^{m} w_i(n)x_i(n) \tag{2.19}$$

$$= \mathbf{w}^T(n)x(n) \tag{2.20}$$

Para $n$ fixo, a equação $w^T x = 0$, traçada num espaço $m$-dimensional traçada para um viés pré-determinado com coordenadas $x_1, x_2, ..., x_m$, define um hiperplano como a superfície entre duas classes diferentes de entradas, quando as classes são linearmente separáveis. Isto quer dizer que existe um vetor de peso w tal que:

$$\begin{cases} \mathbf{w}^T\mathbf{x} > 0, & \text{para todo vetor de entrada } x \text{ pertecente à classe } C_1 \\ \mathbf{w}^T\mathbf{x} \leq 0, & \text{para todo vetor de entrada } x \text{ pertecente à classe } C_2 \end{cases} \tag{2.21}$$

Assim, se o n-ésimo conjunto de treinamento é corretamente classificado pelo vetor de peso $\mathbf{w(n)}$ na n-ésima iteração do algoritmo, então este vetor de pesos do perceptron não é corrigido pela regra de transição:

- $\mathbf{w}(n+1) = \mathbf{w}(n),$ se $\mathbf{w}^T(n)\mathbf{x}(n) > 0$ e $\mathbf{x}$(n) pertence à classe $C_1$
- $\mathbf{w}(n+1) = \mathbf{w}(n),$ se $\mathbf{w}^T(n)\mathbf{x}(n) \leq 0$ e $\mathbf{x}$(n) pertence à classe $C_2$

Caso contrário, se o n-ésimo conjunto de treinamento é corretamente classificado pelo vetor de peso $\mathbf{w(n)}$ na n-ésima iteração do algoritmo, de acordo com a regra:

- $\mathbf{w}(n+1) = \mathbf{w}(n) - \eta(n)\mathbf{x}(n)$, se $\mathbf{w}^T(n)\mathbf{x}(n) > 0$ e $\mathbf{x}$(n) pertence à classe $C_2$
- $\mathbf{w}(n+1) = \mathbf{w}(n) + \eta(n)\mathbf{x}(n)$, se $\mathbf{w}^T(n)\mathbf{x}(n) \leq 0$ e $\mathbf{x}$(n) pertence à classe $C_1$

Onde, $\eta$(n) é um parâmetro que controla o ajuste aplicado ao vetor de peso na iteração n, denominado de taxa de aprendizagem.

### *2.6.4.4 Dilema Taxa de aprendizagem muito lenta versus muito alta*

O comportamento da curva de convergência depende da taxa de aprendizagem, que, se pequena, mais suave será a trajetória no espaço de pesos. Por outro lado, se a taxa de aprendizagem é muito grande para acelerar a aprendizagem, a rede pode se tornar instável (isto é, oscilatória)(GERON, 2019), conforme figura 2.17.

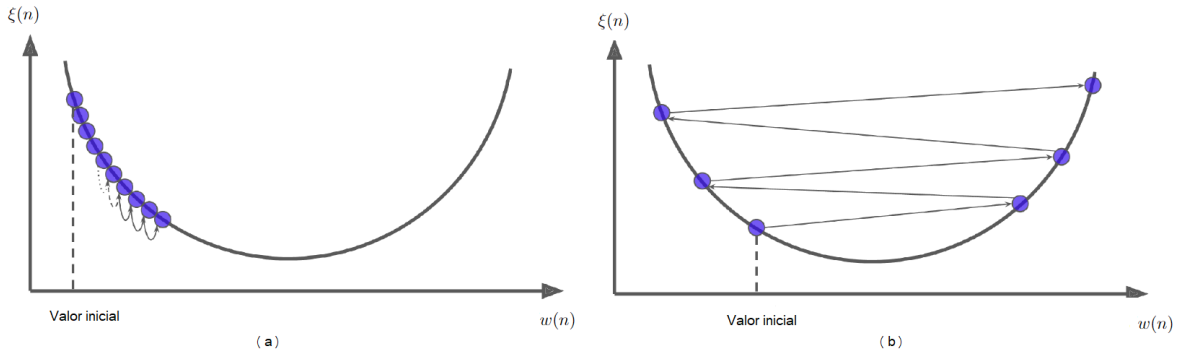


Figura 2.17: Gráfico mostrando busca de convergência com Taxa de aprendizado (a) muito lenta e (b) muito alta. Fonte: (GERON, 2019).

> A incorporação do momento no algoritmo de retropropagação representa uma pequena modificação na atualização de peso; no entanto, pode ter alguns efeitos benéficos no comportamento de aprendizagem do algoritmo. O termo momentum também pode ter o benefício de evitar que o processo de aprendizagem termine em um mínimo local na superfície de erro. (HAYKIN, 2009)

Segundo Haykin (2009) um método de aumentar a taxa de aprendizagem e, ao mesmo tempo, evitar o perigo de instabilidade é incluir um termo momentum, $\alpha$, como mostrado pela equação, que controla o feedback loop agindo em torno de $\Delta w_{ji}(n)$2.22.

$$\Delta w_{ji}(n) = -\eta \sum_{t=0}^{n} \alpha^{n-t} \frac{\delta \xi(t)}{\delta w_{ji}(t)} \tag{2.22}$$

A equação acima, uma série temporal com índice $t$, representa o efeito da sequência de apresentações de padrões na sináptica de pesos devido à constante de momento $\alpha$. O índice $t$ vai do tempo inicial 0 até o tempo atual $n$.

Com base na equação, podemos fazer as seguintes observações:

1. O ajuste atual $w_{ji}(n)$ representa a soma de uma série temporal ponderada exponencialmente que para ser convergente, o valor da constante de momento deve ser restrito ao intervalo $0 < \alpha < 1$. Quando é zero, o algoritmo de retropropagação opera sem momento.
2. Quando, em interações consecutivas, a derivada parcial $\frac{\delta\xi(t)}{\delta w_{ji}(t)}$ tem:
   - o mesmo sinal algébrico, significa que a soma ponderada exponencialmente $\Delta w_{ji}(n)$ cresce conseqüentemente, e , o peso $w_{ji}(n)$ também cresce num passo grande. A inclusão do momento no algoritmo de retropropagação tende a acelerar a descida em descidas com declividade constante.
   - sinais opostos, resulta que, a soma ponderada exponencialmente wji(n) diminui em magnitude e, conseqüentemente, o peso $w_{ji}(n)$ é ajustado em uma pequena quantidade. A inclusão do momento em o algoritmo de retropropagação tem um efeito estabilizador em direções que oscilam em sinal.

É necessária uma certa quantidade de ajustes para acertar a estrutura da rede e alcançar a convergência para algo próximo do ótimo global no espaço de peso (RUSSELL; NORVIG, 2010).

#### *2.6.4.5 Critérios de parada*

A busca da taxa de aprendizagem, termo momento e critério de parada, consiste na obtenção de uma fronteira de decisão que seja capaz de separar as classes. O algoritmo de retropropagação pode ser considerado convergido quando a norma euclidiana do vetor gradiente ou a taxa de variação absoluta do erro médio quadrático for suficiente pequeno, quando a taxa absoluta de a mudança no erro quadrático médio por época é suficientemente pequena (HAYKIN, 2009).

#### *2.6.4.6 Validação cruzada*

Se escolhermos uma rede muito complexa para treinamwento, ela poderá estar sujeita a um sobreajuste, quando ela conseguirá memorizar todos os exemplos formando uma grande busca de todas as condições, mas não necessariamente generalizará bem para entradas que não tenham sido vistas antes (RUSSELL; NORVIG, 2010).

Podemos extrair mais dos dados e ainda obter uma estimativa precisa usando uma técnica chamada validação cruzada múltipla dividindo o conjunto disponível de N exemplos em K subconjuntos, onde K>1; este procedimento assume que K é divisível em N. O modelo é treinado em N-1 subconjuntos, exceto um, e o erro de validação é medido testando-o no subconjunto que é deixado de fora. O desempenho do modelo é avaliado pela média do erro quadrático sob validação em todos os ensaios dos N experimentos (BERRAR, 2018).

## *2.7 Trabalhos correlatos*

Na literatura, vários artigos revisaram o estado da arte com abordagem em PHM. Foi realizada no banco de dados do Scopus[3] uma pesquisa para identificar abordagens relevantes correlatos com o tema no período de 2004 a 2023. Utilizamos as palavras-chaves destacadas na tabela 2.4 para as buscas nos títulos dos periódicos e resumos.

Vemos um crescimento exponencial do número de documentos publicados anualmente com as palavras-chave pesquisadas. A área de pesquisa de PHM, manutenção preditiva e detecção de falhas de máquinas em geral é abundante no que diz respeito ao número de artigos científicos publicados ao longo dos anos. Para aerogeradores, a evolução do número de pesquisas segue uma tendência de conformidade com a evolução anual da potência eólica instalada mundial.

A abordagem de monitoramento da saúde de turbinas eólicas, com a análise de dados operacionais, e utilização de informações do sistema de aquisição de dados e monitoramento (SCADA), Grupo 3, representa quando esses sistemas geralmente já estão implementados na maioria dos aerogeradores. Nesta abordagem, o sistema fornece uma grande quantidade de dados sem a necessidade de sensores extras apresentando significativa relação custo-benefício. É considerada uma solução eficiente para monitoramento do estado de saúde de turbinas eólicas (SREENATHA; MALLIKARJUNA, 2023).

[3] https://www.elsevier.com/solutions/scopus.

| **Objetivo** | **Palavras-chave** | **Grupo** | **Abordagens** |
|---|---|---|---|
| Buscar documentos com abordagem de PHM para deteção de falhas | ("PHM" *OR* "Prognostic and Health Management" *OR* "Condition Based Maintenance" *OR* "Predictive Maintenance") *AND* ("Diagnosis" *OR* "Diagnostic" *OR* "Fault Classification" *OR* "Fault Detection") | **1** | 4626 |
| Identificar documentos com aplicação de deteção de falhas utilizando PHM em turbinas eólicas | ("Wind turbine" *OR* "Eolic" *OR* "Energy") *AND* (**Grupo 1**) | **2** | 620 |
| Identificar as aplicações de deteção de falhas em turbinas eólicas utilizando PHM e SCADA | ("SCADA" *AND* (**Grupo 2**) | **3** | 49 |
| Identificar restrição do universo de busca para modelos de detecção de falhas em turbinas do tipo direct-drive ou PMSG | ("Permanent magnet" *OR* "PMSG" *OR* "Direct-drive") *AND* (**Grupo 2**) | **4** | 8 |
| Identificar as aplicações de modelos de detecção de falhas em turbinas do tipo direct-drive ou PMSG utilizando SCADA | ("SCADA" *AND* (**Grupo 4**) | **5** | 1 |

Tabela 2.4: Pesquisa para identificação de abordagens relevantes no período de 2004 a 2024. Busca atualizada em 10 de abril de 2024.

A figura 2.18 mostra um gráfico resultante de Pesquisa para identificação de abordagens relevantes no período de 2004 a 2024 por grupo de palavras-chave conforme tabela 2.4.

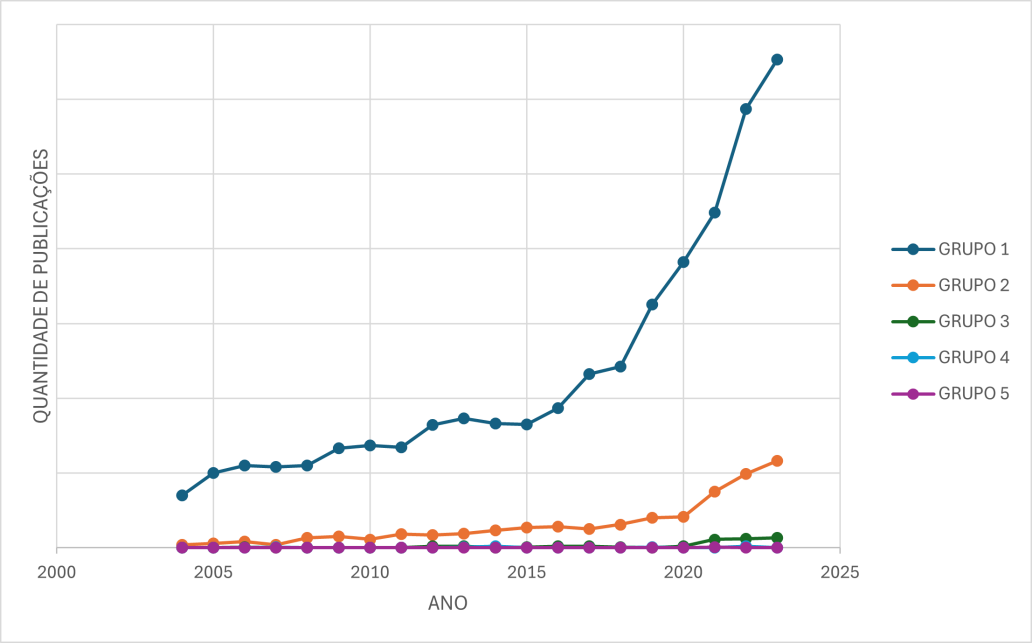


Figura 2.18: Gráfico resultante de Pesquisa para identificação de abordagens relevantes no período de 2004 a 2024 por grupo de palavras-chave.

A inexistência de abordagens do Grupo 3 e a tímida curva no início do período comparativo deve-se ao fato de:

- Os aerogeradores serem instalados em locais remotos, com sensores especiais tendo necessidade de sistema de telemetria para coleta de dados (KONG et al., 2023);
- Há poucos anos, os sistemas SCADA tinham capacidades limitadas de detecção e armazenamento (MUNGUBA et al., 2024);
- Falta de comunicação entre pesquisadores e proprietários de parques eólicos, quando preocupações comerciais e de sigilo propietários resultam na inacessibilidade e compartilhamento de dados operacionais reais causando falta de estudos consistentes na literatura (BERETTA et al., 2021)(LEITE; ARAÚJO; ROSAS, 2018).

A partir do ano de 2012, e conforme evolução do SCADA e tecnologia de telemetria, rede e armazenamento de dados mais extensos (MUNGUBA et al., 2024), os estudos com base de dados coletados do SCADA, passou a ser mais frequente, com 11, 12 e 13 publicações/ano, nos anos de 2021, 2022 e 2023 respectivamente.

Todavia, a escassez de dados continuou a ser um problema em algumas aplicações, as restrições aos dados proprietários continuaram apesar dos esforços em direção a conjuntos de dados de referência abertos (BERETTA et al., 2021) (MUNGUBA et al., 2024).

Assim, a grande maioria dos artigos referentes a aerogeradores apresentam estudo nos componentes mecânicos como mancais, rolamentos e gearbox para os aerogeradores de modo geral. Todavia, estudos e monitoramento da saúde para turbinas eólicas do tipo direct-drive utilizando SCADA não foram citados explicitamente, com exceção de 1. Sendo uma oportunidade de desenvolvimento, a aplicação de modelos de aprendizagem de máquina utilizando dados SCADA para o contexto das turbinas direct-drive do tipo PMSG e contribuição desta pesquisa para a literatura.

Várias abordagens baseadas em dados SCADA, como rede neural artificial, LSTM, KNN, Random Forests foram citadas na literatura. A tabela 2.5 mostra a lista de alguns artigos relacionados ao aprendizado de máquina na detecção de falhas em aerogeradores.

Em métodos inteligentes, o monitoramento de condições e a detecção de anomalias podem ser realizados em duas etapas: extração/seleção de atributos e detecção de falhas (CHEN et al., 2021). Em 100% dos artigos e estudos científicos foi verificada a utilização de algum método de extração/seleção.

Zhang, Robinson e Basu (2023) compararam as abordagens de aprendizado de máquina, como Árvore de Decisão, k-NN (K vizinhos mais próximos), SVM (Suport vector machine), RF (random Forests), RNA (redes neurais artificiais) e GB (Gradient Boost) para diagnóstico de falhas. Random Forests foi o melhor modelo entre todos em termos de melhor precisão (0,98607386), visualizada na figura 2.19.

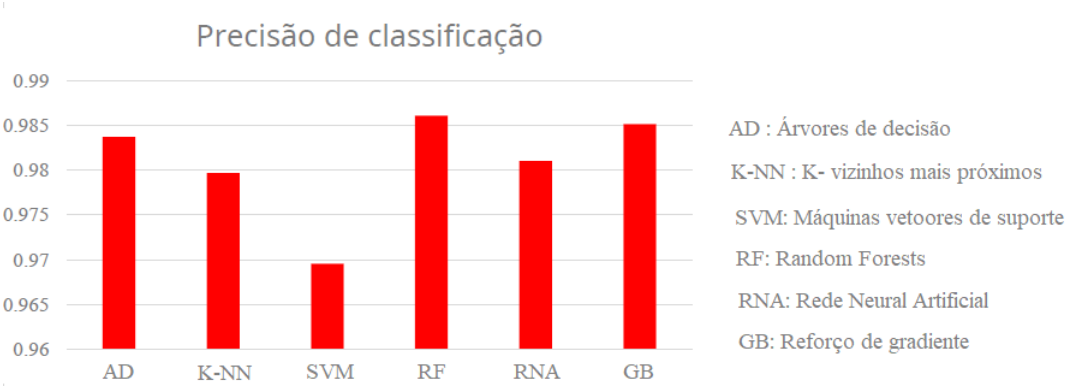


Figura 2.19: Comparação da precisão da classificação. Fonte: (ZHANG; ROBINSON; BASU, 2023).

| **Autor, Ano** | **Técnica** | **Dados de entrada** | **Tipo do aerogerador**[1] |
|---|---|---|---|
| Zhang, Robinson e Basu, 2023 | T2V-LSTM Random Forests | SCADA | Acionamento indireto |
| Sreenatha e Mallikarjuna, 2023 | Rede neural BLSTM | Dados públicos | Gearbox |
| Huanying e Dongsheng, 2023 | Correlação-Clusterização-SVM-SVM | SCADA | Acionamento indireto |
| Xiang et al, 2021 | CNN-LSTM-MA | SCADA | Acionamento indireto |
| Chen et al, 2021 | LSTM-AE | SCADA | Direct-drive |
| Santolamazza, Dadi e Introna, 2021 | CNN, LSTM | SCADA | Acionamento indireto |
| Xiao et al, 2021 | CNN, AOC-ResNet5 | SCADA | DFIG |
| Wang e Liu, 2021 | KNN, memória dinâmica | SCADA | Acionamento indireto |
| Tang et al, 2020 | XLightGBM | SCADA | Gearbox |
| Zhang et al, 2018 | Random Forests, XGBOOST | SIMULINK | Não citado |

[1]Direct-drive, acionamento indireto, não citado ou subsistema do aerogerador, quando especificado.

