



INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DO CEARÁ
IFCE *CAMPUS* MARACANAÚ
PROGRAMA DE PÓS-GRADUAÇÃO EM ENERGIAS RENOVÁVEIS (PPGER)

ISAAC SOUSA DA COSTA

IDENTIFICAÇÃO DE CONDIÇÃO OPERACIONAL EM
SISTEMA FOTOVOLTAICO CONECTADO À REDE ELÉTRICA
UTILIZANDO REDES NEURAS ARTIFICIAIS

MARACANAÚ

2024

ISAAC SOUSA DA COSTA

IDENTIFICAÇÃO DE CONDIÇÃO OPERACIONAL EM SISTEMA FOTOVOLTAICO
CONECTADO À REDE ELÉTRICA UTILIZANDO REDES NEURAIAS ARTIFICIAIS

Dissertação apresentada ao Programa de Pós-Graduação em Energias Renováveis (PPGER) do Instituto Federal de Educação, Ciência e Tecnologia do Ceará (IFCE) - *Campus Maracanaú*, como requisito parcial para obtenção do Título de Mestre em Energias Renováveis. Área de concentração: Energias Renováveis.

Orientador: Prof. Dr. José Renato de Brito Sousa.

Coorientador: Prof. Dr. Cláudio Marques de Sá Medeiros.

MARACANAÚ

2024

Dados Internacionais de Catalogação na Publicação
Instituto Federal do Ceará - IFCE
Sistema de Bibliotecas - SIBI
Ficha catalográfica elaborada pelo SIBI/IFCE, com os dados fornecidos pelo(a) autor(a)

C837i Costa, Isaac Sousa da.

Identificação de condição operacional em sistema fotovoltaico conectado à rede elétrica utilizando redes neurais artificiais / Isaac Sousa da Costa. - 2024.
111 f. : il. color.

Dissertação (Mestrado) - Instituto Federal do Ceará, Mestrado em Energias Renováveis, Campus Maracanaú, 2024.

Orientação: Prof. Dr. José Renato de Brito Sousa.

Coorientação: Prof. Dr. Cláudio Marques de Sá Medeiros.

1. Máquinas de comitê. 2. Perceptron de múltiplas camadas. 3. Redes neurais artificiais. 4. Sistemas fotovoltaicos. I. Título.

CDD 620.91

ISAAC SOUSA DA COSTA

IDENTIFICAÇÃO DE CONDIÇÃO OPERACIONAL EM
SISTEMA FOTOVOLTAICO CONECTADO À REDE ELÉTRICA UTILIZANDO
REDES NEURAIAS ARTIFICIAIS

Dissertação submetida à Coordenação do Curso de Pós-graduação em Energias Renováveis do Instituto Federal de Educação, Ciência e Tecnologia do Ceará, como requisito parcial para a obtenção do título de Mestre em Energias Renováveis, área de concentração Energias Renováveis.

Aprovada em 26/06/2024.

BANCA EXAMINADORA

Prof. Dr. José Renato de Brito Sousa (Orientador)

Instituto Federal de Educação, Ciência e Tecnologia do Ceará - IFCE

Prof. Dr. Cláudio Marques de Sá Medeiros (Coorientador)

Instituto Federal de Educação, Ciência e Tecnologia do Ceará - IFCE

Prof. Dr. Giovanni Cordeiro Barroso

Universidade Federal do Ceará - UFC

Prof. Dr. Guilherme de Alencar Barreto

Universidade Federal do Ceará - UFC

Dedico esse trabalho à minha família, meus professores,
amigos e também a todos que contribuíram
para a sua realização.

AGRADECIMENTOS

Ao Senhor Jesus toda honra, glória, louvor, poder e majestade. Agradeço tudo a Ele, pois me deu forças, sabedoria e entendimento para vencer as lutas e dificuldades da vida e em especial esse momento do mestrado. Agradeço aos meus pais, Alexandre Costa e Maria da Conceição (*in memoriam*) que me educaram e desde cedo me ensinaram a importância dos estudos.

Agradeço à minha esposa Jucélia e meus filhos Isaac Junior e Isabela pelo carinho e compreensão de estar ausente em alguns momentos.

Agradeço ao meu orientador Prof. Dr. José Renato de Brito Sousa, pela disponibilidade e paciência em me guiar no desenvolvimento como pesquisador. Agradeço ao meu coorientador Prof. Dr. Cláudio Marques de Sá Medeiros, pelos inúmeros questionamentos, sugestões e orientações que foram fundamentais para alcançar os objetivos pretendidos neste trabalho.

Agradeço também aos Professores Dr. Giovanni Cordeiro Barroso e Prof. Dr. Guilherme de Alencar Barreto por aceitar o convite para participar da banca examinadora, pelo tempo dedicado e pelas valiosas sugestões que contribuíram significativamente na evolução deste trabalho.

Agradeço também aos meus colegas Professores do IFCE no campus Cedro pelo apoio, e à Direção Geral desse campus, representado pelos Professores Antony Gleydson e Marcos Silvano, que me ajudaram em tudo que precisei. A todos meu muito obrigado.

“Ciência é a disposição para aceitar fatos,
mesmo quando eles se opõem aos desejos”
(Skinner, 1953).

RESUMO

Neste trabalho são apresentados os resultados de uma pesquisa sobre identificação de condição em sistema solar fotovoltaico conectado à rede elétrica. É utilizado um conjunto de dados reais contendo 192.094 amostras dos padrões observados durante o funcionamento normal e também durante a emulação de sujidade e sombreamento sobre essa instalação. O objetivo deste estudo é desenvolver um sistema computacional capaz de identificar o funcionamento normal, bem como condições de sujidade e sombreamento contidas no conjunto de dados utilizado. A metodologia adotada inclui o uso de redes neurais artificiais como ferramenta de análise. O escopo da pesquisa compreende a utilização de vários modelos neurais, como perceptron e perceptron de múltiplas camadas (MLP), e também a exploração de máquinas de comitê com o intuito de determinar qual técnica proporciona alcançar os resultados mais promissores. Os algoritmos utilizados realizam uma classificação binária entre amostras de dados, e, no âmbito das máquinas de comitê, cada conjunto é composto por cinco redes MLP que empregam o processo de votação majoritária para realizar classificações conjuntas das condições identificadas. Adicionalmente, é desenvolvido um algoritmo baseado em máquinas de comitê que classifica um conjunto de k amostras de forma sequencial, em que o rótulo que ocorre com maior frequência é informado pelo sistema de identificação de condições. Neste processo, a cada nova amostra apresentada para classificação, a última já classificada é excluída da quantidade verificada pelo algoritmo, fazendo com que as classificações se desloquem no decorrer do tempo como um tipo de janela deslizante, fazendo o sistema de identificação de condições determinar com maior precisão a real situação do sistema monitorado. Os resultados obtidos com este último algoritmo demonstram a elevada eficiência do sistema de identificação para as condições de normalidade, sujidade e sombreamento. Utilizando todas as amostras disponíveis, o modelo computacional desenvolvido com três máquinas de comitê, cada qual composta por cinco MLP, alcançou uma taxa de acertos de 98,59%. Posteriormente, utilizando o algoritmo de classificações sequenciais e uma janela deslizante para k igual a 120 amostras, a taxa de acertos alcançou 100%. Com base nos resultados, percebe-se que a identificação de condições em sistemas solar fotovoltaico adquire grande robustez quando realiza classificação de dados de forma sequencial com um algoritmo do tipo janela deslizante.

Palavras-chave: Máquinas de comitê. Perceptron de múltiplas camadas. Redes neurais artificiais. Sistemas fotovoltaicos.

ABSTRACT

This work presents the results of a research study on condition identification in a grid-connected photovoltaic solar system. A real dataset containing 192,094 samples of patterns observed during normal operation as well as during the emulation of soiling and shading on the installation is used. The aim of this study is to develop a computational system capable of identifying normal operation, as well as soiling and shading conditions contained in the utilized dataset. The methodology adopted includes the use of artificial neural networks as an analysis tool. The scope of the research includes the use of various neural models, such as perceptron and multi-layer perceptron (MLP), and also the exploration of committee machines in order to determine which technique provides the most promising results. The algorithms used perform binary classification between data samples, and, in the context of committee machines, each set is composed of five MLP networks that employ the majority voting process to jointly classify the identified conditions. Additionally, an algorithm based on committee machines is developed that classifies a set of k samples sequentially, where the label that occurs most frequently is reported by the condition identification system. In this process, with each new sample presented for classification, the last classified sample is excluded from the quantity verified by the algorithm, causing the classifications to shift over time like a sliding window, allowing the condition identification system to more accurately determine the actual situation of the monitored system. The results obtained with this last algorithm demonstrate the high efficiency of the identification system for normal, soiling, and shading conditions. Using all available samples, the computational model developed with three committee machines, each composed of five MLPs, achieved an accuracy rate of 98.59%. Subsequently, using the sequential classification algorithm and a sliding window for k equal to 120 samples, the accuracy rate reached 100%. Based on the results, it can be seen that condition identification in photovoltaic solar systems acquires great robustness when performing data classification sequentially with a sliding window algorithm.

Keywords: Artificial neural networks. Committee machines. Multilayer perceptron. Photovoltaic system.

LISTA DE FIGURAS

| | | |
|-----------|---|----|
| Figura 1 | — Evolução do quantitativo de publicações sobre identificação de condições em sistemas FV envolvendo técnicas de inteligência computacional..... | 22 |
| Figura 2 | — Sistema fotovoltaico isolado..... | 26 |
| Figura 3 | — Sistema fotovoltaico híbrido..... | 27 |
| Figura 4 | — Sistema fotovoltaico conectado à rede..... | 28 |
| Figura 5 | — Representação de módulo fotovoltaico em funcionamento normal..... | 30 |
| Figura 6 | — Representação de módulo fotovoltaico com célula sombreada..... | 30 |
| Figura 7 | — Módulo fotovoltaico real coberto por resíduos..... | 32 |
| Figura 8 | — Classe de amostras. (a) Linearmente separáveis; (b) Não linearmente separáveis..... | 34 |
| Figura 9 | — Perceptron..... | 35 |
| Figura 10 | — Funções de ativação. a) Degrau; b) Linear; c) Sigmóide logística; d) Tangente hiperbólica..... | 36 |
| Figura 11 | — Perceptron de múltiplas camadas com “n” neurônios..... | 37 |
| Figura 12 | — Atuação de uma rede perceptron de múltiplas Camadas. (a) MLP com dois neurônios na camada oculta; (b) Classificação de amostras não linearmente separáveis..... | 38 |
| Figura 13 | — Técnicas computacionais verificadas em pesquisa bibliográfica..... | 51 |
| Figura 14 | — Metodologia..... | 52 |
| Figura 15 | — Sistema fotovoltaico do IFCE..... | 53 |
| Figura 16 | — Sensores utilizados para leitura de dados..... | 53 |
| Figura 17 | — Amostra da base de dados utilizada..... | 54 |
| Figura 18 | — Emulação de condições. (a) Sombreamento; (b) Sujidade..... | 55 |
| Figura 19 | — Verificação de amostras da base de dados..... | 56 |
| Figura 20 | — Pré-processamento dos dados..... | 57 |
| Figura 21 | — Distribuição das amostras..... | 58 |
| Figura 22 | — Seleção de amostras para composição de novo conjunto de dados..... | 58 |
| Figura 23 | — Fluxograma de treinamento e teste..... | 61 |
| Figura 24 | — Classificação binária de dados..... | 62 |
| Figura 25 | — Classificação com três entradas e três saídas | 63 |
| Figura 26 | — Máquina de comitê com “n” especialistas..... | 64 |

| | |
|--|----|
| Figura 27 — Classificação binária com máquina de comitê..... | 65 |
| Figura 28 — Classificação multiclasse. (a) Máquinas de comitê com saídas binárias e com combinador geral; (b) Exemplificação de lógica do combinador geral..... | 66 |
| Figura 29 — Conjunto de amostras emuladas..... | 68 |
| Figura 30 — Histogramas de dados com 4 atributos. (a) Antes do pré-processamento; (b) Após o pré-processamento..... | 71 |
| Figura 31 — <i>Box plot</i> de dados com 4 atributos. (a) Antes do pré-processamento; (b) Após o pré-processamento..... | 71 |
| Figura 32 — Matriz de confusão para perceptron com 4 entradas. (a) Normal x Sujidade; (b) Normal x Sombreamento..... | 74 |
| Figura 33 — Matriz de confusão para perceptron com 3 entradas. (a) Normal x Sujidade; (b) Normal x Sombreamento..... | 76 |
| Figura 34 — Matriz de confusão para MLP com 4 entradas. (a) Normal x Sujidade; (b) Normal x Sombreamento..... | 79 |
| Figura 35 — Matriz de confusão para MLP com 3 entradas. (a) Normal x Sujidade; (b) Normal x Sombreamento..... | 82 |
| Figura 36 — Matrizes de confusão para MLP com 3 saídas. (a) Treinamento; (b) Teste..... | 84 |
| Figura 37 — Matrizes de confusão para máquinas de comitê com 5 MLP. (a) Treinamento - “normal/sujidade”; (b) Teste - “normal/sujidade”; (c) Treinamento - “normal/sombreamento”; (d) Teste - “normal/sombreamento”..... | 90 |
| Figura 38 — Matrizes de confusão de treinamento e teste de máquinas de comitê. (a) Treinamento - “normal/sujidade”; (b) Teste - “normal/sujidade”; (c) Treinamento - “normal/sombreamento”; (d) Teste - “normal/sombreamento”; (e) Treinamento - “sujidade/sombreamento”; (f) Teste - “sujidade/sombreamento”..... | 92 |
| Figura 39 — Matrizes de confusão para conjunto de máquinas de comitê com 3 saídas.... | 93 |
| Figura 40 — Avaliação sequencial com k = 60 amostras. (a) Conjunto com 31535 amostras sequenciais; (b) Classificação individual; (c) Classificação informada..... | 95 |
| Figura 41 — Avaliação sequencial com k = 120 amostras. (a) Conjunto com 31535 amostras sequenciais; (b) Classificação individual; (c) Classificação informada..... | 96 |

| | |
|--|----|
| Figura 42 — Avaliação sequencial com $k = 180$ amostras. (a) Conjunto com 31535 amostras sequenciais; (b) Classificação individual; (c) Classificação informada..... | 97 |
| Figura 43 — Avaliação sequencial com $k = 240$ amostras. (a) Conjunto com 31535 amostras sequenciais; (b) Classificação individual; (c) Classificação informada..... | 98 |
| Figura 44 — Avaliação sequencial com $k = 300$ amostras. (a) Conjunto com 31535 amostras sequenciais; (b) Classificação individual; (c) Classificação informada..... | 99 |

LISTA DE TABELAS E QUADROS

| | | |
|-----------|---|----|
| Tabela 1 | — Trabalhos relacionados a identificação de condições operacionais em SFV..... | 45 |
| Quadro 1 | — Taxas de acertos dos perceptrons para amostras com 4 atributos..... | 73 |
| Quadro 2 | — Taxas de acertos dos perceptrons para amostras com 3 atributos..... | 75 |
| Quadro 3 | — Verificação do número de neurônios ideal na camada oculta..... | 78 |
| Quadro 4 | — Taxas de acertos das redes MLP 4 entradas..... | 79 |
| Quadro 5 | — Taxas de acertos das redes MLP 3 entradas..... | 81 |
| Quadro 6 | — Taxas de acertos da rede MLP com 3 saídas..... | 83 |
| Quadro 7 | — Resultados com topologias diversas para amostras “normal/sujidade”..... | 85 |
| Quadro 8 | — Resultados com topologias diversas para amostras “normal/sombreamento”..... | 86 |
| Quadro 9 | — Taxas de acertos para máquina de comitê com 5 MLP e combinador realizando média dos resultados..... | 87 |
| Quadro 10 | — Resultados para treinamento da máquina de comitê com 5 MLP e amostras “normal/sujidade”..... | 87 |
| Quadro 11 | — Resultados para testes da máquina de comitê com 5 MLP e amostras “normal/sujidade”..... | 88 |
| Quadro 12 | — Resultados para treinamento da máquina de comitê com 5 MLP e amostras “normal/sombreamento”..... | 88 |
| Quadro 13 | — Resultados para testes da máquina de comitê com 5 MLP e amostras “normal/sombreamento”..... | 89 |
| Quadro 14 | — Resultados de treinamento e teste de máquina de comitê com três saídas..... | 91 |

LISTA DE SIGLAS

| | |
|----------|--|
| ABSOLAR | Associação Brasileira de Energia Solar Fotovoltaica |
| ANEEL | Agência Nacional de Energia Elétrica |
| BD | Base de Dados |
| CA | Corrente Alternada |
| CC | Corrente Contínua |
| CNN | <i>Convolutional Neural Network</i> |
| CSV | <i>Comma Separated Values</i> |
| CRESCESB | Centro de Referência para as Energias Solar e Eólica Sérgio de Salvo Brito |
| EL | <i>Ensemble Learning</i> |
| FV | Fotovoltaico |
| GD | Geração Distribuída |
| IA | Inteligência Artificial |
| IOT | <i>Internet of Things</i> |
| IFCE | Instituto Federal de Educação, Ciência e Tecnologia do Ceará |
| INMETRO | Instituto Nacional de Metrologia, Qualidade e Tecnologia |
| KNN | <i>K-Nearest Neighbors</i> |
| kW | Quilo Watt |
| kWp | Quilo Watt Pico |
| MPPT | <i>Maximum Power Point Tracking</i> |
| MFV | Módulo Fotovoltaico |
| MLP | <i>Multilayer Perceptron</i> |
| MW | Mega Watt |
| NTC | <i>Negative Temperature Coefficient</i> |
| PPGER | Programa de Pós-Graduação em Energias Renováveis |
| RNA | Rede Neural Artificial |

| | |
|------|---|
| SFV | Sistema Solar Fotovoltaico |
| SFI | Sistema Solar Fotovoltaico Isolado |
| SFCR | Sistema Solar Fotovoltaico Conectado à Rede |
| SVM | <i>Support Vector Machine</i> |
| XOR | <i>Exclusive OR</i> |
| Wp | Watt Pico |

LISTA DE SÍMBOLOS

| | |
|----------------|----------------|
| α | Alfa |
| β | Beta |
| μ | Micro |
| σ | Sigma |
| Σ | Somatório |
| i | i-ésima |
| = | Igual |
| > | Maior |
| < | Menor |
| \geq | Maior ou igual |
| \leq | Menor ou igual |
| % | Porcentagem |
| °C | Graus Celsius |
| ° | Grau |
| ' | Minuto |
| '' | Segundo |
| n | número |
| m | metro |
| s | segundo |
| h | hora |
| A | Ampére |
| V | Volt |
| W | Watt |
| m ² | Metro quadrado |

SUMÁRIO

| | | |
|--------------|--|-----------|
| 1 | INTRODUÇÃO..... | 18 |
| 1.1 | Justificativa..... | 20 |
| 1.2 | Objetivos..... | 23 |
| <i>1.2.1</i> | <i>Objetivo geral.....</i> | <i>23</i> |
| <i>1.2.2</i> | <i>Objetivos específicos.....</i> | <i>23</i> |
| 1.3 | Produção científica..... | 24 |
| 1.4 | Estrutura do trabalho..... | 24 |
| 1.5 | Conclusões parciais..... | 24 |
| 2 | FUNDAMENTAÇÃO TEÓRICA..... | 25 |
| 2.1 | Sistema solar fotovoltaico..... | 25 |
| <i>2.1.1</i> | <i>Sistema isolado.....</i> | <i>26</i> |
| <i>2.1.2</i> | <i>Sistema híbrido.....</i> | <i>27</i> |
| <i>2.1.3</i> | <i>Sistema conectado à rede.....</i> | <i>28</i> |
| 2.2 | Condições operacionais em sistemas fotovoltaicos..... | 29 |
| <i>2.2.1</i> | <i>Sombreamento.....</i> | <i>29</i> |
| <i>2.2.2</i> | <i>Sujidade.....</i> | <i>31</i> |
| 2.3 | Redes neurais artificiais..... | 33 |
| <i>2.3.1</i> | <i>Perceptron.....</i> | <i>34</i> |
| <i>2.3.2</i> | <i>Perceptron de múltiplas camadas.....</i> | <i>36</i> |
| <i>2.3.3</i> | <i>Máquinas de comitê.....</i> | <i>39</i> |
| 2.4 | Trabalhos relacionados..... | 41 |
| 2.5 | Conclusões parciais..... | 51 |
| 3 | MATERIAIS E MÉTODOS..... | 52 |
| 3.1 | Base de dados..... | 52 |
| 3.2 | Pré-processamento..... | 55 |

| | | |
|------------|---|------------|
| 3.3 | Modelos computacionais..... | 59 |
| 3.4 | Sistema de identificação de condições..... | 66 |
| 3.5 | Conclusões parciais..... | 69 |
| 4 | RESULTADOS E DISCUSSÕES..... | 70 |
| 4.1 | Avaliação dos dados..... | 70 |
| 4.2 | Resultados obtidos com perceptron..... | 72 |
| 4.3 | Resultados obtidos com MLP..... | 77 |
| 4.4 | Resultados obtidos com máquinas de comitê..... | 84 |
| 4.5 | Conclusões parciais..... | 100 |
| 5 | CONCLUSÕES..... | 101 |
| 5.1 | Sugestões de trabalhos futuros..... | 102 |
| | REFERÊNCIAS..... | 103 |
| | ANEXO A — Códigos desenvolvidos..... | 110 |

1 INTRODUÇÃO

A energia solar fotovoltaica (FV) vem sendo cada vez mais utilizada em todo o mundo. Ela é uma das fontes de energia renováveis e inesgotáveis porque utiliza a radiação solar, um recurso abundante e disponibilizado continuamente pelo Sol (Moreira *et al.*, 2021), além de desempenhar em muitos países um papel crucial na transição para uma matriz energética mais limpa e sustentável. A adoção da fonte solar FV para a produção de eletricidade impacta significativamente na sustentabilidade ambiental, pois segundo Madureira *et al.* (2023), embora a fabricação de módulos fotovoltaicos (MFV) envolva a queima de combustíveis fósseis, como carvão, petróleo e gás natural, um sistema solar fotovoltaico (SFV) após instalado, contribui para a redução das emissões de gases de efeito estufa. Conforme Peters *et al.* (2021) a geração de eletricidade a partir da luz solar apresenta inúmeras vantagens, dentre elas a longevidade dos MFV, pois possuem vida útil estimada em 25 anos ou mais. Essa longevidade assegura uma solução sustentável e duradoura para a produção de energia limpa, contribuindo de maneira eficaz para mitigação das mudanças climáticas. Além disso, para Twidell (2021) a energia solar FV é muito versátil, sendo empregada desde iluminação a sistemas de bombeamento de água em regiões remotas, entre outras aplicações, o que faz dela um importante elemento para construção de um futuro melhor.

No Brasil, em 2012 a Agência Nacional de Energia Elétrica (ANEEL) regulamentou a geração distribuída (GD), que se refere à produção de energia elétrica próxima aos consumidores, abrangendo diversas fontes, incluindo a solar FV. Essa regulamentação impulsionou um crescimento expressivo na participação dessa fonte na matriz elétrica brasileira, pois, a partir de então foi possível compensar o excedente de energia gerada por meio de créditos aos consumidores.

Dados da Associação Brasileira de Energia Solar Fotovoltaica (ABSOLAR) informam que existem atualmente cerca de 2,6 milhões de sistemas FV instalados no país, fazendo que essa fonte de energia seja muito importante para a matriz elétrica brasileira (ABSOLAR, 2024). Estudos da ABSOLAR (2024) também apontam que mesmo após a introdução da cobrança de tarifa para a distribuição de energia gerada por fontes FV pelas concessionárias, conhecida como “cobrança do Fio B”, estabelecida pela lei 14.300/2022 (Brasil, 2022), a instalação de novos sistemas FV continua em ascensão. No entanto, apesar desta expansão, Pereira (2018) ressalta que o investimento associado à aquisição de SFV destinados à microgeração ou

minigeração distribuída¹, ainda é muito alto para grande parte da população brasileira. No caso da microgeração, de acordo com Jauris (2023) o período de retorno do investimento (*payback*) pode estender-se até 89 meses, o que equivale a quase 8 anos.

Assim, é imprescindível não apenas o adequado dimensionamento do sistema, mas, também assegurar sua máxima eficiência, a fim de atenuar o impacto financeiro da aquisição. Estratégias como manutenção regular e utilização de sistemas para identificar condições em SFV podem melhorar a confiabilidade, eficiência e segurança deles, uma vez que possibilitam rápida atuação do responsável pela instalação, enquanto a ausência deles podem resultar em perdas significativas na produção de energia elétrica (Al-Katheri *et al.*, 2022).

Segundo Comerio *et al.* (2019) devido os SFV estarem sujeitos a condições climáticas adversas, como ventos fortes, variações extremas de temperatura, tempestades, granizo, poluição entre outros fatores, eles são vulneráveis a diversas situações, portanto, é importante identificar rapidamente uma eventual condição prejudicial para não afetar a geração de energia.

De acordo com Syafaruddin e Zinger (2019) vários métodos podem ser utilizados para detectar condições diversas em SFV, dentre os quais se destacam os métodos computacionais que utilizam técnicas inteligentes para resolver diversos problemas de engenharia, incluindo a identificação e diagnóstico de uma instalação FV.

Desse modo, o estudo realizado na literatura consultada permitiu perceber que a verificação do estado em que se encontra um SFV é importante para garantir o máximo rendimento e confiabilidade de uma planta FV. No entanto, sistemas com essa finalidade são verdadeiras "caixas pretas", ou seja, são sistemas cujo funcionamento interno é desconhecido, não sendo possível compreender como os dados são processados internamente, sendo essa falta de transparência ser intencional, como uma medida de segurança, ou devido à complexidade do sistema. Portanto, é definido como problema de estudo desta pesquisa, a identificação de condição operacional em sistema fotovoltaico conectado à rede elétrica utilizando redes neurais artificiais. Neste trabalho são utilizados dados obtidos por Pereira (2018), aluno egresso do Programa de Pós-Graduação em Energias Renováveis (PPGER) do Instituto Federal de Educação, Ciência e Tecnologia do Ceará (IFCE) e toda a pesquisa é desenvolvida com base em um microssistema solar fotovoltaico conectado à rede elétrica, instalado nas dependências do IFCE na cidade de Fortaleza - CE, Brasil.

¹ Denomina-se microgeração distribuída a central geradora com potência instalada até 75 quilowatts (KW). Já minigeração distribuída é aquela com potência acima de 75 kW e menor ou igual a 3 MW (podendo ser até 5 MW em situações específicas, nos termos dos incisos IX e XIII e do Parágrafo Único do art. 1º da Lei nº 14.300/2022). Ambas são conectadas à rede de distribuição por meio de instalações de unidades consumidoras (Brasil, 2022).

1.1 Justificativa

Atualmente a maioria dos SFV conectados à rede elétrica possuem a funcionalidade de permitir o monitoramento por meio de *websites* dos fabricantes ou mesmo através de aplicativos para *smartphone* conectados à *internet*. Entretanto, a supervisão de usinas FV geralmente limita-se à gestão da geração de energia instantânea e acumulada, não sendo realizado nenhum tipo de verificação da condição operacional na instalação. Assim, uma queda na produção de energia pela usina ou a falta de informação sobre a geração podem levar o proprietário a pensar que houve apenas uma desconexão na comunicação entre equipamentos e *software* de monitoramento. No entanto, pode estar ocorrendo uma condição que potencialmente levará a um problema real, sendo que somente uma inspeção nos componentes poderá confirmar a existência de algum problema. Em microssistemas FV, esse aspecto é agravado pelo fato de que a maioria está localizada em áreas de difícil acesso, como telhados e coberturas de prédios. Portanto, somente pessoas capacitadas ou com experiência conseguem acessar para verificar o estado da instalação e tomar alguma medida para resolver eventuais problemas.

Uma alternativa para efetuar a identificação de condição operacional em usinas FV é a utilização de sistemas comerciais para verificação das ocorrências, mas, segundo López-Vargas *et al.* (2019), embora o custo desses sistemas tenha diminuído nos últimos anos, eles ainda são caros e, em alguns casos, o preço excede o custo do SFV. Portanto, a implementação de sistemas de baixo custo em plataformas de código aberto surge como uma alternativa viável para realização dessa verificação, pois quando bem desenvolvidos têm o potencial de reduzir consideravelmente os custos de manutenção. Esse benefício é ainda mais significativo quando é possível detectar condições a tempo de evitar danos irreparáveis.

Os sistemas para identificação de condições operacionais dos SFV podem prolongar a vida útil dos equipamentos, reduzindo a produção e o descarte de materiais, e consequentemente, contribuindo para a sustentabilidade ambiental. Ao prolongar a vida útil dos equipamentos e minimizar a necessidade de substituições frequentes, esses sistemas ajudam a diminuir a quantidade de lixo e outros resíduos, promovendo uma gestão mais eficiente dos recursos e mitigando os impactos adversos no meio ambiente (Pereira, 2018).

Nesse contexto, as justificativas para desenvolvimento desta pesquisa baseiam-se nas seguintes premissas:

- **Eficiência Energética:** A identificação de condições operacionais podem proporcionar que as usinas FV operem com máxima eficiência. O sombreamento mesmo que parcial e momentâneo, pode causar quedas significativas na produção

de energia, bem como a sujidade sobre os painéis, pois, reduz a geração de energia do sistema. Assim, ao detectar esse tipo de condição, perdas energéticas podem ser mitigadas ou até mesmo eliminadas. Quanto mais eficientes forem os SFV, menor será o *payback* para recuperar o investimento inicial, tornando-os financeiramente mais atraentes para todos.

- Redução de Custos: A intervenção oportuna e precisa pode resultar em custos de manutenção menores. A simples limpeza dos MFV, por exemplo, é mais econômica do que realizar reparos ou substituições deles. Dessa forma, a manutenção preventiva e periódica no SFV pode evitar reparos, reduzindo as despesas gerais associadas ao sistema, tornando-os mais acessível e viável à população em geral.
- Prolongamento da vida útil: Os SFV representam um investimento significativo e de longo prazo. A identificação de condições como sombreamento ou acúmulo de sujidade desempenha um papel fundamental no prolongamento da vida útil desses sistemas. Essas condições podem desencadear danos irreversíveis aos componentes do SFV, levando à degradação prematura dos MFV. O prolongamento da vida útil das usinas FV não apenas protege o investimento realizado, mas também tem implicações ambientais positivas, já que a fabricação e o descarte dos MFV envolvem recursos significativos. Assim, estender a vida útil dos SFV reduz a necessidade de substituição frequente de componentes, diminuindo o impacto ambiental associado.

Conforme Pereira (2018), sistemas para identificação de condições operacionais em SFV podem ser comerciais ou protótipos. Sistemas comerciais são aqueles desenvolvidos por empresas e vendidos em lojas físicas ou mesmo pela *internet*, como por exemplo o detector de arcos elétricos em sistemas FV “TIDA-010231” da Texas Instruments (TI, 2023) e o *software* inteligente “*TrueCapture*” da empresa Nextracker que se utiliza de diversos sensores que fornecem informações sobre sombreamento em tempo real em cada ligação série de MFV a fim de realizar uma otimização no controle do rendimento de uma planta solar (Nextracker, 2023). No entanto, apesar desses sistemas serem eficientes, o custo para aquisição é alto, e a tecnologia por trás do funcionamento deles não é aberta.