Tabela 2.5: Sumário de técnicas e dados de entrada de abordagens relevantes no período de 2004 a 2024. Busca atualizada em 10 de abril de 2024.

Em seu estudo, Zhang, Robinson e Basu (2023) desenvolveram um novo modelo de rede neural de aprendizagem profunda, T2V-LSTM, onde T2V é uma variação do Long Short-Term Memory (LSTM) para não só a detecção, como classificação de falhas com 10 a 210 minutos antes das falhas.

Han e Yang (2023) propuseram um método que primeiramente realiza a análise de correlação para filtrar as variáveis e suprimir dados redundantes com apoio em cálculo da distância entre atributos e clusterização. Em segundo lugar, é proposto um método de maximização da utilidade de dados baseado em duas máquinas de vetores de suporte. Uma antes, para obter a contribuição de cada atributo com a classe e uma depois, para determinar o melhor conjunto de atributos para a classe. Finalmente, uma série de máquinas de vetores de suporte paralelos são usadas para realizar monitoramento de múltiplas condições e diagnóstico de falhas.

Sreenatha e Mallikarjuna (2023) utilizaram um modelo baseado em uma rede neural BLSTM (memória de curto longo prazo bidirecional, que é uma arquitetura de rede neural recorrente projetada para lidar com sequências temporais) junto com um autoencoder,

com a função de identificar falhas baseadas em dados de vibração em gearbox de um aerogerador. O aprendizado e a redução de atributos são alcançados extensivamente por meio do autoencoder. A rede neural BLSTM, com a função de ativação sigmóid alcançou uma precisão de 98,68% na classificação de falhas da caixa de engrenagens da turbina eólica.

Xiang et al. (2021) propuseram um novo método em que uma rede neural convolucional (CNN) que se conecta à rede de memória de longo e curto prazo (LSTM) baseada no mecanismo de atenção (MA). Os dados do SCADA são usados da como variáveis de entrada e constroem a arquitetura CNN para extrair mudanças dinâmicas de dados. O mecanismo de atenção é aplicado para fortalecer o impacto de informações importantes, atribuindo pesos diferentes para concentrar as características do LSTM para aumentar a precisão do aprendizado por meio do mapeamento de peso e aprendizado de parâmetros.

Chen et al. (2021), em seu estudo verificaram o desempenho superior do LSTM sobre a RNA para detecção de anomalias em turbinas do tipo direct-drive. O modelo LSTM-AE (LSTM integrado com autoencoder) melhorou ainda mais a precisão da detecção devido à entrada bruta processada pelo AE e ao recurso de tempo gerenciado pelo LSTM. Com base em sinais de multisensores obtido via SCADA. O modelo detectou e classificou 11 tipos defalhas na roda eólica, rolamento, suporte de rolamento e rotor.

Santolamazza, Dadi e Introna (2021), realizaram uma aplicação de estudo de caso em turbinas eólicas de um parque eólico no sul da Itália. Eles utilizaram redes neurais artificiais, CNN (redes neurais convolucionais) e LSTM (redes neurais recorrentes de para monitoramento de caixas de engrenagens e gerador, utilizando dados aquisitados do SCADA e controle estatístico de processos para comparação entre os resultados encontrados no modelo e os gráficos de controle de qualidade. Os atributos de entrada utilizados foram escolhidos através do estudo da literatura técnica e científica. A variável de saída gerada pelo modelo é então comparada com o valor real medido pelo sistema de medição. O desvio entre os dois valores (real e estimado pelo modelo) é avaliado estatisticamente por meio de gráficos de controle de qualidade para identificar anomalias existentes no sistema.

Segundo Xiao et al. (2021), apresentaram uma abordagem para detecção de falhas em conversores de turbinas eólicas eólicas com Gerador de Indução Duplamente Alimentado, do inglês, (Doubly Fed Induction Generator - DFIG), usando modelos de redes neurais convolucionais, que são desenvolvidos usando dados do sistema SCADA. A abordagem começa com a seleção de variáveis indicadoras de falha, selecionadas com base na análise dos casos de falha existentes e na compreensão dos requisitos para as funções do conversor. Eles propõem uma nova arquitetura de rede convolucional CNN, AOC-ResNet50. A CNN teve sua eficácia avaliada na detecção de falhas em turbinas eólicas por meio de um estudo comparativo sobre a detecção de falhas do conversor de energia de turbinas eólicas com outros modelos de rede neural convolucional, como os modelos de rede ResNet50 e Oct-

ResNet50 para aprendizado profundo.

Em seu artigo, Wang e Liu (2021) propuseram um método baseado em técnicas de estimação de estado multivariado baseado em dados SCADA. Com a aplicação do algoritmo de seleção de características de informação mútua condicional máxima (utilizado para reduzir a redundância de informações nos parâmetros operacionais, permitindo uma seleção mais eficiente e relevante das características) e a construção de uma matriz de memória dinâmica com base no algoritmo, Knn (k-vizinhos mais próximos) e métodos de detecção de falhas baseados em resíduos emite alertas de falha se excederem um limite predefinido, e outro a longo prazo, que analisa os resíduos históricos em níveis diários por meio de gráficos de controle. Comparado às informações de falhas registradas, o método proposto permitiu alertas precoces de falhas com antecedência de 2 a 20 dias, destacando sua eficácia na detecção de problemas nos componentes das turbinas eólicas.

Tang et al. (2020) utilizaram versão adaptável da técnica LightGBM[4] para detectar problemas em caixa de engrenagens em conjunto com dados SCADA. O método para seleção de recursos utilizado foi baseado nos coeficientes máximos de informação, baseado em entropia e proposto para medir a quantidade de informação compartilhada entre dois recursos. Ao inserir o conjunto de recursos original, o método do coeficiente de informação máximo foi usado para seleção de parâmetros e saída do subconjunto de recursos ideal.

[4]Estrutura de Gradient Boosting Decision Tree (GBDT) baseada no algoritmo de árvore de decisão usando amostragem unilateral baseada em gradiente (GOSS) e agrupamento de recursos exclusivos (EFB).

Capítulo Três

# Materiais e Métodos

Neste capítulo é apresentado o desenvolvimento de uma aplicação de manutenção preditiva, que utilize aprendizado de máquina para processar os dados de monitoramento de um aerogerador do tipo direct drive de ímãs permanentes.

Segundo Geron (2019), o aprendizado conjunto, pode ser considerado construindo um modelo de previsão combinando os pontos fortes de modelos básicos mais simples.

Em seu estudo, (ZHANG; ROBINSON; BASU, 2023) compararam diversas abordagens de aprendizado de máquina para seleção de atributos visando diagnóstico de falhas. Random Forests foi o melhor modelo entre todos. Consequentemente, o classificador Random Forests foi um dos escolhidos para conduzir o estudo.

Han e Yang (2023) propuseram um algoritmo baseado em correlação combinado com SVM para filtrar os melhores atributos.

Não exatamente este método, mas também tamum que se desenvolve com métricas de correlação, o CFS subset evaluator foi outra abordagem escolhida.

A aplicação consiste em testar empiricamente um modelo de aprendizado conjunto, construindo uma Rede neural artificial Multilayer perceptron para previsão de classes combinada com algoritmos de seleção de atributos Random Forests e CFS subset evaluator, no banco de dados para geração dos subconjuntos ótimos que serão insumos de partida para a configuração da RNA. Obtendo assim o menor modelo possível para ter a máxima eficiência e desempenho computacional tanto em execução quanto a redução de de consumo de energia.

Finalmente, será avaliada a eficiência da rede neural artificial com base nas métricas de avaliação de modelos.

## *3.1 Banco de dados*

Para este desenvolvimento, foram utilizados dados de monitoramento dos aerogeradores, fornecidos pela Companhia ELETROBRAS. A coleta desses dados foi possível graças ao sistema de Controle de Supervisão e Aquisição de Dados (SCADA), que é alimentado por diversos sensores instalados em um aerogerador de ímãs permanente localizado no Brasil,

mais especificamente no estado da Bahia, em Casa Nova, além dos relatórios de Operação e Manutenção (O&M) disponibilizados pela empresa que presta o serviço de manutenção do parque eólico de Casa Nova A.

A turbina IMPSA IV-82 utiliza o princípio de geração através de imãs permanentes (direct drive). As especificações técnicas dessa turbina estão listadas na Tabela 3.1.

| Especificações Técnicas | |
|---|---|
| Gerador | Síncrono por ímãs permanentes |
| Modelo | IMPSA IV-82 |
| Fabricante | GoldWind |
| Potência nominal | 1500kW |
| Diâmetro do rotor | 82m |
| Altura do cubo | 100m |
| Velocidade de rotação | $9 \sim 21$ rpm |
| Velocidade de conexão do vento | 3m/s |
| Velocidade de rotação | $9 \sim 21$ rpm |
| Tensão nominal (rede) | 690V |
| Frequência nominal (rede) | 60hz (45 a 65hz) |
| Peso da Nacelle (desconsiderando rotor e gerador) | 11 ton |
| Peso do gerador | 44 ton |
| Peso do rotor (incluindo rotor e hub) | 28 ton |

Tabela 3.1: Especificações Técnicas do aerogerador IMPSA IV-82 direct drive de imãs permanentes. Fonte: Projeto P&D.

O SCADA é composto por centenas de sensores, instalados nos aerogeradores, registrando dados em passos de 10 minutos, alarmes e registro de falhas. A Figura 3.1 apresenta uma tela do SCADA com representação e monitoração dos parâmetros de operação dos aerogeradores instalados no parque.

A Tabela 3.2 apresenta apenas alguns dados de processo dos sistemas sensoriados que compõem o monitoramento do aerogerador.

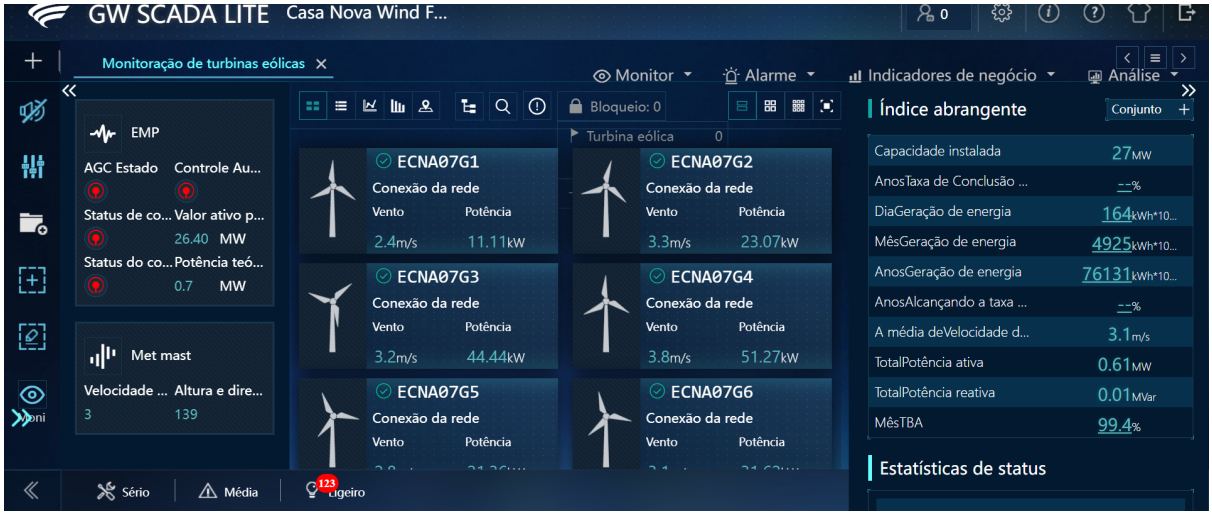


Figura 3.1: Visualização da tela do SCADA.

| Sensores | Aplicação |
|---|---|
| Sensor indutivo 87 graus | Posicionamento da Pá 87 graus |
| Sensor de Temperatura - PT100 | Monitoramento do Motor de Pitch |
| Sensor de Temperatura - PT100 | Monitoramento da temperatura externa ambiente |
| Sensor de Temperatura - PT100 | Monitoramento da temperatura ambiente da Nacelle |
| Sensor de Temperatura - PT100 | Monitoramento da temperatura do freio do motor Yaw |
| Sensor de Temperatura - PT100 | Monitoramento da temperatura do óleo hidráulico |
| Sensor de Temperatura - PT100 | Monitoramento da temperatura do bobinado do gerador |
| Pressostato de acionamento direto | IMonitoramento da Pressão do Sistema Hidráulico |
| Sensor de Vibração - Pêndulo | Monitora vibração da Nacelle |
| Sensor de tensão | Monitoramento positiva do DC link CC |
| Sensor de corrente | Monitoramento de corrente do conversor |
| Sensor de potência | Monitoramentio potência ativa - conversor ativo |
| Sensor de frequência | Monitoramento da frequência da Rede |

Tabela 3.2: Lista de alguns sensores do conjunto turbina IMPSA 1,5MW e conversor GoldWind

Foram foram avaliados os relatórios mensais que compreendem o período de agosto/2021 até março/2023. Nesse período, na turbina de número 18, vemos que as 28 paradas notificadas foram motivadas por 11 falhas diferentes, sendo que a falha de código 442[1](Error IGBT ok Loss) na qual faz referência a perda do sinal do IGBT foi a mais reincidente com um total de 6 notificações. A falha de perda de sinal do IGBT pode ser desencadeada por inúmeros motivos internos e externos à turbina, porém na maioria das vezes em que

[1]Os códigos apresentados nesse documento foram retirados do relatório de falhas da fabricante Goldwind, responsável pela operação e manutenção das turbinas do complexo. Mais informações sobres os códigos e suas respectivas falhas podem ser observadas no documento no anexo I.

essa falha foi notificada nos relatórios se deu pela queima de algum IGBT e, consequentemente, sua troca foi efetuada. Em seguida vieram as falhas de números 84 (Error Gen. Side Capacitor Fuse Feedback Loss), 164 (Pitch Generation Position Sensor Error), 22 (Error Hydraulic) e 95 (Error Pitch Safty Chain Triggered). A Figura 3.2 demonstra as 5 falhas mais recorrentes nesse período.

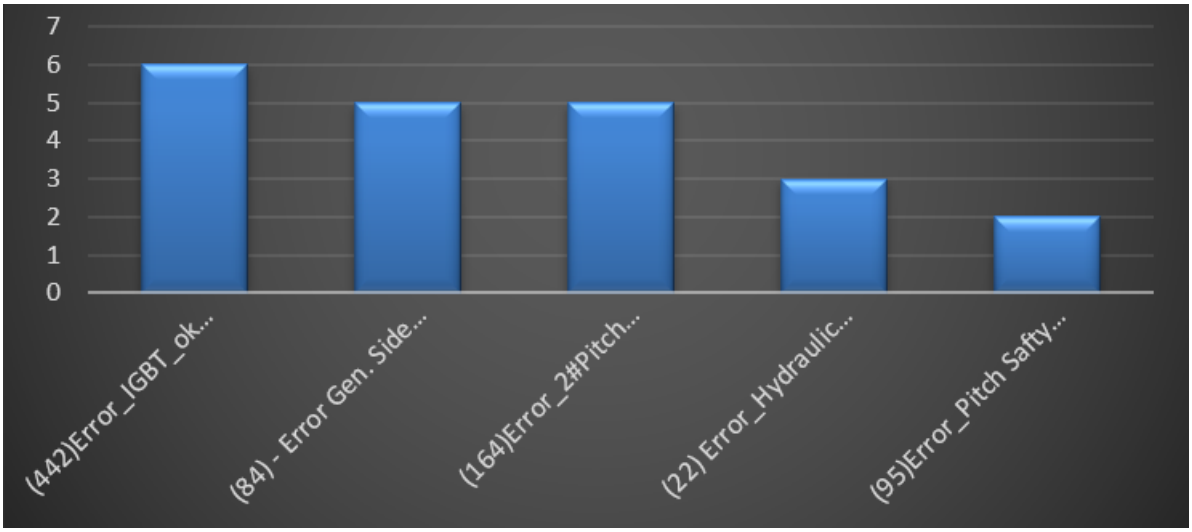


Figura 3.2: Quantidade de notificações de falhas ocorridas no aerogerador 18 no período de agosto/2021 até março/2023. Fonte: Elaborado por P4.3.

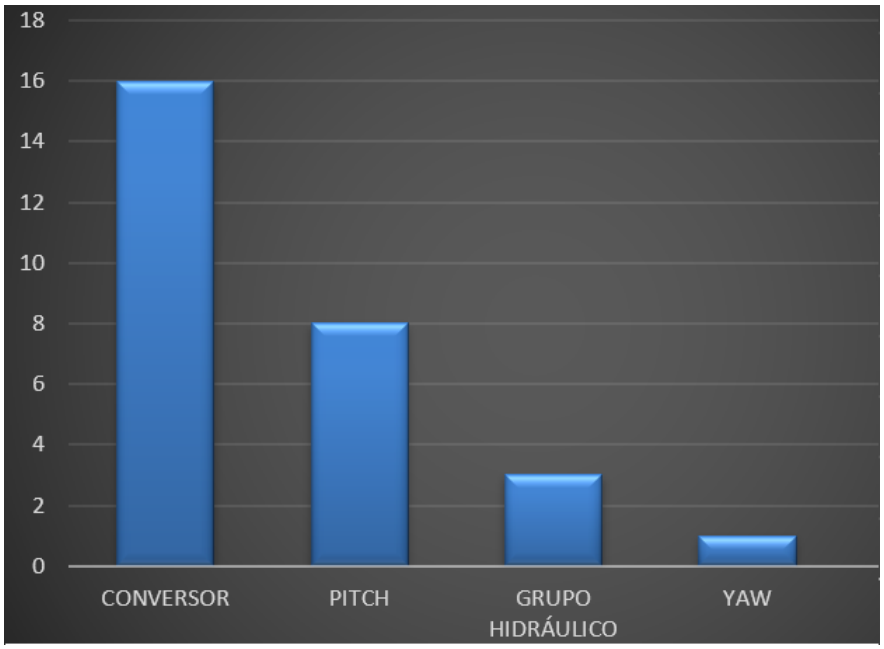


Figura 3.3: Incidência de Notificações de falhas por subsistema na turbina 18 no período de agosto/2021 a março/2023. Fonte: Elaborado por 4.3

Quando observados os locais das falhas dentro da WTG 18, apenas 4 sistemas tiveram notificações de falha, porém seguindo a tendência de todas as outras turbinas do complexo eólico, o local com maior incidência de falhas é o conversor com 16 notificações de falha, seguido do sistema Pitch, grupo hidráulico e Yaw, vide figura 3.3.

O estudo desses relatórios comprovou informações encontradas nas pesquisas realizadas anteriormente, e mostra a grande confiabilidade mecânica das turbinas PMSG, isso porque mais de 80% das falhas identificadas durante o período de avaliação foram ocasionadas por problemas puramente elétricos (curtos-circuitos, mal contatos, sobretensão, subtensão).

## *3.2 Modelo computacional*

Deseja-se distinguir entre duas classes de padrões bidimensionais, rotuladas como falha e não falha.

O modelo computacional, figura 3.4, consiste nas seguintes etapas:

- Realização do pré-processamento dos dados;
- Seleção de atributos, utilizando os algoritmos de seleção CFS subset evaluator e Random Forest, para obtenção de subconjuntos de características mais relevantes;
- Configuração das redes neurais artificial Perceptron de multicamadas, a partir dos melhores atributos obtidos para cada método na etapa anterior. São determinadas a quantidade de neurônios de entrada e camadas ocultas e então a rede é treinada de modo a buscar os hiperparâmetros que resultam numa melhor convergência;
- Validação das redes neurais artificiais;
- Avaliação dos modelos computacionais: onde são avaliadas as eficácias dos algoritmos, através de métricas de avaliação.

A Figura 3.4 descreve a fluxo do estudo.

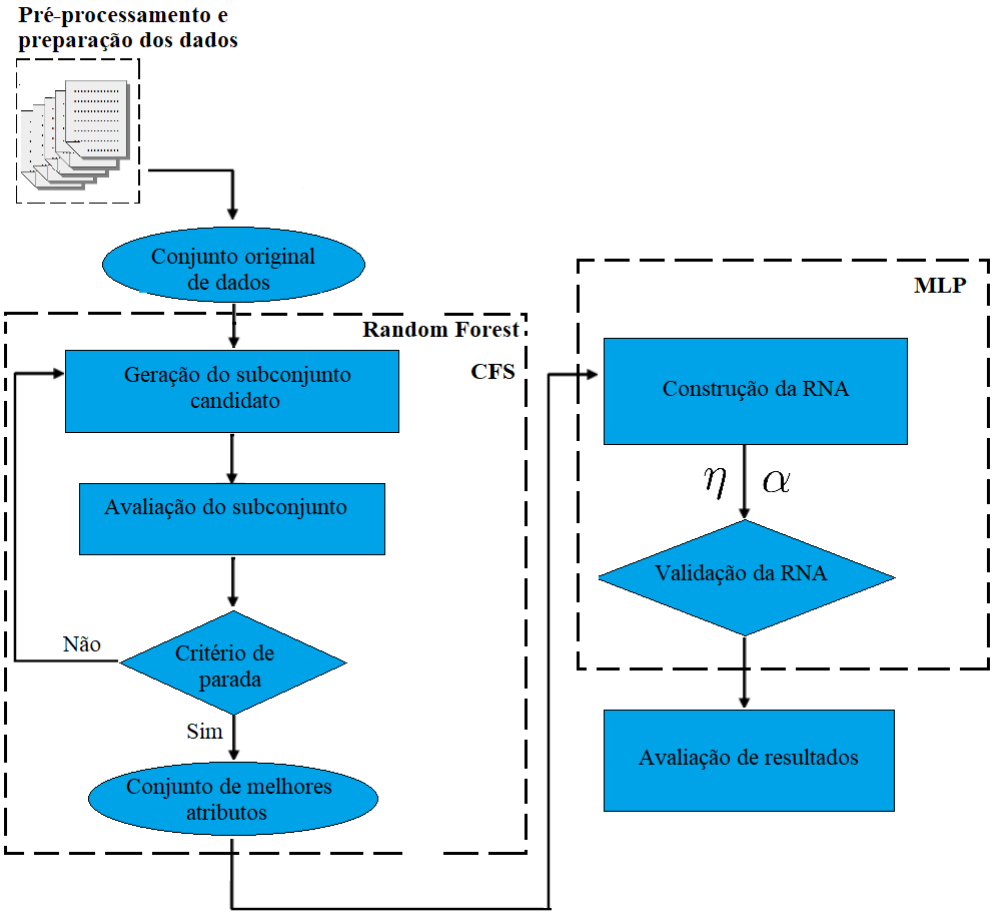


Figura 3.4: Fluxo do estudo.

### *3.2.1 Pré-processamento*

É natural a dificuldade em reunir um grande conjunto de dados representativos de equipamentos reais. Isto porque, para obter os dados é necessária mão de obra especializada para realizar tratamentos e avaliações preliminares dos dados: integridade e aderência dos dados. Existem outros fatores, tais como: segurança da informação, sigilo industrial, avaliação de dados sensíveis e documentação.

A construção de modelos de Aprendizado de Máquina teve início no pré-processamento, que proporcionou uma organização fundamental das informações para facilitar sua análise e viabilizar a sua utilização nos modelos neurais. Esta etapa de consistiu em importar os dados do dataset, provenientes do sistema já instalado no aerogerador (SCADA).

- O controlador lógico programável do processo (instalado na base do aerogerador 18)

recebe as variáveis dos sensores instalados no aerogerador;

- O sistema supervisório troca dados com o controlador lógico programável, sendo esses dados, as variáveis de comunicação e monitoramento requisitadas. As variáveis do sistema existente requisitadas para o desenvolvimento do AM são extraídas manualmente por um usuário habilitado. Um arquivo tipo ".CSV"(separados por vírgula ou comma-separated values) é gerado a partir dos dados obtidos.
- Os registros foram iniciados em 01-01-2021 às 00:00:00 com o último registro em 30-03-2023 às 23:50:00 apresentando passos temporais de 10 minutos.

#### *3.2.1.1 Pré-processamento Primário*

O pré-processamento primário consiste na primeira etapa após o recebimento dos dados fornecidos pela Eletrobras. Os dados são, portanto, enviados em múltiplos arquivos do tipo CSV com colunas que variam a depender da demanda e linhas que indicam a coleta pontual ao longo do tempo. Os tempos entre coletas são variáveis e estão entre 6 e 600 segundos. Os dados faltantes não são representados diretamente, o sistema envia dado sem a coleta no tempo caso haja ausência de dados. Portanto, faz-se necessário um processo de padronização da estrutura de envio dos dados. A Figura 3.5 apresenta o fluxo de dados para execução desse procedimento.

Após o download dos dados, é necessário traduzir os nomes das colunas do chinês para o português. Posteriormente é feita a sincronização dos dados, quando todos dados são transformados para séries temporais com tempo entre coleta de 600 segundos. Para agregar múltiplos valores em cada intervalo de tempo, utilizando métricas baseadas no tipo de variável. Para variáveis categóricas e nominais foi utilizada moda e para variáveis numéricas foi utilizada média.

Posteriormente os dados que estavam em múltiplos arquivos separados são agregados em um único arquivo com extensão do tipo ".CSV"de séries temporais com tempo entre coletas de 600 segundos. Para isso é feito uma combinação em que os nomes das colunas e índices temporais das linhas são agregados e calcula-se a união entre os dataframes de cada arquivo individual. Não há duplicata para as interseções. Esse arquivo é utilizado como entrada para execução dos procedimentos citados nas seções posteriores.

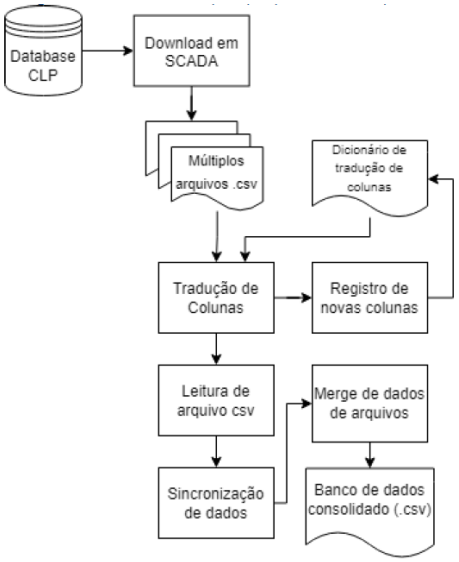


Figura 3.5: Fluxo de dados para pré-processamento primário. Fonte: (BARBOSA, 2023).

Em seguida, conhecer aplicação e os dados e prepará-los para a fase seguinte. Entre as diversas tarefas realizadas nessa fase pode-se citar: Integração de dados, Limpeza de dados, Padronização de dados e atributos. Assim assegura-se que os dados sejam de boa qualidade e apropriados para realizar as etapas seguintes.

Todas essas informações serão utilizadas para treinar os classificadores de seleção de atributos.

#### *3.2.1.2 Pré-processamento Secundário*

Com a conclusão do processamento primário e a garantia da uniformidade das informações brutas, inicia-se a etapa secundária de processamento. O objetivo desta etapa é o refino e estruturação dos dados para viabilizar a aplicação de técnicas estatísticas de análise exploratória, bem como a estruturação básica para viabilizar o emprego das técnicas de AM.

Incialmente o banco de dados apresenta uma alta dimensionalidade representado por 423 atributos que são provenientes de sensores associados ao Aerogerador 18.

A primeira etapa do processamento secundário é feita com a busca por dados faltantes. A sua presença é natural em banco de dados reais. Dados faltantes podem ocorrer por vários fatores, tais como: erros de sistema de coleta, falha em sensores, filtros de dados espúrios etc. As correções mais simples são realizadas pelo preenchimento com valores

médios ou interpolações, entretanto, a aplicação dessas metodologias carece de cuidado. Interpolações em grandes lacunas de dados faltantes podem inserir comportamentos não representativos na série temporal, impactando diretamente no desempenho do treinamento dos modelos. Partindo deste princípio, buscou-se verificar a integridade de todos os atributos do banco de dados.

Os atributos identificados foram automaticamente desconsiderados resultando em 282 atributos para os próximos procedimentos.

Então, a base de dados consolidada contém séries temporais multivariadas com 282 atributos. Dentro desses atributos, há representações temporais de valores máximo, médio e mínimo. Os registros foram iniciados em 01-01-2021 às 00:00:00 com o último registro em 30-03-2023 às 23:50:00 apresentando passos temporais de 10 minutos. Os atributos estão subdivididos em grupos. Os dados mais representativos são os dos sensores de pitch, que representam 73 do total. Em seguida, por representatividade estão os dados de monitoramento dos componentes elétricos, capacitores, diodos e insulated-gate bipolar transist (IGBTs), com 30 atributos. O gerador tem uma representatividade de 22 atributos. As grandezas elétricas tensão, corrente são monitoradas com 18 e 9 atributos respectivamente. Ainda existe o monitoramento da captação do vento, aceleração, yaw, conversor, consumo de energia, dentre outros. Este conjunto de dados é composto por um total de 64.448 registros e abrange as seguintes condições do equipamento: funcionamento normal (32.224 registros) e falha 32.224 (registros).

### *3.2.2 Seleção de atributos*

Nesta etapa do modelo computacional, o próximo e importante passo é determinar quais recursos são relevantes para a construção da rede neural artificial. Teremos a visualização dos padrões extraídos, remoção de padrões irrelevantes ou redundantes, utilizando o algoritmo CFS e Random Forest. O primeiro foi desenvolvido por HALL (1999), e é uma abordagem de filtro que consiste em selecionar recursos para aprendizado de máquina por meio de uma avaliação baseada na relação entre atributos e na correlação entre atributos e classes. Já o segundo, Random Forest, desenvolvido por (BREIMAN, 2001), executa uma seleção características mais relevante baseada em importância para cada característica.

Quanto à eficácia, um critério de avaliação poderia ser o quão semelhantes são o subconjunto selecionado e o subconjunto ideal, mas não temos conhecimento prévio acerca do subconjunto ideal. Pesquisas mostraram que nenhuma abordagem de aprendizagem é claramente superior em todos os casos, mas a qualidade dos dados está em primeiro lugar. Matematicamente não é possível realizar a validação de um algoritmo de seleção de atributos sem estimar sua performance. (LEE, 2005). Após aplicação do método de seleção de

atributos, podemos utilizar um procedimento de validação para verificar se o subconjunto de recursos selecionado é válido (KAREGOWDA; JAYARAM; MANJUNATH, 2011).

A seleção de atributos precedeu a classificação real do processo e foi independente do algoritmo de indução de aprendizagem. Os subconjuntos de atributos selecionados pelos métodos CFS e Random Forest foram utilizados como entrada para o algoritmo classificador Multilayer perceptron para precisão preditiva como uma medida indireta de validação.

### *3.2.3 Configuração da RNA de múltiplas camadas*

Agora, já definidos quais são os melhores atributos que representam o estado de saúde do aerogerador, removendo atributos que não estão contribuindo para o resultado, extraídos pelos métodos CFS subset evaluator e Random Forest, o desempenho de cada método será medido indiretamente, comparando os desempenhos entre si utilizando um algoritmo de aprendizado de máquina.

Para esta comparação, construiu-se um modelo computacional de aprendizagem supervisionada, Multilayer Perceptron (MLP), para classificação dos dados de falha e normalidade.

Nesta etapa busca-se encontrar arquitetura ou topologia de conexão do MLP (número de camadas ocultas e número de neurônios por camada) que ofereça uma boa precisão de previsão nos conjuntos de validação.

#### *3.2.3.1 Número de camadas ocultas*

Não existe regra para definição da quantidade de camadas ocultas. Como a atualização dos hiperparâmetros da rede é feita a partir da retropropagação dos sinais funcionais e de erro, a utilização de um grande número de camadas ocultas torna o processo menos preciso (HAYKIN, 2009).

Os testes empíricos com a rede neural MLP backpropagation atingiram convergência com redução do erro médio quadrático com a utilização de duas camadas intermediárias ocultas.

#### *3.2.3.2 Número de neurônios por camada oculta*

O número de neurônios em cada camada é uma questão empírica, não existindo assim regras explícitas para uma modelagem ideal. Deste modo, foi utilizada a seguinte equação:(HAN; KAMBER; PEI, 2011)

$$N_{ocultos} = 2N_{entrada} + 1$$

onde $N_{entrada}$ é o número de neurônios da camada de entrada e $N_{ocultos}$ representa o número de neurônios da camada oculta.

### *3.2.4 Método de validação de modelos*

Uma vez contruída a rede neural, sua precisão e capacidade de generalização foram testadas utilizando os dados dos sensores do aerogerador, que compõe o conjunto de amostras para treinamento e testes do modelo. A técnica de validação cruzada com 10 partições (10-fold cross-validation) delineada nos estudos de (PAL; PATEL, 2020) foi utilizada para validação do modelo, que foi treinado em N-1 subconjuntos, exceto um, e o erro de validação é medido testando-o no subconjunto que é deixado de fora (BERRAR, 2018).

### *3.2.5 Métricas de avaliação de resultados das RNAs*

Os resultados foram avaliados por meio de métricas de avaliação estatísticas por classe e verificada a maior convergência com a utilização de um novo conjunto balanceado de dados contendo 1.894 amostras, que representam 2,93% do conjunto de dados.

Desse modo, a precisão preditiva é comumente utilizada como uma medida indireta para avaliar a qualidade dos atributos selecionados, comparando os rótulos dos dados reais com os dados preditos.

Usualmente utiliza-se as denominações:

- VP - Verdadeiro Positivo: quando o rótulo predito é coincidente com o rótulo real.
- VN - Verdadeiro Negativo: quando o rótulo predito é coincidente com o rótulo real.
- FP - Falso Positivo: quando o rótulo predito é divergente do rótulo real. O modelo previu um resultado positivo mas o valor real é negativo.

- FN - Falso Negativo: quando o rótulo predito é coincidente com o rótulo real. O modelo previu positivo mas o valor real é negativo. O modelo previu um resultado negativo mas o valor real é positivo.

Neste estudo para a avaliação deste modelo MLP, optou-se pelo emprego das seguintes abordagens:

- Matriz de confusão: É uma ótima representação visual entre as classes reais (num eixo) e preditas (no outro eixo). Nela estão dispostos os quadrantes no formato de matriz ou tabela nxn, onde para o caso de uma classificação binária é uma matriz 2x2.

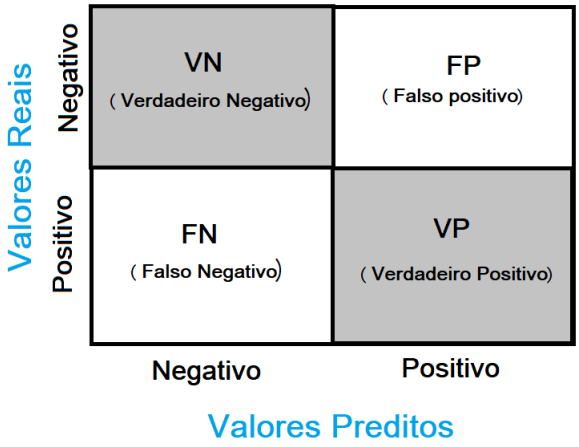


Figura 3.6: Matriz de confusão.