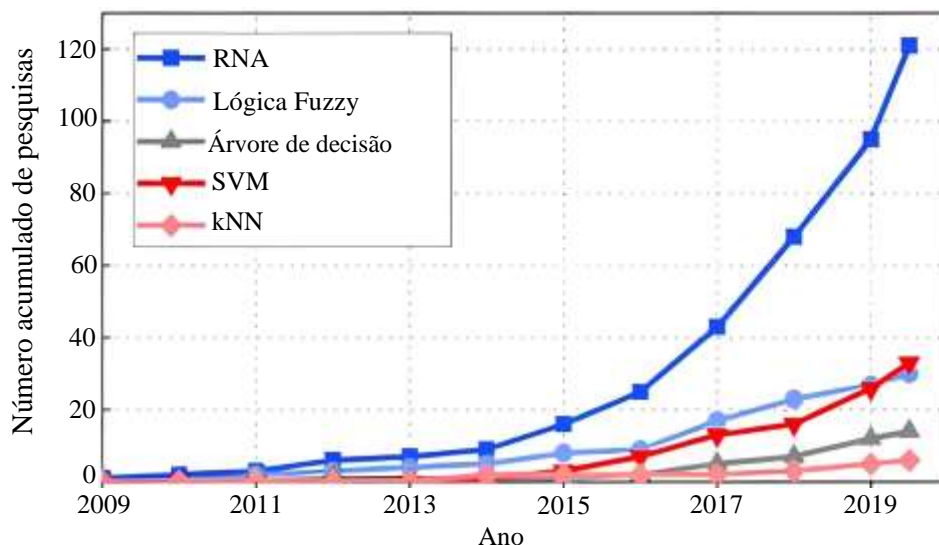
Os sistemas protótipos são aqueles que estabelecem uma versão inicial construída em um curto período de tempo e aprimorada em várias iterações para testar e avaliar a eficácia para resolver um problema específico (CIN, 2023). A literatura científica apresenta diversos tipos de sistemas protótipos desenvolvidos durante pesquisas realizadas em laboratórios e instituições de ensino para identificar condições em SFV. Há sistemas que utilizam técnicas

estatísticas como análise de componentes principais (Rouani *et al.*, 2021; Maia 2019), lógica fuzzy (Syafaruddin; Zinger, 2019) e também aqueles que trabalham com redes neurais artificiais (RNA) em SFV (Prakash; Kumar, 2022).

Conforme Ahmadi *et al.* (2020) e Abdolrasol *et al.* (2021) a utilização de RNA para identificação de condições operacionais em SFV é eficiente e confiável, pois são técnicas computacionais que tem a capacidade de aprender de forma supervisionada e não supervisionada, além de possuir uma grande capacidade de lidar com problemas complexos (não lineares), alta adaptabilidade e rápido processamento de informações.

Pesquisas de Costa *et al.* (2019) destacam que as RNA vem sendo bastante utilizadas para realizar identificação de condições em SFV. Essas técnicas computacionais são estudadas e utilizadas em diferentes áreas de ciências e engenharia, sendo que elas podem ser utilizadas em várias aplicações de inteligência computacional, sobretudo em reconhecimento de padrões (Haykin, 2001). As RNA são classificadores não lineares, robustos e capazes de aprender a partir de amostras (padrões) que representam o comportamento de um sistema (Li *et al.*, 2021). Logo, após a rede ter sido treinada, ela aprende a relação entre as entradas e saídas, e, portanto, torna-se capaz de generalizar uma saída para o caso de novas amostras sensibilizarem sua entrada (Faceli *et al.*, 2021). Na Figura 1 é apresentado um gráfico relacionando as pesquisas com protótipos para criação de sistemas de identificação de condições em SFV. Neste gráfico é possível verificar que esse campo de estudo está em evidência com diversas técnicas computacionais sendo utilizadas, no entanto, as RNA vem se destacando em relação a outros modelos computacionais.

Figura 1 - Evolução do quantitativo de publicações sobre identificação de condições em sistemas FV envolvendo técnicas de inteligência computacional.



Fonte: Adaptado de Li *et al.* (2019)

Em relação às condições que podem ocorrer em SFV, o sombreamento e a sujidade têm recebido muito atenção (Shariah; Al-Ibrahim, 2023). Estudos de Silva (2023) relatam que a sujidade é uma das situações que possuem grande potencial para reduzir a eficiência de uma instalação FV. Pesquisas de Pillai e Rajasekar (2018) abordam o sombreamento como uma condição de curta duração, mas que deve ser detectado rapidamente para evitar a queda na produção de energia e preservar a vida útil do sistema. O sombreamento prolongado pode resultar na formação de pontos quentes (*hot spots*) nos MFV, que são áreas na qual as células FV sofrem com um aumento excessivo de temperatura. Quando uma célula do MFV é sombreada, ela não consegue gerar eletricidade eficientemente, o que leva à formação de uma alta impedância. Isso faz com que a energia produzida pelas outras células no módulo seja dissipada na forma de calor nessa célula sombreada. Esse aquecimento localizado pode causar danos irreparáveis à célula afetada, comprometendo a eficiência e a durabilidade do SFV.

1.2 Objetivos

Para solucionar o problema abordado, é necessário estabelecer objetivos a serem alcançados. Dessa forma, são definidos os seguintes objetivos geral e específicos para orientar o desenvolvimento desta pesquisa.

1.2.1 Objetivo geral

Desenvolver metodologia baseada em redes neurais artificiais para identificar condições de sombreamento e sujidade em um sistema fotovoltaico conectado à rede elétrica.

1.2.2 Objetivos específicos

- Determinar o conjunto de atributos relevantes para solução do problema;
- Determinar a topologia do classificador neural;
- Desenvolver metodologia de classificação robusta e confiável;
- Validar a metodologia proposta.

1.3 Produção científica

Durante o mestrado em energias renováveis do Programa de Pós-Graduação em Energias Renováveis (PPGER) do IFCE foi publicado um artigo.

- COSTA, I. S; SOUSA, J. R. B; JUCÁ, S. C. S; PEREIRA, R. I. S; ALEXANDRIA, A. R. Monitoramento IoT de planta de bombeamento fotovoltaico utilizando sistema embarcado Linux. Enciclopédia Biosfera, v.18 n.37; p. 349-363, 2021.

1.4 Estrutura do trabalho

A presente dissertação encontra-se organizada em cinco capítulos. O plano de desenvolvimento do texto é descrito a seguir:

No Capítulo 2 é apresentada a parte conceitual desse estudo, em que são explicados alguns dos principais fundamentos teóricos ao longo do trabalho e também traz o estado da arte apresentando alguns trabalhos relacionados diretamente ao tema em questão.

No Capítulo 3 é abordada a parte de materiais e métodos utilizados na pesquisa para desenvolvimento do algoritmo de identificação de condições operacionais em SFV usando RNA. Também são abordadas as máquinas de comitê e como elas podem proporcionar melhoria na avaliação dos resultados das classificações.

No Capítulo 4 são apresentados e discutidos os resultados obtidos nos vários testes realizados ao longo da pesquisa.

O Capítulo 5 apresenta as conclusões, bem como sugestões de trabalhos futuros envolvendo o tema em questão.

1.5 Conclusões parciais

Neste capítulo é discutido que a energia solar FV tem um papel muito importante para a sustentabilidade ambiental e transição energética. Verificou-se também que a eficiência energética e a redução de custos são pontos importantes a serem considerados atualmente, dessa forma, sistemas para identificação de condições podem maximizar o retorno do investimento na aquisição de um SFV e ainda prolongar sua vida útil.

No Capítulo 2 será explorada parte conceitual deste estudo e se delineará o estado da arte com apresentação de trabalhos correlatos que tratam da identificação de condições em SFV.

2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo são apresentadas definições importantes sobre SFV, sombreamento e sujidade, classificadores perceptron, perceptron de múltiplas camadas e também máquinas de comitê. Além disso, é apresentado um estudo bibliográfico sobre trabalhos relacionados ao tema proposto encontrados na literatura científica, a fim de nortear o desenvolvimento desta pesquisa.

2.1 Sistema solar fotovoltaico

Os SFV desempenham um papel crucial na obtenção de energia elétrica limpa e renovável, contribuindo significativamente para a redução da dependência de combustíveis fósseis e para mitigação dos impactos ambientais das emissões de carbono (Fernandes *et al.*, 2020). Eles podem ser implantados em residências, empresas e instalações industriais para gerar eletricidade a partir da energia solar, resultando em redução de custos com eletricidade.

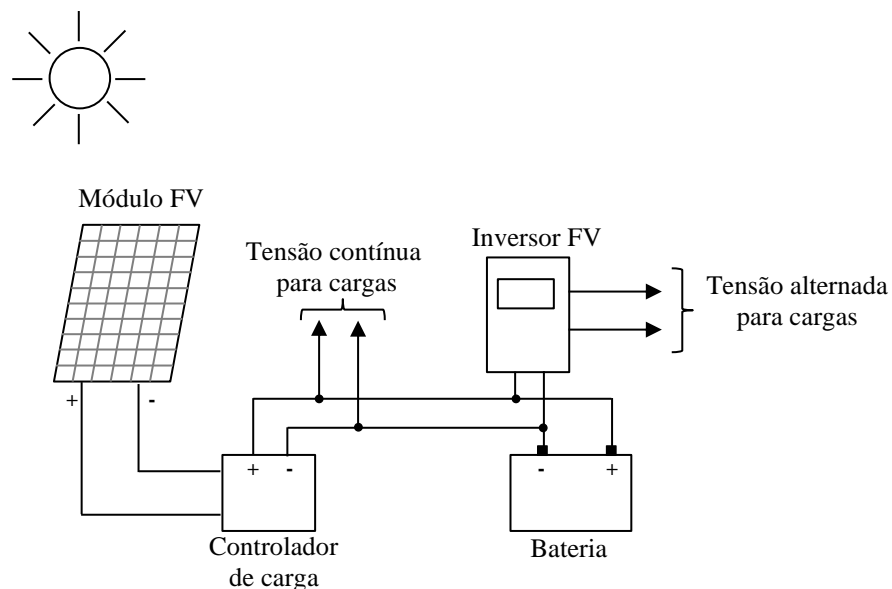
De acordo com Yahyaoui (2018), um SFV é composto por um conjunto interconectado de equipamentos que formam uma usina fotovoltaica (FV). Essa instalação é responsável por converter diretamente a radiação solar em eletricidade. Entre os componentes fundamentais dos SFV, Reeves (2018) destaca o MFV como responsável por converter a energia solar em elétrica e o inversor, componente responsável por converter a tensão e corrente contínua (CC) gerada pelos MFV em tensão e corrente alternada (CA) para alimentar cargas ou ser injetada na rede da concessionária local.

Os SFV podem ser isolados, híbridos e conectados à rede elétrica (CRESESB, 2023). Quando eles operam de forma isolada, funcionando de maneira autônoma, sem conexão com a rede elétrica das distribuidoras de energia elétrica, neste caso eles geram e armazenam toda a energia necessária para funcionamento, são conhecidos como sistemas *off-grid* (do inglês, “fora da rede”). Por outro lado, quando são conectados à rede elétrica, são chamados de sistemas *on-grid*, (do inglês, “ligado à rede”). Já os sistemas híbridos, são caracterizados por ter várias fontes de alimentação, além dos MFV, podendo inclusive trabalhar desconectados ou conectados à rede elétrica (Elbaset *et al*, 2019). Dentre as fontes de energia utilizadas por esses equipamentos, destacam-se os bancos de baterias que fornecem autonomia ao sistema em períodos sem luz solar, como durante à noite.

2.1.1 Sistema isolado

Os sistemas fotovoltaicos isolados (SFI) geralmente são utilizados em locais de difícil acesso e longe da rede elétrica. Segundo o CRESESB (2023) os SFI utilizam alguma forma de armazenamento de energia. Este armazenamento pode ser feito através de baterias, quando se deseja utilizar aparelhos elétricos ou armazena-se na forma de energia gravitacional quando se bombeia água para tanques em sistemas de abastecimento. Esta última condição ocorre, por exemplo, quando se armazena água em reservatórios suspensos, para consumo ou irrigação. Na Figura 2 é ilustrado um SFI no qual podem ser observados um MFV, controlador de carga, bateria e inversor FV.

Figura 2 - Sistema fotovoltaico isolado.



Fonte: Autoria própria (2024)

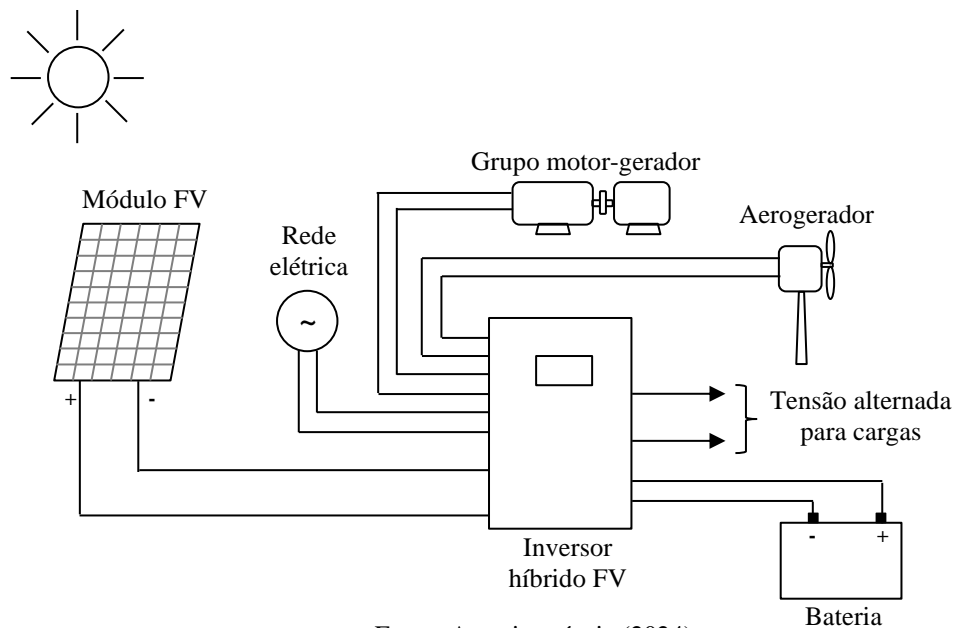
De acordo com Fonseca *et al.* (2020), os SFI têm como principais elementos: Os módulos FV, controlador de carga, cabos, bateria e inversor. O controlador de carga é o dispositivo responsável por controlar a carga e a descarga das baterias, componentes responsáveis por propiciar autonomia à instalação. Dessa forma, a energia elétrica produzida é armazenada e utilizada para conversão em tensão e corrente alternada (CA) pelo inversor. Para SFI operando em residências, Fonseca *et al.* (2020) também relatam que eles devem ter autonomia de dois a três dias. Porém, em aplicações especiais ou de emergência como sistemas de telecomunicações ou em locais remotos, a autonomia deve ser de aproximadamente cinco a

sete dias, pois esse tipo de instalação não pode ficar sem atendimento de energia elétrica em nenhum momento.

2.1.2 Sistema híbrido

Conforme pesquisas de Yahyaoui (2018) um SFV é denominado híbrido quando gera energia elétrica em conjunto com uma ou mais fontes adicionais. Esse tipo de sistema utiliza o conceito “*all-in-one*” (em inglês, “*todos em um*”), pois o inversor empregado pode receber energia de várias fontes simultaneamente, como solar, eólica, de um gerador elétrico (síncrono ou assíncrono) e até mesmo da rede elétrica. Além disso, ele é capaz de alimentar diversas cargas e, simultaneamente, armazenar energia em uma bateria ou banco de baterias para uso posterior. Na Figura 3 é ilustrado um sistema híbrido.

Figura 3 - Sistema fotovoltaico híbrido.



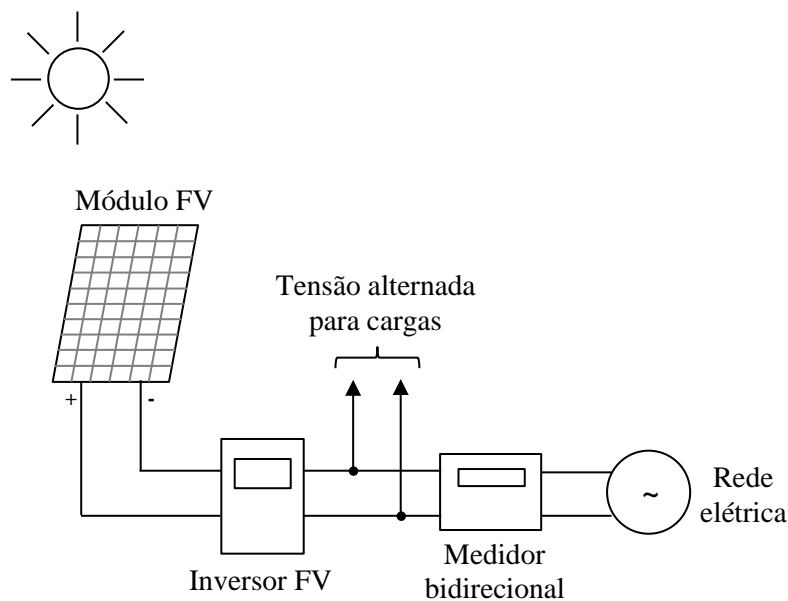
Fonte: Autoria própria (2024)

No Brasil, o Instituto Nacional de Metrologia, Qualidade e Tecnologia (INMETRO) publicou em 21 de março de 2022, a Portaria nº. 140/2022 que regulariza a utilização de inversores híbridos com potência de até 75 kW. Segundo o INMETRO (2022) antes da publicação desse documento, não existia nenhum regulamento para a utilização dos inversores híbridos conectados à rede em território brasileiro. Dessa forma, esse tipo de sistema pode trabalhar atualmente de forma isolada ou conectada à rede elétrica, desde que autorizado pela concessionária local.

2.1.3 Sistema conectado à rede

Os sistemas fotovoltaicos conectados à rede elétrica (SFCR) não requerem baterias para funcionar. Em SFCR a rede elétrica se comporta como sistema de acumulação de energia (CRESESB, 2023). Esses sistemas aproveitam a energia solar captada pelos MFV durante o dia e a transformam em energia elétrica. A tensão e corrente contínua (CC) gerada é convertida pelo inversor FV em tensão e corrente alternada (CA) com as mesmas características da rede elétrica local. Na Figura 4 é ilustrado um SFCR em operação.

Figura 4 - Sistema fotovoltaico conectado à rede.



Fonte: Autoria própria (2024)

Uma característica importante desse tipo de sistema é que a energia elétrica gerada é utilizada para atender as cargas de uma instalação e o excedente, ou seja, o que não foi consumido é enviado à rede elétrica, sendo essa energia contabilizada por um medidor bidirecional. Assim, a energia tanto pode vir da concessionária nos momentos em que a produção do sistema não é suficiente para atender toda a instalação, como também ir para a rede elétrica, nesse caso gerando créditos, que podem ser utilizados nos momentos em que não há produção de energia ou quando ela é insuficiente (Pereira, 2018).

Dados da ABSOLAR (2024) comprovam que os SFCR são os tipos de SFV mais utilizados atualmente no Brasil. Os principais elementos que compõem eles são: módulos FV, inversor, cabos, conectores, estruturas de fixação e medidor bidirecional, sendo este último de responsabilidade da concessionária. Os demais são ser adquiridos pelos usuários.

2.2 Condições operacionais em sistemas fotovoltaicos

Conforme Al-katheri *et al.* (2022) diversas situações podem impactar a operação de um SFV. Segundo Dupont (2021), causas físicas, ambientais e elétricas, tanto no lado de corrente contínua (CC), quanto no lado de corrente alternada (CA) são as responsáveis por interferir no funcionamento normal de uma instalação FV. Em relação às condições do lado CC, influências ambientais como sombreamento e sujidade (provocada pelo acúmulo de poeira, excrementos de animais, folhas de árvores e até neve sobre a superfície dos módulos) podem reduzir significativamente a energia gerada pelo sistema e até mesmo provocar graves problemas. Além disso, perdas nos condutores e o envelhecimento do sistema também são fatores que afetam os índices de geração. Estudos de Maia (2019) apontam que tais condições têm o potencial de reduzir a potência gerada pelo SFV em média 18,9% ao ano, enquanto que erros no rastreamento do ponto de potência máxima (do lado CA) realizado pelo inversor produzem perdas menores. No entanto, podem ocorrer também do lado CC condições envolvendo erros na ligação dos módulos e até mesmo a danificação dos cabos, o que pode resultar em perdas no sistema e interromper a produção de energia.

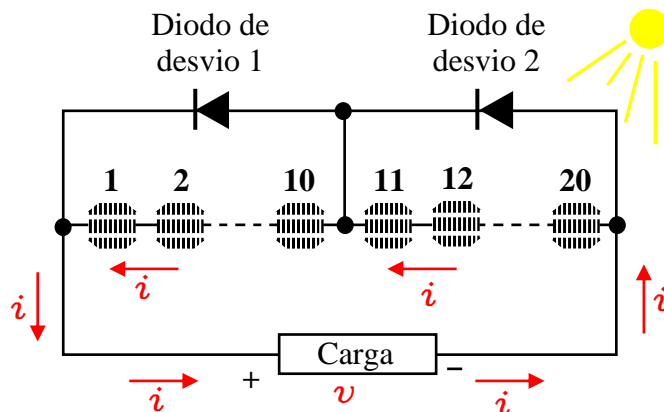
Dentre as possíveis condições que podem surgir nos SFV, são abordados neste trabalho somente a sujidade e o sombreamento.

2.2.1 Sombreamento

É uma das condições mais simples e corriqueiras que podem ocorrer em SFV, pois, pode surgir de forma natural, ou até mesmo por erro no projeto. Ela surge quando parte ou todo MFV é encoberto, podendo ocorrer apenas em um MFV ou em vários ao mesmo tempo. Conforme Pignaton (2022) o sombreamento pode ser estático ou dinâmico. Ele é chamado estático quando a sombra sobre a superfície dos módulos é constante, sendo geralmente causada por pequenos objetos depositados sobre um ou mais MFV. Já o sombreamento dinâmico é causado pelo ângulo do sol ao incidir seus raios sobre edificações, árvores, torres, entre outros, tendo como característica o deslocamento da sombra ao longo do tempo. Desta forma, em ambas as situações, a irradiação solar não consegue incidir plenamente sobre as células FV dos MFV, afetando a geração do sistema. Segundo Madeti e Singh (2017) o sombreamento é a condição que propicia a maior perda de energia elétrica gerada, pois, impacta diretamente a conversão de energia solar em elétrica.

Com o objetivo de mitigar as perdas oriundas dessa condição, os fabricantes de MFV desenvolveram uma caixa de conexão nesses componentes contendo diodos de desvios ou *bypass*, conectados em antiparalelo com as extremidades de um conjunto de células FV (Peroza e Rampinelli, 2018). Na Figura 5 é ilustrado um MFV em funcionamento normal, composto por 20 células conectadas em série, com um diodo de desvio entre cada metade das células e uma carga conectada aos seus terminais. Na Figura 6, é exemplificado o funcionamento do mesmo painel da Figura 5, porém com uma de suas células FV coberta, ilustrando o sombreamento.

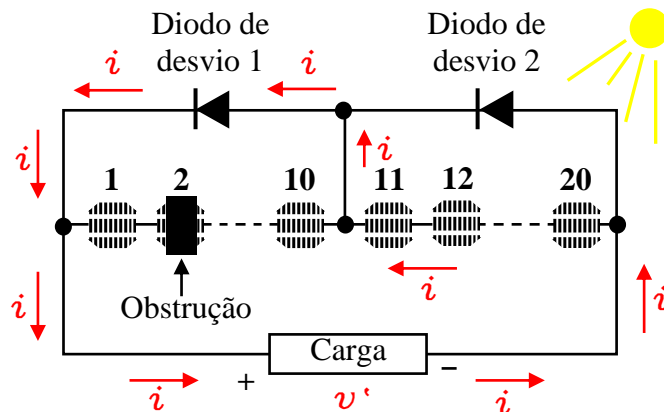
Figura 5 - Representação de módulo fotovoltaico em funcionamento normal.



Fonte: Autoria própria (2024)

Analisando o funcionamento do painel ilustrado na Figura 5, quando a luz solar incide sobre as células FV, cada uma delas gera uma certa tensão elétrica. Como todas estão ligadas em série, o somatório delas é uma tensão (v) que é aplicada aos terminais da carga. As setas indicam o sentido da corrente (i) que percorre as células FV e a carga. Nesse caso, como nenhuma célula FV é coberta, os diodos de desvios não são polarizados e consequentemente não há circulação de corrente por eles.

Figura 6 - Representação de módulo fotovoltaico com uma célula sombreada.



Fonte: Autoria própria (2024)

Conforme ilustrado na Figura 6, na ocorrência de uma obstrução (sombreamento) na célula FV 2, esta passa a se comportar como uma carga de alta impedância, dificultando a passagem da corrente elétrica. Nesse momento, ocorre a polarização do diodo de desvio 1, que está em paralelo com a *string* da célula sombreada, permitindo assim que a corrente elétrica circule através dele. No entanto, a tensão produzida pelo MFV diminui, pois agora, apenas metade das células FV contribuem para a geração de energia elétrica que alimenta a carga.

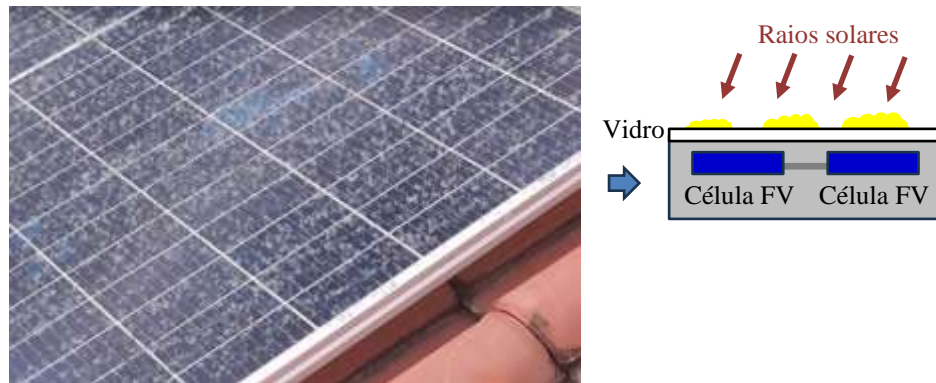
Em alguns casos, outro problema pode surgir em consequência do sombreamento, é o ponto quente ou *hot spot*. Esta condição ocorre quando uma célula FV permanece obstruída por um longo período, resultando com isso o surgimento de uma tensão entre seus terminais, que tende a aquecê-la (Pereira, 2021). Este aquecimento pode não só danificar essa célula FV, mas também outras próximas e, em casos extremos, comprometer ou até mesmo incendiar o MFV, bem como todo o SFV.

2.2.2 Sujidade

A camada de sedimentos acumulada sobre a superfície dos MFV é chamada de sujidade. Ela pode surgir devido ao acúmulo de poeira, poluição e até mesmo excrementos de aves depositados sobre o vidro dos MFV. Estudos conduzidos por Nurjanah *et al.* (2021) indicam que a sujidade absorve ou dispersa a irradiância incidente sobre os MFV, sendo essa deposição influenciada por fatores como chuva, temperatura, umidade, vento e inclinação dos módulos.

Segundo as pesquisas de Hickel (2017) e Silva (2023), a sujidade reduz a capacidade de geração do sistema, provocando perdas econômicas que podem variar entre 1,5% e 6,2% da energia gerada anualmente. A redução na capacidade de produção de energia ocasionada pela sujidade deve ser considerada no projeto de viabilidade financeira do sistema, pois a implantação de uma instalação FV depende também dessa variável (Comerio *et al.*, 2019). Além de impactar negativamente na eficiência energética, a sujidade pode contribuir para a degradação dos MFV, pois os elementos constituintes da sujidade podem causar corrosão, enquanto partículas abrasivas provocam desgaste físico, dessa forma, esses fatores combinados diminuem a vida útil e o desempenho dos MFV. Na Figura 7 é mostrado um MFV real que apresenta uma fina camada de resíduos, e ao lado uma ilustração do efeito da sujidade sobre a superfície dele.

Figura 7 - Módulo fotovoltaico coberto por resíduos.



Fonte: Autoria própria (2024)

A magnitude do impacto da sujeidade sobre o rendimento dos SFV varia significativamente conforme a localização geográfica, em função das condições ambientais específicas do local de instalação e das variações sazonais dessas condições. Esses fatores influenciam diretamente a natureza e a quantidade das partículas de sujeidade depositadas na superfície dos MFV. Conforme relatado por Araújo, Carvalho e Dupont (2019) esta deposição varia não apenas em volume, mas também em tipo e tamanho das partículas, o que afeta substancialmente a eficiência dos MFV.

De acordo com Alves (2018), a sujeidade também pode ser proveniente de partículas minerais, matéria orgânica derivada de biofilmes, suspensões químicas emitidas por escapamentos automotivos e material particulado como salinidade e minérios. Pesquisas realizadas por Romanholo (2019) indicam que substâncias orgânicas transportadas pelo vento e produzidas por microrganismos também contribuem para a sujeidade. Adicionalmente Araújo, Carvalho e Dupont (2019) definem que a sujeidade pode ser composta por uma infinidade de substâncias, sendo a mais comum a poeira, que pode ser qualquer material particulado com menos de 500 μm de diâmetro. A poeira entra na atmosfera por diversas fontes, como vento, escapamento de veículos, erupções vulcânicas e poluição do ar, podendo conter pequenas quantidades de pólen, fungos, bactérias, vegetação, microfibras e, mais comumente, minerais orgânicos, como areia, argila e calcário erodido.

Segundo estudos de Ilse *et al.* (2018) a avaliação detalhada dos efeitos da sujeidade em SFV é um problema complexo devido a muitos fatores influenciadores, como parâmetros ambientais, concentração de poeira no ar, umidade relativa, temperatura ambiente, velocidade e direção do vento. Sem contar que há também propriedades inerentes aos MFV como as propriedades do vidro, características do local de instalação, ângulo de inclinação e orientação do sistema.

2.3 Redes neurais artificiais

As RNA são uma das técnicas de inteligência computacional mais antigas que se tem conhecimento. A primeira publicação relacionada a esse tema é datada de 1943 através do artigo elaborado por Warren Sturgis McCulloch e Walter Pitts intitulado “*A logical calculus of the ideas immanente in nervous activity*”, no qual pesquisadores desenvolveram um modelo matemático inspirado no neurônio biológico que deu partida nos estudos relacionados às RNA que são modelos lógicos-matemáticos que combinam funções aritméticas a fim de solucionar problemas diversos em ciências e engenharia através da imitação do comportamento cognitivo do cérebro humano a partir de dados fornecidos. Haykin (2001) destaca que a propriedade primordial das RNA é sua habilidade de aprender e melhorar seu desempenho através da aprendizagem. No entanto, para isso ocorrer ela precisa ser inicialmente treinada com dados do sistema a ser verificado, só a partir de então, ela passará a desempenhar de forma autônoma seu trabalho.