- Acurácia: representa o quanto o modelo acertou (previsões corretas) em relação a todas as classes.

$$Acuracia = \frac{VP + VN}{VP + VN + FP + FN} \tag{3.1}$$

- Revocação: Mede a quantidade de exemplos que o modelo previu como positivos e acertou dividido pelo número total de exemplos que realmente são positivos (que ele deveria ter acertado).

$$Revocacao = \frac{VP}{VP + FN} \tag{3.2}$$

- Precisão : Mede a quantidade de vezes que o modelo acerta em relação ao total de vezes que ele tenta acertar.

$$Precisao = \frac{VP}{VP + FP} \tag{3.3}$$

- F1-Score

$$F1 - Score = 2\frac{Precisao * Revocacao}{Precisao + Revocacao} \tag{3.4}$$

Capítulo Quatro

# Resultados e Discussão

Este capítulo tem como objetivo apresentar os resultados experimentais e desempenho do modelo de classificação utilizando os algoritmos de seleção de recursos Filtragem de atributos relevantes (CFS) e Random Forests com o algoritmo de aprendizado de máquina MLP, utilizando os dados SCADA.

Essa etapa é o resultado de pesquisas, aprimoramentos de processamento, exploração e interpretação dos dados, além do aperfeiçoamento da técnica de Aprendizado de máquina aplicada em dados reais de operação dos aerogeradores situados no parque eólico de Casa Nova A. Esta seção também irá contemplar discussões e conclusões sobre os resultados alcançados, bem como suas vantagens e vulnerabilidades frente aos dados obtidos até o presente momento.

## *4.1 Filtragem de atributos relevantes*

Chegou-se na etapa de seleção de atributos. Os principais recursos foram selecionados pelas estratégias de avaliação de subconjuntos CFS subset evaluator e Random Forest.

### *4.1.1 CFS subset evaluator*

A estratégia de avaliação de subconjunto CFS subset evaluator combinada com o método de geração de subconjunto Best-First foi implementada para reduzir a dimensionalidade.

Após submeter o banco de dados original, com 282 atributos, ao algoritmo CFS subset evaluator, foram selecionados 7 atributos mais relevantes para entrada na RNA MLP, proporcionando classificação de falhas, conforme ilustrados na Figura 4.1. São eles: temperatura máxima ambiente, aceleração média eixo y, corrente mínima da fase A grid side, ângulo médio dos valores mínimos de pitch, aceleração média rms, temperatura máxima do reator grid side 1 e ângulo mínimo de pitch da pá 3.

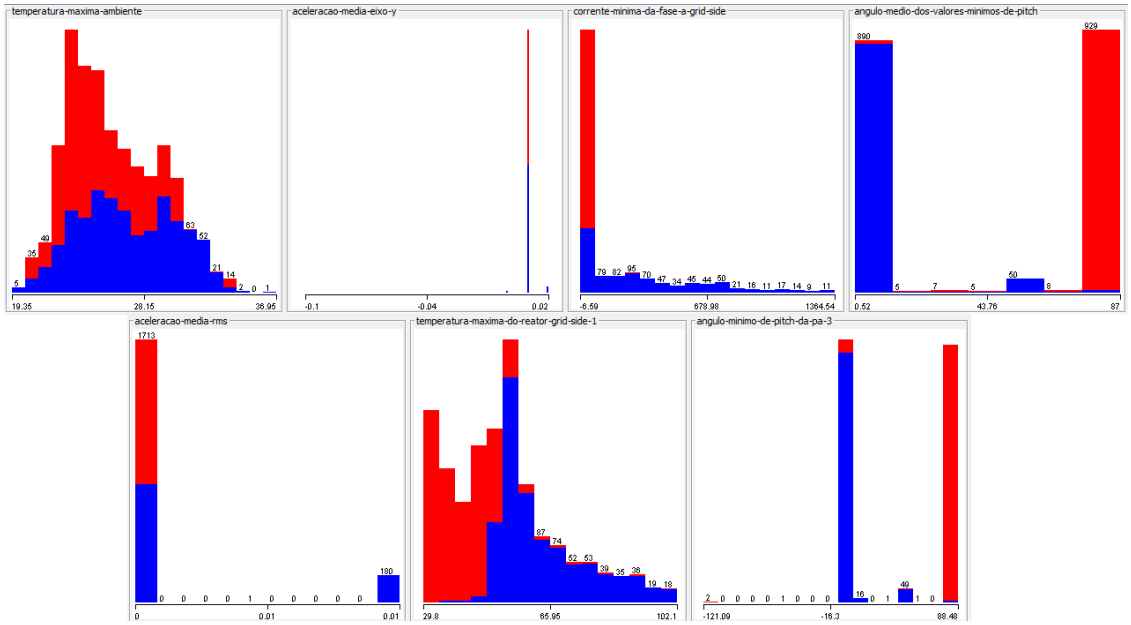


Figura 4.1: Subconjunto dos 7 (sete) atributos mais relevants extraídos pelo método CFS subset evaluator.

### *4.1.2 Random Forests*

O mesmo banco de dados original, com 282 atributos, ao foi submetido ao algoritmo Random Forest para redução de dimensionalidade. O resultado foi de 6 atributos com características mais discriminativas no conjunto de dados com base na diminuição média de impurezas em 100 árvores de decisão com bagging, para entrada na RNA MLP, proporcionando classificação de falhas, conforme ilustrados na Figura 4.2. São eles: temperatura inicial de falha grid side, tempo inicial de trabalho hidráulico, corrente máxima da fase C grid side, tempo final de operação do modo motor de yaw 2, ângulo máximo de pitch da pá 2 e velocidade média do vento.

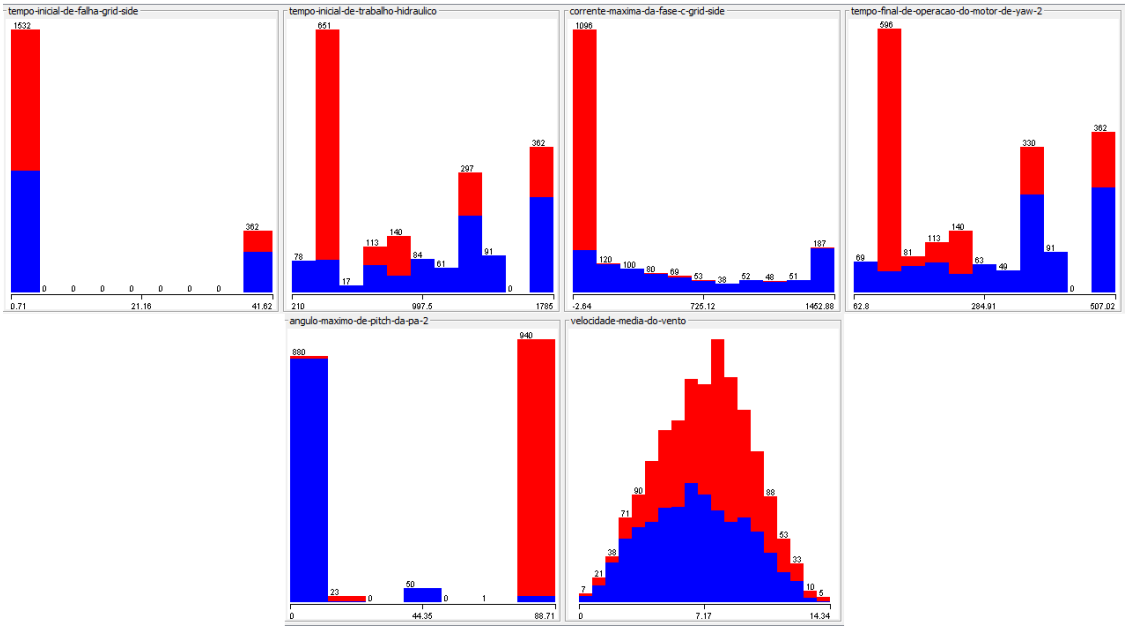


Figura 4.2: Subconjunto dos 6 (seis) atributos mais relevantes extraídos pelo método Random Forest.

## 4.2 *Configuração das Redes Neurais Artificiais de múltiplas camadas*

Duas redes neurais perceptron tendo como entrada os atributos extraídos do CFS evaluator subset e Random Forest foram configuradas. Suas composições foram de quatro camadas: uma camada de entrada composta pelos neurônios (atributos mais relevantes), duas camadas ocultas e uma camada de saída composta por dois neurônios que representa as classes de funcionamento normal (1) ou falha (0).

Para a MLP com atributos extraídos pelo CFS a camada de entrada foi composta pelos 7 neurônios (atributos mais relevantes), e as camadas ocultas contendo 15 neurônios cada. Já a MLP com atributos extraídos pelo Random Forest, a camada de entrada foi composta pelos 6 atributos mais relevantes, e as camadas ocultas contendo 13 neurônios cada.

A arquitetura final da rede neural CFS-MLP é ilustrada na Figura 4.3 e a da rede neural RF-MLP é ilustrada na Figura 4.4

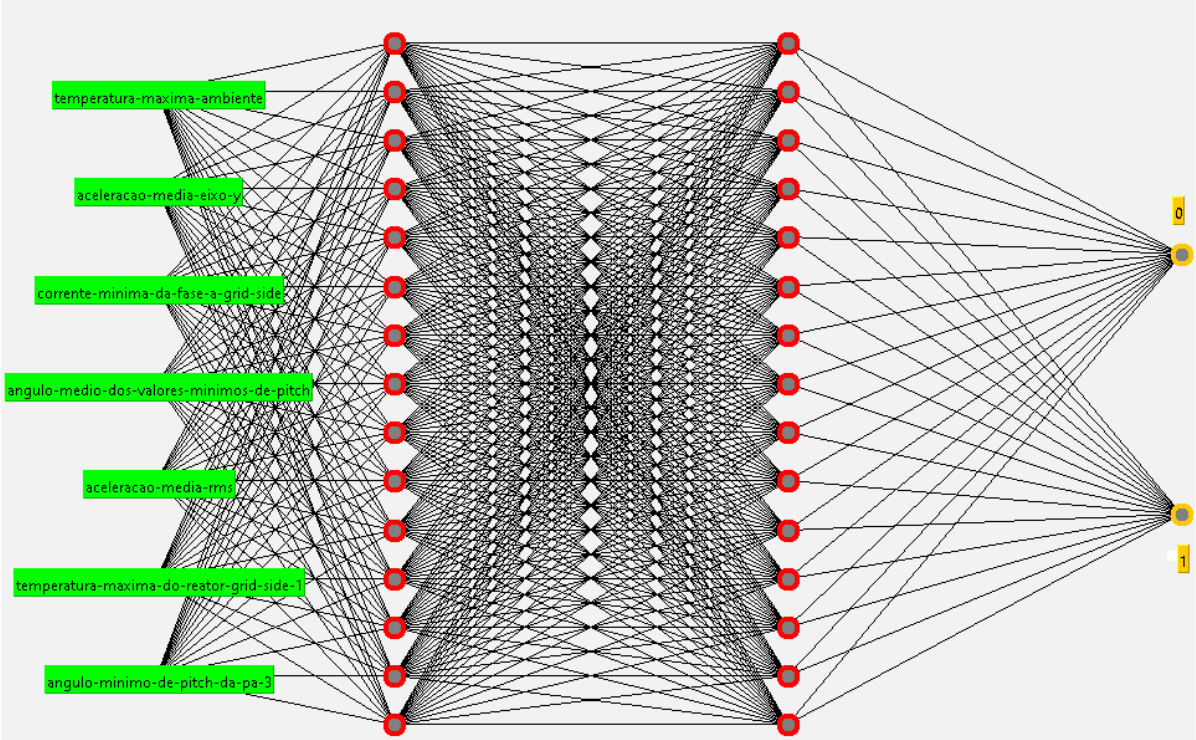


Figura 4.3: Arquitetura da RNA MLP utilizada com os 7 atributos extraídos pelo método CFS subset evaluator. Fonte: (MERGULHãO et al., 2024).

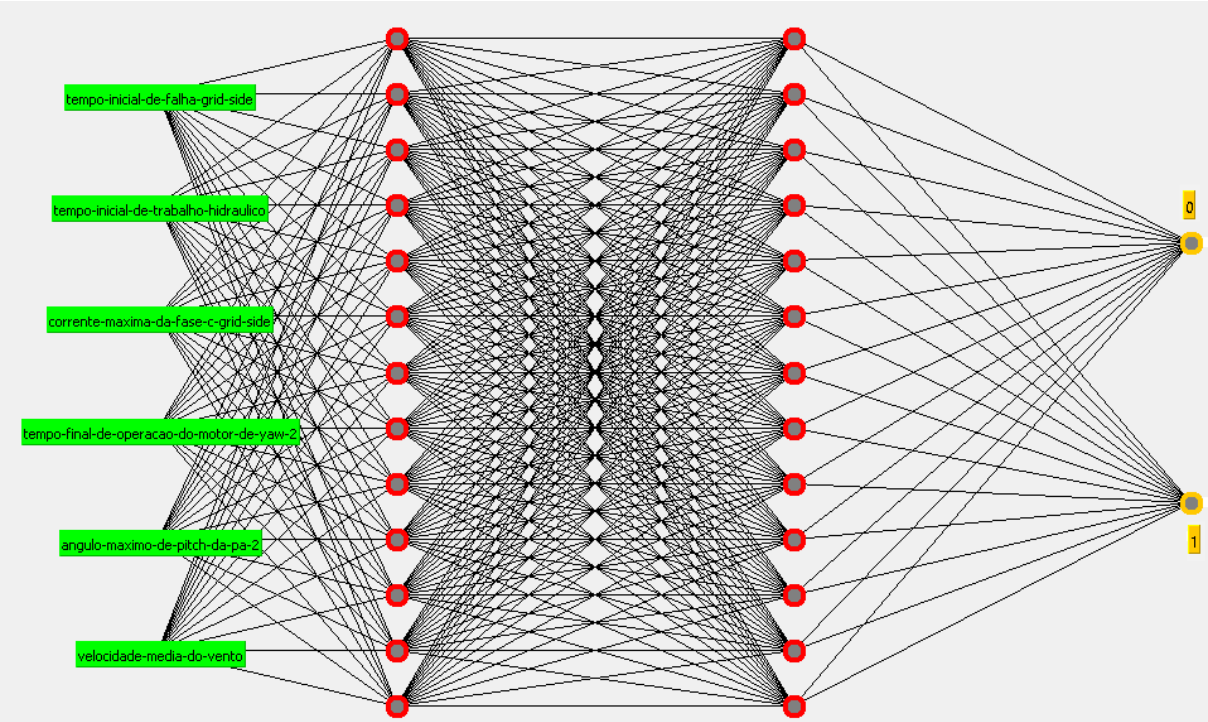


Figura 4.4: Arquitetura da RNA MLP utilizada com os 6 atributos extraídos pelo método Random Forest.

A função de ativação utilizada foi sigmóide, sua saída foi interpretada como a probabilidade de uma determinada instância pertencer à classe 1 (normalidade), $\varphi(v) = P(y = 1|x; w)$, dadas os seus atributos, $\mathbf{x}$, parametrizado pelos pesos, $\mathbf{w}$. Por exemplo, se calcularmos $\varphi(v) = 0,7$ para uma determinada instância, significa que a chance dessa instância ser uma condição de normalidade é de 70%. Portanto, a probabilidade de que esta instância seja uma condição de falha pode ser calculada como $P(y = 0|x; w) = 1 - P(y = 1|x; w) = 0,3$ ou 30%. A probabilidade prevista foi então convertida em um binário resultado por meio de uma função de limite:

$$\begin{cases} 1, & \text{se } \varphi(v) \geq 0,5 \\ 0, & \text{caso contrario} \end{cases} \tag{4.1}$$

## *4.3 Validação das RNA Multilayer perceptron*

Uma vez contruída a rede neural, sua precisão e capacidade de generalização foram testadas utilizando a técnica de validação cruzada com 10 partições nos dados dos sensores do aerogerador disponibilizados pela ELETROBRAS que compõe o conjunto de amostras para treinamento e testes do modelo. O treinamento do modelo é baseado na retropropagação. De acordo com os gradientes de peso da estrutura da rede, Foram simuladas combinações do parâmetro da taxa de aprendizagem (0.01, 0.1, 0.5, 0.9) e da constante de momentum (0.0, 0.1, 0.5, 0.9) para observar o efeito da convergência da RNA. Cada combinação foi treinada até 300 épocas, após o que ele foi encerrado. Esta extensão de treinamento foi considerada adequada para o algoritmo da retropropagação alcançar um mínimo local na superfície de erro. O Erro quadrático médio (RMSE) foi utilizada como função de perda do modelo. O objetivo do treinamento do modelo é minimizar a função de perda por meio de otimização. Neste estudo, a função sigmóide foi usada para atualizar os pesos da rede sob determinada taxa de aprendizagem $\eta$ e ao longo das épocas.

O treinamento do Perceptron, assim, consiste em encontrar a taxa de aprendizagem e constante de momento, $\eta$ e o $\alpha$ respectivamente, ótimas, que em média, produzem convergência para o mínimo local ou global na superfície de erro com o menor número de épocas. Ou seja, não se busca o menor valor da variação da função de custo, mas a curva com melhor acentuação de convergência.

### *4.3.1 Validação da RNA CFS-MLP*

As curvas de aprendizagem médias da RNA CFS-MLP são ilustradas nas Figuras 4.5 a 4.8. Agrupadas pela taxa de aprendizagem ($\eta$) em função das constantes de momentum ($\alpha$).

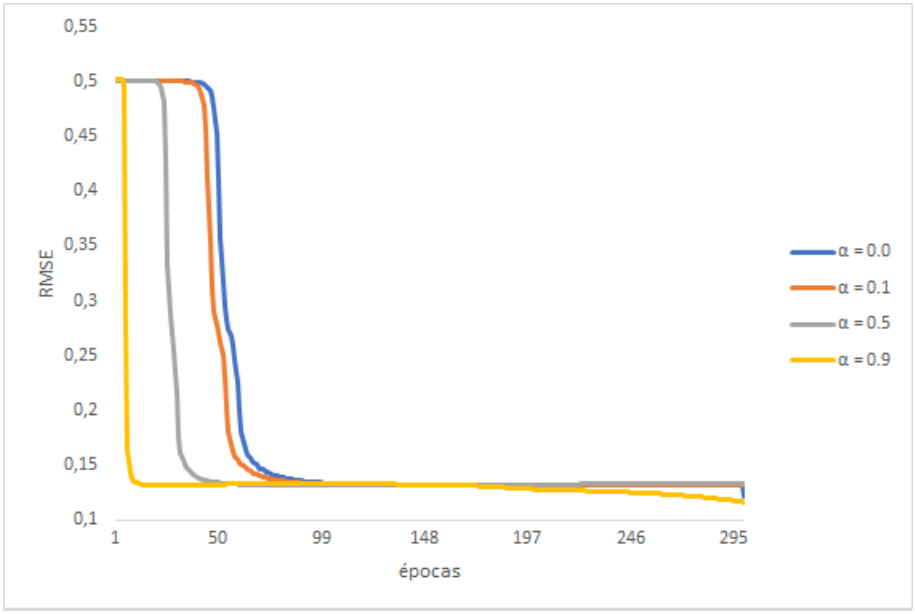


Figura 4.5: Curvas de aprendizagem médias para $\eta = 0.01$ com atributos extraídos pelo método CFS. Fonte: (MERGULHãO et al., 2024).

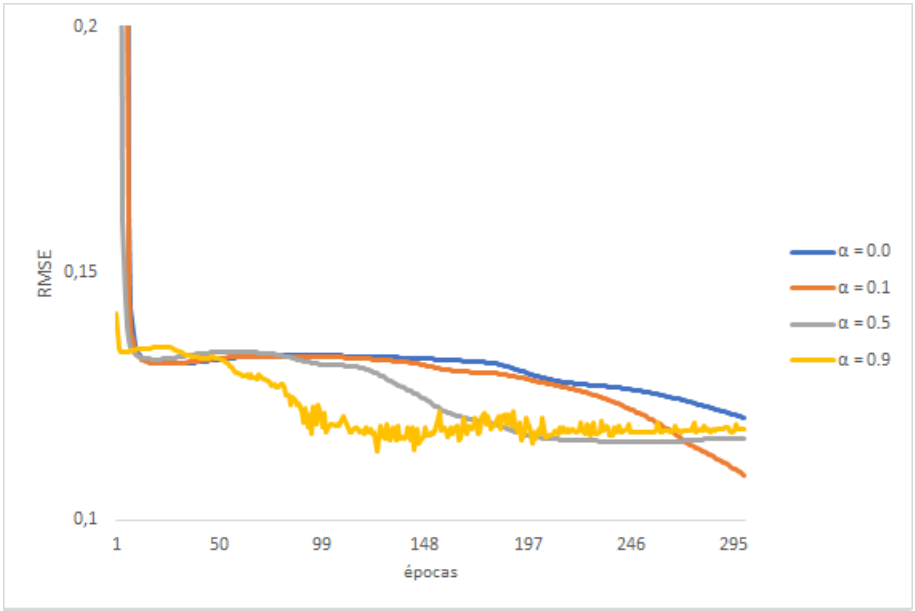


Figura 4.6: Curvas de aprendizagem médias para $\eta = 0.1$ com atributos selecionados pelo método CFS. Fonte: (MERGULHãO et al., 2024).

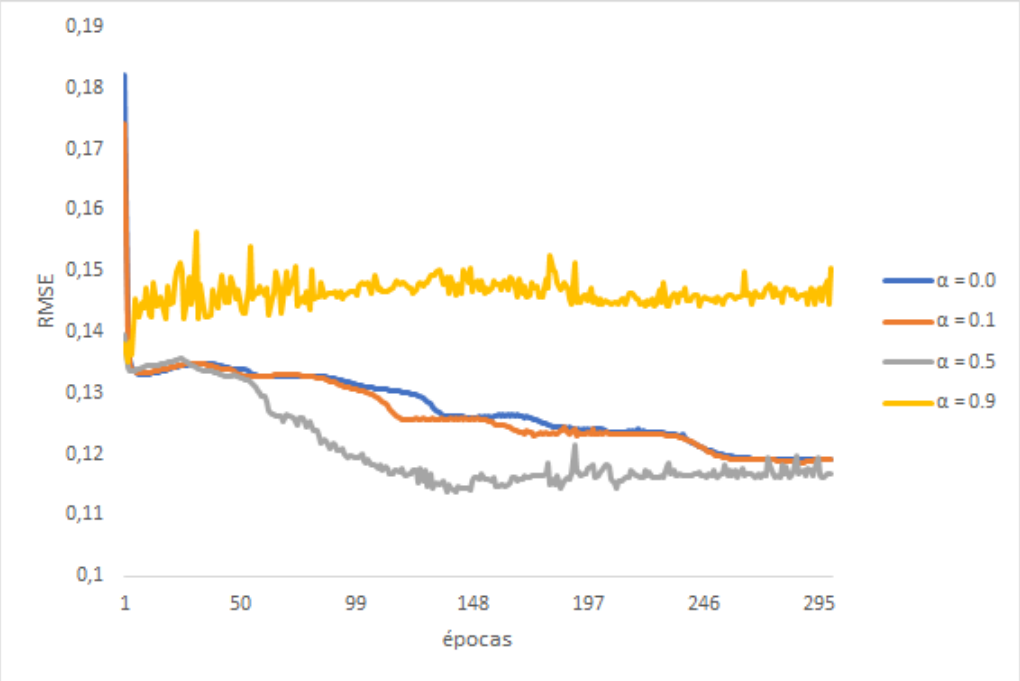


Figura 4.7: Curvas de aprendizagem médias para $\eta = 0.5$ com atributos selecionados pelo método CFS. Fonte: (MERGULHãO et al., 2024).

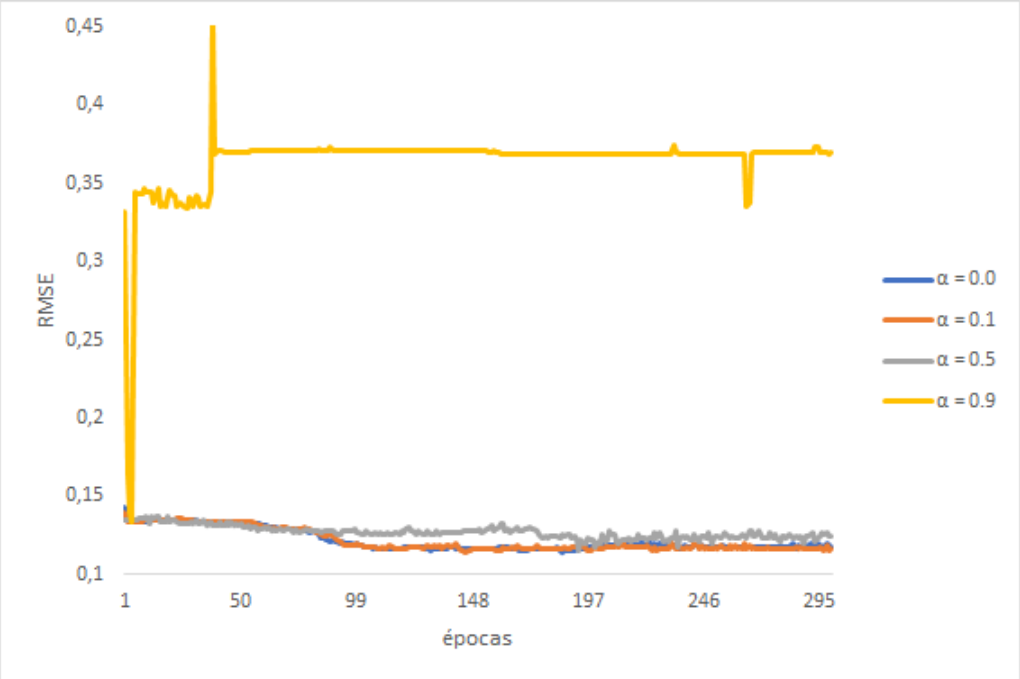


Figura 4.8: Curvas de aprendizagem médias para $\eta = 0.9$ com atributos selecionados pelo método CFS. Fonte: (MERGULHãO et al., 2024).

Observa-se, a partir dos resultados, ilustrados nas Figuras 4.5 a 4.8, que em geral, para um

valor pequeno (0.01), atribuído à taxa de aprendizagem ($\eta$), resulta em uma convergência mais lenta. O uso das taxas de aprendizagem $\eta = 0.5$ e $\eta = 0.9$, sendo $\alpha = 0.9$, observam-se oscilações bastante ruidosas no erro médio quadrado durante o processo de aprendizagem, ocasionando um valor mais alto do erro na convergência da rede neural, sendo ambos efeitos indesejáveis. A Figura 4.9 apresenta as melhores curvas de aprendizagem para cada grupo das curvas ilustradas nas Figuras 4.5 a 4.8, para determinar a melhor curva de aprendizagem global.

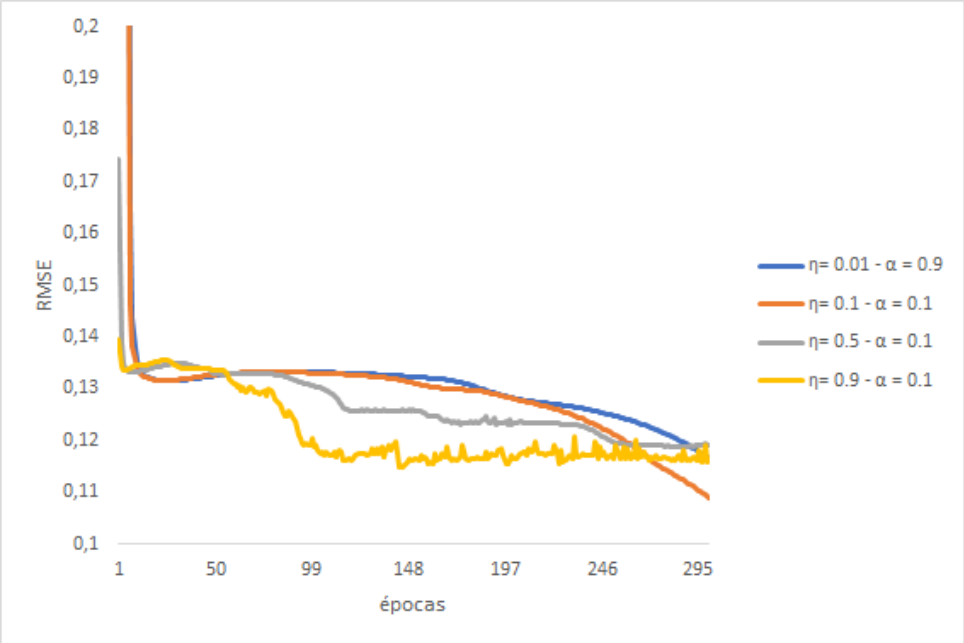


Figura 4.9: Melhores curvas de aprendizagem selecionadas utilizando método CFS. Fonte: (MERGULHãO et al., 2024).

Observa-se, a partir da Figura 4.9, que a curva melhor acentuada foi a que possui parâmetro da taxa de aprendizagem $\eta = 0.1$ e constante de momentum $\alpha = 0.1$. Nesse contexto, o fato de o erro médio quadrado da curva não variar muito, sugere que os parâmetros são os mais adequados para esse problema. A Tabela 4.1 apresenta um resumo dos parâmetros otimizados para a RNA.

Tabela 4.1: Configuração Final da rede Neural utilizando CFS.

| Parâmetro | Valor |
|---|---|
| Número de camadas | 4 |
| Número de camadas ocultas | 2 |
| Número de neurônios (por cada camada oculta) | 15 |
| Épocas | 300 |
| Função de ativação | Sigmóid |
| Taxa de aprendizado | 0.1 |
| Constante de momentos | 0.1 |

Os dados para obtenção das curvas de convergência podem ser verificados a partir do link: <https://docs.google.com/spreadsheets/d/1IT1h0YbqmS_D6Z1ZCwGKrJGQwfjD0PJV/edit?usp=sharing&ouid=106791200257114486467&rtpof=true&sd=true>

### *4.3.2 Validação da RNA RF-MLP*

As curvas de aprendizagem médias da RNA RF-MLP são ilustradas nas Figuras 4.10 a 4.13. Agrupadas pela taxa de aprendizagem ($\eta$) em função das constantes de momentum ($\alpha$).

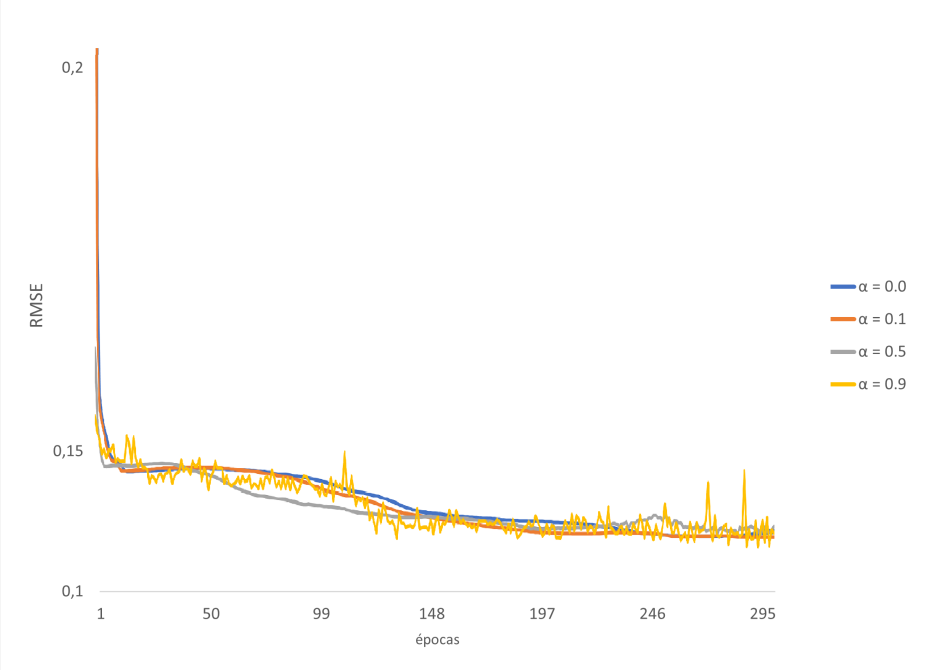


Figura 4.10: Curvas de aprendizagem médias para $\eta = 0.01$ com atributos selecionados pelo método Random Forest.

A partir dos resultados, ilustrados na figura 4.13, observa-se que para um valor de taxa de aprendizagem $\eta = 0.9$, resultou em muitas oscilações no erro médio quadrado durante o processo de aprendizagem ou até não convergência. A combinação com o $\alpha = 0.9$ resultou num elevado valor de erro médio quadrático e portanto numa não convergência da rede neural.

Já com a taxa de aprendizagem $\eta = 0.5$, figura 4.12, sendo $\alpha = 0.5$ e 0.9, observam-se muitas oscilações no erro médio quadrado durante o processo de aprendizagem.

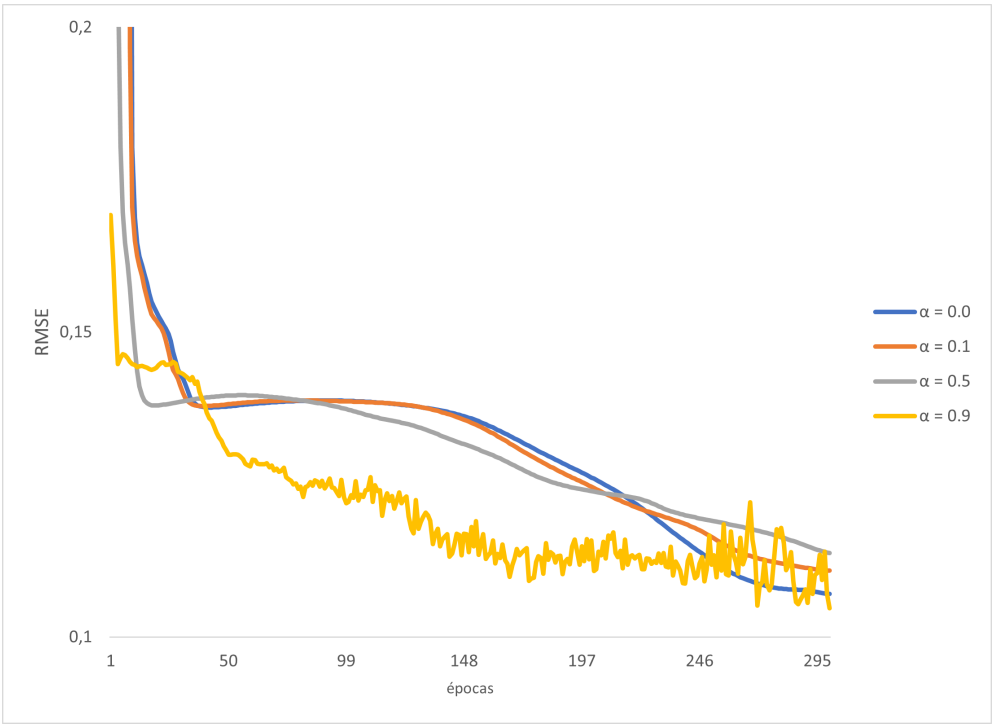


Figura 4.11: Curvas de aprendizagem médias para $\eta = 0.1$ com atributos selecionados pelo método Random Forest.