Nas RNA o neurônio é o elemento central e atua como uma unidade fundamental para o processamento de informações (Manfredini *et al.*, 2021). Dentro da rede, as variáveis externas são tratadas como um conjunto de sinais de entrada, e a cada um é atribuído um peso específico que são valores numéricos utilizados para ponderar a importância de cada entrada. Assim, quando uma rede neural é treinada, o que basicamente ocorre é um processo de ajuste de pesos para que a rede possa mapear corretamente as entradas para as saídas desejadas.

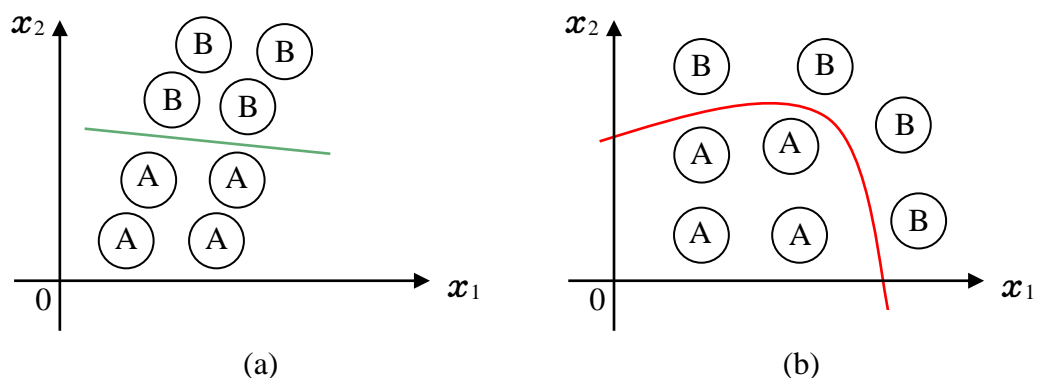
A aprendizagem de uma rede neural pode ser supervisionada; não supervisionada e também por reforço (Faceli *et al.*, 2021). Na forma supervisionada, a rede é treinada com dados rotulados que representam as classes às quais os dados pertencem. Esse processo permite que a rede aprenda a associar entradas específicas a saídas desejadas, facilitando a classificação precisa de novos dados com base no conhecimento adquirido. Na forma não supervisionada, por outro lado, os dados não são rotulados. Assim, a rede categoriza os dados organizando-os com base em semelhanças e padrões identificados durante o processo de treinamento. Essa abordagem permite a descoberta de estruturas ocultas e agrupamentos naturais nos dados, sem a necessidade de intervenção humana para fornecer rótulos. Já na aprendizagem por reforço, também chamado semi-supervisionado, o treinamento é realizado por tentativa e erro, pois a única resposta disponível para uma determinada entrada é se esta é satisfatória ou não. Portanto, quando uma entrada é considerada satisfatória, ajustes graduais são feitos nos pesos e *bias*, para reforçar a associação entre a entrada e o comportamento desejado no sistema.

2.3.1 Perceptron

Em 1957 Frank Rosenblatt idealizou o perceptron, um tipo de RNA baseada no neurônio biológico e utilizado para reconhecer padrões que gerou grandes expectativas entre vários pesquisadores da época, pois esse modelo computacional tinha a capacidade de “aprender”. Esta criação foi considerada um grande avanço tecnológico, mas segundo Haykin (2001), em 1969 pesquisas de Marvin Minsky e Seymour Papert provaram a limitação desse modelo computacional em solucionar a operação lógica “ou exclusivo”, também chamada de “XOR”, além de demonstrarem também que esse modelo computacional não consegue realizar a correta classificação de padrões para classes não linearmente separáveis. Contudo, no final da década de 1980 as RNA voltaram a ser fortemente pesquisadas visto que nesse período surgiram computadores com maior capacidade de processamento, maiores conhecimentos sobre o sistema nervoso biológico e também foram desenvolvidos algoritmos de otimização mais eficientes e robustos (Silva; Spatti; Flauzino, 2016).

Na Figura 8 é ilustrado um conjunto de amostras com classes linearmente separáveis; ou seja, que pode ser separado por uma função linear, como uma reta ou um plano, e também como conjunto de amostras não linearmente separáveis, na qual somente funções complexas conseguem separar as amostras.

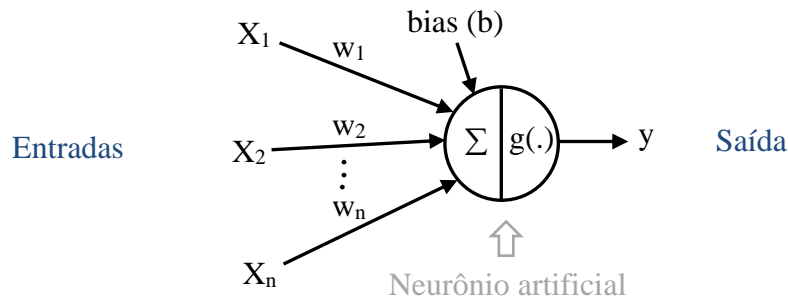
Figura 8 - Classe de amostras. (a) linearmente separáveis; (b) não linearmente separáveis.



Fonte: Autoria própria (2024)

De acordo com Faceli *et al.* (2021), o perceptron é constituído por apenas uma camada de entrada e uma camada de saída, sendo, portanto, a configuração mais simples de uma RNA. Na Figura 9 é ilustrado um perceptron com i entradas e apenas uma saída (y_s).

Figura 9 - Perceptron.



Fonte: Autoria própria (2024)

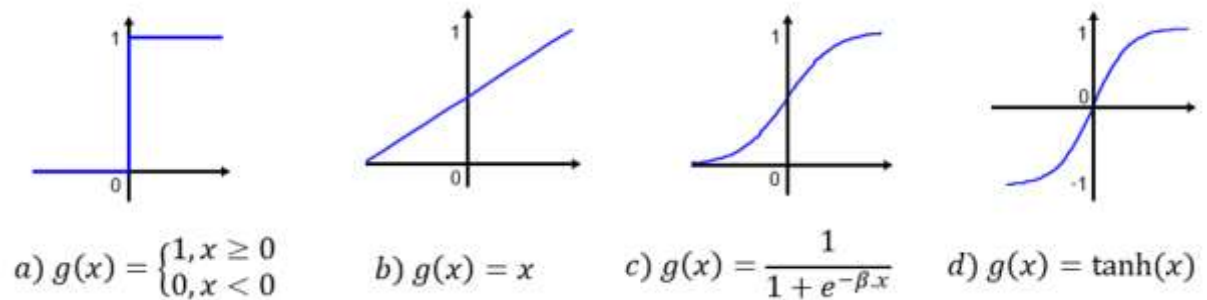
Os valores (x_1, x_2, \dots, x_n) representados na Figura 9 são os sinais externos (dados de entrada) aplicados na entrada do Perceptron, com “ i ” indicando a i -ésima entrada, onde $i = 1, 2, \dots, n$. Os valores (w_1, w_2, \dots, w_n) , são os pesos ou memórias do modelo, que são ajustados durante o processo de treinamento ou aprendizagem. Assim, quando sinais são aplicados nas entradas desse modelo, os produtos $x_i \cdot w_i$ emulam as sinapses do neurônio artificial. O resultado desse produto em cada entrada passa por uma junção aditiva (Σ) que realiza a somatório deles juntamente com o valor do *bias* (b), que de acordo com Haykin (2001) é um parâmetro usado para ajustar a função de ativação e conseqüentemente a saída do modelo. O *bias* (b) também é ajustado durante o treinamento e por conseguinte aplica essa soma numa função de ativação $g(\cdot)$ que por sua vez compara o cálculo com um valor limite (*threshold*) e caso o resultado seja maior ou igual o perceptron dispara. A saída desse modelo é dada conforme as Equação 1.

$$y = g \left(\sum_{i=1}^n x_i \cdot w_i + b \right) \quad (1)$$

Segundo Feltrin (2020) nas RNA uma função de ativação é na realidade uma função matemática que determina se um neurônio deve ser ativado ou não com base na soma ponderada das suas entradas. Adicionalmente, Faceli *et al.* (2021) destacam que a função de ativação introduz não linearidades nas RNA permitindo que elas aprendam relações complexas através dos dados de entrada. Na Figura 10 são ilustradas algumas funções de ativação comumente utilizadas e suas representações matemáticas.

Conforme Faceli *et al.* (2021), a escolha da função de ativação depende do problema que está sendo resolvido e da arquitetura da rede neural utilizada. A função degrau atribui 1 na saída da rede se a ativação for maior ou igual a um certo limite (*threshold*), e 0 caso contrário. Sua forma gráfica é uma transição abrupta de 0 para 1 como ilustrada na Figura 10a.

Figura 10 – Funções de ativação. a) Degrau; b) Linear; c) Sigmóide logística; d) Tangente hiperbólica



Fonte: Autoria própria (2024)

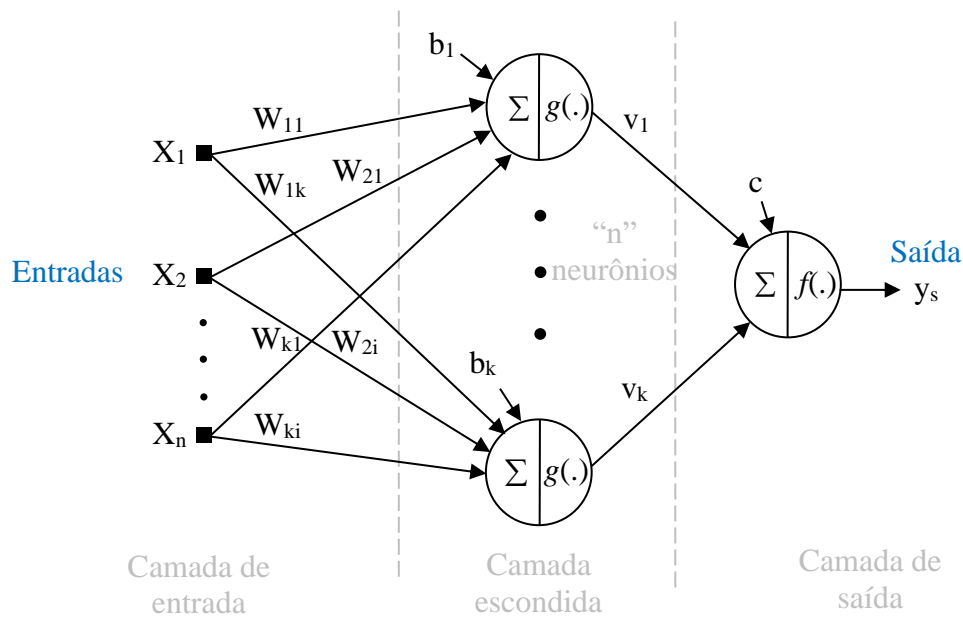
Na função linear a saída (ativação) do neurônio é igual à sua entrada ponderada, sem qualquer transformação não linear como visto na Figura 10b. A função logística vista na Figura 10c produz valores contínuos entre 0 e 1 na saída da rede, sendo uma função suave e diferenciável. Já a função tangente hiperbólica produz valores contínuos entre -1 e 1, sendo semelhante à função logística, mas com saídas centradas em zero, conforme ilustrado na Figura 10d.

2.3.2 Perceptron de múltiplas camadas

Conforme Medeiros (2008), dentre os vários tipos de RNA existentes, a rede perceptron de múltiplas camadas ou *multilayer perceptron* (MLP), tem recebido muita atenção, pois vários pesquisadores provaram que esse tipo de rede com apenas uma camada escondida é um aproximador universal de função. Portanto, permitindo que esses modelos de RNA possam ser utilizadas para resolver diversos problemas em engenharias e ciências aplicadas.

Diferentemente do perceptron, a MLP possui uma ou mais camadas ocultas, cada qual composta por um ou mais neurônios. Nesse tipo de RNA, todos os neurônios de uma camada se conectam com todos os da camada seguinte, sem haver conexão entre neurônios da mesma camada ou de camadas anteriores, conforme ilustrado na Figura 11. Em relação ao perceptron, a MLP também possui maior flexibilidade, podendo ser aplicada numa ampla variedade de tarefas de aprendizado de máquina, como classificação, regressão, previsão, reconhecimento de padrões, entre outras, o que faz desse modelo computacional ser bastante utilizado por toda comunidade científica.

Figura 11 - Perceptron de múltiplas camadas com “n” neurônios.



Fonte: Autoria própria (2024)

Similarmente ao perceptron, os valores (x_1, x_2, \dots, x_n) , com $i = 1, 2, \dots, n$ mostrados na Figura 11, são os sinais aplicados na entrada da MLP que representam as entradas x_i onde i é o índice que percorre todas as entradas do vetor de entrada. Os valores representados pelas letras w_{ki} são os pesos do modelo, e os produtos $x_i \cdot w_{ki}$ emulam as sinapses do neurônio artificial, com o resultado desses produtos sendo somados pelas junções aditiva (Σ) que realizam a somatório deles juntamente com os bias. Nessa MLP, b_k e $g(\cdot)$ são respectivamente o bias e a função de ativação da camada escondida, enquanto, c e $f(\cdot)$ são os da camada de saída. O termo v_k representa os pesos dos neurônios da camada escondida para o neurônio de saída, e por fim, o termo y_s representa a saída da rede, sendo dado conforme disposto na Equação 2.

$$y_s = f \left\{ \sum_{k=1}^j v_k \cdot g \left(\sum_{i=1}^n x_i \cdot w_{ki} + b_k \right) + c \right\} \quad (2)$$

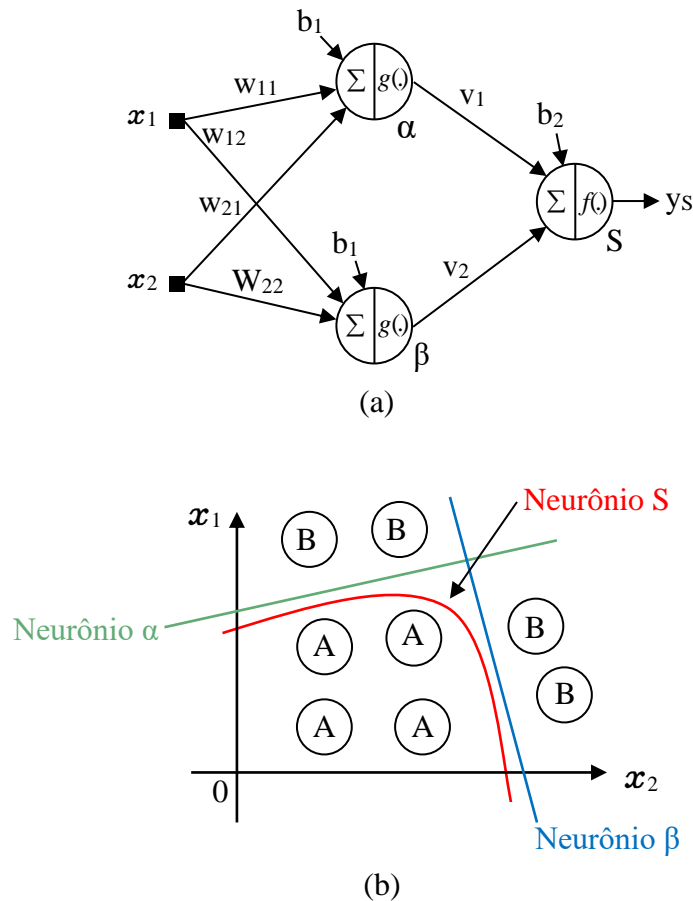
Durante a fase de treinamento da MLP, uma questão muito importante é a estipulação de quantos neurônios devem ser utilizados nas camadas ocultas dessa rede. Conforme Medeiros (2008) as inúmeras regras existentes para determinar o número inicial de neurônios ocultos ignoram o número de exemplos de treinamento, a quantidade de ruído nos dados² e a complexidade da função de mapeamento. Desta forma, o número de neurônios e pode ser

² Qualquer tipo de informação irrelevante ou aleatória que não contribui para a análise dos dados, mas pode distorcer os resultados e reduzir a precisão de um modelo computacional.

definido de maneira empírica, testando várias combinações de número de camadas e de neurônios. Para Haykin (2001) uma MLP com apenas uma camada escondida é suficiente para resolver muitos problemas com não linearidades, como por exemplo, classificação de dados não linearmente separáveis e modelagem de funções não lineares. Deste modo, conclui-se que ao utilizar uma rede MLP pode-se realizar a separação das classes A e B vistas na Figura 8b, pois ela ilustra de dados não linearmente separáveis.

Na Figura 12a é ilustrada a atuação de uma MLP realizando a separação das amostras de classes A e B vistas na Figura 8b. As retas destacadas representam a atuação dos neurônios α e β dessa rede.

Figura 12 – Atuação de um perceptron de múltiplas camadas. (a) MLP com dois neurônios na camada oculta e uma saída; (b) Classificação de amostras não linearmente separáveis.



Fonte: Autoria própria (2024)

Os neurônios α e β são os responsáveis por traçar as retas no espaço de amostras da Figura 12b. O neurônio S tem ação de não linearidade sobre a separação das amostras. Assim, todas as amostras que estão abaixo da reta do neurônio α à esquerda da reta representada pelo neurônio β são rotuladas como Classe A. As outras amostras são rotuladas como Classe B. O

neurônio S recebe as informações da camada oculta e rotula as amostras como Classe A ou Classe B.

Estudos de Feltrin (2020) e Faceli *et al.* (2021) destacam que a não linearidade das redes MLP que é fundamental para sua capacidade de resolver problemas complexos, permitindo assim que a rede aprenda e modele relações complexas e não lineares entre as entradas e saídas, algo que não seria impossível somente com funções lineares.

2.3.3 Máquinas de comitê

Conforme Haykin (2001), a combinação de classificadores especialistas constitui uma máquina de comitê. Essa técnica computacional é uma abordagem que visa melhorar o desempenho de modelos individuais combinando várias previsões de modelos mais simples ou complexos, a fim de realizar generalização dos modelos com maior precisão (Ganaie *et al.*, 2021). Pesquisas de Divina *et al.* (2018) relatam que máquinas de comitê são amplamente utilizadas em problemas de classificação e regressão. Dentre os diferentes modelos computacionais existentes, as máquinas de comitê podem ser compostas por: Árvores de Decisão, KNN (“*K-Nearest Neighbors*”, do inglês, “K-Vizinhos Mais Próximos”), SVM (“*Support Vector Machine*”, do inglês, “Máquina de Vetores de Suporte”) e MLP (*Multilayer Perceptron*, do inglês, “Perceptron de múltiplas camadas”), dentre outras, podendo operar sozinhos ou combinados entre si.

As máquinas de comitê são também conhecidas como *Ensembles*. Elas podem ser classificadas em dois tipos: comitês de estrutura estática ou dinâmica (Barbosa, 2009). Nos modelos estáticos as respostas dos diferentes componentes são combinadas por um mecanismo que não considera o sinal de entrada do sistema. Por outro lado, nos comitês de estrutura dinâmica, a combinação dos componentes leva em consideração a observação dos dados de entradas, de modo que os componentes mais adequados são selecionados e combinados para gerar a saída.

De acordo com estudos de Pelisson (2021) as máquinas de comitê são algoritmos excelentes, pois geralmente superam modelos individuais, fornecendo previsões mais precisas e robustas, especialmente em tarefas complexas e com dados ruidosos. A utilização dessa técnica computacional também reduz o sobreajuste (*overfitting*) do sistema, pois, a combinação de múltiplos modelos tende a fazer com que os erros de uns durante a operação, possam ser compensados pelos acertos de outros. As máquinas de comitê também são muito boas para lidar

com a diversidade de técnicas computacionais para capturar aspectos diferentes dos dados, e com isso explorar o melhor de cada um, aumentando o desempenho. No entanto, para Ganaie *et al.* (2021), devido à alta complexidade dessas técnicas computacionais as vezes a interpretabilidade dos modelos individualmente pode ser sacrificada em favor do desempenho, tornando assim mais difícil entender como as decisões são tomadas no interior dos algoritmos. Portanto, encontrar os parâmetros ideais de ajustes pode ser um desafio.

Haykin (2001) apresenta várias abordagens para máquinas de comitê, mas apesar das diferentes características, todos tem o mesmo objetivo, melhorar o desempenho do processo. Porém, a aplicação de cada um vai depender da natureza do problema e dos tipos de dados que serão colocados como entradas desses algoritmos. Dentre essas abordagens, destaca-se:

- *Bagging*: Envolve o treinamento de várias técnicas computacionais em conjuntos de dados de treinamento diferentes, obtidos por meio de amostragem. Assim, cada modelo produz uma previsão e, no final, as previsões são agregadas, muitas vezes por votação majoritária, para determinar a previsão final. Um exemplo de algoritmos do tipo *Bagging* é o *Random Forest*.
- *Boosting*: É uma abordagem em que os diferentes modelos computacionais são treinados sequencialmente, com cada modelo tentando corrigir os erros do modelo anterior. Dessa forma, os modelos são treinados dando mais peso às instâncias que foram classificadas incorretamente e menos peso às instâncias que foram classificadas corretamente, e após serem treinados todos os modelos utilizados, as previsões individuais deles são combinadas para se obter a previsão final.
- *Voting*: Combina as previsões de vários modelos base por meio de alguma regra de votação, porém deve-se utilizar um mesmo conjunto de dados para treinamento. Pode ser classificado em *Voting Hard* (voto majoritário) e *Voting Soft* (média das probabilidades). Essa abordagem é muito eficaz para aproveitar a diversidade dos modelos, ou até mesmo um único modelo para tomar a melhor decisão final.
- *Stacking*: É uma abordagem em que diferentes modelos são treinados para prever o mesmo resultado, e suas saídas são usadas como entrada para um outro, chamado de modelo de filtragem, que é treinado para aprender a melhor combinação das previsões dos modelos base, permitindo assim, obter o melhor de cada modelo base e melhorar o desempenho computacional.

Para Hu e Hwang (2000); Yang e Ismail (2022) todas essas abordagens para máquinas de comitê são relevantes, pois combinam previsões de múltiplos modelos para reduzir variância, prevenir *overfitting* e corrigir erros iterativamente.

2.4 Trabalhos relacionados

Na literatura científica diversos trabalhos possuem como objetivo principal a realização da identificação de condições operacionais em SFV. Estudos realizados por Dupont (2021) citam pesquisas nessa área envolvendo análise de imagens termográficas, fluorescência ultravioleta, fotoluminescência e eletroluminescência, mas há também trabalhos que utilizam métodos matemáticos baseados na análise de dados históricos de desempenho, métodos estatísticos, e também os que utilizam inteligência computacional.

De acordo com Syafaruddin e Zinger (2019) os métodos computacionais que utilizam técnicas inteligentes para identificar condições operacionais em SFV vem crescendo e ganhando muita popularidade. Além disso, a integração dessas diversas técnicas melhoram a confiabilidade e eficiência dos SFV.

Com o propósito de investigar as técnicas mais prevalentes relacionadas a identificação de condições operacionais em SFV e possíveis lacunas de pesquisa, este estudo fundamenta-se na análise de trabalhos científicos publicados nos últimos cinco anos. É conduzida uma revisão bibliográfica abrangente em diversos repositórios na *internet*, tais como *Web of Science*, *Scopus*, *IEEE Xplore* e *Scielo* por meio do portal de periódicos da Coordenação de aperfeiçoamento de pessoal de nível superior (CAPES). A *string* de busca utilizada é “*fault AND detection AND in AND photovoltaic AND systems*”, um termo em inglês, pois a maioria dos periódicos e conferências internacionais de alto impacto publicam as pesquisas em língua inglesa. Após essa consulta são selecionados 18 trabalhos com acesso aberto para compor a base teórica desta dissertação.

No primeiro trabalho consultado, verificou-se que os pesquisadores Kurukuru *et al.* (2019) utilizaram RNA para analisar imagens obtidas por meio de termografia realizada sobre os MFV, a fim de identificar eventuais problemas nos mesmos, tais como degradação e pontos quentes (*hot spot*), e com isso identificar também quais MFV apresentam essas condições. A rede MLP desenvolvida tem 5 neurônios na camada de entrada, 12 neurônios na camada oculta e 8 neurônios na camada de saída, sendo a seleção dos neurônios da camada oculta e a otimização da configuração da rede alcançadas com a ajuda de um algoritmo de poda. Nesse estudo, é realizada a identificação de sete tipos de condições previamente estabelecidas e

segundo os pesquisadores, foi alcançada acurácia de 92,80%. Pesquisas de Maia (2019) propõem o desenvolvimento de um sistema baseado no algoritmo KNN (“*K-Nearest Neighbors*”, do inglês, “K-Vizinhos Mais Próximos”) para diagnosticar e identificar os módulos afetados pelo sombreamento, tirando vantagem do efeito de descasamento (*mismatch*) dos módulos, que os tornam eletricamente diferentes entre si. Os resultados demonstram ser possível a identificação individual ou em grupos dos módulos sob efeito de sombreamento e do tipo de sombra (parcial ou total) sobre os mesmos. De acordo com o pesquisador a taxa de acertos ficou entre 90,00% e 96,40% a depender do nível de irradiância. Lu *et al.* (2019) propõem um método para identificar curto-circuito, circuito aberto, diodo de desvios abertos ou em curto, problemas de conexão e sombreamento sobre os MFV. O método proposto utiliza redes neurais convolucionais ou *convolutional neural networks* (CNN) para analisar dados sequenciais de transientes no domínio do tempo e identificar características dos *arrays* FV. As correntes e tensões do SFV são transformadas em gráficos bidimensionais (2D) de séries temporais elétricas e submetidas a uma CNN com nove camadas convolucionais, nove camadas de *max-pooling* e uma camada totalmente conectada é utilizada para extração de características e classificação. A CNN é treinada e testada com dados reais e simulados do SFV para validar a precisão e eficácia do modelo. De acordo com os pesquisadores a rede alcançou uma acurácia média superior a 99,00% nos estudos de caso realizados, ressaltando que o sistema proposto é eficaz e confiável em comparação com outros métodos de diagnóstico de SFV.

Estudos de Aziz *et al.* (2020) propõem o desenvolvimento de um sistema de identificação de problemas nas conexões nos MFV, na alimentação dos inversores e também na detecção do ponto de máxima potência ou *maximum power point tracking* (MPPT) do SFV, utilizando uma rede CNN. É utilizada a Transformada Contínua de *Wavelet* (CWT) para converter dados de uma dimensão (como irradiância, temperatura, corrente de curto-circuito, tensão de circuito aberto, corrente do sistema no lado CC, corrente no ponto de máxima potência, tensão no ponto de máxima potência e potência no ponto de máxima potência) em escalogramas em formato bidimensional (2D) que são representações visuais que capturam as características de sinais ao longo do tempo combinando tanto a informação temporal quanto a frequência. Segundo os pesquisadores, o sistema desenvolvido tem alta precisão na identificação das condições elencadas e possui grande capacidade de lidar com dados de entrada de alta dimensionalidade. A acurácia encontrada nesta pesquisa foi de 98,50%, que é uma taxa de acerto muito alta e indica que a abordagem proposta é muito eficaz. Já Basnet, Chun e Bang (2020) apresentam um modelo inteligente de identificação de condições nos MFV, como problemas nas conexões e arcos elétricos, além do desempenho geral do SFV usando análise

de assinatura elétrica, termografia, reflectometria, MLP e redes neurais probabilísticas. O modelo proposto utiliza técnicas de mineração de dados e redes neurais probabilísticas para treinar um modelo para prever e classificar com precisão as condições verificadas. Estudos de Costa (2020) propõem o desenvolvimento de um método de identificação de curto-circuito, degradação, circuito aberto e sombreamento em SFV. Neste trabalho, é utilizado o algoritmo KNN, pois, segundo o autor, é um dos classificadores fundamentais de aprendizado de máquina, e por se tratar de um algoritmo simples, permite que se avalie se é ou não necessário o investimento em classificadores mais complexos para a classificação realizada. Neste caso, a acurácia do teste atinge 92,08%.

Pesquisas de Alves *et al.* (2021) apresentam o desenvolvimento de um sistema para identificar e classificar condições de sombreamento nos MFV utilizando RNA do tipo CNN. Nesse estudo são classificados até 11 anomalias diferentes em MFV por meio de imagens termográficas em um conjunto de dados. É obtida uma precisão de 78,85% para classificar defeitos em oito classes. Gao (2021) desenvolveu um algoritmo com duas redes neurais diferentes, uma RNA do tipo MLP e outra do tipo CNN para detectar e distinguir diferentes tipos de condições num modelo de SFV. No trabalho desenvolvido por Patil *et al.* (2021) foram desenvolvidas redes neurais convolucionais para identificar arcos fotovoltaicos em microrredes CC. Os autores afirmam que a metodologia proposta apresenta desempenho satisfatório na identificação desses problemas. Estudos de Soffiah, Manoharan e Deepamangai (2021) abordam o desenvolvimento de uma rede neural artificial MLP com um neurônio de entrada, um neurônio de saída e uma camada oculta com nove neurônios no *software* Matlab para identificar curto-circuito em MFV. Nesse estudo os autores destacam a importância do monitoramento e manutenção eficientes dos MFV para garantir seu desempenho ideal. A acurácia do treinamento é 93,70%, enquanto no teste alcança 93,30%, já na validação o índice chega a 96,30 %.

Estudos de Prakash e Kumar (2022) apresentam uma RNA do tipo MLP para previsão e identificação de condições operacionais em SFV. A RNA é treinada utilizando um algoritmo de aprendizado por retropropagação dos erros e os parâmetros do sistema foram otimizados por meio de experimentação. A saída da RNA foi utilizada para classificar o sistema como defeituoso ou normal. O artigo também discute o uso de classificadores como máquinas de vetores de suporte (SVM) para categorizar os sistemas defeituosos e não defeituosos com base em atributos de limite. Segundo os autores foram obtidas acurácias superior a 95,00%. Pesquisas de Yang e Ismail (2022) propõem desenvolver um algoritmo do tipo *random forest* para analisar componentes de um SFV e identificar a ocorrência de sombreamento. Segundo os

autores as taxas de precisão foram de 99,88% e 99,43% para os cenários analisados. Já Hong e Pula (2022) desenvolveram uma CNN do tipo 3D, ou seja, onde os sinais de corrente contínua e corrente alternada são convertidos em imagens 3D, a fim de identificar essas condições de sombreamento em SFV. Em termos de resultados, segundo os autores a rede supera outros métodos de inteligência computacional como KNN, *random forest*, árvore de decisão e máquina de vetores de suporte (SVM). Al-katheri *et al.* (2022) desenvolveram uma RNA para identificar condições de sombreamento parcial, circuito aberto e problemas na caixa de diodos de desvios do SFV. Segundo os autores a rede teve acurácia superior a 95,00%.

Pesquisas de Dhimish e Tyrrell (2023) apresentam o desenvolvimento de redes MLP para identificar problemas nos diodos de desvios dos MFV. Segundo os pesquisadores, a rede desenvolvida obteve acurácia de 96,40% e 92,60%, respectivamente, para diodo em curto-circuito e aberto. Estudos de Eskandari, Milimonfared e Aghaei (2023) apresentam uma máquina de comitê para identificar faltas no lado de CC de um SFV utilizando as características da curva corrente/tensão (I-V). Nessa pesquisa são utilizados os algoritmos Naive Bayes, SVM e KNN, resultando em uma acurácia de 99,00%, conforme relatado pelos autores. Já Mellit *et al.* (2023) apresentam um sistema embarcado para monitoramento remoto e diagnóstico de condições operacionais como sujidade e sombreamento em SFV usando máquina de comitê. De acordo com os pesquisadores, o resultado é satisfatório e eles pretendem no futuro testar esse sistema em outros países. Por fim, trabalhos de Obaidi e Derbel (2023) apresentam o desenvolvimento e implementação de um algoritmo de aprendizado profundo (*deep learning*) para identificar condições de sombreamento em SFV usados para bombeamento de água. Nesse trabalho são utilizados sensores de tensão e corrente CC, temperatura e radiação solar para monitorar um sistema de bombeamento FV. Os valores medidos por esses sensores são convertidos em gráficos de séries temporais, em seguida é utilizada a transformada contínua de *Wavelet* que decompõe o sinal original em uma função *wavelet* para diferentes escalas (frequências) e tempos, produzindo coeficientes que representam a correlação entre a *wavelet* e o sinal, em seguida os coeficientes resultantes são organizados em uma matriz 3D (três dimensões), onde um eixo representa o tempo, outro eixo representa a frequência (ou escala) e o terceiro eixo representa a amplitude dos coeficientes da transformada *wavelet*. Segundo os pesquisadores, o sistema alcançou 100% de acertos na identificação de sombreamento.

Na Tabela 1 é mostrado um resumo das principais características observadas nos diversos trabalhos citados neste capítulo.

Tabela 1– Trabalhos relacionados a identificação de condições operacionais em SFV.

| Trabalho | Autor(res) | Título | Objetivo da pesquisa | Pontos interessantes | Crítica(s) | Resultados |
|----------|-------------------------------|--|---|---|--|--|
| 1 | Kurukuru <i>et al.</i> (2019) | Fault Classification for Photovoltaic modules using Thermography and Image Processing. | Desenvolver RNA MLP (5E/12CE/8S) combinada com a técnica de termografia para identificar MFV com problemas. | A combinação da termografia infravermelha, detecção de bordas e Transformada de Hough para reconhecer e classificar problemas nos módulos FV é algo inovador. | O trabalho não especifica se o sistema foi testado e validado com uma variedade de condições ambientais. | A pesquisa obteve uma acurácia de treinamento e teste de aproximadamente 94,0% e 93,1% |
| 2 | Maia (2019) | Estudo de modelos e técnicas de detecção e diagnóstico da condição do SFV. | Desenvolver algoritmo KNN para diagnosticar e identificar os módulos afetados pelo sombreamento e mismatch. | O estudo aborda especificamente as condições de sombreamento, um aspecto crítico na eficiência dos SFV, e explora o efeito de descasamento entre os módulos. | A pesquisa não detalha as dificuldades enfrentadas durante a implementação do modelo KNN, nem discute as limitações ou potenciais melhorias. | Taxa de acertos variando entre 90% e 96,4% a depender do nível de irradiância |
| 3 | Lu <i>et al.</i> (2019) | Fault diagnosis for photovoltaic array based on convolutional neural network and electrical time series graph. | Desenvolver um sistema utilizando redes neurais convolucionais para identificar arrays FV com curto-circuito, circuito aberto, diodo abertos, conexão com problemas e sombreamento. | A pesquisa inova ao transformar dados sequenciais em 1D de arrays FV em gráficos de séries temporais elétricas em 2D para diagnóstico do SFV. | Faltou detalhar comparar o sistema desenvolvido com mais modelos computacionais, a fim de se comprovar a eficácia. | O sistema alcançou uma acurácia média superior a 99% nos estudos de caso realizados. |

| Trabalho | Autor(res) | Título | Objetivo da pesquisa | Pontos interessantes | Crítica(s) | Resultados |
|----------|----------------------------|--|---|---|---|--|
| 4 | Aziz <i>et al.</i> (2020) | A Novel Convolutional Neural Network-Based Approach for Fault Classification in Photovoltaic Arrays. | Propor uma abordagem baseada em redes neurais convolucionais para a identificar condições operacionais em SFV com o objetivo de melhorar a eficiência e a vida útil dos sistemas FV, permitindo a detecção precoce e precisa de condições diversas. | A pesquisa se destaca pelo uso de redes neurais convolucionais bidimensionais para analisar curto-circuito, inversão do diodo by-pass, circuito aberto, inversão de módulo, sombreamento, alta resistividade nas conexões dos cabos e problemas de aterramento. | Sistemas que realizam muitas atividades ao mesmo tempo podem ter sua eficiência reduzida, como é o caso nessa pesquisa. | A rede proposta alcançou uma alta acurácia de 73,53%. |
| 5 | Basnet, Chun e Bang (2020) | An Intelligent Fault Detection Model for Fault Detection in Photovoltaic Systems. | Desenvolver um modelo inteligente de identificação de condições para evitar problemas como conexões folgadas e arcos elétricos, usando análise de assinatura elétrica, termografia, MLP e redes neurais probabilísticas. | O sistema verificou as condições operacionais do SFV pesquisado em diferentes cenários ambientais, especialmente no inverno. | O estudo é limitado pela especificidade dos dados de entrada. (São poucos dados informados). | O sistema obteve acurácia de 100 % no estudo proposto. |
| 6 | Costa (2020) | Classificação de falhas em plantas fotovoltaicas usando aprendizado de máquina. | Desenvolver algoritmo KNN para identificar curto-circuito, degradação, circuito aberto e sombreamento | Foi bem inovador o desenvolvimento do algoritmo KNN para identificação de condições operacionais em SFV. | Não é possível avaliar a robustez e a aplicabilidade prática do trabalho. | Acurácia do teste foi de 92,08%. |

| Trabalho | Autor(res) | Título | Objetivo da pesquisa | Pontos interessantes | Crítica(s) | Resultados |
|----------|---|--|--|---|--|--|
| 7 | Gao (2021) | PV Array Fault Detection Based on Deep Neural Network. | Desenvolver algoritmo de duas redes neurais diferentes, uma RNA do tipo MLP e outra do tipo Convolutacional (CNN) para identificar condições operacionais num SFV. | Trabalharam com MLP e CNN ambas com duas entradas (Potência e Irradiância) e 4 saídas. | Utilizou MLP com muitos neurônios (3 camadas ocultas com 500, 300, 200 neurônios). A CNN tem 80, 160 camadas convolucionais. | Acurácia de 97,16% para a MLP e 99,12% para a CNN. |
| 8 | Patil <i>et al.</i> (2021) | Arc Fault Detection in DC Microgrid Using Deep Neural Network. | Desenvolver redes neurais MLP e Convolucionais, e compará-las para identificar arcos voltaicos em microrredes CC. | O desempenho das CNN é superior em ambientes ruidosos. | Não é informado dados adicionais sobre a implementação e a validação do modelo, especialmente em condições de campo reais. | Os valores de acurácia para diferentes tipos de condições estão entre 72% e 86% para a MLP e CNN, com a convolutacional ficando à frente da MLP. |
| 9 | Soffiah, Manoharan e Deepamangai (2021) | Fault Detection in Grid Connected PV System using Artificial Neural Network. | Desenvolver RNA MLP para identificar curto-circuito em MFV. | A MLP é treinada com o Algoritmo de Levenberg-Marquardt (LMA) com 2709 dados amostrais, abrangendo condições normais e defeituosas. | A suposição de que todo o módulo está em curto-circuito pode não representar sempre as condições do mundo real, onde podem ocorrer outras situações. | A acurácia do treinamento foi de 93,7%, no teste foi 93,3% e na validação foi 96,3%. |

| Trabalho | Autor(res) | Título | Objetivo da pesquisa | Pontos interessantes | Crítica(s) | Resultados |
|----------|----------------------------|--|---|--|--|---|
| 10 | Alves <i>et al.</i> (2021) | Automatic fault classification in photovoltaic modules using Convolutional Neural Networks. | Desenvolver RNA tipo Convolutacional (CNN) para classificar anomalias, entre até onze classes diferentes, em MFV por meio de imagens termográficas. | O modelo foi capaz de classificar com precisão razoável defeitos como problemas no diodo de desvio, ponto quente e rachaduras nos painéis. | A acurácia média de teste de 59,53% para a classificação de defeitos específicos sugere que o modelo pode ser melhorado. | A acurácia de teste foi estimada em 92,5% para a detecção de anomalias nos módulos e 78,85% para a classificação dos oito tipos de classes de defeitos estipulados. |
| 11 | Prakash e Kumar (2022) | Artificial Neural Network Based Fault Prediction and Detection in Grid Computing. | Desenvolveram redes MLP e SVM para identificar condições operacionais em SFV. | São utilizadas as redes MLP e SVM. | O estudo fala sobre identificar condições operacionais, mas não detalha quais são elas. | Foram obtidas previsões e detecções precisas e confiáveis. |
| 12 | Yang e Ismail (2022) | Voting-Based Ensemble Learning Algorithm for Fault Detection in Photovoltaic Systems under Different Weather Conditions. | Desenvolver um algoritmo do tipo <i>Random Forest</i> para analisar componentes de um SFV e identificar o sombreamento. | Trabalharam com Máquinas de Comitê realizando votação simples. | Faltou maiores detalhes do desenvolvimento da rede neural, da base de dados e do treinamento. | Taxas de precisão de 99,88% e 99,43% para os cenários analisados |

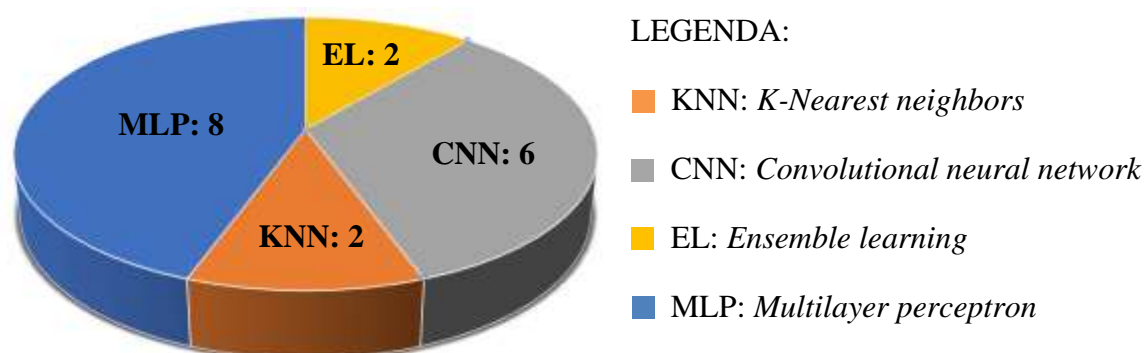
| Trabalho | Autor(res) | Título | Objetivo da pesquisa | Pontos interessantes | Crítica(s) | Resultados |
|----------|---|---|--|--|---|--|
| 13 | Hong e Pula (2022) | Detection and classification of faults in photovoltaic arrays using a 3D convolutional neural network. | Desenvolver RNA tipo Convolutacional (CNN) do tipo 3D, ou seja, os sinais de corrente contínua e corrente alternada são convertidos em imagens 3D, para identificar condições operacionais em SFV. | É realizada comparação do desempenho do modelo proposto com outras técnicas computacionais. | Poderia ter sido incluso na pesquisa uma análise mais detalhada sobre o impacto no desempenho do sistema. | Supera outros métodos de inteligência computacional como, KNN, Random Forest, Árvore de decisão e Máquina de vetores de suporte (SVM), pois atinge 100% de Acurácia. |
| 14 | Al-katheri <i>et al.</i> (2022) | Artificial Neural Network Application for Faults Detection in PV Systems. | Desenvolver rede neural Convolutacional (CNN) para identificar sombreamento parcial, curto circuito, circuito aberto, degradação, diodos de desvio abertos. | Abordagem é focada na identificação de uma variedade de condições. | A identificação de degradação e híbridas (várias condições momentaneamente) teve baixa eficiência, logo, requer aprimoramento do algoritmo. | A identificação de condições de sombreamento, circuito aberto e diodo de desvio aberto são de 100% e degradação 42,6% e híbridas 55,2%. |
| 15 | Dhimish e Tyrrell (2023) | Photovoltaic Bypass Diode Fault Detection Using Artificial Neural Networks. | Desenvolver redes neurais MLP para identificar diodos de desvios abertos. | Trabalharam com MLP (3E/9CE/6S) Informaram como foram encontrados neurônios intermediários (Testes: 1-100) | Faltou maiores detalhes do desenvolvimento da rede neural, da base de dados e treinamento | A MLP obteve acurácia de 96,4% e 92,6%, respectivamente, para diodo em curto e aberto. |
| 16 | Eskandari, Milimonfared e Aghaei (2023) | Line-line fault detection and classification for photovoltaic systems using ensemble learning model based on I-V characteristics. | Desenvolver Máquina de comitê para identificar faltas na linha utilizando as características da curva I-V. Os algoritmos de aprendizagem usados foram SVM, Naive Bayes e KNN. | Trabalharam com Máquinas de Comitê realizando Votação baseada em média. | Faltou maiores detalhes do desenvolvimento da rede neural, da base de dados e treinamento. | Foi obtida na pesquisa uma Acurácia de 99 %. |

| Trabalho | Autor(res) | Título | Objetivo da pesquisa | Pontos interessantes | Crítica(s) | Resultados |
|----------|-----------------------------|--|--|---|--|--|
| 17 | Mellit <i>et al.</i> (2023) | An embedded system for remote monitoring and fault diagnosis of photovoltaic arrays using machine learning and the internet of things. | Desenvolver um sistema embarcado para monitoramento remoto e identificação de condições em SFV usando máquina de comitê. | A pesquisa combina inteligência artificial com a Internet das Coisas (IoT) para monitoramento remoto e diagnóstico do SFV. | Faltou realizar testes do sistemas em diferentes condições ambientais. | A taxa de acertos é de 97,5%, enquanto a taxa de classificação é de 96,8% |
| 18 | Obaidi e Derbel (2023) | IoT-based monitoring and shading faults detection for a PV water pumping system using deep learning approach. | Desenvolver e implementar um algoritmo de aprendizado profundo (<i>deep learning</i>) para identificar condições de sombreamento em um SFV de bombeamento de água. | Foram utilizados sensores de tensão e corrente CC, temperatura e radiação solar para monitorar e detectar sombreamento em um sistema de bombeamento FV. | A pesquisa não especifica se o sistema foi testado em condições climáticas diversas. | A abordagem proposta mostra 100% de eficiência na identificação de sombreamento. |

Fonte: Autoria própria (2024)

Com base na análise detalhada dos trabalhos listados na Tabela 1, é desenvolvido o gráfico ilustrado na Figura 13. Nele é apresentado uma síntese das abordagens computacionais empregadas nas 18 pesquisas examinadas. Observa-se que as redes neurais do tipo perceptron de múltiplas camadas predominam no uso, seguidas pelas redes neurais convolucionais, e pelos algoritmos KNN e máquinas de comitê.

Figura 13 – Técnicas computacionais verificadas em pesquisa bibliográfica.



Fonte: Autoria própria (2024)

Este estudo revela que a utilização de RNA do tipo MLP constitui uma tendência metodológica predominante no campo de pesquisa voltado para a identificação de condições operacionais em SFV. Ademais, identifica-se como uma lacuna significativa a necessidade de robustez dos sistemas para diagnóstico em SFV, especialmente considerando a diversidade de variações climáticas que podem afetar seu desempenho. Neste contexto, é proposto o desenvolvimento de uma metodologia computacional baseada em redes MLP, constituindo máquinas de comitês dedicadas à análise de desempenho que ofereçam diagnósticos precisos e confiáveis acerca das condições enfrentadas por um SFV conectado à rede elétrica.

2.5 Conclusões parciais

Neste capítulo é apresentada a fundamentação teórica sobre sistemas fotovoltaicos, abordando conceitualmente o que são sistemas isolados, híbridos e conectados à rede, bem como condições de sujidade e sombreamento. Este capítulo também explora a aplicação de redes neurais e outras metodologias computacionais para a identificação de condições operacionais em SFV.

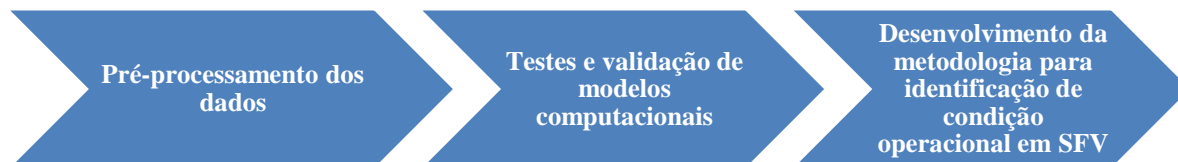
No capítulo 3 será abordado a parte de materiais e métodos utilizados/desenvolvidos para realizar identificação de condições de sujidade e sombreamento no SFV pesquisado.

3 MATERIAIS E MÉTODOS

Neste capítulo é apresentada a metodologia para realizar identificação de condição operacional em SFV conectado à rede elétrica.

A metodologia utilizada é desenvolvida em três etapas: Na primeira é realizado o pré-processamento dos dados disponíveis para melhorar a qualidade deles, visto que um conjunto de dados pode conter valores incorretos, inconsistentes, duplicados ou até mesmo ausentes (Faceli *et al.*, 2021). Na segunda etapa são utilizados e testados diferentes modelos computacionais baseados em RNA, a fim de verificar qual apresenta melhor desempenho com o conjunto de dados pré-processados, em seguida é realizada a validação com todos os dados disponíveis. Na terceira etapa é desenvolvida uma metodologia de classificação sequencial dos dados e a identificação da condição do SFV por meio da verificação majoritária de todas as amostras por um algoritmo do tipo janela deslizante, assim, o modelo proposto pode rejeitar classificações indevidas de curta duração com o objetivo de identificar de forma confiável os rótulos das amostras apresentadas ao sistema de identificação de condições em SFV. Na Figura 14 é mostrado uma visão geral das três etapas da metodologia proposta.

Figura 14 – Metodologia.



Fonte: Autoria própria (2024)

Nas subseções seguintes são detalhadas as etapas da metodologia proposta, mas antes é explicado como foi realizada a obtenção dos dados utilizados.

3.1 Base de dados

A análise das características de uma base de dados viabiliza a identificação de padrões e tendências contidas neles. Nesta pesquisa é utilizado um *dataset* (do inglês, “conjunto de dados”) disponibilizado por Pereira (2018), ex-aluno do Programa de Pós Graduação do IFCE (PPGER). Esse conjunto de dados consiste de 192.094 amostras coletadas em dias que apresentaram similaridade em relação às condições climáticas. Esse conjunto de dados expressa diversas situações de funcionamento de um SFV de 2,16 kWp vinculado ao Laboratório de

Energia Solar Fotovoltaica (LABSOLAR) e ao Laboratório de Processamento de Energia (LPE) do Instituto Federal do Ceará (IFCE), na cidade de Fortaleza-CE, Brasil, coordenadas geográficas com latitude de $3^{\circ}44'39''S$ e longitude de $38^{\circ}32'12''W$.

Esta planta FV é constituída por oito módulos fotovoltaicos (MFV) com potência individual de 270 Wp e interligados em série, formando uma *string* que alimenta um inversor *on-grid* de 3 kW, monofásico, tensão nominal de saída 220 V ~ CA e frequência 60 Hz. Na Figura 15 é ilustrado esse sistema juntamente com seu inversor.

Figura 15 – Sistema fotovoltaico do IFCE.



Fonte: Adaptado de Pereira (2018)

De acordo com Pereira (2018) a coleta das amostras de dados para compor o *dataset* disponibilizado ocorreu por meio da leitura dos sensores de irradiância solar (piranômetro), tensão e corrente contínua (CC) e também vários termistores do tipo *Negative Temperature Coefficient* (NTC) para medição das temperaturas em todos os painéis e caixas de junção desses MFV. Na Figura 16 são mostrados alguns desses sensores.

Figura 16 - Sensores utilizados para leitura de dados.



Fonte: Adaptado de Pereira (2018)

Os registros das amostras foram feitos de forma automática através de um sistema desenvolvido por Pereira (2018) a fim de coletar esses dados e gravar os sinais dos sensores em um arquivo eletrônico no formato “CSV”, com intervalos fixos de um segundo entre cada amostra. Ao final da coleta dessas informações, é constituída uma base de dados com doze

atributos e um rótulo, o qual assume valores de 0 a 5. Na Figura 17 são apresentadas algumas amostras coletadas durante o processo de montagem do *dataset*.

Figura 17 – Amostras da base de dados utilizada.

| Atributos de entrada | | | | | | | | | | | | Classe |
|----------------------|------|-----|-----|------|------|------|------|------|------|------|------|--------|
| Irr | Icc | Vcc | TPM | TD01 | TD02 | TD03 | TD04 | TD05 | TD06 | TD07 | TD08 | RÓTULO |
| 53,67 | 0,40 | 214 | 19 | 19 | 19 | 19 | 19 | 19 | 19 | 18 | 18 | 0 |
| 446,49 | 3,25 | 223 | 27 | 22 | 21 | 22 | 22 | 22 | 24 | 23 | 23 | 1 |
| 805,34 | 7,50 | 214 | 38 | 30 | 27 | 29 | 28 | 26 | 34 | 28 | 28 | 2 |
| 992,03 | 8,50 | 73 | 45 | 34 | 33 | 33 | 32 | 35 | 46 | 34 | 34 | 3 |
| 104,61 | 1,50 | 115 | 25 | 24 | 22 | 23 | 23 | 22 | 24 | 23 | 23 | 4 |
| 724,93 | 4,70 | 235 | 37 | 29 | 24 | 28 | 29 | 31 | 31 | 31 | 31 | 5 |

Fonte: Adaptado de Pereira (2018)

Os atributos de entrada presentes na base de dados originalmente constituída são:

- Irr - Irradiância solar, medida em watts por metro quadrado (W/m^2);
- Icc - Corrente elétrica contínua produzida nos MFV, medida em amperes (A);
- Vcc - Tensão elétrica contínua produzida pela associação série dos 08 (oito) módulos do sistema SFV, medida em volts (V);
- TPM – Temperatura média dos MFV, medida em graus celsius ($^{\circ}C$);
- TD01 a TD 08: Temperatura da caixa de diodos de desvio dos MFV 01 a 08, respectivamente, medida em graus celsius ($^{\circ}C$);

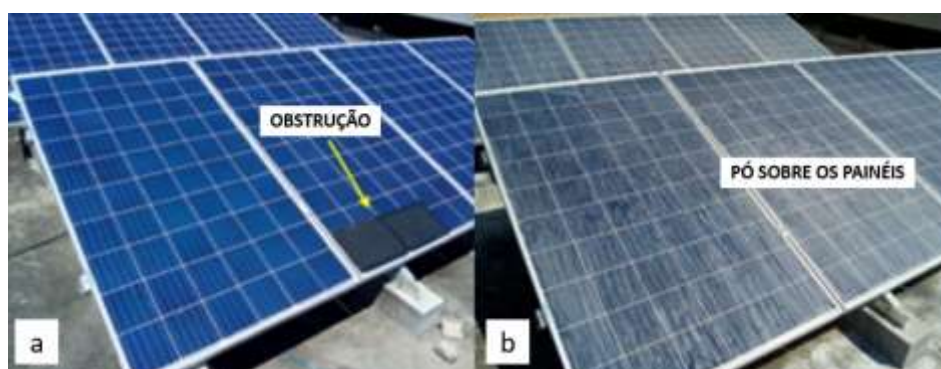
A classe dos rótulos representa a condição acometida pelo SFV durante a coleta de cada amostra de dados, ou seja, a situação dos atributos verificadas no instante em que essas informações são gravadas. Neste caso, cada linha da base de dados corresponde a uma amostra de atributos com seu referente rótulo. A classe de rótulo “0” representa a condição normal de funcionamento do SFV. É a situação do sistema com módulos limpos e sem nenhum sombreamento. Os rótulos de “1” a “4” representam o nível de severidade de sombreamento sofrida pelos MFV durante o funcionamento do SFV. As amostras com rótulo “5”, indicam a presença de sujidade sobre a superfície dos MFV.

As amostras com rótulos de “1” a “4” (sombreamento) e “5” (sujidade) foram obtidas por meio da emulação de diversas situações (Pereira, 2018). Para emular essas amostras que indicam a condição de sombreamento, foi colocada uma obstrução sobre a face dos módulos para se obter dados com essas características. Assim, os rótulos de “1” a “4” representam uma obstrução crescente das células dos MFV. O rótulo “1” foi obtido por meio da emulação do

sombreamento ou células FV com defeitos afetando até um terço de um MFV. O rótulo “2” representa a mesma condição, porém afetando até dois terços de um MFV ou parcialmente dois. No caso de dois MFV, a condição deve afetar até um terço de cada um. O rótulo “3” é a condição de sombreamento ou células FV com defeitos afetando completamente um MFV ou parcialmente três. Para três MFV, a condição deve afetar até um terço de cada um deles. O rótulo “4” representa emulação do sombreamento em quantidades superiores aos níveis atendidos pelo rótulo “3” e representa a condição no final do dia quando os MFV começam a sombrear uns aos outros.

A emulação da condição de sujidade, representada pelo rótulo “5”, foi realizada com a deposição de fécula de mandioca sobre todos os MFV do SFV. Na Figura 18 são ilustradas algumas dessas emulações de condições realizadas na usina FV pesquisada.

Figura 18 - Emulação de condições. (a) Sombreamento; (b) Sujidade.



Fonte: Adaptado de Pereira (2018)

3.2 Pré-processamento

Após análise da base de dados fornecida por Pereira (2018), são identificadas amostras de dados que exibem repetições sistemáticas nos valores dos atributos, bem como outras que apresentam valores inconsistentes. Como exemplo dessas inconsistências, verificou-se a existência de amostras nas quais o atributo de corrente (I_{cc}) exibe valores negativos, sendo que o SFV é constituído somente por um conjunto de MFV ligados em série, portanto essa é uma condição que não ocorre para esse tipo de arranjo FV. Também é observado a presença de amostras que não deveriam integrar a base de dados, como nas situações que não há produção de energia elétrica, conforme evidenciado pelas amostras que exibem um valor nulo para o atributo corrente (I_{cc}) e/ou tensão de entrada abaixo do limiar de partida do inversor FV. Diante dessas considerações, torna-se necessário realizar um processo de pré-processamento dos dados

a fim de aprimorá-los. Na Figura 19 são apresentadas algumas amostras mencionadas com inconsistências.

Figura 19 – Verificação de algumas amostras da base de dados original.

→ Amostras com dados inconsistentes

| Irr | Icc | Vcc | TPM | TD01 | TD02 | TD03 | TD04 | TD05 | TD06 | TD07 | TD08 | RÓTULO |
|--------|--------|-----|-----|------|------|------|------|------|------|------|------|--------|
| 30,74 | 0,00 | 206 | 19 | 19 | 19 | 18 | 19 | 19 | 19 | 18 | 18 | 0 |
| 30,74 | 0,00 | 197 | 19 | 19 | 19 | 18 | 19 | 19 | 19 | 18 | 18 | 0 |
| 30,74 | 0,00 | 196 | 19 | 19 | 19 | 18 | 19 | 19 | 19 | 18 | 18 | 0 |
| 53,67 | 0,30 | 219 | 19 | 19 | 19 | 19 | 19 | 19 | 19 | 18 | 18 | 0 |
| 53,67 | 0,40 | 214 | 19 | 19 | 19 | 19 | 19 | 19 | 19 | 18 | 18 | 0 |
| 873,94 | - 0,30 | 264 | 42 | 33 | 30 | 32 | 31 | 30 | 38 | 31 | 31 | 2 |
| 873,94 | - 0,20 | 264 | 42 | 33 | 30 | 32 | 31 | 30 | 38 | 31 | 31 | 2 |
| 872,30 | - 0,20 | 264 | 42 | 33 | 30 | 32 | 32 | 30 | 38 | 30 | 30 | 2 |
| 992,03 | 8,50 | 73 | 45 | 34 | 33 | 33 | 32 | 35 | 46 | 34 | 34 | 3 |
| 992,03 | 8,60 | 193 | 45 | 34 | 33 | 33 | 32 | 35 | 46 | 34 | 34 | 3 |
| 104,61 | 1,50 | 115 | 25 | 24 | 22 | 23 | 23 | 22 | 24 | 23 | 23 | 4 |
| 873,94 | - 0,09 | 249 | 43 | 33 | 30 | 31 | 31 | 30 | 38 | 31 | 31 | 2 |
| 872,31 | - 0,09 | 248 | 43 | 32 | 30 | 31 | 31 | 30 | 38 | 31 | 31 | 2 |

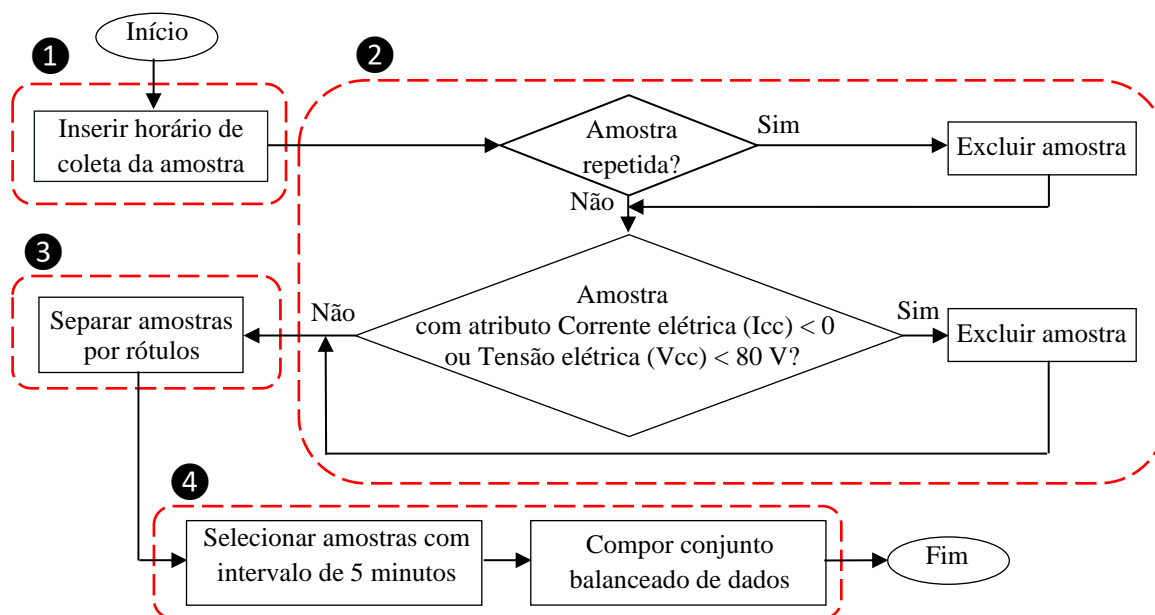
Fonte: Adaptado de Pereira (2018)

Pesquisas realizadas por Lichao *et al.* (2022) conceituam o pré-processamento de dados como uma série de procedimentos executados em um conjunto de amostras de informações com o objetivo de aprimorar sua qualidade. Tais procedimentos englobam a inclusão de rótulos ausentes, a exclusão de informações irrelevantes ou duplicadas, entre outros. O tratamento de dados representa uma etapa de suma importância na garantia da qualidade de qualquer análise preditiva, especialmente quando esta se fundamenta em registros de medições (Pessanha *et al.*, 2020).