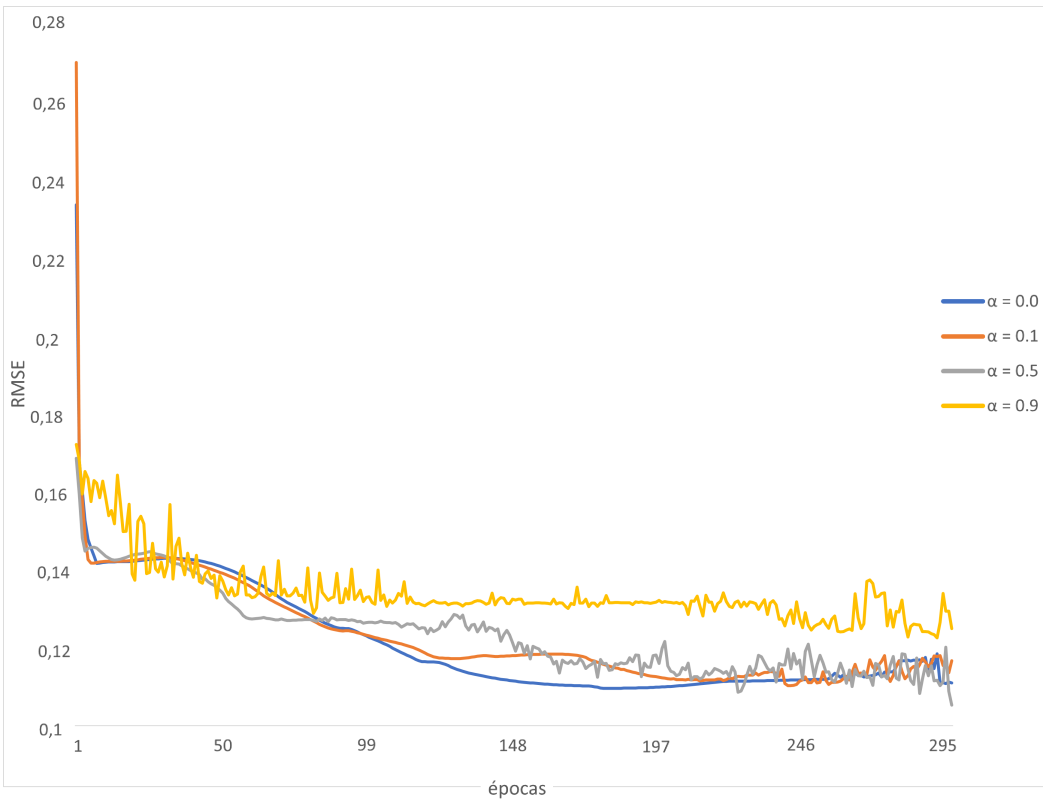


Figura 4.12: Curvas de aprendizagem médias para $\eta = 0.5$ com atributos selecionados pelo método Random Forest.

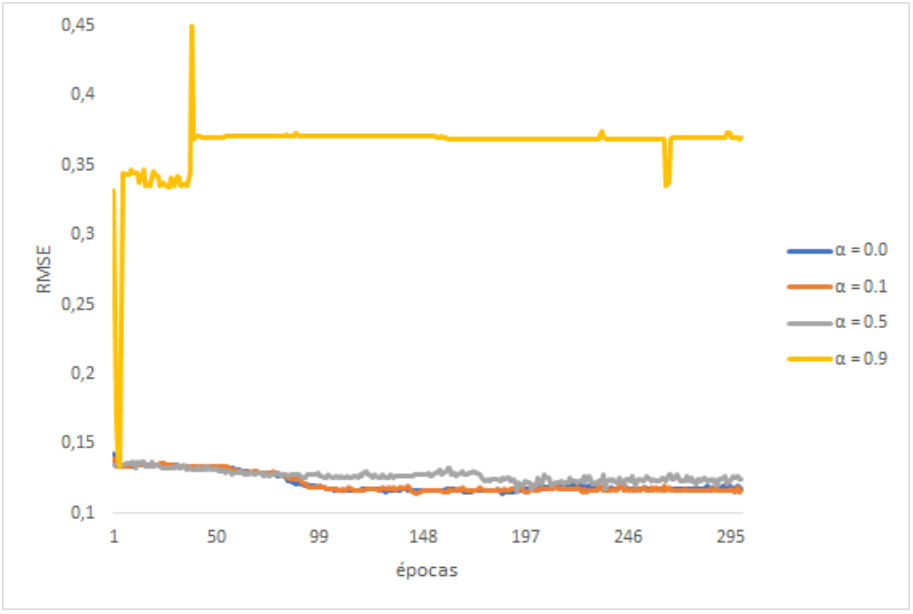


Figura 4.13: Curvas de aprendizagem médias para $\eta = 0.9$ com atributos selecionados pelo método Random Forest.

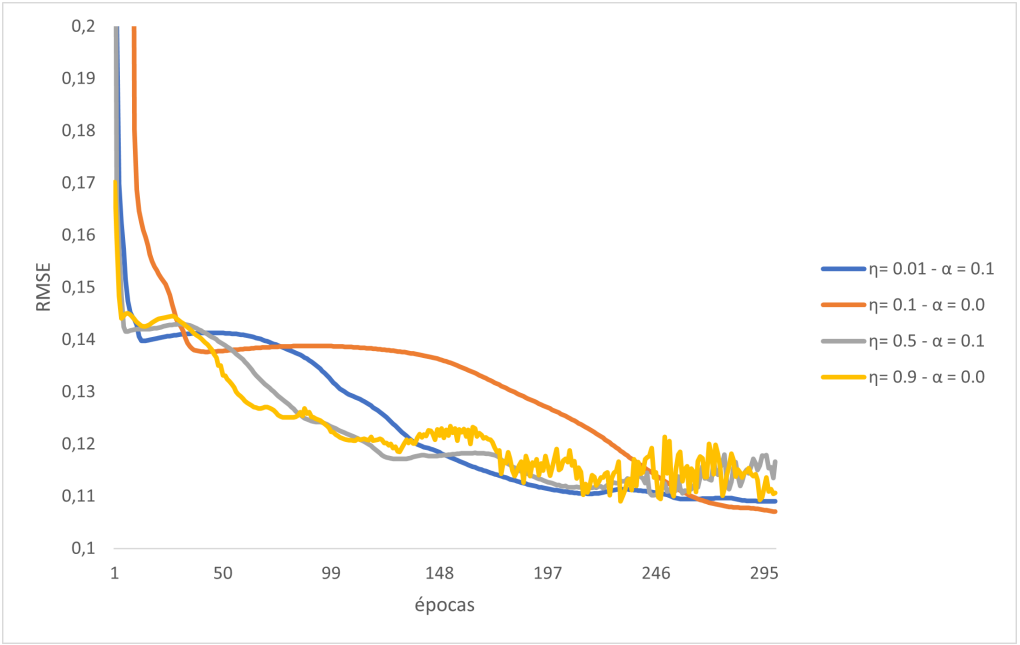


Figura 4.14: Melhores curvas de aprendizagem selecionadas utilizando método Random Forest.

A Figura 4.14 apresenta as melhores curvas de aprendizagem para cada grupo das curvas ilustradas nas Figuras 4.10 a 4.13, para determinar a melhor curva de aprendizagem global. Observa-se, que a curva laranja apresentou menor erro ao final do treinamento, mas a curva azul, com parâmetro da taxa de aprendizagem $\eta = 0.01$ e a constante de momentum $\alpha = 0.0$ foi a curva que apresentou maior uniformidade e estagnou, sugerindo

que os parâmetros são os mais adequados para esse problema. A Tabela 4.2 apresenta um resumo dos parâmetros otimizados para a rede neural artificial RF-MLP.

| Parâmetro | Valor |
|---|---|
| Número de camadas | 4 |
| Número de camadas ocultas | 2 |
| Número de neurônios (por cada camada oculta) | 13 |
| Épocas | 300 |
| Função de ativação | Sigmóid |
| Taxa de aprendizado | 0.01 |
| Constante de momentos | 0.1 |

Tabela 4.2: Configuração Final da rede Neural utilizando Random Forest.

Os dados para obtenção das curvas de convergência podem ser verificados a partir do link: <https://docs.google.com/spreadsheets/d/1GS5-ccoGSJf8zT0U7VLOCUoaSwRE_tRF/edit?usp=sharing&ouid=106791200257114486467&rtpof=true&sd=true>

## *4.4 Avaliação dos resultados das RNAs*

### *4.4.1 Avaliação dos resultados da RNA CFS-MLP*

Nesse contexto, as redes neurais foram avaliadas para determinar a sua precisão e capacidade de generalização com a utilização de um novo conjunto balanceado de dados contendo 1.894 amostras, que representam 2,93% do conjunto de amostras.

Findado o experimento, a RNA demonstrou um percentual de instâncias classificadas corretamente igual a 98,73%. O número de instâncias classificadas corretamente foi de 1.870, contra 24 instâncias classificadas incorretamente.

Ao analisar a matriz de confusão, na Tabela 4.3, observa-se, que a classe 0 (falha), obteve 933 instâncias classificadas corretamente, contra 14 instâncias classificadas incorretamente. Já a classe 1 (funcionamento normal), obteve 937 instâncias classificadas corretamente, contra 10 instâncias classificadas incorretamente.

| Classes | 0 | 1 |
|---|---|---|
| 0 | 933 | 14 |
| 1 | 10 | 937 |

Tabela 4.3: Matriz de Confusão da Classificação de dados da MLP utilizando CFS.

O método apresentou uma precisão média de 98,90% para a classe 0, ou seja, a RNA demostrou boa capacidade de classicação correta de ocorrência de falhas. Consequentemente, a classe 1 apresentou uma precisão media de 98,50% para classicação correta de funcionamento normal do aerogerador conforme apresentado na tabela 4.4.

| Acurácia | Classes | | média |
|---|---|---|---|
| | 0 | 1 | |
| TP RATE | 0,985 | 0,989 | 0,987 |
| FP RATE | 0,011 | 0,015 | 0,013 |
| Precisão | 0,989 | 0,985 | 0,987 |
| Revocação | 0,985 | 0,989 | 0,987 |
| F1-Score | 0,987 | 0,987 | 0,987 |

Tabela 4.4: Resumo estatístico da da rede neural artificial CSF-MLP detalhada por classe e acurácia média.

### *4.4.2 Avaliação dos resultados da RNA RF-MLP*

O mesmo conjunto de dados contendo 1.894 amostras foi utilizado para a avaliação dos resultados da RF-MLP.

A RNA RF-MLP demonstrou um percentual de instâncias classificadas corretamente igual a 97,88%. O número de instâncias classificadas corretamente foi de 1.854, contra 40 instâncias classificadas incorretamente.

Ao analisarmos a matriz de confusão, na Tabela 4.5, observa-se, que a classe 0 (falha), obteve 922 instâncias classificadas corretamente, contra 25 instâncias classificadas incorretamente. Já a classe 1 (funcionamento normal), obteve 932 instâncias classificadas corretamente, contra 15 instâncias classificadas incorretamente.

| Classes | 0 | 1 |
|---|---|---|
| 0 | 922 | 25 |
| 1 | 15 | 932 |

Tabela 4.5: Matriz de Confusão da Classificação de dados da MLP utilizando Random Forest.

O método apresentou uma precisão média de 98,40% para a classe 0, ou seja, a RNA demostrou boa capacidade de classicação correta de ocorrência de falhas. Consequentemente, a classe 1 apresentou uma precisão media de 97,40% para classicação correta de funcionamento normal do aerogerador conforme apresentado na tabela 4.6.

| Acurácia | Classes | | média |
|---|---|---|---|
| | 0 | 1 | |
| TP RATE | 0,974 | 0,984 | 0,979 |
| FP RATE | 0,016 | 0,026 | 0,021 |
| Precisão | 0,984 | 0,974 | 0,979 |
| Revocação | 0,974 | 0,984 | 0,979 |
| F1-Score | 0,979 | 0,979 | 0,979 |

Tabela 4.6: Resumo estatístico da da rede neural artificial RF-MLP detalhada por classe e acurácia média.

## *4.5 Discussão*

Neste estudo, para classificação de estado de saúde, o treinamento, apresentou uma convergência dos erros para valores relativamente baixos e estabilizou ao longo das épocas. É importante ressaltar que o número de épocas foi selecionado empiricamente durante o desenvolvimento dos algoritmos. No entanto, verificou-se que, o número de épocas escolhido foi adequado para alcançar estabilidade no treinamento, sem criar impactos significativos no tempo de treinamento.

O modelo arquitetural desenvolvido para fins de classificação foi fundamentado em um Perceptron Multicamadas (MLP), caracterizado por uma arquitetura de natureza simplificada, com uma camada de saída empregando a função sigmóide.

| Algoritmo | |
|---|---|
| CFS | Random Forest |
| temperatura máxima ambiente | tempo inicial de falha grid side |
| aceleração média eixo y | tempo inicial de trabalho hidráulico |
| corrente mínima da fase A grid side | corrente máxima da fase C grid side |
| angulo médio dos valores mínimos de pitch | tempo final de operação do motor de yaw 2 |
| ângulo mínimo de pitch da pá 3 | ângulo máximo de pitch da pá 2 |
| temperatura máxima do reator grid side 1 | velocidade média do vento |
| aceleração média rms | |

Tabela 4.7: Atributos selecionados pelos métodos CFS e Random Forest

A tabela 4.7, faz menção aos atributos selecionados pelos métodos de seleção de atributos CFS e Random Forest, os quais reduziram consideravelmente o número de atributos do banco de dados original para 7 e 6 atributos respectivamente, porém, curiosamente, sem interseções, uma vez que ambos os métodos utilizam critérios diferentes. Há atributos coletados de sensores que medem as mesmas grandezas mas com referenciais e fases diferentes. Como é o caso do ângulo mínimo de pitch da pá 3 e ângulo máximo de pitch da pá 2, assim como corrente mínima da fase A grid side e corrente máxima da fase C grid side.

O conjunto de dados selecionados pelo Random Forest podem conter atributos que possuem alguns valores que são fortemente preditivos localmente (em uma pequena área do espaço de amostras), enquanto os valores restantes podem ter baixo poder preditivo, ou ser parcialmente correlacionado com outros recursos. Uma vez que o CFS avalia o mérito de um atributo de forma global, sua tendência para pequenos subconjuntos de atributos pode impedir que algum atributo (selecionado pelo RF) seja incluído, principalmente se for preterido por outros atributos mais preditivos.

O método CFS filtrou dois atributos de temperatura dentre do subconjunto de 7, o que nos faz relembrar o capítulo 2, quando na literatura foi abordada a importância do monitoramento desta grandeza para detecção efetiva de falhas em aerogeradores do tipo PMSG.

As RNA configuradas a partir dos dois métodos atingiram bons níveis de precisão. A RNA com atributos extraídos pelo CFS apresentou melhor precisão para o conjunto de dados. O resumo da acurácia das RNAs estão representados na tabela 4.8, assim como um gráfico comparativo na figura 4.15, que mostra a taxa de instâncias classificadas corretamente e incorretamente, através das métricas estatísticas utilizadas no estudo.

| Algoritmo | Sensores | Acuracia | Precisão | |
|---|---|---|---|---|
| | | | Falha | Normal |
| CFS-MLP | 7 | 98.7% | 98.9% | 98.5% |
| RF-MLP | 6 | 97.9% | 98.4% | 97.4% |

Tabela 4.8: Resumo da acurácia das RNAs

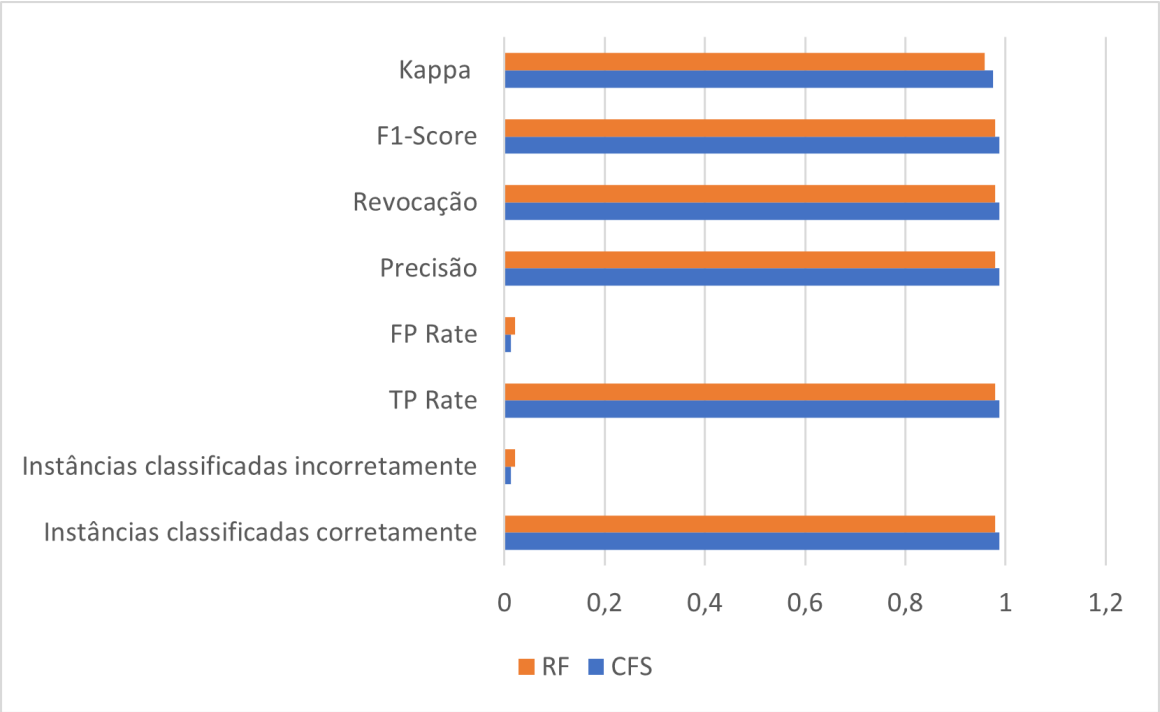


Figura 4.15: Comparativo entre resultados das redes neurais CFS-MLP e Random Forest-MLP.

Capítulo Cinco

# Considerações finais

A Política energética de implantação de fontes de geração de energia levam em consideração uma série de fatores e impactos climáticos, econômicos, culturais, ambientais, tecnológicos sociais. No quesito disponibilidade, a manutenção tem papel preponderante para a sustentabilidade da solução e melhor fonte energética eleita para cada situação.

A energia renovável eólica também tem seus estais climáticos, tecnológicos e econômicos. Reduzindo o downtime, significa dizer que em com um maior controle e monitoramento disponível para aerogeradores de tecnologia específica pode-se atingir níveis ótimos de disponibilidade.

Atualmente vivenciamos uma era em que, com a utilização de variados sensores e tipos de monitoramento, grandes volumes de dados e de altas dimensões estão sendo gerados, portanto, mais do que de ser um desafio processá-los, temos que ter estes dados como nossos aliados, saber extrair as melhores informações que estes dados podem nos transmitir.

Quando trabalhamos com um banco de dados nesses moldes, extenso e com grande quantidade de informações e atributos irrelevantes, temos a oportunidade de utilizar técnicas para definição de quais atributos utilizar, tem-se um propósito duplo, selecionar atributos relevantes e simultaneamente descartar atributos redundantes.