Assim, é elaborado um fluxograma com o propósito de conduzir o pré-processamento com todas as amostras de dados disponíveis. A realização desta etapa envolve a utilização do *software* Microsoft Excel em sua versão de 2021, para se realizar a filtragem dos dados, bem como organizá-los, a fim de submetê-los aos modelos computacionais utilizados.

Na Figura 20 é apresentado o fluxograma desenvolvido para realização do pré-processamento de dados, destacando os diversos passos seguidos.

Figura 20 - Pré-processamento dos dados.



Fonte: Autoria própria (2024)

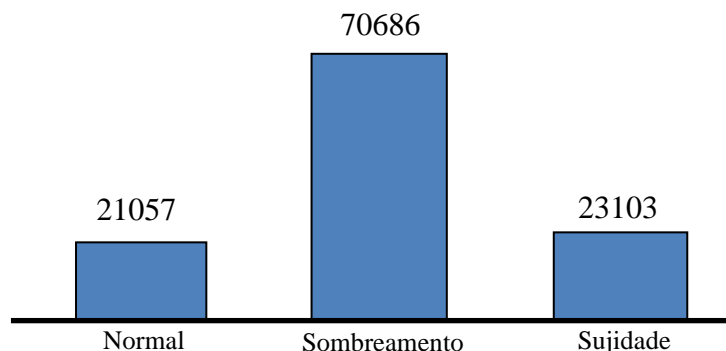
Segundo Reinders *et al.* (2019), Pessanha *et al.* (2020) e Meyers *et al.* (2020), a análise de dados com o seu respectivo horário permite melhor contextualizar e interpretar as informações contidas neles. Assim, seguindo essa etapa, optou-se por realizar no primeiro passo do pré-processamento a inserção do horário de coleta de cada amostra.

De acordo com Lichao *et al.* (2022) amostras redundantes podem comprometer a precisão das análises dos dados. Dessa forma, no passo 2 são excluídas as amostras repetidas e irrelevantes, ou seja, amostras que contém o atributo corrente (I_{cc}) negativa ou nula, e ainda amostras com o atributo tensão (V_{cc}) menor que 80 V, pois, de acordo com o manual do inversor fotovoltaico utilizado (Fronius, 2018), esse é o valor da tensão de partida do referido equipamento. Logo, amostras com tensão abaixo de 80 V não devem influenciar o estado do SFV, pois nesse momento ele está em *stand-by* (do inglês, “estado de espera”).

Após a execução dos passos 1 e 2, verifica-se que das 192.094 amostras contidas no conjunto de dados originalmente repassado por Pereira (2018), restam 114.846 amostras exclusivas. Sendo 21.057 amostras da condição de normalidade (rótulo 0), 22.993 amostras de de sombreamento (rótulo 1), 19.325 amostras de sombreamento (rótulo 2), 19.163 amostras de sombreamento (rótulo 3), 9.205 amostras de sombreamento (rótulo 4) e 23.103 amostras de sujidade (rótulo 5). No entanto, durante esse passo do pré-processamento, opta-se por agregar os dados das classes de “1” a “4”, em um único rótulo, sem fazer distinção entre o nível de sombreamento sobre o SFV. Contudo, ao se fazer isso, surge uma nova demanda, as classes de dados ficam muito desbalanceadas e isso é uma questão importante a ser resolvida, pois

segundo Faceli *et al.* (2023), um algoritmo de classificação quando alimentado com dados desbalanceados, tende a favorecer dados da classe majoritária, e conseqüentemente fornecer respostas errôneas sobre a situação do sistema analisado. Na Figura 21 é ilustrado esse desbalanceamento do quantitativo de amostras de dados, onde percebe-se que agora há muitas amostras referente ao rótulo de sombreamento em relação às classes normal e sujidade.

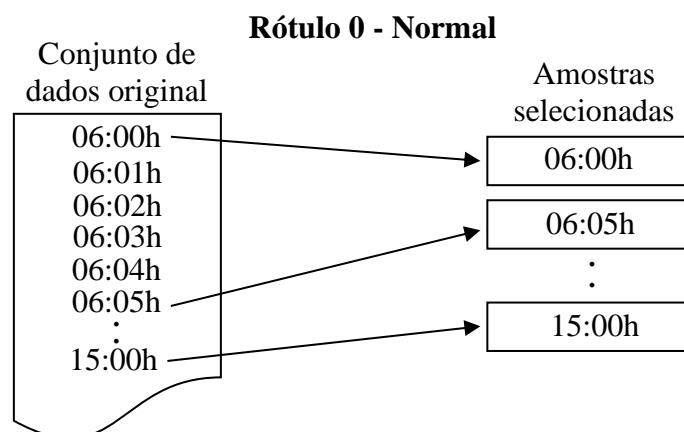
Figura 21 – Distribuição das amostras.



Fonte: Autoria própria (2024)

Dessa forma, antes de agregar os dados das classes de “1” a “4”, seguindo método similar às pesquisas de Reinders *et al.* (2019) e Pessanha *et al.* (2020), respectivamente, no passo 3 as amostras são separadas por rótulos para facilitar o tratamento e no passo 4 são selecionadas com intervalos de cinco minutos para compor um novo conjunto de dados com classes de rótulos balanceadas. Assim, analisando inicialmente os dados da classe normal, verifica-se que elas são coletadas entre o horário de 05:46h e 15:11h de um único dia, logo, resolve-se selecionar amostras dessa classe apenas entre 06:00h e 15:00h, totalizando com isso 109 amostras com rótulos da classe normal. Na Figura 22 é ilustrado como foi realizada a seleção das amostras nesse horário escolhido.

Figura 22 – Seleção de amostras para composição de novo conjunto de dados.



Fonte: Autoria própria (2024)

Seguindo a mesma estratégia, procede-se de forma semelhante com as amostras de sujidade e sombreamento. No entanto, no caso das amostras de sombreamento, a coleta foi realizada de forma escalonada, obtendo-se amostras de cada uma das classes de 1 a 4, mantendo um intervalo constante de cinco minutos entre cada coleta. Dessa forma, são agrupadas 109 amostras rotuladas como normais, 109 amostras rotuladas como sujidade e 109 amostras rotuladas como sombreamento, que constituem um conjunto de dados equilibrado totalizando 327 amostras. Ainda no passo 4, a última atividade consiste em reduzir a dimensionalidade desse novo agrupamento de dados, isto é, linhas de informações. A redução do número de atributos pode aprimorar o desempenho do modelo de classificação utilizado, reduzir o custo computacional e tornar os resultados obtidos mais compreensíveis (Faceli *et al.*, 2023).

De acordo com o estudo realizado por Lichao *et al.* (2022), a eliminação de atributos redundantes ou irrelevantes pode ser conduzida de maneira manual. Assim, inicialmente, procede-se à exclusão manual de todos os atributos relacionados às temperaturas das caixas de diodos de desvio no conjunto de dados, representadas pelos atributos TD01, TD02, TD03, TD04, TD05, TD06, TD07 e TD08, com o objetivo de investigar se medições de temperatura influenciam significativamente na análise dos dados, pois do ponto de vista de engenharia a medição de vários sinais de temperatura torna a instalação e manutenção muito onerosa. Assim, o conjunto de dados pré-processados fica com apenas quatro atributos, Irradiância solar (Irr), Corrente CC (Icc), Tensão CC (Vcc) do SFV e Temperatura média dos módulos fotovoltaicos (TPM). Esses quatro atributos são representados por três classes distintas, a saber: rótulo 1 para amostra “normal” do SFV, rótulo 2 para “sujidade” e rótulo 3 para “sombreamento”, seja ele estático ou dinâmico.

Posteriormente, após realização de testes com esse novo conjunto de dados, opta-se por retirar também o atributo TPM, para reduzir a quantidade de sensores instalados, pois apesar de tratar-se de um único atributo, para se obter os valores de cada amostra é necessário medir a temperatura de todos os MFV. Portanto, o conjunto de dados pré-processado e utilizado em toda a pesquisa passa a ser constituído apenas por dados de Irradiância solar (Irr), Corrente CC (Icc) e Tensão CC (Vcc) do SFV pesquisado.

3.3 Modelos computacionais

Nesta etapa são pesquisados e implementados algoritmos para realizar classificação de dados. Opta-se por utilizar nesse processo a linguagem de programação Python, a qual tem

conquistado um lugar de destaque no meio acadêmico e profissional contemporâneo, especialmente nas áreas de aprendizado de máquina e ciência de dados (Da Silva e Silva, 2019).

O editor de códigos utilizado é o *Google Colaboratory* (Colab) que é uma plataforma *web* gratuita, criada pela área de pesquisas científicas do *Google* e permite qualquer usuário com uma conta ativa, escrever e executar códigos Python num navegador *web* conectado à *internet* (Google, 2021). Dentre as vantagens de executar códigos no Colab, pode-se mencionar: fácil configuração/acesso gratuito ao sistema, integração com o Google Drive e recursos de alto desempenho, pois não é necessário baixar pacotes de atualizações, o que poupa tempo e recursos computacionais do computador utilizado. Todos os algoritmos desenvolvidos durante a pesquisa são armazenados em uma pasta no Drive que pode ser acessada através do link disponibilizado no Apêndice A.

Seguindo o princípio da navalha de Occam, o qual postula que “uma máquina de aprendizado deve ser o mais simples possível para obter um bom desempenho de generalização”, foram implementados inicialmente dois perceptrons de acordo com pesquisas de Rosenblatt (1962). Cada um desses classificadores neurais realiza classificação binária, entre amostras com rótulo “normal” e “sujidade”, bem como entre as dados de rótulo “normal” e “sombreamento”. Na prática, a saída de cada rede pode ser comparada a um circuito lógico que apresenta dois níveis de atuação, um “alto” ou “baixo”, indicando uma ou outra classe respectivamente.

Os perceptrons desenvolvidos foram configuradas com uma taxa de aprendizado de 0,1. A função de ativação escolhida foi do tipo sigmoide logística, pois como a atualização dos pesos depende do gradiente da função de ativação e este tipo de função têm derivadas definidas em todos os pontos, facilita o treinamento da rede. Seguindo abordagem sugerida por Géron (2019), o treinamento realizado foi limitado a 100 épocas, sendo cada perceptron submetido a um processo de treinamento com 70% do conjunto de dados pré-processados e posteriormente testadas com os 30% restantes.

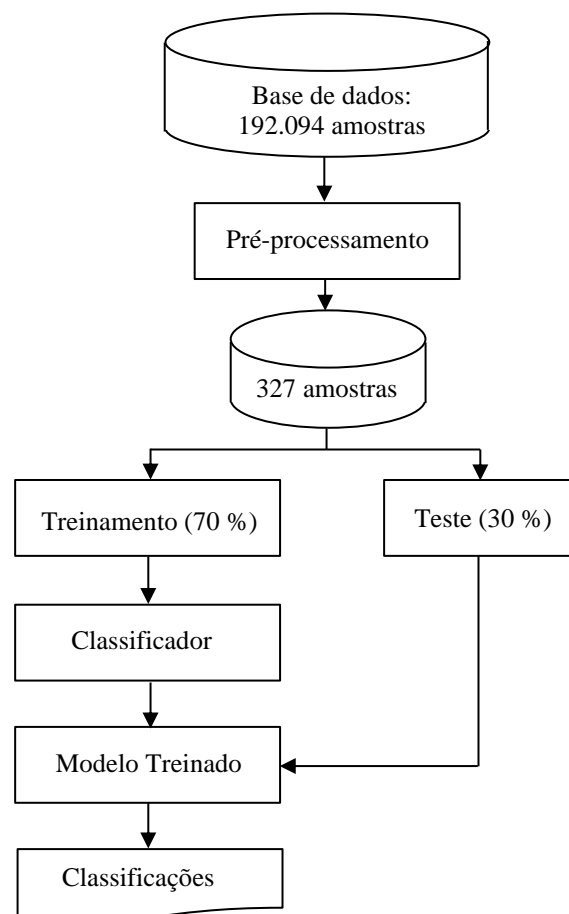
Antes da realização do treinamento e teste, as amostras do conjunto de dados pré-processados são submetidas a um processo de normalização. Essa ação é necessária, pois em termos numéricos, os atributos podem ter amplitudes diferentes uns dos outros, com isso um atributo pode predominar e influenciar a análise dos dados (Faceli *et al.*, 2021). De acordo com Silva; Spatti e Flauzino (2016) os valores padronizados podem ser positivos ou negativos, dependendo se estão acima ou abaixo da média original dos dados. Logo, a padronização mantém a estrutura relativa dos dados, mas ajusta a escala para que todos os atributos tenham

a mesma média e o mesmo desvio padrão, tornando-os comparáveis e adequados. Assim, a normalização realizada nesta pesquisa padroniza os dados de todos os atributos dentro de uma faixa de valores com média zero e desvio padrão um. Este processo de padronização dos dados segue a Equação 3, em que x_i é o valor original do atributo, μ é a média e σ é o desvio padrão e z_i é valor normalizado.

$$z_i = \frac{x_i - \mu}{\sigma} \quad (3)$$

Na Figura 23 é mostrado o fluxograma de realização do treinamento e teste para identificação de condições operacionais no SFV pesquisado. Esta rotina é seguida pelos demais modelos computacionais desenvolvidos.

Figura 23 – Fluxograma de treinamento e teste.

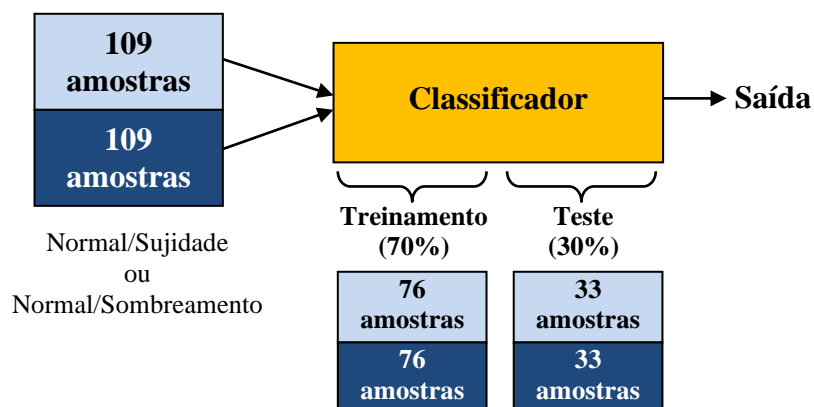


Fonte: Autoria própria (2024)

Para cada classificação com amostras de classes rotuladas como “normal” e “sujidade”, bem como amostras rotuladas como “normal” e “sombreamento” são empregadas 109 amostras de cada classe, totalizando, portanto, 218 amostras no conjunto de dados. Dentre essas amostras, 76 de cada classe são designadas para o conjunto de treinamento, enquanto as 33 amostras restantes de cada classe são alocadas para o conjunto de teste. Isso resulta em um conjunto de

treinamento composto por um total de 152 amostras, acompanhado por 66 amostras no conjunto de teste. Os resultados obtidos a partir das previsões feitas com base nesses dados são minuciosamente analisados e discutidos no Capítulo 4. Na Figura 24 é ilustrado a atuação de um classificador genérico conforme proposto inicialmente neste trabalho, ou seja, com a leitura de dois tipos de amostras de dados e com apenas uma saída.

Figura 24 – Classificação binária de dados.



Fonte: Autoria própria (2024)

Conforme os estudos verificados na Seção de 2.4 desta pesquisa, a qual trata sobre “Trabalhos relacionados”, após testes com perceptrons, são projetadas duas redes perceptrons de múltiplas camadas (MLP) para também realizar classificação binária de dados, seguindo metodologia semelhante à empregada com o perceptron. A utilização da rede MLP possibilita a captura de relações intrincadas nos dados, tornando-a adequada para tarefas de classificação (Silva; Spatti; Flausino, 2016). Deste modo, cada MLP desenvolvida é composta por uma camada de entrada, uma camada oculta com “n” neurônios e apenas um neurônio de saída.

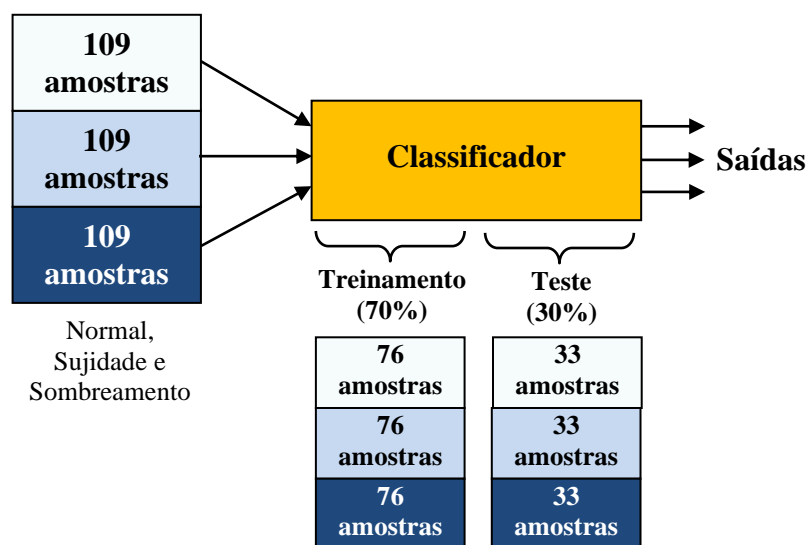
Na camada de entrada são lidos os atributos da base de dados. Já na camada oculta ou escondida, para determinação do quantitativo de neurônios utilizados optou-se por conduzir testes experimentais para determinar empiricamente o número mais apropriado de neurônios para essa aplicação. No capítulo 4 é apresentada uma tabela com os resultados obtidos para determinação do valor de “n” para os neurônios da camada oculta.

Findados os testes com redes perceptrons simples e MLP com uma saída, é projetada uma única rede MLP, para leitura de três amostras distintas, com o mesmo número “n” de neurônios na camada escondida para as MLP anteriores, e com três neurônios na camada de saída. O objetivo dessa nova abordagem é realizar uma classificação multiclasse com todas as 327 amostras do conjunto de dados pré-processado, ou seja, fazer a classificação simultânea

com amostras de todas as três classes no pré-processamento dos dados, a saber: Classe “normal”, “sujidade” e “sombreamento”.

Com esse tipo de classificador com três saídas, cada uma delas representa uma determinada classe, e quando uma amostra de dados é classificada, somente uma dessas saídas vai para nível lógico “alto”, enquanto as outras permanecem em nível lógico “baixo”. Conforme abordagens anteriores, esse modelo também é submetido a treinamento com 70% do conjunto de dados pré-processados e testados com os 30% restantes. Os resultados obtidos com esse classificador são verificados e discutidos no Capítulo 4. Na Figura 25 é apresentada a MLP realizando classificação multiclasse.

Figura 25 – Classificação com três entradas e três saídas.



Fonte: Autoria própria (2024)

Baseado nos resultados obtidos após testes com a MLP da Figura 29, buscou-se implementar outro modelo computacional capaz de melhor generalizar os dados pré-processados para identificação das condições em SFV. Assim, utilizando um princípio muito utilizado em engenharia, “dividir e conquistar” (Haykin, 2001), desenvolveu-se duas máquinas de comitê para verificar o comportamento da generalização dos dados com esse tipo de técnica computacional. Segundo Hu e Hwang (2001), a utilização de máquinas de comitê permite alcançar um desempenho inalcançável por um membro individual do comitê por conta própria. A razão disso decorre no fato de que os erros dos membros individuais do comitê anulam-se, até certo ponto, quando suas previsões são combinadas.

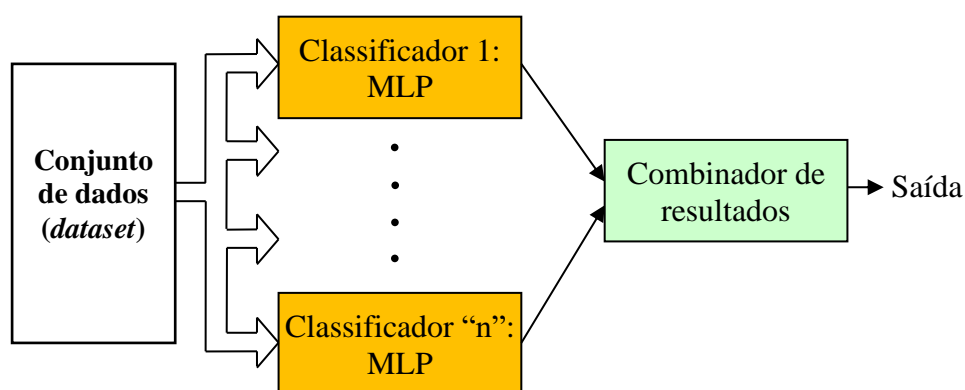
As máquinas de comitê podem ser constituídas de diferentes tipos de classificadores, também chamados de especialistas, como por exemplo, “árvore de decisão”, “MLP”, “*support*

vector machine” (SVM), entre outros, porém, também seguindo o estudo realizado na seção de “trabalhos relacionados”, optou-se por utilizar apenas perceptrons de múltiplas camadas. Desse modo, foram criadas máquinas de comitê com 3 (três), 5 (cinco) e 10 (dez) MLPs para se verificar qual delas tem a melhor generalização dos dados. No capítulo 4 é apresentado esse comparativo.

Independentemente do número de especialistas utilizados, no funcionamento de qualquer máquina de comitê, elas são treinadas com os mesmos dados, porém, os pesos de todas são diferentes entre si, o que possibilita cada modelo generalizar suas respostas de formas diferentes entre si (Hu; Hwang, 2001). Para as máquinas de comitês criadas escolheu-se utilizá-las com a abordagem “*Voting*”, na qual o resultado fornecido por cada classificador é agregado por meio de um elemento combinador, assim, ao término do processo de classificação, o sistema informa como resultado final o rótulo que obtém maior contagem de votos entre os classificadores. Durante os testes também foi idealizado um combinador que calcula a média das previsões de probabilidade dos modelos individuais para cada classe. Um comparativo sobre essa questão é abordado no Capítulo 4.

Na Figura 26 é ilustrada uma máquina de comitê com “n” especialistas e seu combinador de resultados para classificação de dados.

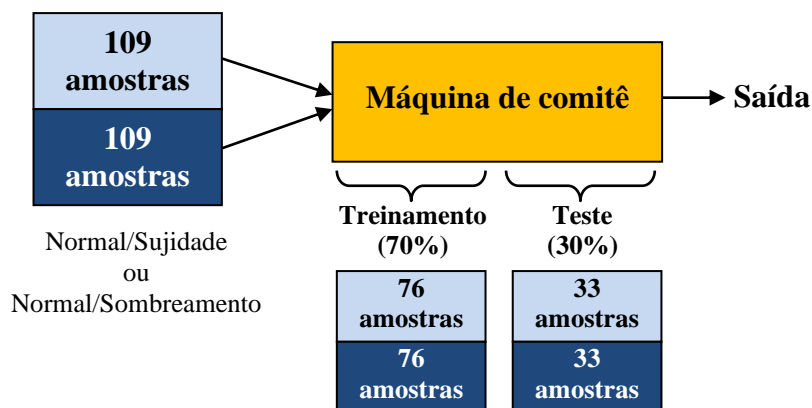
Figura 26 – Máquina de comitê com “n” especialistas.



Fonte: Autoria própria (2024)

Após definição da quantidade de MLP que devem compor as máquinas de comitês utilizadas, são realizados 10 testes com os dados pré-processados para observar a generalização obtida por esses modelos computacionais realizando classificação binária, similarmente ao já realizado com a rede perceptron e MLP atuando sozinho. Na Figura 27 é exemplificada uma máquina de comitê genérica, realizando a classificação entre as classes “normal” e “sujidade”, bem como “normal” e “sombreamento”.

Figura 27 – Classificação binária com máquinas de comitê.



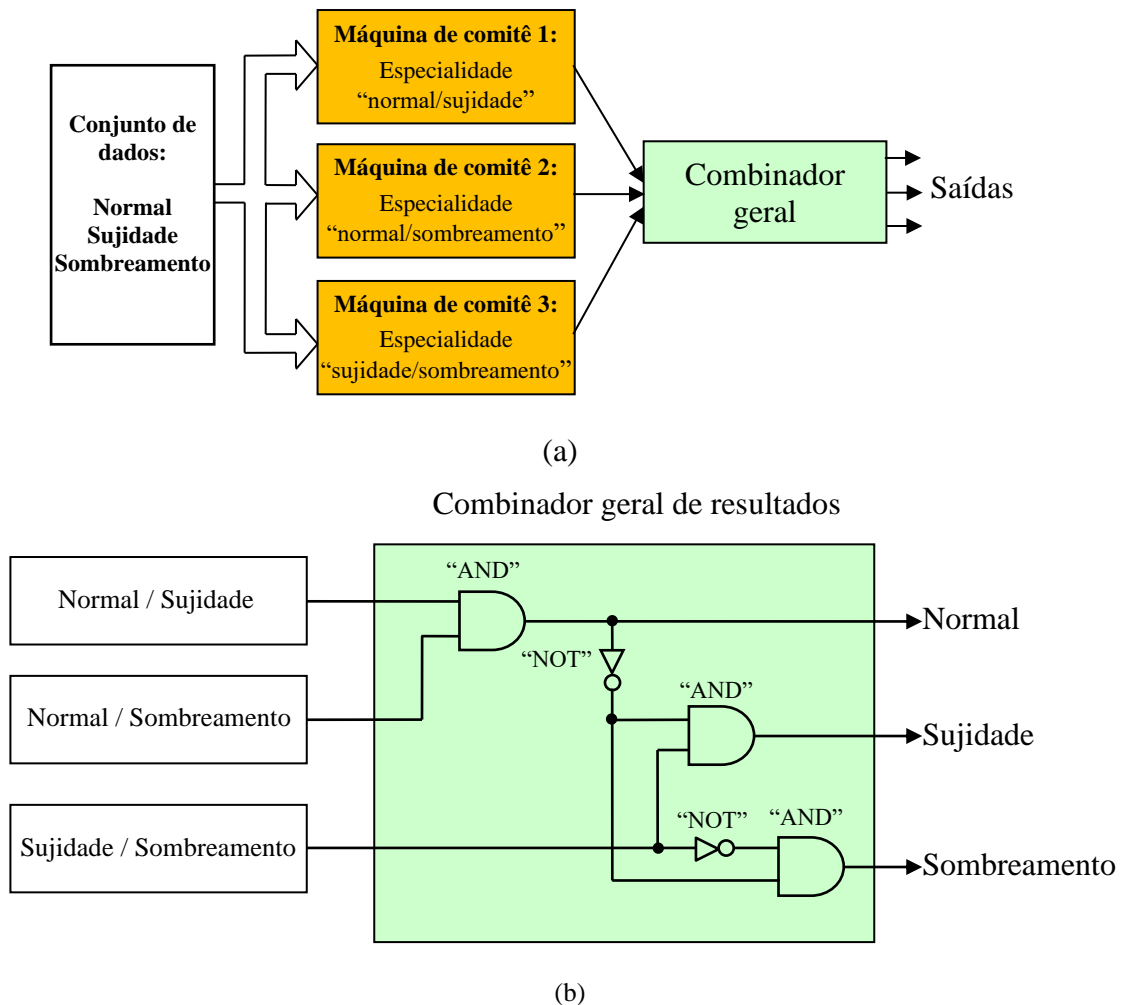
Fonte: Autoria própria (2024)

Realizados testes para classificar amostras anteriores, uma terceira máquina de comitê, similar às anteriores, também é desenvolvida, porém, para trabalhar apenas com as amostras das classes “sujidade” e “sombreamento”. Posteriormente, procede-se à união de todas elas, a fim de constituir um único modelo computacional capaz de classificar todas as amostras do conjunto de dados pré-processado de maneira simultânea. Desta forma, cada máquina de comitê é treinada com os dados relativos à sua especialidade, mas o teste é realizado com amostras das três classes.

A classificação de múltiplas classes de dados, conhecida também como classificação multiclasse, é realizada por meio de comitês de máquinas que realizam classificação binária, seguindo a seguinte lógica de execução: O combinador geral avalia os resultados de cada um dos comitês. Quando os comitês, são treinados com amostras das classes "normal" e “sujidade”, e aqueles treinados com amostras das classes “normal” e “sombreamento”, concordam que a amostra em questão pertence à classe “normal”, o combinador geral classifica a amostra como "normal", no entanto, quando eles discordam quanto à classificação da amostra avaliada, a classe da amostra é definida pelo comitê treinado com amostras das classes “sujidade” e “sombreamento”.

Na Figura 28a são ilustradas três máquinas de comitê trabalhando em conjunto. Para ilustrar de uma outra forma o funcionamento do combinador geral, na Figura 28b é exemplificado na forma de circuito lógico digital, a rotina realizada por esse bloco para classificar uma amostra de dados em qualquer das três possíveis classes do conjunto de dados pré-processado. Para compreendê-lo, basta-se considerar uma das classes informadas no retângulo como sendo a entrada do circuito.

Figura 28 – Classificação multiclasse. (a) Máquinas de comitês binárias com combinador geral. (b) Exemplificação da lógica realizada pelo combinador geral.



Fonte: Autoria própria (2024)

No capítulo 4 são apresentados os resultados obtidos com as máquinas de comitê binárias realizando classificação das amostras com as três classes escolhidas no pré-processamento (classificação multiclasse).

3.4 Sistema de identificação de condições

Realizada a definição do modelo computacional a ser utilizado neste estudo, busca-se desenvolver uma lógica que atue juntamente com a técnica de classificação dos dados, a fim de realizar identificação de condições mais precisa e confiável no SFV pesquisado. Como resultado, é implementada uma rotina que efetua classificação das amostras de forma sequencial, ou seja, à medida que as amostras vão sendo colocadas na entrada da rede elas vão

sendo classificadas, diferentemente de abordagens anteriores, pois a classificação ocorre com amostras aleatórias do conjunto de dados.