Existem métodos para seleção de recursos que utilizam medidas de distância, otimização de ganho de informação, entropia, correlação dentre outros e até variação de mais de um método combinado com outros. A partir daí, permite-se que seja obtido um conjunto que melhor represente os variados estados do conjunto original. Neste sentido, este estudo utiliza os métodos CFS subset evaluator e Random Forests para seleção de recursos para melhorar a precisão da rede neural artificial, os quais mostraram-se eficientes, sem necessidade de utilização de métodos de seleção de recursos mais sofisticados e requintados.

Algumas etapas essenciais neste estudo foram implementadas: coleta do conjunto de dados do sistema SCADA, pré-processamento primário e secundário de dados, seleção de atributos mais relevantes, configuração das arquiteturas das rede neutrais artificiais (RNAs) Multilayer Perceptron - MLP para classificação das condições de falha e normalidade do aerogerador e validação do processo de seleção de atributos.

Em busca dos melhores parâmetros para a convergência da RNAs, foi realizado treinamento usando o método de validação cruzada com 10 partições. O estudo apresentou

curvas de aprendizagem construídas através da combinação de parâmetros de taxa de aprendizagem e constante de momentum buscando encontrar a combinação que possa reduzir as oscilações no erro médio quadrado durante o processo de aprendizagem e avaliar a convergência das redes neurais.

Em seguida, foi realizada uma avaliação de desempenho com base na precisão, tendo como base resultados do teste com um novo conjunto de dados.

Os resultados computacionais relatados no estudo, demonstraram que, na base de dados utilizada, e para o método estudado de melhor acurácia, CFS, as falhas podem ser previstas com uma precisão média de 98,90% para classificação correta de ocorrência de falhas, enquanto a classificação correta de funcionamento normal do gerador apresentou uma precisão média de 98,50%. Tão importante quanto o resultado da MLP, é importante destacar que com a modelagem utilizando apenas sete atributos selecionadas num universo de 282 possíveis, foi possível obter este nível de precisão, o que poderia sugerir num estudo mais aprofundado, com base de dados mais extensa, abrangendo um período de monitoramento também mais extenso para um determinado parque eólico, em quais sensores são indispensáveis para assertivo e econômico monitoramento da saúde das turbinas eólicas deste parque.

## *5.1 Conclusões*

As modelagens de aprendizado de máquina são reféns da qualidade e tratamento dos dados utilizados. Os métodos de seleção de atributos foram capazes de lidar efetivamente com múltiplas entradas, facilmente aplicados sem ajustes complexos de parâmetros e também não foram superajustados ao lidarem com dados em grande escala.

Apesar de os resultados obtidos neste estudo serem promissores, deve-se ter em mente que os subconjuntos selecionados não devem ser tratados como definitivos, mas representam uma boa indicação dos atributos mais importantes de um conjunto de dados.

Deste modo, não se deve considerar que este nível de redução de atributos é válido para todos os aerogeradores tipo direct drive de imãs permanentes, muito menos para toda a gama de aerogeradores fabricados mundialmente. Mas, que com um maior controle e monitoramento e banco de dados disponível e extraído de um parque eólico em estudo, para aerogeradores de tecnologia específica podemos ter suporte utilizando de métodos de seleção de atributos mais relevantes em conjunto com aprendizado de máquina de indução, como uma MLP, otimizando assim os níveis de disponibilidade.

## 5.2 Contribuições

Esta pesquisa é parte integrante do Subprojeto - P4.3, do P&D ANEEL PD-0048-0217/ 2020 - Sistema inteligente com aerogerador integrado às fontes de energia eólica, solar e storage, como plataforma de desenvolvimento visando melhorias contínuas no processo de geração de energia elétrica. O estudo, contribui com resultados de experimentos reais com base de dados do monitoraramento de elementos de um aerogerador de imãs permanentes coletada de sistema de aquisição de dados SCADA, ao longo de mais de dois anos, combinado com técnicas de aprendizado de máquina. Amparando para um melhor monitoramento da condição e saúde de aerogeradores, reduzindo o número de falhas, e dando subsídios para manutenção preditiva, principalmente para turbinas eólicas do tipo PMSG, para a qual há uma escassez de estudos com este direcionamento. Assim, poderá ser utilizado por pesquisadores e comunidade acadêmica para outras implementações.

## 5.3 Atividades Futuras de Pesquisa

A partir dos resultados obtidos e conclusões, é possível um estudo mais aprofundado, com base de dados e período de coleta mais extensos não apenas do aerogerador 18, mas com outros aerogeradores do parque eólico estudado.

Da forma como a as idéias estão estruturadas, para um determinado parque eólico, a partir de uma coleta de dados estruturada pode-se estudar e determinar quais sensores são indispensáveis para um assertivo e econômico monitoramento da saúde das turbinas eólicas deste parque.

- Ampliar a base de dados analisada, buscar informações de mais equipamentos e outros parques eólicos da ELETROBRAS;
- Testar outros modelos para classificação dos dados e métodos e ensembles para seleção de atributos ;
- Estender estudo para detecção de estado de pré-falha de aerogeradores do tipo PMSG;
- A partir de estado de pré-falha, classificar tipos de falhas mapeáveis em aerogeradores do tipo PMSG;
- Estender estudo para determinação de prognóstico e estimação de vida útil remanescente aerogeradores do tipo PMSG;
- Realizar estudo semelhante com base de dados de aerogeradores de tecnologia diferente da de imãs permanentes.

Apêndice A

# Documentos

Aqui foram acrescentados alguns dados auxiliares, mas que são importantes para complementar as informações neste documento.

## *A.1 Falhas com parada ocorridas no aerogerador 18 no período de Agosto/21 até Outubro/23*

| Data da parada por falha | Código e descrição da falha | Componente afetado | Tempo de parada | Componente substituido? | Categoria da falha |
|---|---|---|---|---|---|
| 20/10/2021 | (164)Error_2#Pitch Generation Position Sensor Error (Falha sensor de posição da pá) | SISTEMA PITCH | 1,45 | NÃO | PITCH |
| 21/10/2021 | (164)Error_2#Pitch Generation Position Sensor Error (Falha sensor de posição da pá) | SISTEMA PITCH | 1,45 | NÃO | PITCH |
| 22/10/2021 | (164)Error_2#Pitch Generation Position Sensor Error (Falha sensor de posição da pá) | SISTEMA PITCH | 2,92 | NÃO | PITCH |
| 29/11/2021 | (163)Error_1#Pitch Generation Position Sensor Error (Falha sensor de posição da pá N°01) | SENSOR INDUTIVO PITCH N°01 | 18,75 | SIM | PITCH |
| 26/01/2022 | (84) - Error Gen. Side Capacitor Fuse Feedback Loss (Perda de feedback do fusível do capacitor lateral do gerador) | FUSÍVEL 100A DO CONVERSOR | 1,82 | SIM | CONVERSOR |
| SEM DADOS | (422)Error_Converter IGBT ok Lost (Sinal IGBT do converter desativado) | CONVERSOR | 8,28 | SIM | CONVERSOR |
| 21/04/2022 | (84) - Error Gen. Side Capacitor Fuse Feedback Loss (Perda de feedback do fusível do capacitor lateral do gerador) | FUSÍVEL 100A DO CONVERSOR | 9,28 | SIM | CONVERSOR |
| 26/04/2022 | (84) - Error Gen. Side Capacitor Fuse Feedback Loss (Perda de feedback do fusível do capacitor lateral do gerador) | FUSÍVEL 100A DO CONVERSOR | 14,67 | SIM | CONVERSOR |
| 08/05/2022 | (84) - Error Gen. Side Capacitor Fuse Feedback Loss (Perda de feedback do fusível do capacitor lateral do gerador) | FUSÍVEL 100A DO CONVERSOR | 0,1 | SIM | CONVERSOR |
| 24/06/2022 | (175) - Grid LVRT (Investigar Falha de Low Voltage Ride Through) | FUSÍVEL NH 100A DO CONVERSOR | 14,2 | SIM | CONVERSOR |
| 19/07/2022 | (164)Error_2#Pitch Generation Position Sensor Error (Falha sensor de posição da pá N°02) | PITCH N°02 | 4,42 | SIM | PITCH |
| 04/11/2022 | (442)Error_IGBT_ok Loss (Sinal de IGBT OK desativado) | MODULO IGBT | 3,25 | SIM | CONVERSOR |

| Data da parada por falha | Código e descrição da falha | Componente afetado | Tempo de parada | Componente substituido? | Categoria da falha |
|---|---|---|---|---|---|
| 29/11/2022 | (442)Error_IGBT_ok Loss (Sinal de IGBT OK desativado) | MODULO IGBT | 4,8 | SIM | CONVERSOR |
| 23/12/2022 | (87) Error_Yawing Speed Out of Limit (Erro velocidade de Guinada) | SISTEMA YAW | 19,19 | NÃO | YAW |
| 24/12/2022 | (442)Error_IGBT_ok Loss (Sinal de IGBT OK desativado) | MODULO IGBT | 26,65 | SIM | CONVERSOR |
| 27/12/2022 | (84) - Error Gen. Side Capacitor Fuse Feedback Loss (Perda de feedback do fusível do capacitor lateral do gerador) | FUSÍVEL 100A DO CONVERSOR | 14,35 | SIM | CONVERSOR |
| 27/01/2023 | (447) Error_converter grid side IGBT over current (Sobrecorrente no IGBT lado da rede) | FUSÍVEL 100A DO CONVERSOR | 13,98 | SIM | CONVERSOR |
| 27/01/2023 | (442)Error_IGBT_ok Loss (Sinal de IGBT OK desativado) | MODULO IGBT | 13,98 | SIM | CONVERSOR |
| 04/02/2023 | (442)Error_IGBT_ok Loss (Sinal de IGBT OK desativado) | MODULO IGBT | 5,01 | SIM | CONVERSOR |
| 04/02/2023 | (442)Error_IGBT_ok Loss (Sinal de IGBT OK desativado) | MODULO IGBT | 3,86 | SIM | CONVERSOR |
| 11/02/2023 | (470)Error_Boost Chopper DC Voltage High (Erro de sobretensão no Step up/Link DC) | CHOPPER | 9,75 | NÃO | CONVERSOR |
| 13/02/2023 | (470)Error_Boost Chopper DC Voltage High (Erro de sobretensão no Step up/Link DC) | CHOPPER | 2,41 | NÃO | CONVERSOR |
| 31/08/2023 | (22) Error_Hydraulic System Oil Level Low (Baixo nível de Óleo Grupo Hidrúalico) | GRUPO HIDRÁULICO (NACELLE) | 7,02 | NÃO | GRUPO HIDRÁULICO |
| 09/09/2023 | (174)Error_3#Pitch Generation Position Sensor Error (Falha sensor de posição da pá N°03) | PITCH N°03 | 8,09 | SIM | PITCH |
| 07/09/2023 | (95)Error_Pitch Safty Chain Triggered (Série de emergência das pás ativada) | SISTEMA PITCH | 9,44 | NÃO | PITCH |
| 11/09/2023 | (95)Error_Pitch Safty Chain Triggered (Série de emergência das pás ativada) | SISTEMA PITCH | 4,84 | SIM | PITCH |
| 09/10/2023 | (22) Error_Hydraulic System Oil Level Low (Baixo nível de Óleo Grupo Hidrúalico) | GRUPO HIDRÁULICO (NACELLE) | 14,6 | NÃO | GRUPO HIDRÁULICO |
| 20/10/2023 | (22) Error_Hydraulic System Oil Level Low (Baixo nível de Óleo Grupo Hidrúalico) | GRUPO HIDRÁULICO (NACELLE) | 11,73 | NÃO | GRUPO HIDRÁULICO |

# Referências Bibliográficas


ABEEóLICA. n. Acesso em 04 de abril de 2024, 2024. Disponível em: <https://abeeolica.org.br/>. (document), 2.3

AHMAD, R.; KAMARUDDIN, S. An overview of time-based and condition-based maintenance in industrial application. *Computers  Industrial Engineering*, v. 63, n. 1, p. 135–149, 2012. ISSN 0360-8352. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0360835212000484>. 2.5

ALLA, S.; ADARI, S. K. *Beginning anomaly detection using python-based deep learning*. [S.l.]: Springer, 2019. 2.6.1

ATAMURADOV, V.; MEDJAHER, K.; DERSIN, P.; LAMOUREUX, B.; ZERHOUNI, N. Prognostics and health management for maintenance practitioners-review, implementation and tools evaluation. *International Journal of Prognostics and Health Management*, PHM Society, v. 8, n. 3, p. 1–31, 2017. 1

AYODELE, T. Types of machine learning algorithms. In: _____. [S.l.: s.n.], 2010. ISBN 978-953-307-034-6. 2.6.2

BAKDI, A.; KRISTENSEN, N. B.; STAKKELAND, M. Multiple instance learning with random forest for event logs analysis and predictive maintenance in ship electric propulsion system. *IEEE Transactions on Industrial Informatics*, v. 18, n. 11, p. 7718–7728, 2022. 2.6.2.2

BAKIR, I.; YILDIRIM, M.; URSAVAS, E. An integrated optimization framework for multi-component predictive analytics in wind farm operations  maintenance. *Renewable and Sustainable Energy Reviews*, v. 138, p. 110639, 2021. ISSN 1364-0321. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1364032120309230>. 2.5

BARBOSA, C. F. de O.; PINHO, J. T.; GALHARDO, M. A. B.; PEREIRA, E. J. da S. Conceitos sobre sistemas híbridos de energia para produção de eletricidade. In: *Anais Congresso Brasileiro de Energia Solar-CBENS*. [S.l.: s.n.], 2016. p. 1–8. 2.2

BARBOSA, N. *Estudo preliminar para desenvolvimento de um sistema de manutenção em turbinas eólicas de uma planta híbrida, aplicando técnicas de IA, para clusterizar e classificar dados.* Dissertação (Mestrado em Modelagem Computacional e Tecnologia Industrial) — Centro Universitário SENAI CIMATEC, Salvador, 2023. (document), 3.5

BARROS, C. M. V.; BARROS, L. S. Modeling and analysis of stator interturn faults in permanent magnet synchronous machine. In: *2017 IEEE Power  Energy Society General Meeting*. [S.l.: s.n.], 2017. p. 1–5. 2.4

BERETTA, M.; VIDAL, Y.; SEPULVEDA, J.; PORRO, O.; CUSIDó, J. Improved ensemble learning for wind turbine main bearing fault diagnosis. *Applied Sciences*, v. 11, n. 16, 2021. ISSN 2076-3417. Disponível em: <https://www.mdpi.com/2076-3417/11/16/7523>. 2.7

BERRAR, D. Cross-validation. In: _____. [S.l.: s.n.], 2018. ISBN 9780128096338. 2.6.4.6, 3.2.4

BONELLI, A. F. *Modelagem e simulação de unidade eólica para estudos de indicadores de qualidade da energia elétrica.* Dissertação (Mestrado em Ciências.) — Universidade Federal de Uberlândia, Uberlândia/MG, 2010. (document), 2.5, 2.6

BREIMAN, L. Random forests. *Machine Learning*, v. 45, p. 5–32, 10 2001. 1, 3.2.2

BURTON, T.; JENKINS, N.; SHARPE, D.; BOSSANYI, E. *Wind energy handbook.* [S.l.]: John Wiley & Sons, 2011. 2.2

CALVERT, K. *Energy and Society.* Second edition. Oxford: Elsevier, 2015. 615-620 p. ISBN 978-0-08-097087-5. Disponível em: <https://www.sciencedirect.com/science/article/pii/B9780080970868910018>. 2.1

CHAMBY-DIAZ, J. C.; RECAMONDE-MENDOZA, M.; BAZZAN, A. L. Dynamic correlation-based feature selection for feature drifts in data streams. In: IEEE. *2019 8th Brazilian Conference on Intelligent Systems (BRACIS).* [S.l.], 2019. p. 198–203. (document), 2.10

CHATTERJEE, J.; DETHLEFS, N. Scientometric review of artificial intelligence for operations maintenance of wind turbines: The past, present and future. *Renewable and Sustainable Energy Reviews*, v. 144, p. 111051, 2021. ISSN 1364-0321. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1364032121003403>. 1, 2.5

CHEN, H.; LIU, H.; CHU, X.; LIU, Q.; XUE, D. Anomaly detection and critical scada parameters identification for wind turbines based on lstm-ae neural network. *Renewable Energy*, v. 172, p. 829–840, 2021. ISSN 0960-1481. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0960148121004341>. 2.7, 2.7

COUTO, A.; FERREIRA, P.; ESTANQUEIRO, A. Centrais híbridas: caracterização da complementaridade eólica e solar fotovoltaica em portugal. In: LNEG-LABORATÓRIO NACIONAL DE ENERGIA E GEOLOGIA. *CIES2020-XVII Congresso Ibérico e XIII Congresso Ibero-americano de Energia Solar.* [S.l.], 2020. p. 57–65. 2.2

CRESESB. *Energia eólica: princípios e aplicações. [S.l.].* [S.l.]: Editora, 2008. (document), 2.4

CUTLER, A.; CUTLER, D.; STEVENS, J. Random forests. In: _____. [S.l.: s.n.], 2011. v. 45, p. 157–176. ISBN 978-1-4419-9325-0. 2.6.2.2

DANGETI, P. *Statistics for machine learning.* [S.l.]: Packt Publishing Ltd, 2017. 2.6.2.2

DYKES, K. et al. *Opportunities for research and development of hybrid power plants.* [S.l.], 2020. 2.2

ELIJORDE, F.; KIM, S.; LEE, J. A wind turbine fault detection approach based on cluster analysis and frequent pattern mining. *KSII Transactions on Internet and Information Systems*, v. 8, p. 664–677, 02 2014. 1

EPE. Nota técnica : Usinas híbridas no contexto do planejamento energético. *N. EPE-DEE-NT-029/2019-r0*, 2019. 2.2

FISCHER-KOWALSKI, M.; SCHAFFARTZIK, A. Energy availability and energy sources as determinants of societal development in a long- term perspective. 2015. 2.1