Em conjunto com a classificação sequencial, também é utilizado um algoritmo tipo janela deslizante, semelhante à abordagem observada em Suresh *et al.* (2020) com o objetivo de melhorar o desempenho na classificação das amostras. Nesse algoritmo, uma janela de “k” amostras são classificadas antes do sistema de identificação de condições informar o estado do SFV. Após a conclusão da análise dessas “k” amostras, a janela de classificação avança, deslocando-se temporalmente à medida que mais amostras são classificadas. Nessa abordagem a classe majoritária entre o conjunto de amostras verificadas, determina qual a indicação deve ser feita pelo sistema. São escolhidos “k” valores para o quantitativo de amostras a serem classificadas pela janela deslizante, a saber: 60, 120, 180, 240 e 300. Tendo em vista que o intervalo de tempo entre cada amostra do conjunto de dados utilizado é um segundo, os valores utilizados para “k” também representam na realidade o tempo em segundos, para que o sistema de identificação de condições em SFV informe o estado real dele, assim, a depender do valor de “k” escolhido, os classificadores terão até 5 minutos para informar a situação do SFV, que pode ser “normal”, “sujidade” ou “sombreamento”.

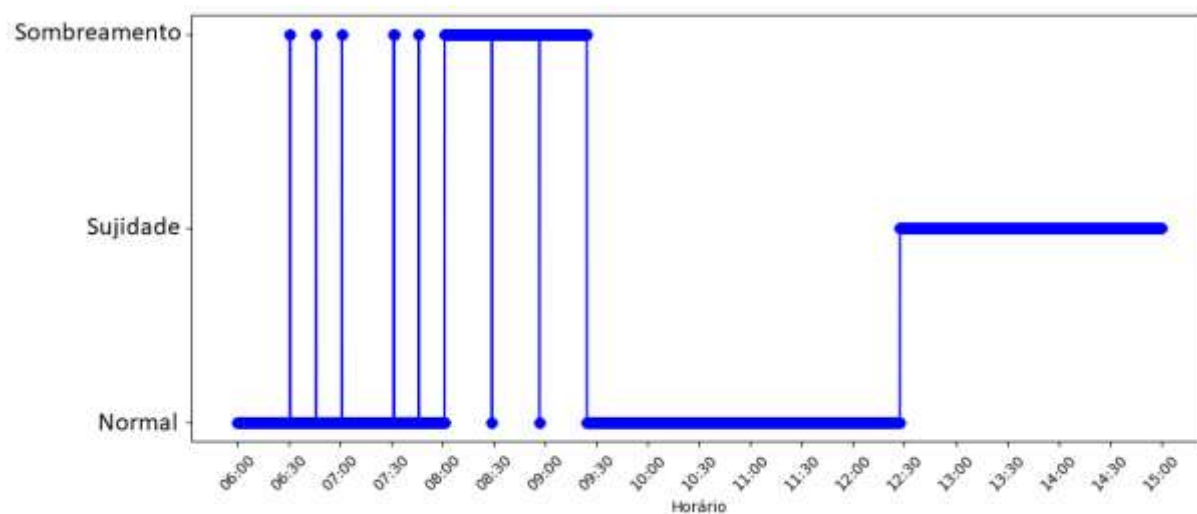
Com o intuito de se realizar testes e validar esse sistema de classificação sequencial de amostras, é emulado um novo conjunto de dados dentre as 192.094 amostras disponibilizadas por Pereira (2018). Nesta nova abordagem, as amostras selecionadas, são mescladas e delimitam o período entre 06:00h e 15:00h, abrangendo um espectro diversificado das potenciais situações que o SFV pode enfrentar durante o decorrer do dia, que varia desde a condição “normal”, até condições de “sombreamento” e “sujidade”.

A representação visual deste conjunto de amostras no decorrer do tempo é apresentada no gráfico ilustrado na Figura 29, em que o eixo horizontal exibe os horários originais de registro das amostras, enquanto o eixo vertical representa os rótulos das classes atribuídas a cada amostra.

Analisando o gráfico ilustrado na Figura 29, verifica-se entre o horário de 06:00h às 06:30h apenas amostras que indicam que o SFV está trabalhando em estado de normalidade. Essa é a condição normal do SFV, porém, entre o horário de 06:30h às 07:00h, existem três picos, que vão até a indicação de sombreamento. Estes picos representam situações muito breves de sombreamento sofrida pelo SFV. Cada pico é composto por 10 amostras com intervalo de um segundo entre eles, e há um intervalo de 15 minutos entre cada um deles. A

presença desses picos no gráfico emula, por exemplo, a ocorrência momentânea de sombreamento causada pela passagem de um pássaro sobre os MFV.

Figura 29 – Conjunto de amostras emuladas.



Fonte: Autoria própria (2024)

Assim, como o período de sombreamento emulado é muito pequeno, durando apenas 10 segundos, o sistema de identificação de condições não deve indicar que o SFV está sombreado, uma vez que se trata de uma condição passageira. Outros dois picos de sombreamento ocorrem entre as 07:30h e 07:45h, reproduzindo condições similares aos três primeiros picos registrados.

Entre as 08:00h e 09:30h, ocorre uma inversão da situação, com a maioria das amostras indicando que o SFV está sombreado. No entanto, destacam-se dois picos com uma rápida transição para a condição de normalidade, ocorrendo às 08:30h e 09:00h, com duração de apenas 10 segundos cada. Essa situação foi pensada para emular a condição em que o sistema experimenta um sombreamento dinâmico, por exemplo, devido a um galho de árvore que, devido a um vento forte, é deslocado, fazendo com que o sistema retorne rapidamente à condição normal. Desse modo, o sistema de identificação de condições deve continuar a indicar a presença de sombreamento no SFV.

No período compreendido entre as 09:30h e as 12:30h, o SFV opera em condições normais, e, portanto, o sistema não deve apontar qualquer anomalia. A partir das 12:30h até as 15:00h, o cenário emulado passa a indicar a presença de sujidade afetando o SFV. Na prática, é sabido que a sujidade não aparece tão rapidamente como mostrado no gráfico ilustrado na Figura 33. No entanto, essa condição é colocada apenas a título de emulação, a fim de verificar atuação do sistema de identificação de condições em SFV com amostras desse tipo.

São apresentados e discutidos no Capítulo 4 os resultados obtidos com a rotina de verificação sequencial desse conjunto de amostras emuladas.

3.5 Conclusões parciais

Neste capítulo verificou-se a base de dados utilizada neste trabalho para treinar e testar os modelos computacionais desenvolvidos. Também é descrito o pré-processamento dos dados e apresentado a metodologia desenvolvida juntamente com os modelos computacionais utilizados para desenvolvimento do sistema de identificação de condições de sujidade e sombreamento no SFV proposto.

No Capítulo 4 são discutidos e analisados os resultados obtidos com os testes envolvendo perceptron, MLP e máquinas de comitê.

4 RESULTADOS E DISCUSSÕES

No presente capítulo são apresentados os principais resultados obtidos no desenvolvimento deste trabalho. Inicialmente é apresentada uma avaliação sobre os dados pré-processados. Em seguida, são discutidos os resultados alcançados com a implementação de redes neurais artificiais (RNA) do tipo perceptron e perceptron de múltiplas camadas. Posteriormente, é avaliado o desempenho alcançado com a utilização de máquinas de comitê. Por fim, é analisada a identificação de condições em SFV utilizando o algoritmo de classificação com amostras sequenciais e janela deslizando.

4.1 Avaliação dos dados

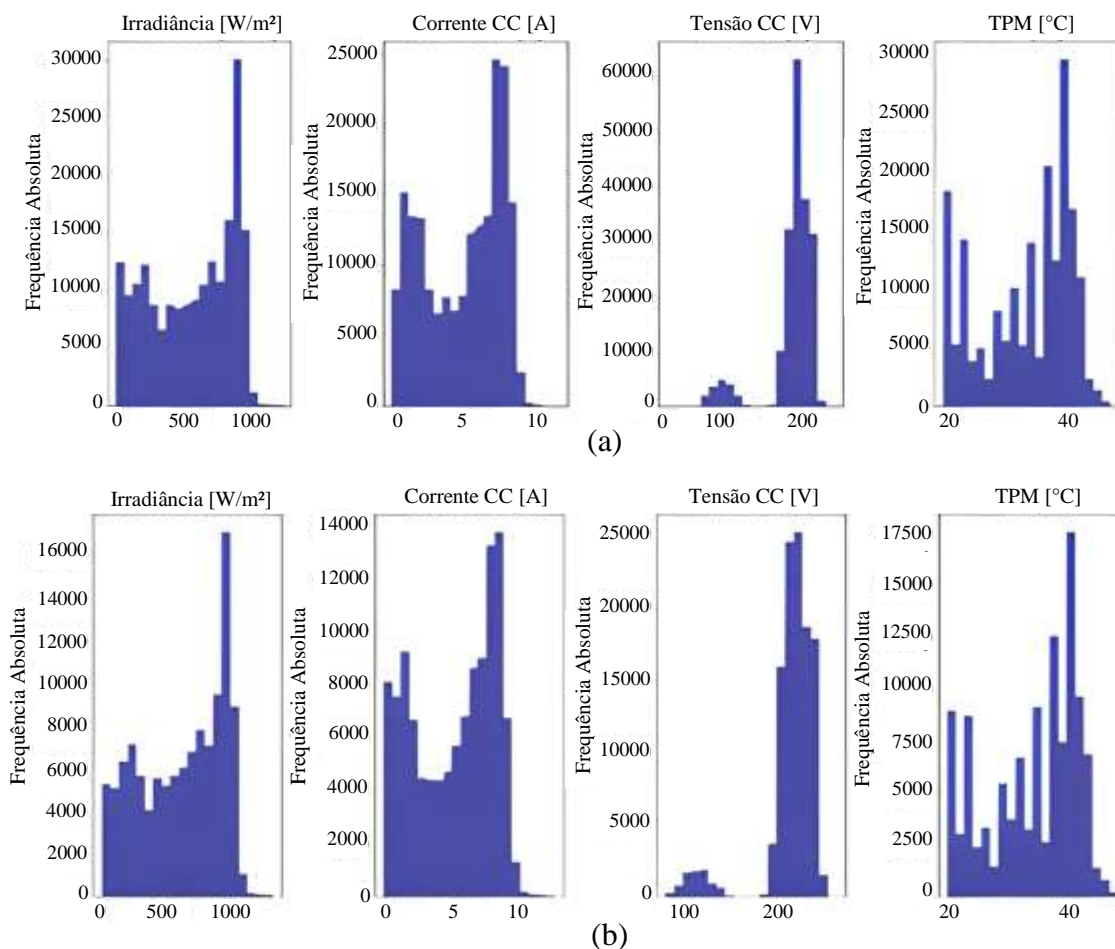
Conforme destacado por Morettin e Singer (2023), o histograma desempenha um papel crucial como uma representação gráfica que ilustra a distribuição das amostras em um conjunto de dados. Esta forma de visualização gráfica é fundamental na compreensão da natureza intrínseca dos dados, proporcionando uma análise profunda que permite identificar tendências, padrões e até mesmo a comparação com outras bases de dados.

Neste contexto, na Figura 30 são apresentados os histogramas que exibem as frequências de distribuição das amostras contidas no conjunto de dados originalmente desenvolvido por Pereira (2018). Estes histogramas consideram uma redução na dimensionalidade do *dataset*, focalizando-se em apenas quatro atributos específicos: a irradiância solar, denominada “Irradiância”, a corrente elétrica no lado de corrente contínua, conhecida como “Corrente CC”, a tensão gerada pelos MFV, identificada como “Tensão CC”, e a temperatura média dos MFV, designada como “TPM”.

É importante ressaltar que esta análise é conduzida tanto antes da implementação do processo de pré-processamento quanto após essa etapa. O objetivo subjacente a essa análise consiste em avaliar se a distribuição das frequências das amostras sofre alterações significativas durante o pré-processamento.

Essa avaliação reveste-se de considerável importância, uma vez que permite verificar se a integridade e a fidedignidade dos dados foram substancialmente afetadas por essa etapa de preparação e tratamento dos dados, e desempenha um papel essencial no entendimento e na validação deles, contribuindo assim para aumentar a robustez e a confiabilidade das informações deste estudo.

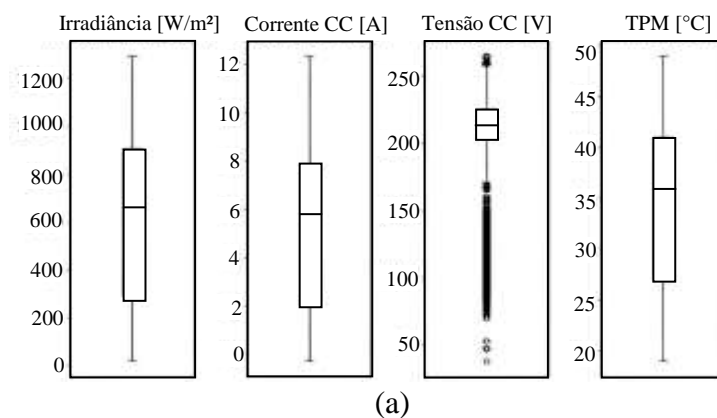
Figura 30 – Histogramas de dados com 4 atributos.
(a) Antes do pré-processamento; (b) Após o pré-processamento.

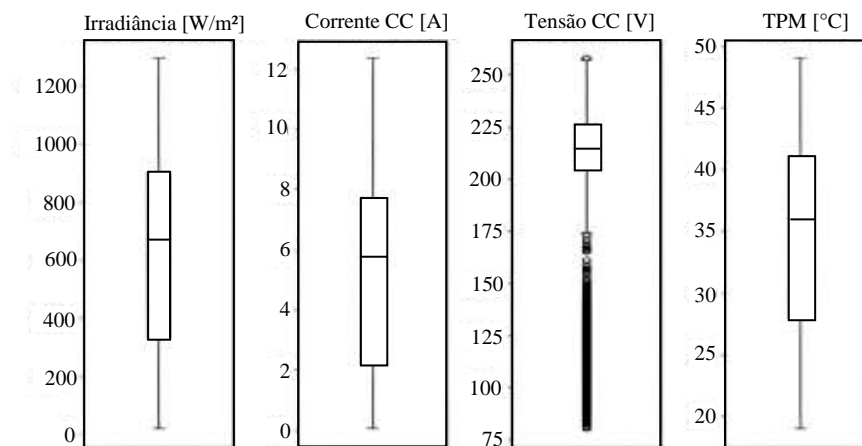


Fonte: Autoria própria (2024)

Observando os gráficos das Figuras 30a e 30b percebe-se que após a realização da etapa de pré-processamento dos dados, a distribuição das frequências não sofre grandes alterações porque transformações lineares não alteram a distribuição de uma variável aleatória. Outra forma de comparar as distribuições dos dados antes e após a realização dessa etapa, é através de gráficos do tipo *box plot*, conforme mostrado na Figura 31.

Figura 31 – *Box plot* de dados com 4 atributos.
(a) Antes do pré-processamento; (b) Após o pré-processamento.





(b)

Fonte: Autoria própria (2024)

A análise dos gráficos das Figuras 31a e 32b revelam que a amplitude dos atributos e mediana de cada *box plot* não diverge significativamente após o pré-processamento, reforçando também as observações previamente constatadas por meio da análise dos histogramas correspondentes. Portanto, espera-se que a redução na dimensionalidade dos dados para três atributos não influencie a generalização deles pelos algoritmos classificadores.

4.2 Resultados obtidos com perceptron

Inicialmente os dados pré-processados são empregados no treinamento e teste de dois perceptrons. Esses modelos computacionais são desenvolvidos com quatro entradas e uma saída com a finalidade de realizar classificação binária entre amostras de dados rotuladas como “normal” e “sujidade”, bem como entre amostras rotuladas como “normal” e “sombreamento”. As entradas de cada perceptron são designadas para verificar a “Irradiação solar” no local de instalação do SFV, a “corrente contínua” que percorre a *string* de MFV, a “tensão contínua” gerada pelo SFV e a informação da “temperatura média” dos MFV. Com cada um desses perceptrons são realizados dez testes utilizando amostras do conjunto de 327 amostras pré-processadas. Desta forma, um perceptron utiliza 109 amostras rotuladas como “normal” e 109 amostras rotuladas como “sujidade”, enquanto que o outro utiliza 109 amostras rotuladas como “normal” e 109 amostras rotuladas como “sombreamento”, totalizando 218 amostras para ambos, sendo que as amostras rotuladas como “normal” são as mesmas para os dois perceptrons.

Nesta etapa são utilizados conjuntos equilibrados de dados (composta por 109 amostras de cada rótulo), onde 70% deles ou 152 amostras são utilizados para treinamento dos perceptrons, enquanto os demais 30% ou 66 amostras são designadas para testes. Os resultados dessas avaliações estão resumidos no Quadro 1.

Quadro 1 - Taxas de acerto dos perceptrons para amostras com 4 atributos.

| Amostras – “normal/sujidade” | | | Amostras – “normal/sombreamento” | | |
|------------------------------|----------------------|--------------|----------------------------------|----------------------|--------------|
| Treinamento (%) | Teste (%) | | Treinamento (%) | Teste (%) | |
| 84,21 | 1 | 75,76 | 85,53 | 1 | 72,73 |
| | 2 | 78,79 | | 2 | 83,33 |
| | 3 | 75,76 | | 3 | 84,85 |
| | 4 | 81,82 | | 4 | 75,76 |
| | 5 | 69,70 | | 5 | 80,30 |
| | 6 | 77,27 | | 6 | 83,33 |
| | 7 | 75,76 | | 7 | 78,79 |
| | 8 | 83,33 | | 8 | 68,18 |
| | 9 | 78,79 | | 9 | 84,85 |
| | 10 | 81,82 | | 10 | 83,33 |
| | Média | 77,88 | | Média | 79,55 |
| | Desvio Padrão | 3,78 | | Desvio Padrão | 5,39 |

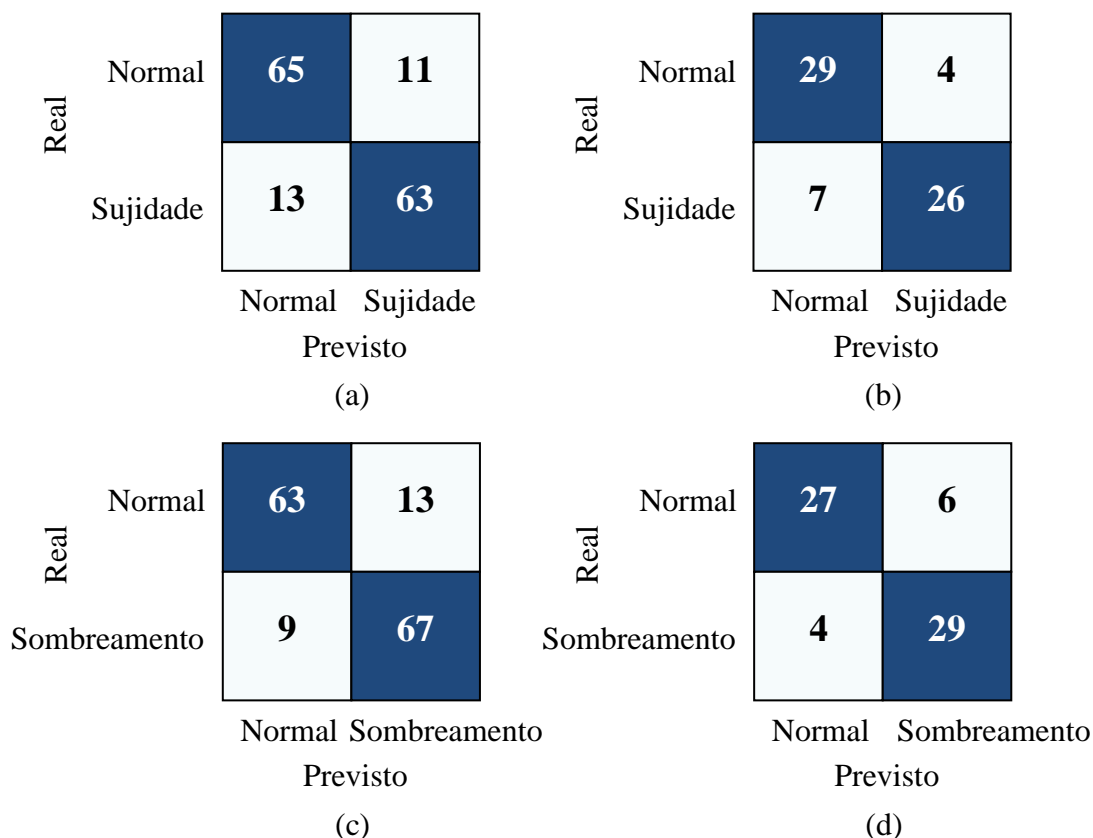
Fonte: Autoria própria (2024)

Após execução do algoritmo de treinamento com cada perceptron são realizados 10 testes, tanto na detecção de amostra do tipo “normal” e “sujidade” quanto para “normal” e “sombreamento”, os resultados obtidos são respectivamente de 83,33% e 84,85%. Ao avaliar esses resultados do ponto de vista da eficiência, observamos que eles são interessantes para um classificador extremamente simples, como o perceptron. No entanto, ao considerar o desenvolvimento de um sistema robusto para a identificação de condições em SFV, espera-se alcançar taxas como menor desvio padrão entre os resultados e com taxas de acertos superiores. Portanto, visando avaliar com maior propriedade o desempenho dessas classificações e seus possíveis pontos de melhorias, a análise das matrizes de confusão é imprescindível. Na Figura 32 são apresentadas as matrizes de confusão obtidas nos treinamentos e testes que tiveram as melhores taxas de acertos.

A Figura 32a ilustra a matriz de confusão dos resultados do treinamento entre amostras de rótulo “normal” e “sujidade”. Por sua vez, a Figura 32b apresenta a matriz de confusão dos resultados do teste. Analisando as informações do teste sobre a detecção da condição “normal” e “sujidade”, percebe-se que esse modelo não encontra dificuldade em classificar as amostras do tipo “normal”, pois somente quatro amostras são classificadas equivocadamente. Porém em relação a identificar amostras de “sujidade”, ele confunde algumas amostras do tipo “normal”

com “sujidade”, pois nesse caso ele errou sete amostras. Portanto, esse fato deve ser levado em consideração para se escolher o melhor classificador que vai constituir o sistema proposto no SFV.

Figura 32 – Matriz de confusão para perceptron com 4 entradas.
 (a) Treinamento - “normal/sujidade”; (b) Teste - “normal/sujidade”;
 (c) Treinamento - “normal/sombreamento”; (d) Teste - “normal/sombreamento”.



Fonte: Autoria própria (2024)

Por outro lado, a Figura 32c ilustra a matriz de confusão dos resultados do treinamento, desta vez com rótulos “normal” e “sombreamento”. A Figura 32d, por sua vez, apresenta a matriz de confusão dos resultados do teste. Analisando os resultados obtidos, percebe-se que o perceptron comete menos erros na classificação das amostras de “normal” e “sombreamento”. No entanto, ainda ocorre confusão entre este tipo de amostra e aquelas rotuladas apenas como “normal”, resultando em uma leve queda na taxa de acertos.

Com base nos resultados desses classificadores utilizando dados com quatro atributos, optou-se por desenvolver também mais dois perceptrons, porém, utilizando amostras com apenas três atributos.

Devido a medição de temperatura não é simples de ser realizada, dependendo do quantitativo de módulos fotovoltaicos da instalação, as medições e a manutenção do sistema tendem a ser muito cara.

Desta forma, as entradas de cada perceptron verificam agora somente a “Irradiação solar”, a “corrente contínua” do sistema e a “tensão contínua” gerada pelo SFV e têm uma única saída, para também realizar classificação binária entre amostras de dados rotuladas como “normal” e “sujidade”, bem como entre amostras rotuladas como “normal” e “sombreamento”. Assim, é verificado experimentalmente se a utilização de um conjunto de dados com menor dimensionalidade influencia de forma significativa nos resultados obtidos.

Da mesma forma como nos dois testes iniciais, após realização de um treinamento são realizados também dez testes utilizando conjuntos equilibrados de dados (composto por 109 amostras de cada rótulo), onde 70% deles ou 152 amostras são utilizados para treinamento, enquanto os demais 30% ou 66 amostras são designadas para testes. O Quadro 2 resume os resultados obtidos.

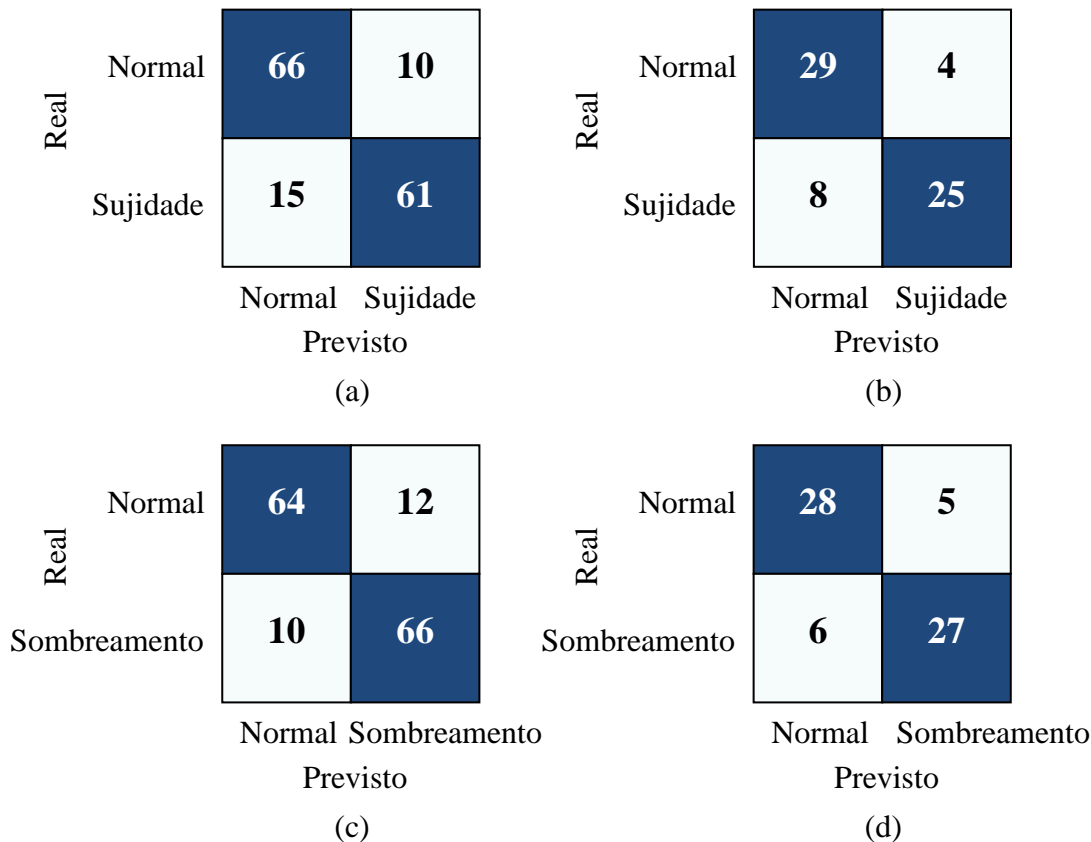
Quadro 2 - Taxas de acerto dos perceptrons para amostras com 3 atributos.

| Amostras – “normal/sujidade” | | | Amostras – “normal/sombreamento” | | |
|------------------------------|----------------------|--------------|----------------------------------|----------------------|--------------|
| Treinamento (%) | Teste (%) | | Treinamento (%) | Teste (%) | |
| 83,55 | 1 | 74,24 | 85,53 | 1 | 78,79 |
| | 2 | 81,82 | | 2 | 83,33 |
| | 3 | 74,24 | | 3 | 71,21 |
| | 4 | 75,76 | | 4 | 83,33 |
| | 5 | 77,27 | | 5 | 78,79 |
| | 6 | 74,24 | | 6 | 81,82 |
| | 7 | 81,82 | | 7 | 80,30 |
| | 8 | 80,30 | | 8 | 77,27 |
| | 9 | 75,76 | | 9 | 81,82 |
| | 10 | 81,82 | | 10 | 81,82 |
| | Média | 77,73 | | Média | 79,55 |
| | Desvio Padrão | 3,18 | | Desvio Padrão | 3,46 |

Fonte: Autoria própria (2024)

Realizados os testes as melhores taxas de acertos obtidas, ou seja, as acurácias, tanto na identificação de amostras “normal” e “sujidade” quanto para “normal” e “sombreamento”, são respectivamente de 81,82% e 83,33%. Avaliando esses resultados do ponto de vista quantitativo, observa-se que eles são muito próximos dos testes realizados com o conjunto de dados com quatro atributos, contudo em termos de eficiência de acertos não ocorreu melhorias significativas. Portanto, visando também avaliar com maior propriedade o desempenho dessas classificações e seus possíveis pontos de melhorias, a análise das matrizes de confusão também é necessária. Na Figura 33 são apresentadas as matrizes de confusão obtidas nos treinamentos e testes que tiveram as melhores taxas de acertos.

Figura 33 – Matriz de confusão para perceptron com 3 entradas.
 (a) Treinamento - “normal/sujidade”; (b) Teste - “normal/sujidade”;
 (c) Treinamento - “normal/sombreamento”; (d) Teste - “normal/sombreamento”.



Fonte: Autoria própria (2024)

A Figura 33a ilustra a matriz de confusão dos resultados do treinamento entre amostras de rótulo “normal” e “sujidade”. Por sua vez, a Figura 33b apresenta a matriz de confusão dos resultados do teste. Percebe-se por meio dos resultados dessa matriz de confusão de teste que este classificador não encontra dificuldade em identificar amostras do tipo “normal”, pois neste caso somente quatro amostras são classificadas equivocadamente, mas em relação a confundir a amostras de “sujidade” com a condição “normal”, ele comete um erro a mais em comparação com o perceptron trabalhando com amostras de quatro atributos, já que nesse caso oito amostras são classificadas como sendo da classe “normal”, mas de fato essas amostras são pertencentes à classe de “sujidade”.

Por outro lado, a Figura 33c ilustra a matriz de confusão dos resultados do treinamento, desta vez com rótulos “normal” e “sombreamento”. A Figura 33d, por sua vez, apresenta a matriz de confusão dos resultados do teste. Os resultados obtidos nos testes indicam que o perceptron tem uma leve tendência a errar menos a classificação de amostras com 4 entradas”,

comparada à classificação das amostras de “sujidade”, mas ele ainda confunde muitas amostras de “sombreamento” com as do tipo “normal”.

Nesse contexto, verifica-se que a utilização de um conjunto de dados com três atributos pouco influência nos resultados obtidos, porém, em relação ao tipo de classificador, percebe-se limitações do perceptron em identificar com precisão a condição “normal”, “sujidade” e “sombreamento”, portanto, como a proposta desta pesquisa é desenvolver um sistema que realize identificação de condições em SFV, os resultados dos testes apontam que este classificador é inviável para essa abordagem de estudo.

4.3 Resultados obtidos com MLP

Considerando as limitações percebidas com a utilização do classificador perceptron, buscou-se implementar outro modelo computacional. Assim, são desenvolvidas duas redes perceptron de múltiplas camadas, com o objetivo de aprimorar a identificação de condições no SFV pesquisado.

Seguindo a mesma abordagem metodológica empregada no desenvolvimento dos perceptrons, tanto na estruturação de sua arquitetura quanto na condução dos processos de treinamento e teste, inicialmente cada MLP é configurada com quatro variáveis de entrada. Entretanto, diferentemente dos perceptrons, as redes MLP incorporaram uma camada intermediária composta por “n” neurônios, mas continuam a apresentar uma única saída. O propósito dessa configuração é possibilitar também a realização de classificações binárias entre amostras classificadas como “normal” e aquelas identificadas como “sujidade” ou “sombreamento”.