GERON, A. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. 2nd. ed. [S.l.]: O'Reilly Media, Inc., 2019. ISBN 1492032646. (document), 1, 2.6, 2.9, 2.6, 2.6.1, 2.6.2, 2.6.2.2, 2.6.3, 2.6.4.1, 2.6.4.4, 2.17, 3

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep Learning*. [S.l.]: MIT Press, 2016. (Adaptive computation and machine learning). ISBN 9780262035613. 2.6, 2.6.1, 2.6.2.2, 2.6.3

GRIFFITH, S. *Energy and Society: Toward a Sustainable Future*. Cham: Springer International Publishing, 2022. 181–203 p. ISBN 978-3-030-99031-2. Disponível em: <https://doi.org/10.1007/978-3-030-99031-2_9>. 1, 2.1

GRUS, J. *Data Science do Zero. Noções Fundamentais com Python*. 2nd. ed. [S.l.]: O'Reilly Media, Inc., 2021. 2.6

HALL, M. A. *Correlation-based feature subset selection for machine learning*. Tese (Thesis submitted in partial fulfillment of the requirements of the degree of Doctor of Philosophy.) — University of Waikato, Hamilton - New Zealand, 1999. 1, 2.6.2.1, 3.2.2

HAN, H.; YANG, D. Correlation analysis based relevant variable selection for wind turbine condition monitoring and fault diagnosis. *Sustainable Energy Technologies and Assessments*, v. 60, p. 103439, 2023. ISSN 2213-1388. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2213138823004320>. 2.6.2.1, 2.7, 3

HAN, J.; KAMBER, M.; PEI, J. *Data Mining: Concepts and Techniques*. 3. ed. Amsterdam: Morgan Kaufmann, 2011. (Morgan Kaufmann Series in Data Management Systems). ISBN 978-0-12-381479-1. Disponível em: <http://www.sciencedirect.com/science/book/9780123814791>. 3.2.3.2

HAYKIN, S. S. *Neural networks and learning machines*. Third. Upper Saddle River, NJ: Pearson Education, 2009. (document), 2.6.3, 2.6.4.1, 2.13, 2.6.4.1, 2.14, 2.6.4.1, 2.15, 2.6.4.1, 2.6.4.1, 2.6.4.2, 2.16, 2.6.4.4, 2.6.4.5, 3.2.3.1

HOSSEINPOUR-ZARNAQ, M.; OMID, M.; BIABANI-AGHDAM, E. Fault diagnosis of tractor auxiliary gearbox using vibration analysis and random forest classifier. *Information Processing in Agriculture*, Elsevier, v. 9, n. 1, p. 60–67, 2022. 2.6.2.1

INGOLE, O. et al. Investigation of different regression models for the predictive maintenance of aircraft's engine. In: *2022 Fourth International Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT)*. [S.l.: s.n.], 2022. p. 1–6. 2.6.2.2

IZBICKI, R.; SANTOS, T. M. dos. *Aprendizado de máquina: uma abordagem estatística*. [S.l.: s.n.], 2020. ISBN 978-65-00-02410-4. 2.6.4, 2.6.4.1

JO, T. Deep learning foundations /. Springer International Publishing AG,, Cham :, p. 1 online resource (xx, 426 pages), 2023. Disponível em: <http://library.usi.edu/record/1472169>. 2.6.1, 2.6.4.1, 2.6.4.2, 2.6.4.2

KAREGOWDA, A. G.; JAYARAM, M.; MANJUNATH, A. Feature subset selection using cascaded ga & cfs: a filter approach in supervised learning. *International Journal of Computer Applications*, Citeseer, v. 23, n. 2, p. 1–10, 2011. 2.6.2, 3.2.2

KHAN, P. W.; BYUN, Y.-C. A review of machine learning techniques for wind turbine's fault detection, diagnosis, and prognosis. *International Journal of Green Energy*, Taylor & Francis, v. 21, n. 4, p. 771–786, 2023. Disponível em: <https://doi.org/10.1080/15435075.2023.2217901>. 1, 2.5

KONG, K.; DYER, K.; PAYNE, C.; HAMERTON, I.; WEAVER, P. M. Progress and trends in damage detection methods, maintenance, and data-driven monitoring of wind turbine blades – a review. *Renewable Energy Focus*, v. 44, p. 390–412, 2023. ISSN 1755-0084. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1755008422000606>. 2.7

KOTSIANTIS, S. Supervised machine learning: A review of classification techniques. *Informatica (Slovenia)*, v. 31, p. 249–268, 01 2007. 1

KUCHENBECKER, W. E.; ROSA, W. M.; TEIXEIRA, J. C. Pmsg fault identification applied to wind power. *Electric Power Systems Research*, Elsevier, v. 165, p. 102–109, 2018. 2.5

KUCHENBECKER, W. E.; TEIXEIRA, J. Procedures to determine inductances of permanent magnet generators. *IEEE Latin America Transactions*, IEEE, v. 13, n. 8, p. 2646–2652, 2015. 2.4

LAWSON, J. Which wind turbine generator will win? Artigo de imprensa, 2012. Disponível em: <https://www.renewableenergyworld.com/wind-power/which-wind-turbine-generator-will-win/>. 2.4

LEE, H. D. *Seleção de atributos importantes para extração de conhecimento de bases de dados*. Tese (Doutor em Ciências – Ciências de Computação e Matemática Computacional.) — IC C-USP, São Carlos - SP, 2005. 2.6.2, 3.2.2

LEE, J.; ZHAO, F. Gwec global wind report. *Glob Wind Energy Counc*, v. 75, 2022. 2.2

LEITE, G. d. N. P.; ARAÚJO, A. M.; ROSAS, P. A. C. Prognostic techniques applied to maintenance of wind turbines: a concise and specific review. *Renewable and Sustainable Energy Reviews*, Elsevier, v. 81, p. 1917–1925, 2018. 1, 2.5, 2.5, 2.7

MARTI-PUIG, P.; BLANCO-M, A.; CáRDENAS, J. J.; CUSIDó, J.; SOLé-CASALS, J. Effects of the pre-processing algorithms in fault diagnosis of wind turbines. *Environmental Modelling Software*, v. 110, p. 119–128, 2018. ISSN 1364-8152. Special Issue on Environmental Data Science and Decision Support: Applications in Climate Change and the Ecological Footprint. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1364815217302104>. 2.5

MELO, E. C. d. S.; ARAGÃO, M. R. d. S.; CORREIA, M. d. F. Regimes do vento à superfície na área de petrolina, submédio são francisco. *Revista Brasileira de Meteorologia*, SciELO Brasil, v. 29, p. 229–241, 2014. 2.2

MENDONçA, M. O.; NETTO, S. L.; DINIZ, P. S.; THEODORIDIS, S. Chapter 13 - machine learning: Review and trends. In: DINIZ, P. S. (Ed.). *Signal Processing and Machine Learning Theory*. Academic Press, 2024. p. 869–959. ISBN 978-0-323-91772-8. Disponível em: <https://www.sciencedirect.com/science/article/pii/B9780323917728000193>. 2.6

MERGULHãO, H. G. et al. Fault detection in wind turbines: a supervised learning approach with multilayer perceptron neural network: Detecção de falhas em turbinas eólicas: uma abordagem de aprendizagem supervisionada com rede neural perceptron multicamadas. *Concilium*, v. 24, p. 348–361, mar. 2024. Disponível em: <https://clium.org/index.php/edicoes/article/view/2951>. (document), 4.3, 4.5, 4.6, 4.7, 4.8, 4.9

MORALES, R.; MÉNDEZ, J. *Inteligencia artificial. Técnicas, métodos y aplicaciones*. [S.l.]: McGraw-Hill Interamericana de España S.L., 2008. ISBN 9788448156183. 2.6.2, 2.6.3, 2.6.4, 2.6.4.1

MUNGUBA, C. F. de L. et al. Ensemble learning framework for fleet-based anomaly detection using wind turbine drivetrain components vibration data. *Engineering Applications of Artificial Intelligence*, Elsevier, v. 133, p. 108363, 2024. 2.5, 2.7

OLIVEIRA, J. P. J. *Geradores síncronos a imãs permanentes aplicados a aerogeradores: modelagem, obtenção de parâmetros e validação laboratorial.* Dissertação (Mestrado em Energia Elétrica) — Universidade de Brasília, Brasília/DF, 2018. 2.4

OSISANWO, F. et al. Supervised machine learning algorithms: Classification and comparison. *International Journal of Computer Trends and Technology*, v. 48, p. 128–138, 2017. Disponível em: <https://api.semanticscholar.org/CorpusID:55362795>. 2.6.1

PAL, K.; PATEL, B. V. Data classification with k-fold cross validation and holdout accuracy estimation methods with 5 different machine learning techniques. p. 83–87, 2020. 3.2.4

PALMA-MENDOZA, R.-J.; MARCOS, L. de; RODRIGUEZ, D.; ALONSO-BETANZOS, A. Distributed correlation-based feature selection in spark. *Information Sciences*, v. 496, p. 287–299, 2019. ISSN 0020-0255. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0020025518308697>. 2.6.2.1, 2.6.2.1

PANDIT, R. K.; INFIELD, D. Scada-based wind turbine anomaly detection using gaussian process models for wind turbine condition monitoring purposes. *IET Renewable Power Generation*, Wiley Online Library, v. 12, n. 11, p. 1249–1255, 2018. 1

PECHT, M. G.; KANG, M. *Prognostics and Health Management of Electronics. 1st ed.* [S.l.]: Wiley, 2018. (document), 2.5, 2.7

PINTO, M. *Fundamentos de energia eólica Brasil.* Rio de Janeiro: LTC, 2013. (document), 2.2, 2.4, 2.5, 2.6, 2.4, 2.5

POHAN, F.; SAPUTRA, I.; TUA, R. Scheduling preventive maintenance to determine maintenance actions on screw press machine. *urnal Riset Ilmu Teknik, 1(1), 1–14.*, 2023. 2.5

QIAO, W.; LU, D. A survey on wind turbine condition monitoring and fault diagnosis—part ii: Signals and signal processing methods. *IEEE Transactions on Industrial Electronics*, IEEE, v. 62, n. 10, p. 6546–6557, 2015. 1

RASCHKA, S.; MIRJALILI, V. *Python Machine Learning.* Second. Livery Place 35 Livery Street Birmingham B3 2PB, UK: Packt Publishing Ltd., 2017. 2.6.3, 2.6.4

RASCHKA, S.; MIRJALILI, V. *Python machine learning: Machine learning and deep learning with Python, scikit-learn, and TensorFlow 2*. [S.l.]: Packt publishing ltd, 2019. 2.6.2.2, 2, 2.6.4, 2.6.4.1

ROCHA, A. V.; RODRIGUES, L. K. d. O.; OLIVEIRA, O. C.; JACOMINI, R. V. *Fundamentos de Energia Eólica*. [S.l.]: Editora IFRN, 2023. v. 1. 2.4, 2.4

RUIZ, A. E.; MÁRQUEZ, A. C.; CASTELLANO, E.; FERNÁNDEZ, J. F. G. A dynamic opportunistic maintenance model to maximize energy-based availability while reducing the life cycle cost of wind farms. *Value Based and Intelligent Asset Management: Mastering the Asset Management Transformation in Industrial Plants and Infrastructures*, Springer, p. 259–287, 2020. 2.5

RUSSELL, S.; NORVIG, P. *Artificial Intelligence: A Modern Approach*. 3. ed. [S.l.]: Prentice Hall, 2010. 2.6.4.4, 2.6.4.6

SÁ, B. A.; BARROS, C. M.; SIEBRA, C. A.; BARROS, L. S. A multilayer perceptron-based approach for stator fault detection in permanent magnet wind generators. In: IEEE. *2019 IEEE PES Innovative Smart Grid Technologies Conference-Latin America (ISGT Latin America)*. [S.l.], 2019. p. 1–6. 2.5

SAARI, P.; EEROLA, T.; LARTILLOT, O. Generalizability and simplicity as criteria in feature selection: Application to mood classification in music. *IEEE Transactions on Audio, Speech, and Language Processing*, v. 19, n. 6, p. 1802–1812, 2011. 2.6.2.1

SAIDI, L.; Ben Ali, J.; BENBOUZID, M.; BECHHOFER, E. An integrated wind turbine failures prognostic approach implementing kalman smoother with confidence bounds. *Applied Acoustics*, v. 138, p. 199–208, 2018. ISSN 0003-682X. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0003682X18301063>. 2.5

SANTOLAMAZZA, A.; DADI, D.; INTRONA, V. A data-mining approach for wind turbine fault detection based on scada data analysis using artificial neural networks. *Energies*, MDPI, v. 14, n. 7, p. 1845, 2021. 1, 2.7

SI, Y.; QIAN, L.; MAO, B.; ZHANG, D. A data-driven approach for fault detection of offshore wind turbines using random forests. In: IEEE. *IECON 2017-43rd Annual Conference of the IEEE Industrial Electronics Society*. [S.l.], 2017. p. 3149–3154. 1, 2.5, 2.6.2.2

SREENATHA, M.; MALLIKARJUNA, P. A fault diagnosis technique for wind turbine gearbox: An approach using optimized blstm neural network with undercomplete autoencoder. *Engineering, Technology & Applied Science Research*, v. 13, n. 1, p. 10170–10174, 2023. 2.7, 2.7

TANG, B.; SONG, T.; LI, F.; DENG, L. Fault diagnosis for a wind turbine transmission system based on manifold learning and shannon wavelet support vector machine. *Renewable Energy*, Elsevier, v. 62, p. 1–9, 2014. 1

TANG, M. et al. An improved lightgbm algorithm for online fault detection of wind turbine gearboxes. *Energies*, v. 13, n. 4, 2020. ISSN 1996-1073. Disponível em: <https://www.mdpi.com/1996-1073/13/4/807>. 2.7

TAUTZ-WEINERT, J.; WATSON, S. J. Using scada data for wind turbine condition monitoring – a review. *IET Renewable Power Generation*, v. 11, n. 4, p. 382–394, 2017. Disponível em: <https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/iet-rpg.2016.0248>. 1

TAVNER, P. J.; BUSSEL, G. J. W. van; SPINATO, F. Machine and converter reliabilities in wind turbines. In: . [s.n.], 2006. Disponível em: <https://api.semanticscholar.org/CorpusID:55566495>. (document), 2.5, 2.8

TURNBULL, A.; CARROLL, J.; MCDONALD, A. Combining scada and vibration data into a single anomaly detection model to predict wind turbine component failure. *Wind Energy*, v. 24, n. 3, p. 197–211, 2021. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1002/we.2567>. 1

URRY, J. The problem of energy. *Theory, Culture  Society*, 2014. 2.1

VATHOOPAN, M.; JOHNY, M.; ZOITL, A.; KNOLL, A. Modular fault ascription and corrective maintenance using a digital twin. *IFAC-PapersOnLine*, v. 51, n. 11, p. 1041–1046, 2018. ISSN 2405-8963. 16th IFAC Symposium on Information Control Problems in Manufacturing INCOM 2018. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2405896318315982>. 2.5

VELMURUGAN, T.; ANURADHA, C. Performance evaluation of feature selection algorithms in educational data mining. *Performance Evaluation*, v. 5, n. 02, 2016. 2.6.2.1

VIAN, Â.; TAHAN, C. M. V.; AGUILAR, G. J. R.; GOUVEA, M. R.; GEMIGNANI, M. M. F. *Energia Eólica: Fundamentos Tecnologia e Aplicaçoes*. [S.l.]: Editora Blucher, 2021. 1, 2.3, 2.3, 2.4

VINH, T. Q.; HUY, N. T. Predictive maintenance iot system for industrial machines using random forest regressor. In: *2022 International Conference on Advanced Computing and Analytics (ACOMPA)*. [S.l.: s.n.], 2022. p. 86–91. 2.6.2.2

WANG, Z.; LIU, C. Wind turbine condition monitoring based on a novel multivariate state estimation technique. *Measurement*, Elsevier, v. 168, p. 108388, 2021. 1, 2.7

WANG, Z.; WANG, Y.; JI, Z. *Advances in fault detection and diagnosis using filtering analysis*. [S.l.]: Springer, 2022. 2.5

WENYI, L.; ZHENFENG, W.; JIGUANG, H.; GUANGFENG, W. Wind turbine fault diagnosis method based on diagonal spectrum and clustering binary tree svm. *Renewable Energy*, Elsevier, v. 50, p. 1–6, 2013. 1

WWEA, W. W. E. A. World energy report. 2022. Disponível em: <https://wwindea.org/wp-content/uploads/2023/03/WWEA_WPR2022.pdf>. (document), 2.3, 2.2

XIANG, L.; WANG, P.; YANG, X.; HU, A.; SU, H. Fault detection of wind turbine based on scada data analysis using cnn and lstm with attention mechanism. *Measurement*, Elsevier, v. 175, p. 109094, 2021. 2.7

XIAO, C.; LIU, Z.; ZHANG, T.; ZHANG, X. Deep learning method for fault detection of wind turbine converter. *Applied Sciences*, MDPI, v. 11, n. 3, p. 1280, 2021. 1, 2.7

YANG, W.; TAVNER, P.; WILKINSON, M. Condition monitoring and fault diagnosis of a wind turbine synchronous generator drive train. *IET Renewable Power Generation*, IET, v. 3, n. 1, p. 1–11, 2009. 2.5

YANG, W.; TAVNER, P. J.; CRABTREE, C. J.; FENG, Y.; QIU, Y. Wind turbine condition monitoring: technical and commercial challenges. *Wind Energy*, v. 17, n. 5, p. 673–693, 2014. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1002/we.1508>. 1

ZHANG, S.; ROBINSON, E.; BASU, M. Wind turbine predictive fault diagnostics based on a novel long short-term memory model. *Algorithms*, v. 16, n. 12, 2023. ISSN 1999-4893. Disponível em: <https://www.mdpi.com/1999-4893/16/12/546>. (document), 1, 2.5, 2.6.2, 2.6.2.2, 2.11, 2.7, 2.19, 2.7, 3

*Análise, avaliação e validação do uso de técnicas de aprendizado de máquina para detecção de falhas em turbinas eólicas do tipo PMSG - Gerador síncrono de ímãs permanentes*

Henrique Gomes Mergulhão

Salvador, Maio de 2024.