As redes MLP propostas efetuam a análise de quatro variáveis de entrada, que são: a “Irradiação solar” local, a “corrente elétrica” que circula na instalação fotovoltaica no lado de corrente contínua (CC), a “tensão” total gerada pelos painéis e a “temperatura média” dos MFV.

Com o intuito de estabelecer a topologia mais apropriada para as MLP propostas, são realizados treinamento e testes, fazendo uso das 327 amostras previamente pré-processadas. Durante esses ensaios, são avaliadas diversas configurações que abrangem diferentes quantidades de neurônios na camada oculta, conforme indicado no Quadro 3.

Quadro 3 – Verificação do número de neurônios ideal na camada oculta.

| Amostras – “normal/sujidade” | | | Amostras – “normal/sombreamento” | | |
|-------------------------------|---------------------|---------------------|----------------------------------|---------------------|---------------------|
| Neurônios na camada escondida | Treinamento | Teste | Neurônios na camada escondida | Treinamento | Teste |
| | Acurácia Obtida (%) | Acurácia Obtida (%) | | Acurácia Obtida (%) | Acurácia obtida (%) |
| 1 | 84,21 | 83,33 | 1 | 85,53 | 84,85 |
| 2 | 85,53 | 84,85 | 2 | 87,50 | 86,36 |
| 4 | 86,84 | 86,36 | 4 | 88,16 | 87,88 |
| 6 | 88,82 | 87,88 | 6 | 89,47 | 89,39 |
| 8 | 90,79 | 89,39 | 8 | 91,45 | 90,91 |
| 10 | 91,45 | 90,91 | 10 | 93,42 | 92,42 |
| 12 | 90,79 | 89,39 | 12 | 91,45 | 90,91 |
| 14 | 90,13 | 89,39 | 14 | 90,79 | 90,02 |
| 16 | 89,47 | 87,88 | 16 | 91,4 5 | 90,91 |
| 18 | 88,82 | 86,36 | 18 | 90,13 | 89,39 |
| 20 | 89,47 | 87,88 | 20 | 89,47 | 87,88 |
| 25 | 88,82 | 87,32 | 25 | 88,82 | 86,36 |
| 30 | 88,16 | 87,88 | 30 | 88,16 | 87,88 |
| 35 | 89,47 | 89,39 | 35 | 88,16 | 87,88 |
| 40 | 90,13 | 89,39 | 40 | 87,50 | 86,36 |
| 45 | 89,47 | 87,88 | 45 | 88,16 | 89,39 |
| 50 | 88,16 | 86,36 | 50 | 89,47 | 89,39 |
| 100 | 90,79 | 89,39 | 100 | 88,82 | 87,88 |

Fonte: Autoria própria (2024)

Nesta etapa é realizado um treinamento com cada topologia proposta conforme disposto na coluna ‘Neurônios na camada escondida’ e em seguida um teste. Os melhores resultados nesses testes revelam taxas de acerto de 90,91% e 92,42% na identificação de condições associadas à “sujidade” e ao “sombreamento”, respectivamente. Essas acurácias apontam a utilização de 10 neurônios na camada oculta como a configuração mais apropriada para as redes MLP propostas. Na Figura 34 é ilustrada a MLP com dez neurônios e uma saída

Após a escolha do número de neurônios na camada oculta, são realizados 10 novos testes com as 327 amostras pré-processadas, utilizando também conjuntos equilibrados de dados (composto por 109 amostras de cada rótulo), onde 70% deles ou 152 amostras são utilizados para treinamento, enquanto os demais 30% ou 66 amostras são designadas para testes. O Quadro 4 apresenta o desempenho de classificador nos treinamentos e testes.

Quadro 4 - Taxas de acerto das redes MLP com 4 entradas.

| Amostras – “normal/sujidade” | | | Amostras – “normal/sombreamento” | | |
|------------------------------|----------------------|--------------|----------------------------------|----------------------|--------------|
| Treinamento (%) | Teste (%) | | Treinamento (%) | Teste (%) | |
| 91,45 | 1 | 87,88 | 93,42 | 1 | 90,91 |
| | 2 | 89,39 | | 2 | 89,39 |
| | 3 | 87,88 | | 3 | 89,39 |
| | 4 | 90,91 | | 4 | 92,42 |
| | 5 | 89,39 | | 5 | 90,91 |
| | 6 | 89,39 | | 6 | 89,39 |
| | 7 | 90,91 | | 7 | 92,42 |
| | 8 | 90,91 | | 8 | 89,39 |
| | 9 | 89,39 | | 9 | 90,91 |
| | 10 | 90,91 | | 10 | 87,88 |
| | Média | 89,70 | | Média | 90,30 |
| | Desvio Padrão | 1,13 | | Desvio Padrão | 1,39 |

Fonte: Autoria própria (2024)

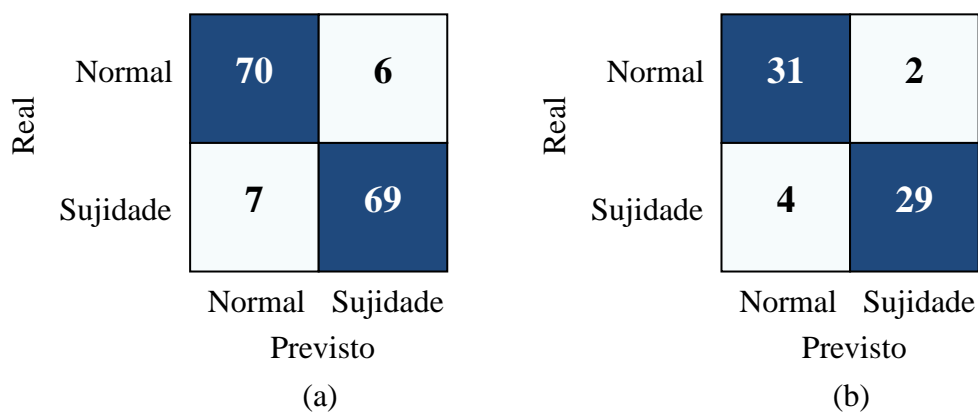
Os resultados apresentados no Quadro 4 indicam que as taxas de acerto durante o treinamento alcançam respectivamente 91,45% e 93,42% para a identificação de amostras categorizadas como “normal” e “sujidade”, bem como entre amostras “normal” e “sombreamento”. Durante os testes, as taxas de acerto mais elevadas são de 90,91% para a identificação de amostras do tipo “normal/sujidade” e 92,42% para amostras do tipo “normal” e “sombreamento”.

De forma preliminar, observa-se uma melhoria de aproximadamente 10% no desempenho das redes MLP em comparação com os perceptrons. Entretanto, espera-se alcançar resultados ainda mais satisfatórios com um tipo de rede tão robusta quanto a MLP.

Na Figura 34 são apresentadas as matrizes de confusão obtidas nos treinamentos e testes que tiveram as melhores taxas de acertos.

Figura 34 – Matriz de confusão para MLP com 4 entradas.

(a) Treinamento - “normal/sujidade”; (b) Teste - “normal/sujidade”;
(c) Treinamento - “normal/sombreamento”; (d) Teste - “normal/sombreamento”.



| | | | | | | | |
|------|--------------|----------|--------------|------|--------------|----------|--------------|
| Real | Normal | 72 | 4 | Real | Normal | 31 | 2 |
| | Sombreamento | 6 | 70 | | Sombreamento | 3 | 30 |
| | | Normal | Sombreamento | | | Normal | Sombreamento |
| | | Previsto | | | | Previsto | |
| | | (c) | | | | (d) | |

Fonte: Autoria própria (2024)

Na Figura 34a é ilustrada a matriz de confusão obtida do treinamento da MLP, analisando a distinção entre as amostras “normal” e “sujidade”, enquanto na Figura 34b, encontra-se a matriz de confusão correspondente ao teste para a mesma rede neural. Os resultados deste treinamento e teste indicam que a MLP exibe um padrão de comportamento similar ao verificado com o perceptron, demonstrando uma leve tendência, porém com uma taxa de acertos mais elevada, em confundir amostras “normal” e “sujidade”.

Na Figura 34c é exibida a matriz de confusão para o treinamento da MLP designada para identificar as condições “normal” e “sombreamento”. Na Figura 34d encontra-se a matriz de confusão correspondente ao teste dessa mesma rede neural. Analisando os resultados obtidos verifica-se também um comportamento similar ao perceptron por parte dessa MLP, no entanto, observa-se uma tendência menos acentuada em confundir amostras do tipo “normal” e “sombreamento”, embora ainda persista certa dificuldade na classificação de amostras rotuladas como “sombreamento”.

De forma a ampliar a investigação sobre a identificação de condições em SFV, foram desenvolvidas mais duas redes MLP para examinar amostras da base de dados pré-processada, no entanto, nesse contexto, opta-se por excluir o atributo “temperatura média dos módulos fotovoltaicos” dos dados, eliminando assim a necessidade de realizar medições individuais da temperatura em cada MFV, facilitando a identificação de condições no SFV pesquisado. Desse modo, as amostras agora consistem exclusivamente nos atributos “Irradiação solar”, “corrente contínua” e “tensão contínua” gerada pelo SFV. Consequentemente, cada MLP passou a apresentar somente três entradas, uma camada oculta composta por 10 neurônios e uma única saída. Essa configuração continua sendo destinada à classificação binária, tanto entre amostras rotuladas como “normal” e “sujidade”, quanto entre amostras rotuladas como “normal” e “sombreamento”.

Após o treinamento dessas novas redes com o conjunto de 327 amostras previamente pré-processadas, utilizando também conjuntos equilibrados de dados (composto por 109 amostras de cada rótulo), onde 70% deles ou 152 amostras são utilizados para treinamento, enquanto os demais 30% ou 66 amostras são designadas para testes. O Quadro 5 apresenta os resultados obtidos nos treinamentos e testes.

Quadro 5 - Taxas de acerto das redes MLP com 3 entradas.

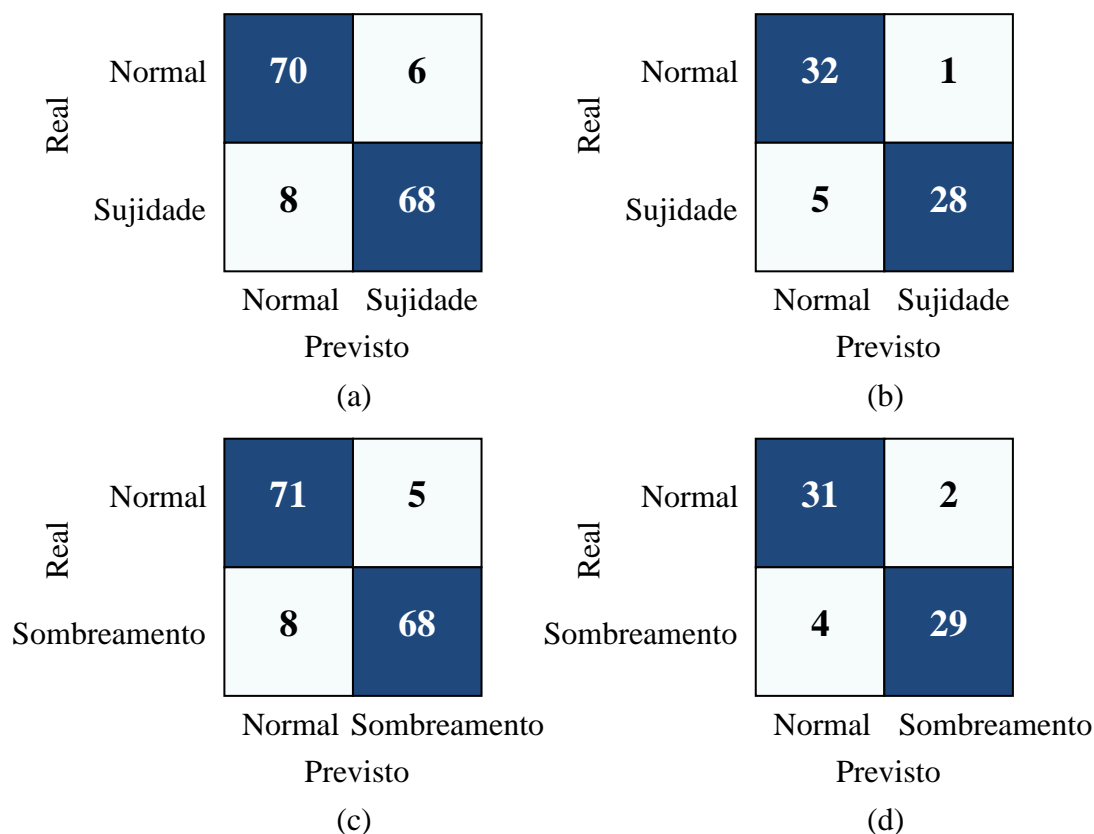
| Amostras “normal/sujidade” | | | Amostras “normal/sombreamento” | | |
|----------------------------|----------------------|--------------|--------------------------------|----------------------|--------------|
| Treinamento (%) | Teste (%) | | Treinamento (%) | Teste (%) | |
| 90,79 | 1 | 87,88 | 91,45 | 1 | 89,39 |
| | 2 | 86,36 | | 2 | 90,91 |
| | 3 | 87,88 | | 3 | 89,39 |
| | 4 | 89,39 | | 4 | 90,91 |
| | 5 | 89,39 | | 5 | 90,91 |
| | 6 | 86,36 | | 6 | 86,36 |
| | 7 | 89,39 | | 7 | 90,91 |
| | 8 | 87,88 | | 8 | 87,88 |
| | 9 | 89,39 | | 9 | 90,91 |
| | 10 | 89,39 | | 10 | 90,91 |
| | Média | 88,33 | | Média | 89,85 |
| | Desvio Padrão | 1,18 | | Desvio Padrão | 1,52 |

Fonte: Autoria própria (2024)

Ao examinar o Quadro 5, verifica-se que as taxas de acerto durante o treinamento atingem 90,79% para a verificação de amostras rotuladas como “normal” e “sujidade” e 91,45% para amostras classificadas como “normal” e “sombreamento”. Durante os testes, o desempenho mais destacado na identificação das amostras “normal” e “sujidade”, bem como entre amostras “normal” e “sombreamento” alcança, respectivamente, 89,39% e 90,91%. Estes percentuais de acerto indicam que a redução da dimensionalidade exerce um impacto mínimo nos resultados preliminares. Assim, opta-se por utilizar exclusivamente dados contendo três atributos para a identificação de condições de sombreamento e sujidade em toda a pesquisa.

Na Figura 35 são apresentadas as matrizes de confusão obtidas nos treinamentos e testes que tiveram os melhores desempenhos. Na Figura 35a, é ilustrada a matriz de confusão resultante do treinamento da MLP, focalizando na distinção entre as condições “normal” e “sujidade”. Na Figura 35b, encontra-se a matriz de confusão correspondente ao teste realizado com a mesma rede neural. Durante a avaliação das amostras “normal” e “sujidade”, nota-se que a MLP ainda apresenta uma leve inclinação para confundir amostras de rótulos “normal” e “sujidade”.

Figura 35 – Matriz de confusão para MLP com 3 entradas.
 (a) Treinamento - “normal/sujidade”; (b) Teste - “normal/sujidade”;
 (c) Treinamento - “normal/sombreamento”; (d) Teste - “normal/sombreamento”.



Fonte: Autoria própria (2024)

Na Figura 35c, é exibida a matriz de confusão referente ao treinamento da MLP destinada a identificar as condições "normal" e "sombreamento", enquanto na Figura 35d, encontra-se a matriz de confusão correspondente ao teste dessa mesma rede neural. Ao analisar os resultados dos testes, percebe-se também uma tendência da MLP em confundir amostras do tipo "normal" e "sombreamento", embora de forma menos pronunciada em comparação com a verificação das amostras anteriores. Contudo, persiste a dificuldade em classificar amostras rotuladas como "sombreamento".

Assim, diante dos resultados obtidos com as redes MLP, ao utilizar amostras contendo apenas três atributos, os quais indicam que não ocorrem variações significativas nas taxas de eficiência observadas em comparação com amostras contendo quatro atributos, opta-se por adotar exclusivamente MLP com três entradas ao longo da pesquisa. Essa simplificação na dimensionalidade dos dados proporciona benefícios adicionais, como maior rapidez no processamento das informações, redução nos custos computacionais e uma simplificação no processo de identificação das condições em SFV. Além disso, elimina a necessidade de medição das temperaturas de todos os MFV, conforme mencionado anteriormente.

Visando aprimorar as taxas de acerto utilizando as redes perceptrons de múltiplas camadas, é implementada uma nova MLP capaz de realizar a classificação das amostras rotuladas como “normal”, “sombreamento” e “sujidade” simultaneamente, efetuando, assim, uma classificação multiclasse. Esta nova arquitetura opera com três entradas, as quais monitoram exclusivamente os atributos “Irradiação solar”, “corrente contínua” e “tensão contínua”. A camada intermediária é composta por 10 neurônios, e a rede dispõe de três saídas para classificar o conjunto de 327 amostras previamente pré-processadas formadas por um conjunto equilibrados de dados (composto por 109 amostras de cada rótulo), onde 70% deles ou 228 amostras são utilizados para treinamento, enquanto os demais 30% ou 99 amostras são designadas para testes. O Quadro 6 apresenta os resultados obtidos no treinamento e em 10 testes realizados.

Quadro 6 - Taxas de acerto da rede MLP com 3 saídas.

| Amostras de dados Normal/Sujidade/Sombreamento | | |
|---|----------------------|--------------|
| Treinamento (%) | Teste (%) | |
| 89,47 | 1 | 87,88 |
| | 2 | 86,87 |
| | 3 | 84,85 |
| | 4 | 87,88 |
| | 5 | 87,88 |
| | 6 | 86,87 |
| | 7 | 85,86 |
| | 8 | 88,89 |
| | 9 | 84,85 |
| | 10 | 86,87 |
| | Média | 86,87 |
| | Desvio Padrão | 1,28 |

Fonte: Autoria própria (2024)

A análise do Quadro 6 permite perceber que a eficiência dessa nova arquitetura é menor que a das redes MLP realizando classificação binária. Verifica-se que tanto o treinamento quanto teste possuem acertos percentuais menores. O melhor resultado nos testes realizados alcançou apenas 88,89% de acertos. Na Figura 36 é apresentada a matriz de confusão obtida com o treinamento e teste com melhor desempenho observado.

Examinando a matriz de confusão da Figura 36a, percebe-se que a MLP com três saídas também confunde algumas amostras, no entanto na verificação entre amostras de rótulo “normal” ela não confunde nenhuma amostra como sendo de “sombreamento”, o que é um bom resultado para esse tipo de classificação, fato este que se repete durante o teste, conforme

ilustrada na matriz de confusão da Figura 36b, onde novamente amostras de rótulo “normal” ela não são confundidas como sendo de amostras de “sombreamento”, e também se estende para amostras de “sujidade”, o que é um ótimo resultado para este classificador.

Figura 36 - Matriz de confusão do classificador MLP com 3 saídas.
(a) Treinamento; (b) Teste.

| | | | | | | | | | |
|------|--------------|-----------|-----------|--------------|------|--------------|-----------|-----------|--------------|
| Real | Normal | 69 | 7 | 0 | Real | Normal | 31 | 2 | 0 |
| | Sujidade | 8 | 67 | 1 | | Sujidade | 5 | 28 | 0 |
| | Sombreamento | 5 | 3 | 68 | | Sombreamento | 3 | 1 | 29 |
| | | Normal | Sujidade | Sombreamento | | | Normal | Sujidade | Sombreamento |
| | | Previsto | | | | | Previsto | | |
| | | (a) | | | | | (b) | | |

Fonte: Autoria própria (2024)

Na verificação entre amostras do tipo “normal” e “sujidade”, o classificador confunde apenas duas amostras do tipo “normal” como sendo “sombreamento”, enquanto na classificação entre amostras de “sujidade” e “normal” apenas cinco amostras são classificadas equivocadamente. Para a classificação das amostras de “sombreamento” verifica-se que apenas três são classificadas como sendo de rótulo “normal” e apenas uma como sendo de “sujidade”, o que demonstra ser um bom resultado para esse tipo de classificação. No entanto, de um modo geral percebe-se que este classificador apresenta uma leve tendência em confundir amostras de “sujidade” e “sombreamento” com aquelas da classe “normal”. Este comportamento assemelha-se ao observado em outras redes desenvolvidas para classificação binária, resultando em uma eficiência global reduzida por parte desta rede MLP para classificação de várias classes simultaneamente (multiclasse).

4.4 Resultados obtidos com máquinas de comitê

Com o intuito de aprimorar as acurácias obtidas nos testes com perceptron e MLP, busca-se um modelo computacional que generalize de maneira mais eficaz os dados pré-processados. Nesse contexto, opta-se pela utilização de máquinas de comitê constituídas de redes MLP para realizar a classificação binária entre amostras do tipo “normal” e “sujidade” e

entre amostras “normal” e “sombreamento”. Dessa forma, as máquinas de comitê são projetadas para operar mediante votação simples, no qual cada classificador constituinte desse modelo computacional indica a classe ou rótulo de uma amostra avaliada e ao término da votação, um combinador de resultados contabiliza todos os votos e determina a classe majoritária como sendo a saída da máquina de comitê. Para evitar empates durante as votações, é utilizado um número ímpar de classificadores. Assim, são arquitetadas seis máquinas de comitê, sendo duas compostas por três MLP, duas com cinco MLP e outras duas por sete MLP, nas quais são treinadas e posteriormente testadas dez vezes para se averiguar experimentalmente aquelas que apresentam os melhores resultados.

O conjunto de dados utilizado pelas máquinas de comitê é equilibrado e composto por 109 amostras rotuladas como “normal”, 109 de rótulo “sujidade” e outras 109 com rótulo “sombreamento”, totalizando as 327 amostras. No entanto, para um par de máquinas de comitê, uma delas utiliza 109 amostras rotuladas como “normal” e 109 como “sujidade”, enquanto a outra utiliza 109 amostras rotuladas como “normal” e 109 como “sombreamento”. Logo, cada máquina de comitê utiliza apenas 218 amostras do conjunto de dados, mas em todas, as amostras rótulo “normal” são as mesmas. O treinamento de cada máquina de comitê é realizado com 70 % dos dados ou 152 amostras, enquanto os 30 % restantes ou 66 amostras são utilizados para teste.

No Quadro 7 são apresentados os resultados obtidos durante treinamento e testes para as máquinas de comitê com diferentes números de MLP utilizando as amostras rotuladas como “normal” e “sujidade”, enquanto no Quadro 8 são apresentados os resultados para amostras rotuladas como “normal” e “sombreamento”.

Quadro 7 - Resultados com topologias diversas para amostras “normal/sujidade”.

| 3 MLP | | | 5 MLP | | | 7 MLP | | |
|-----------------|----|--------------|-----------------|----|--------------|-----------------|----|--------------|
| Treinamento (%) | | Teste (%) | Treinamento (%) | | Teste (%) | Treinamento (%) | | Teste (%) |
| 96,71 | 1 | 96,97 | 97,37 | 1 | 95,45 | 96,71 | 1 | 93,94 |
| | 2 | 93,94 | | 2 | 93,94 | | 2 | 95,45 |
| | 3 | 92,42 | | 3 | 96,97 | | 3 | 92,42 |
| | 4 | 95,45 | | 4 | 95,45 | | 4 | 95,45 |
| | 5 | 92,42 | | 5 | 92,42 | | 5 | 96,97 |
| | 6 | 93,94 | | 6 | 93,94 | | 6 | 95,45 |
| | 7 | 95,45 | | 7 | 95,45 | | 7 | 92,42 |
| | 8 | 92,42 | | 8 | 96,97 | | 8 | 93,94 |
| | 9 | 95,45 | | 9 | 95,45 | | 9 | 92,42 |
| | 10 | 93,94 | | 10 | 96,97 | | 10 | 95,45 |
| Média | | 94,24 | Média | | 95,30 | Média | | 94,39 |
| Desvio Padrão | | 1,49 | Desvio Padrão | | 1,43 | Desvio Padrão | | 1,52 |

Fonte: Autoria própria (2024)

Os Quadros 7 e 8 visam oferecer uma visão comparativa dos desempenhos obtidos com as diferentes topologias para as máquinas de comitê propostas. No Quadro 7 é possível perceber que a máquina de comitê com cinco MLP alcança a maior média de acertos atingindo 95,30% e também o menor desvio padrão, com um valor de 1,43%. Estes resultados indicam que essa topologia apresenta o melhor desempenho na classificação de amostras com rótulo “normal” e “sujidade”.

Quadro 8 – Resultados com topologias diversas para amostras “normal/sombreamento”.

| 3 MLP | | | 5 MLP | | | 7 MLP | | |
|----------------------|----|--------------|----------------------|----|--------------|----------------------|----|--------------|
| Treinamento (%) | | Teste (%) | Treinamento (%) | | Teste (%) | Treinamento (%) | | Teste (%) |
| 97,37 | 1 | 96,97 | 98,68 | 1 | 95,45 | 97,37 | 1 | 92,42 |
| | 2 | 95,45 | | 2 | 96,97 | | 2 | 96,97 |
| | 3 | 92,42 | | 3 | 92,42 | | 3 | 92,42 |
| | 4 | 93,94 | | 4 | 96,97 | | 4 | 93,94 |
| | 5 | 92,42 | | 5 | 98,48 | | 5 | 93,94 |
| | 6 | 93,94 | | 6 | 95,45 | | 6 | 96,97 |
| | 7 | 96,97 | | 7 | 96,97 | | 7 | 96,97 |
| | 8 | 92,42 | | 8 | 93,94 | | 8 | 95,45 |
| | 9 | 96,97 | | 9 | 96,97 | | 9 | 93,94 |
| | 10 | 93,94 | | 10 | 96,97 | | 10 | 96,97 |
| Média | | 94,54 | Média | | 96,06 | Média | | 95,00 |
| Desvio Padrão | | 1,82 | Desvio Padrão | | 1,69 | Desvio Padrão | | 1,80 |

Fonte: Autoria própria (2024)

Analisando os dados do Quadro 8, constata-se que a máquina de comitê com cinco MLP alcança a maior média de acertos com 96,06% e também o menor desvio padrão, chegando no máximo a 1,69%. Estes resultados demonstram que essa topologia apresenta o melhor desempenho na classificação de amostras com rótulo “normal” e “sombreamento”.

Com o objetivo de validar a utilização de máquinas de comitê com cinco MLP e operando mediante votação simples com seus classificadores, essa mesma topologia é testada utilizando um combinador de resultados que calcula a média das previsões de probabilidade indicadas por cada classificador. Assim, a classe com a maior média de probabilidade é selecionada como a previsão final da máquina de comitê. No Quadro 9 são apresentados os resultados obtidos para treinamento e testes com as amostras verificadas.

Considerando os dados apresentados no Quadro 9, para a classificação das amostras “normal” e “sujidade”, obtêm-se 92,91% para média de acertos e um desvio padrão de 1,61%. Para as amostras “normal” e “sombreamento”, obtêm-se 94,08% de média de acertos com um desvio padrão de 1,82%. Tendo em vista que esses resultados não superam os exibidos nos Quadros 7 e 8, portanto, conclui-se que a abordagem que utiliza o combinador de resultados

com votação simples é a mais apropriada para realizar as classificações propostas.

Quadro 9 – Taxas de acertos para máquina de comitê com 5 MLP e combinador realizando média dos resultados.

| Amostras “normal/sujidade” | | | Amostras “normal/sombreamento” | | |
|----------------------------|----------------------|--------------|--------------------------------|----------------------|--------------|
| Treinamento (%) | Teste (%) | | Treinamento (%) | Teste (%) | |
| 96,05 | 1 | 93,94 | 97,37 | 1 | 93,94 |
| | 2 | 90,91 | | 2 | 92,88 |
| | 3 | 92,88 | | 3 | 95,45 |
| | 4 | 90,91 | | 4 | 90,91 |
| | 5 | 92,88 | | 5 | 96,97 |
| | 6 | 95,45 | | 6 | 92,88 |
| | 7 | 92,88 | | 7 | 92,88 |
| | 8 | 90,91 | | 8 | 93,94 |
| | 9 | 92,88 | | 9 | 96,97 |
| | 10 | 95,45 | | 10 | 93,94 |
| | Média | 92,91 | | Média | 94,08 |
| | Desvio Padrão | 1,61 | | Desvio Padrão | 1,82 |

Fonte: Autoria própria (2024)

Determinada a topologia das máquinas de comitê utilizadas para realizar a classificação binária entre amostras do tipo “normal” e “sujidade” e entre amostras “normal” e “sombreamento”, são realizados novos testes para avaliar o desempenho individual de cada MLP nesse processo, bem como o resultado global fornecido pelo modelo computacional utilizado. Nos Quadros 10 e 11 são apresentados, respectivamente, os desempenhos obtidos durante o treinamento e testes para as amostras “normal” e “sujidade”. Por sua vez, nos Quadros 12 e 13 são apresentados, respectivamente, os resultados para treinamento e teste das taxas de acerto nas classificações das amostras “normal” e “sombreamento”.

Quadro 10 – Resultados percentuais para treinamento da máquina de comitê com 5 MLP e amostras “normal/sujidade”.

| MLP 1 | MLP 2 | MLP 3 | MLP 4 | MLP 5 | Máquina de comitê |
|-------|-------|-------|-------|-------|-------------------|
| 93,42 | 94,08 | 93,42 | 93,42 | 94,74 | 97,37 |

Fonte: Autoria própria (2024)

Conforme apresentado no Quadro 10, já no treinamento é possível perceber que a taxa de acertos global da máquina de comitê é superior às taxas de acertos de todas as redes MLP operando isoladamente. De acordo com Haykin (2001) a combinação de especialistas (classificadores) funde o conhecimento adquirido por cada um deles (durante o treinamento), fazendo a decisão global da máquina de comitê atingir resultados não alcançáveis por qualquer um dos classificadores atuando isoladamente.

Quadro 11 – Resultados percentuais para testes da máquina de comitê com 5 MLP e amostras “normal/sujidade”.

| Teste | MLP 1 | MLP 2 | MLP 3 | MLP 4 | MLP 5 | Máquina de comitê | Média | Desvio Padrão |
|----------------------|--------------|--------------|--------------|--------------|--------------|-------------------|--------------|---------------|
| 1 | 90,91 | 89,39 | 89,39 | 90,91 | 89,39 | 95,45 | 90,91 | 1,66 |
| 2 | 89,39 | 92,42 | 89,39 | 90,91 | 92,42 | 93,94 | 91,41 | 1,30 |
| 3 | 90,91 | 90,91 | 90,91 | 87,88 | 90,91 | 96,97 | 91,42 | 2,11 |
| 4 | 89,39 | 89,39 | 89,39 | 89,39 | 87,88 | 95,45 | 90,15 | 1,89 |
| 5 | 87,88 | 90,91 | 89,39 | 90,91 | 89,39 | 92,42 | 90,15 | 1,12 |
| 6 | 90,91 | 86,36 | 89,39 | 87,88 | 90,91 | 93,94 | 89,90 | 1,88 |
| 7 | 89,39 | 90,91 | 89,39 | 89,39 | 92,42 | 95,45 | 91,16 | 1,72 |
| 8 | 92,42 | 89,39 | 90,91 | 90,91 | 90,91 | 96,97 | 91,92 | 1,88 |
| 9 | 89,39 | 92,42 | 89,39 | 89,39 | 89,39 | 95,45 | 90,91 | 1,79 |
| 10 | 89,39 | 89,39 | 90,91 | 90,91 | 90,91 | 93,94 | 90,91 | 1,17 |
| Média | 90,00 | 90,15 | 89,85 | 89,85 | 90,45 | 95,00 | | |
| Desvio Padrão | 1,21 | 1,69 | 0,70 | 1,22 | 1,36 | 1,36 | | |

Fonte: Autoria própria (2024)

Em relação à classificação de amostras “normal” e “sujidade”, os resultados indicam acertos individuais de cada MLP na ordem de 90%. Em contrapartida, a máquina de comitê tem eficiência média de 95%. Apesar do desvio padrão não ser o menor entre os diversos classificadores, a melhoria na taxa de acerto também é atribuída à combinação de múltiplas previsões individuais, que contribuem para uma decisão mais precisa e confiável na classificação dos dados, tal como observado durante o treinamento.

Quadro 12 – Resultados percentuais para treinamento da máquina de comitê com 5 MLP e amostras “normal/sombreamento”.

| MLP 1 | MLP 2 | MLP 3 | MLP 4 | MLP 5 | Máquina de comitê |
|-------|-------|-------|-------|-------|-------------------|
| 94,08 | 93,42 | 92,76 | 94,74 | 93,42 | 98,68 |

Fonte: Autoria própria (2024)

Analisando os dados do Quadro 12, também é possível perceber que a máquina de comitê possui desempenho superior ao das redes MLP operando isoladamente na classificação de amostras com rótulo “normal” e “sombreamento”.

Conforme observado na classificação de amostras “normal” e “sujidade”, os resultados apresentados no Quadro 13 também indicam acertos individuais para cada MLP na ordem de

90% e para a máquina de comitê os acertos tem média de 95%.

Quadro 13 – Resultados percentuais para testes da máquina de comitê com 5 MLP e amostras “normal/sombreamento”.

| Teste | MLP 1 | MLP 2 | MLP 3 | MLP 4 | MLP 5 | Máquina de comitê | Média | Desvio Padrão |
|----------------------|--------------|--------------|--------------|--------------|--------------|-------------------|--------------|---------------|
| 1 | 90,91 | 90,91 | 89,39 | 90,91 | 87,88 | 95,45 | 90,91 | 2,31 |
| 2 | 92,42 | 89,39 | 90,91 | 92,42 | 90,91 | 96,97 | 92,17 | 2,38 |
| 3 | 90,91 | 90,91 | 92,42 | 89,39 | 92,42 | 92,42 | 91,41 | 1,13 |
| 4 | 90,91 | 92,42 | 92,42 | 90,91 | 90,91 | 96,97 | 92,42 | 2,14 |
| 5 | 89,39 | 89,39 | 90,91 | 92,42 | 89,39 | 98,48 | 91,66 | 3,24 |
| 6 | 90,91 | 92,42 | 90,91 | 92,42 | 90,91 | 95,45 | 92,17 | 1,61 |
| 7 | 92,42 | 90,91 | 89,39 | 89,39 | 92,42 | 96,97 | 91,92 | 2,58 |
| 8 | 92,42 | 92,42 | 87,88 | 90,91 | 89,39 | 93,94 | 91,16 | 2,04 |
| 9 | 90,91 | 92,42 | 90,91 | 92,42 | 87,88 | 96,97 | 91,92 | 2,72 |
| 10 | 89,39 | 90,91 | 92,42 | 92,42 | 92,42 | 96,97 | 92,42 | 2,32 |
| Média | 91,06 | 91,21 | 90,76 | 91,36 | 90,45 | 96,06 | | |
| Desvio Padrão | 1,06 | 1,13 | 1,43 | 1,18 | 1,67 | 1,69 | | |

Fonte: Autoria própria (2024)

Comparando os desempenhos da máquinas de comitê nos Quadros 11 e 13, é possível perceber que para a classificação de amostras “normal” e “sombreamento”, as taxas de acertos são levemente superior à classificação de amostras “normal” e “sujidade”. Isso ocorre devido às amostras de sujidade se assemelharem às amostras normal. No entanto, ambas os acertos são muito próximos.

Na Figura 37 são apresentas as matrizes de confusão para os treinamentos e testes com melhor desempenho observado para classificação de amostras “normal/sujidade” e “normal/sombreamento” com máquinas de comitê.

Na Figura 37a, é ilustrada a matriz de confusão dos resultados do treinamento entre amostras rotuladas como “normal/sujidade”. Nesta etapa, observa-se que a máquina de comitê apresenta apenas uma classificação incorreta para as amostras “normal”, enquanto para as amostras “sujidade” são registradas três classificações equivocadas. Na Figura 37b é apresentada a matriz de confusão dos resultados de teste, onde não são observados erros na classificação das amostras “normal”, contudo, ocorrem dois erros para as amostras “sujidade”.

Figura 37 – Matrizes de confusão para máquinas de comitê com 5 MLP.
 (a) Treinamento - “normal/sujidade”; (b) Teste - “normal/sujidade”;
 (c) Treinamento - “normal/sombreamento”; (d) Teste - “normal/sombreamento”.

| | | | | | | | |
|------|--------------|-----------|--------------|------|--------------|-----------|--------------|
| Real | Normal | 75 | 1 | Real | Normal | 33 | 0 |
| | Sujidade | 3 | 73 | | Sujidade | 2 | 31 |
| | | Normal | Sujidade | | | Normal | Sujidade |
| | | Previsto | | | | Previsto | |
| | | (a) | | | | (b) | |
| Real | Normal | 76 | 0 | Real | Normal | 33 | 0 |
| | Sombreamento | 2 | 74 | | Sombreamento | 1 | 32 |
| | | Normal | Sombreamento | | | Normal | Sombreamento |
| | | Previsto | | | | Previsto | |
| | | (c) | | | | (d) | |

Fonte: Autoria própria (2024)

Por outro lado, na Figura 37c é ilustrada a matriz de confusão dos resultados do treinamento entre amostras rotuladas como “normal/sombreamento”. Verifica-se nessa etapa que a máquina de comitê não comete erros nas classificações de amostras “normal”, mas para as amostras “sombreamento” são registradas duas classificações equivocadas. Por sua vez, na Figura 37d é apresentada a matriz de confusão dos resultados de teste, onde também não são observados erros na classificação das amostras “normal”. No entanto, para as amostras “sombreamento” é verificado um erro.

De modo geral, com base nos resultados de treinamento e testes utilizando máquinas de comitê para classificar amostras “normal/sujidade” e “normal/sombreamento”, percebe-se o excelente desempenho desse modelo computacional. No entanto, de maneira semelhante ao perceptron e à MLP, a máquina de comitê ainda demonstra uma leve tendência a confundir amostras de “sujidade” com “normal”, ocasionando erros na classificação dos dados para ambos os rótulos. Por outro lado, não se evidenciam dificuldades significativas na classificação das

amostras do tipo “normal”, em comparação com identificação de amostras de “sombreamento”.

Embora nenhuma das classificações dos testes alcance uma taxa de acerto de 100%, o que é desejável, tal ocorrência pode indicar que o modelo está memorizando os dados de treinamento em vez de aprender os padrões gerais, resultando em classificações errôneas e falta de generalização para novos dados.

Os resultados verificados demonstram que esse modelo computacional apresenta bom desempenho, principalmente em comparação com o perceptron e a MLP quando operam de forma individualizada.

Após a realização dos testes iniciais com as máquinas de comitê e verificação de bons resultados com a utilização desse modelo, é arquitetada uma nova metodologia com objetivo de realizar a classificação simultânea de todas as amostras disponíveis. Dessa forma, é elaborada uma estrutura constituída por três máquinas de comitê, cada uma com cinco MLP, que realizam classificação binária, de amostras com rótulos “normal/sujidade”, “normal/sombreamento” e “sujidade/sombreamento”, pois baseados nos resultados anteriores, é o modelo que apresenta o melhor desempenho. Assim, cada uma das máquinas de comitê constituintes desse novo modelo computacional também é treinada e testada utilizando as 327 amostras pré-processadas utilizadas anteriormente. No Quadro 14 são apresentados os resultados obtidos para treinamento e teste dessas três máquinas de comitê.

Quadro 14 – Resultados percentuais de treinamento e teste para máquina de comitê com três saídas.

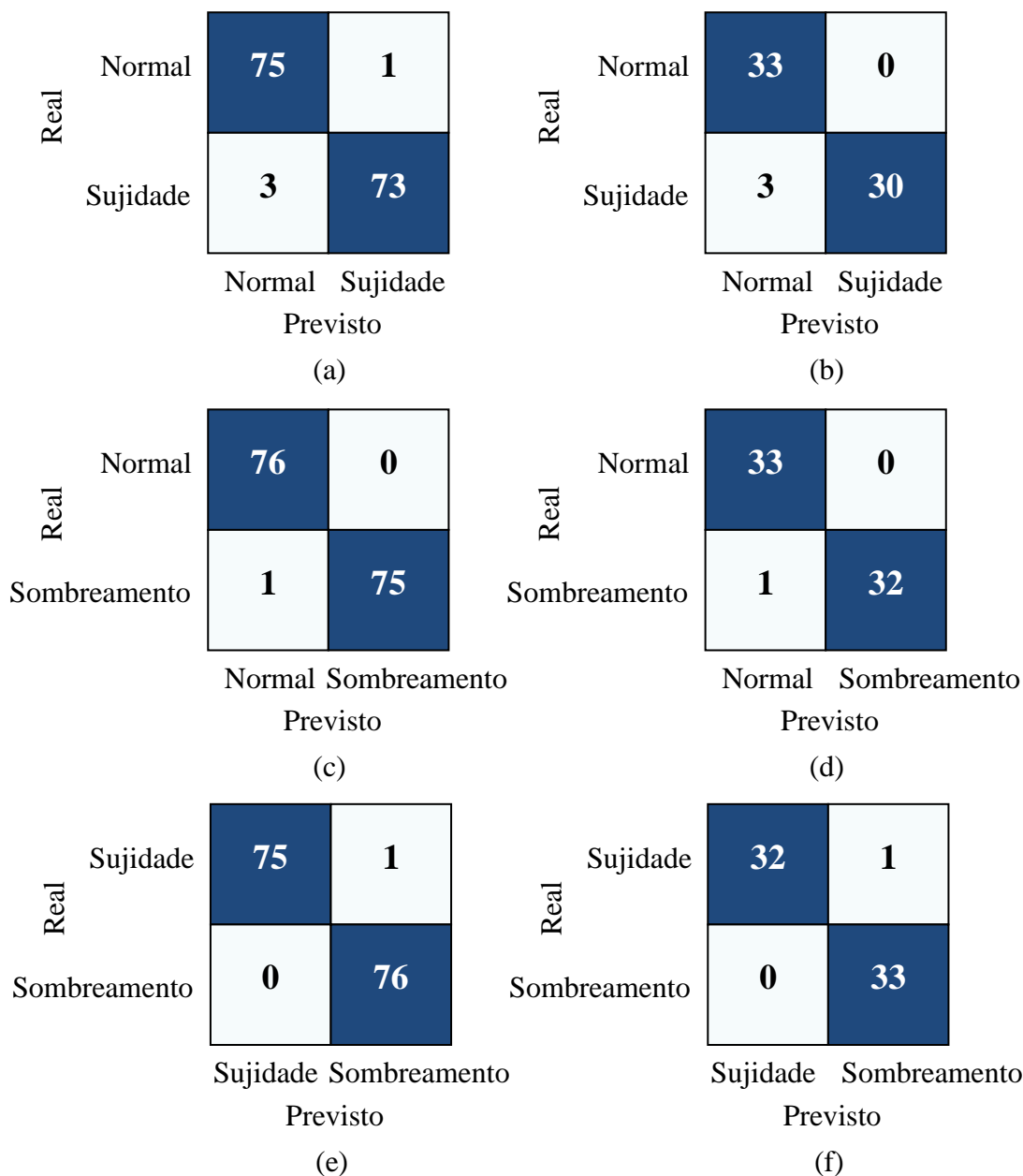
| Amostras “normal/sujidade” | | Amostras “normal/sombreamento” | | Amostras “sujidade/sombreamento” | |
|-------------------------------|-------|-----------------------------------|-------|-------------------------------------|-------|
| Treinamento | Teste | Treinamento | Teste | Treinamento | Teste |
| 97,37 | 95,45 | 99,34 | 98,48 | 99,34 | 98,48 |

Fonte: Autoria própria (2024)

Na Figura 38 são ilustradas as matrizes de confusão para treinamento e teste de cada umas dessas máquinas de comitê.

Após o treinamento e teste dessas três últimas máquinas de comitê, desenvolve-se um combinador geral para concentrar as classificações de cada uma delas e fornecer, na saída deste algoritmo, a classificação para o rótulo da amostra verificada, que pode ser “normal”, “sujidade” ou “sombreamento”. Dessa forma, tem-se um modelo computacional capaz de realizar a classificação de múltiplas classes (multiclasse) de dados simultaneamente.

Figura 38 – Matrizes de confusão de treinamento e teste de máquinas de comitê.
 (a) Treinamento – “normal/sujidade”; (b) Teste – “normal/sujidade”;
 (c) Treinamento – “normal/sombreamento”; (d) Teste – “normal/sombreamento”;
 (e) Treinamento – “sujidade/sombreamento”; (f) Teste – “sujidade/sombreamento”.



Fonte: Autoria própria (2024)

Em seguida, submete-se este novo modelo computacional, composto por três máquinas de comitê, para classificar todas as 192.094 amostras disponibilizadas por Pereira (2018), a fim de validar sua capacidade de generalização. Após este teste, obtém-se uma taxa de acertos de 98,59 % na classificação das amostras de dados. Na Figura 39, é apresentada a matriz de confusão desse teste.

Figura 39 – Matriz de confusão para conjunto de Máquinas de Comitê com três saídas.

| | | | | |
|------|--------------|--------------|--------------|---------------|
| Real | Normal | 33483 | 409 | 56 |
| | Sujidade | 1422 | 35079 | 35 |
| | Sombreamento | 337 | 1778 | 120822 |
| | | Normal | Sujidade | Sombreamento |
| | | Previsto | | |

Fonte: Autoria própria (2024)

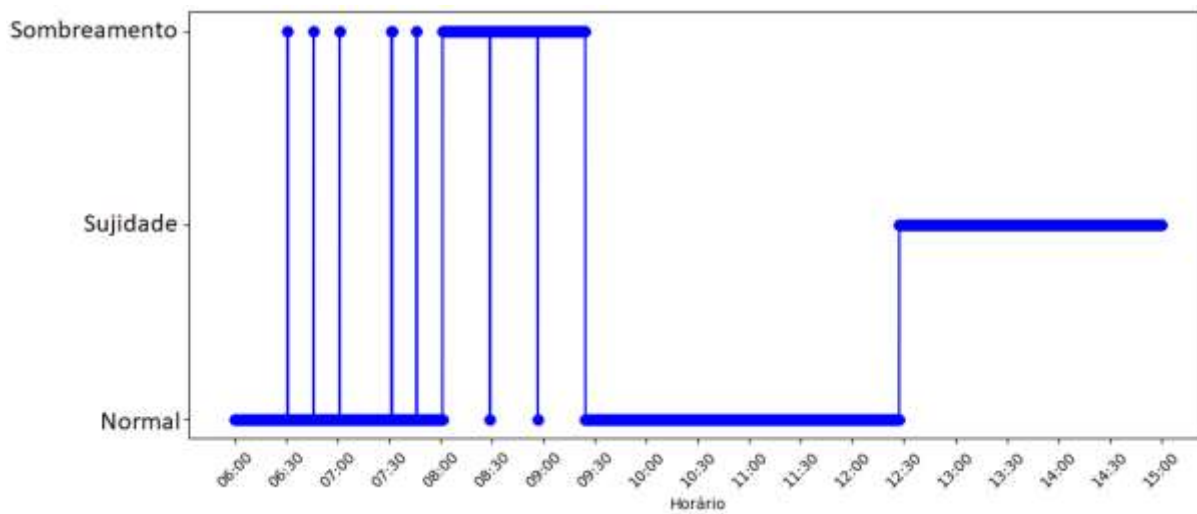
Analisando a matriz de confusão obtida com o conjunto de máquinas de comitê percebe-se que ela generaliza as amostras de treinamento e consegue classificar com bastante eficácia as amostras submetidas à avaliação. Observa-se também que mesmo sendo esse conjunto de dados com 192.094 amostras sendo desbalanceado, a dificuldade de classificação ficou somente com as amostras de sujidade, pois nota-se que elas são levemente confundidas com amostras da classe normal.

Com o êxito na classificação dos dados utilizando conjunto de máquinas de comitê, foi desenvolvida uma nova lógica de atuação desse modelo computacional a fim de realizar a classificação de amostras em série, ou seja, a cada amostra colocada na entrada do classificador, ela automaticamente é classificada, mas não é informado na saída essa indicação. Somente após a classificação de um conjunto de “k” amostras é informado pelo modelo computacional o estado do SFV, que é na realidade a classe majoritária dentre esse conjunto de amostras classificadas. Essa lógica também realiza o deslocamento temporal dessa quantidade “k” de amostras classificadas. Assim, esse conjunto de máquinas de comitê conseguem avaliar com precisão e robustez as amostras submetidas à classificação.

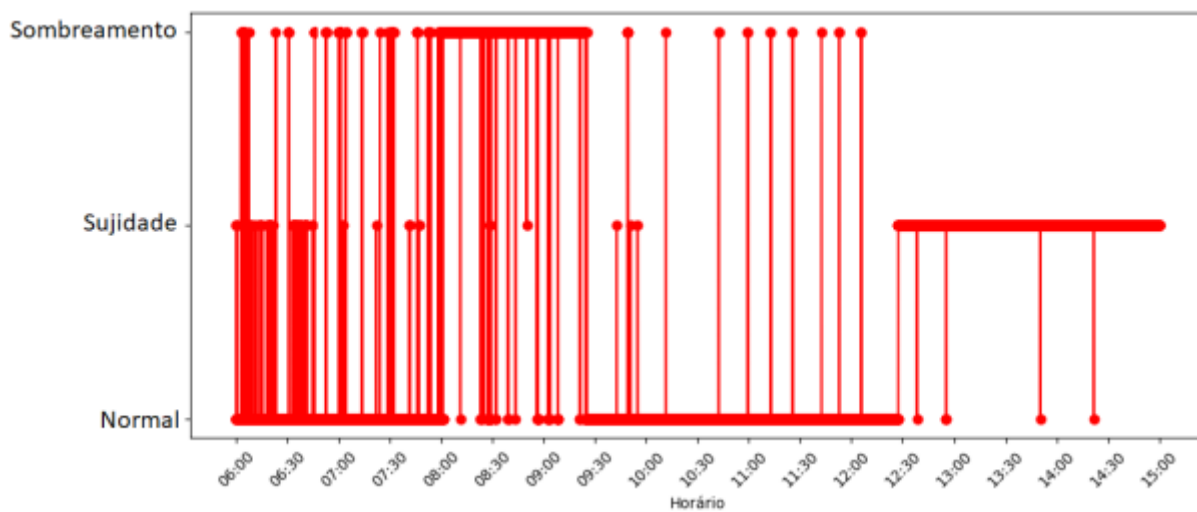
Um exemplo dessa aplicação é quando um pássaro ou avião sobrevoa os MFV. No classificador atuando esse tipo de lógica sequencial de amostras, a resposta obtida pelo usuário é de que o sistema está sombreado, já que parte dos MFV estão encobertos, mas na prática entende-se que esse tipo de situação não deve ser visto como uma condição do SFV, pois trata-se de um evento de curta duração.

Para validar esse tipo de abordagem, são selecionadas amostras sequenciais entre 06:00h da manhã e 15:00h da tarde, já que é o horário que contém amostras disponíveis, dentre as 192.094 amostras disponibilizadas por Pereira (2018) para emular as condições de situação “normal”, “sombreamento” e “sujidade”. Nas Figuras 40a, 41a, 42a, 43a e 44a é mostrado o conjunto de amostras emuladas para fins de classificação sequencial. Por outro lado, as Figuras 40b, 41b, 42b, 43b e 44b ilustram as classificações instantâneas efetuadas pelos classificadores, enquanto as Figuras 40c, 41c, 42c, 43c e 44c representam as classificações fornecidas pelo sistema de identificação de condições em SFV para um conjunto de “k” amostras.

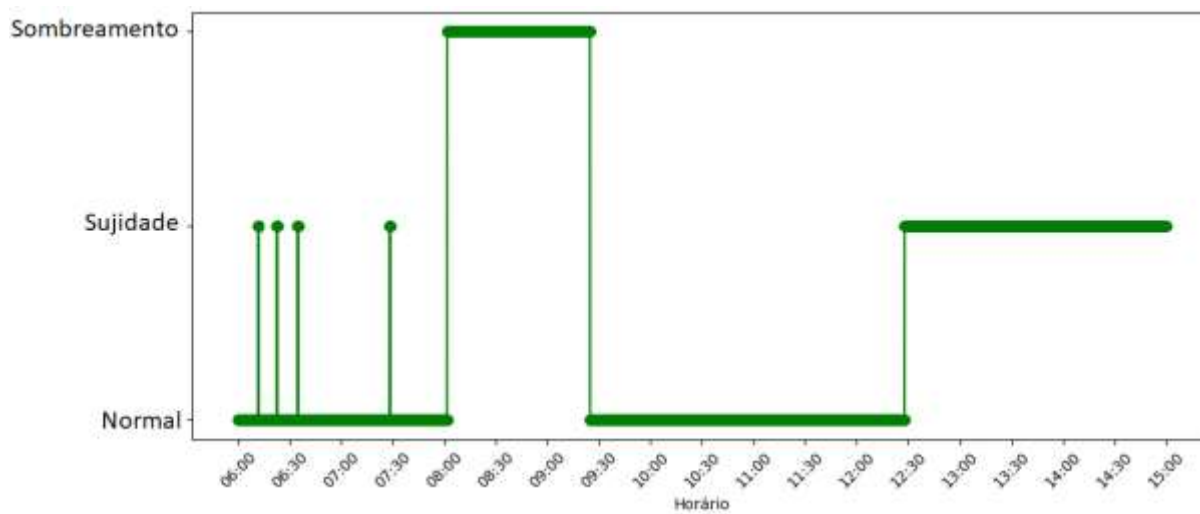
Figura 40 – Avaliação sequencial de $k = 60$ amostras. (a) Conjunto de 31535 amostras sequenciais; (b) Classificação instantânea; (c) Classificação informada.



(a)

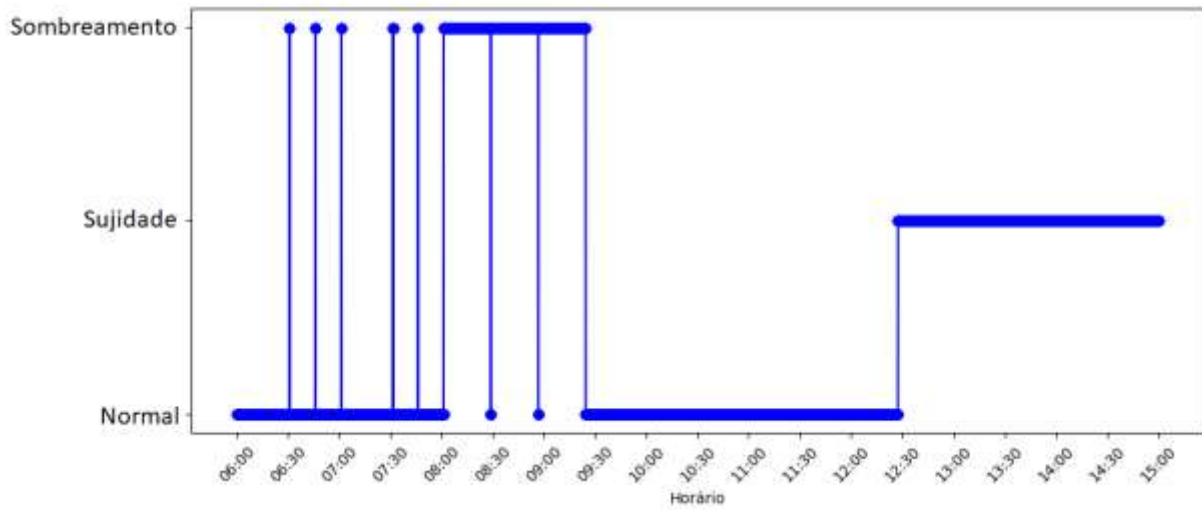


(b)

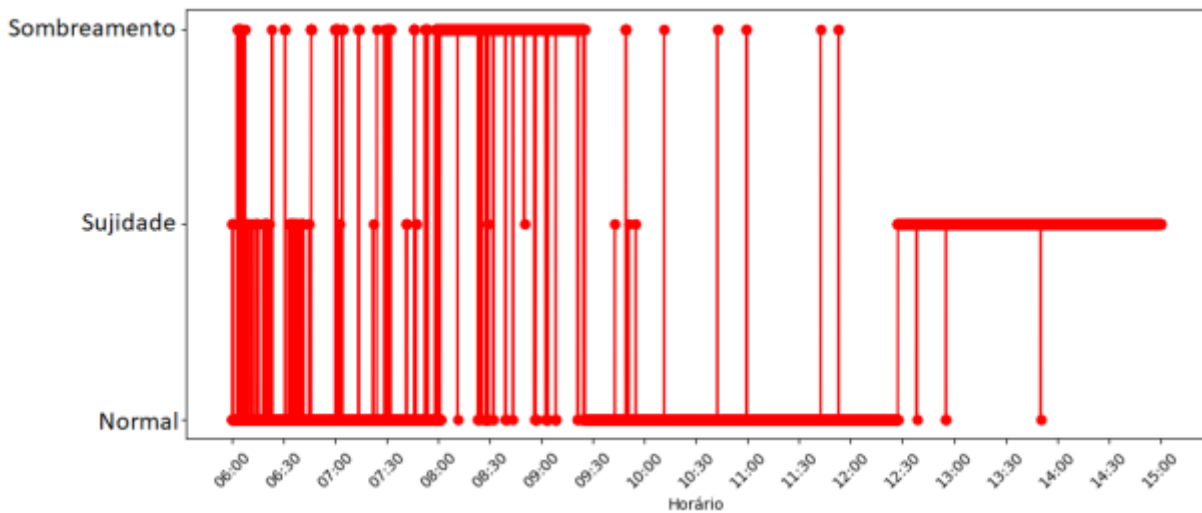


(c)

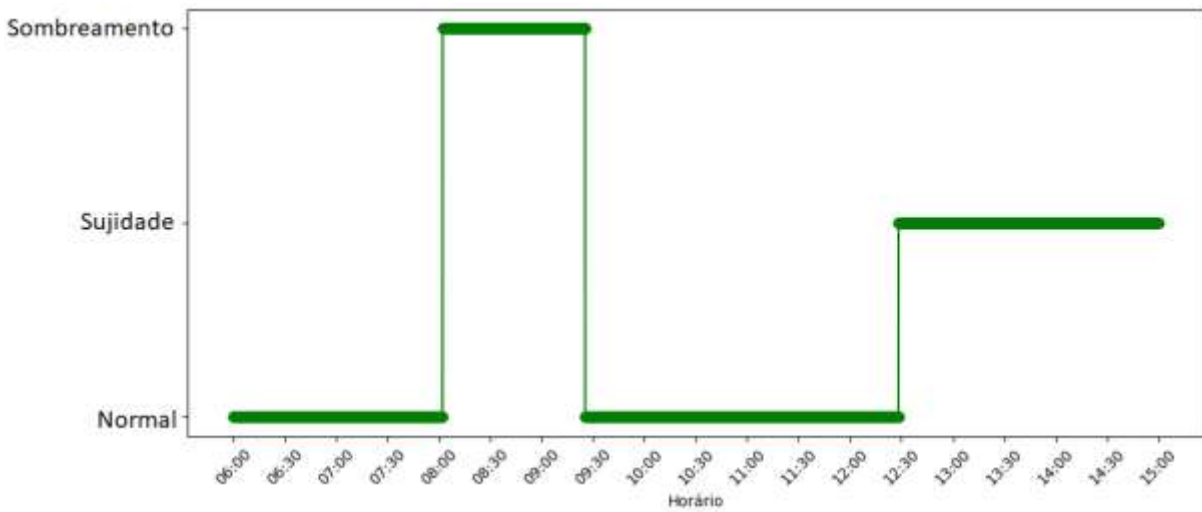
Figura 41 – Avaliação sequencial de $k = 120$ amostras. (a) Conjunto de 31535 amostras sequenciais; (b) Classificação instantânea; (c) Classificação informada.



(a)

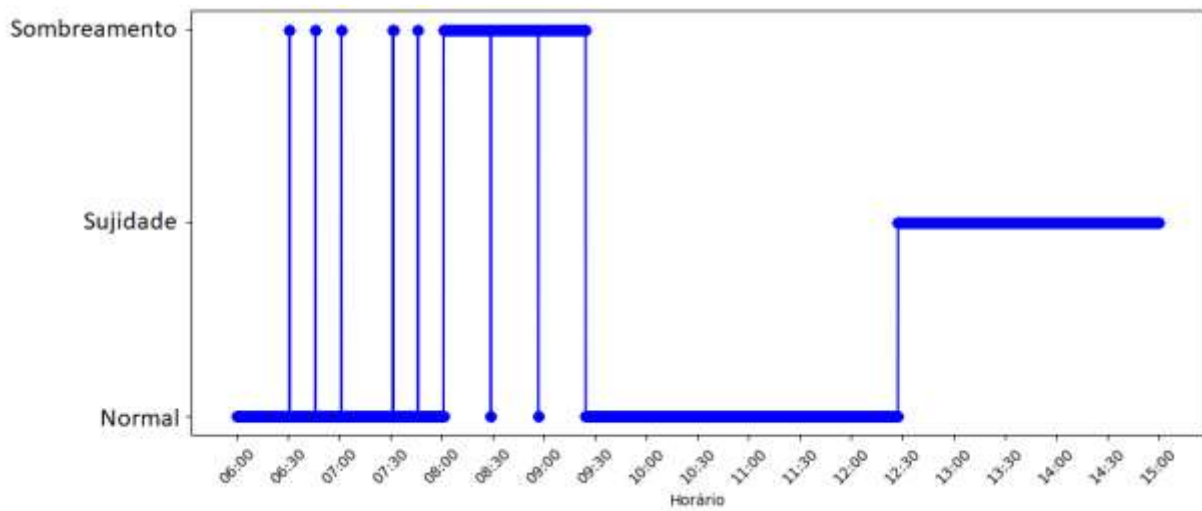


(b)

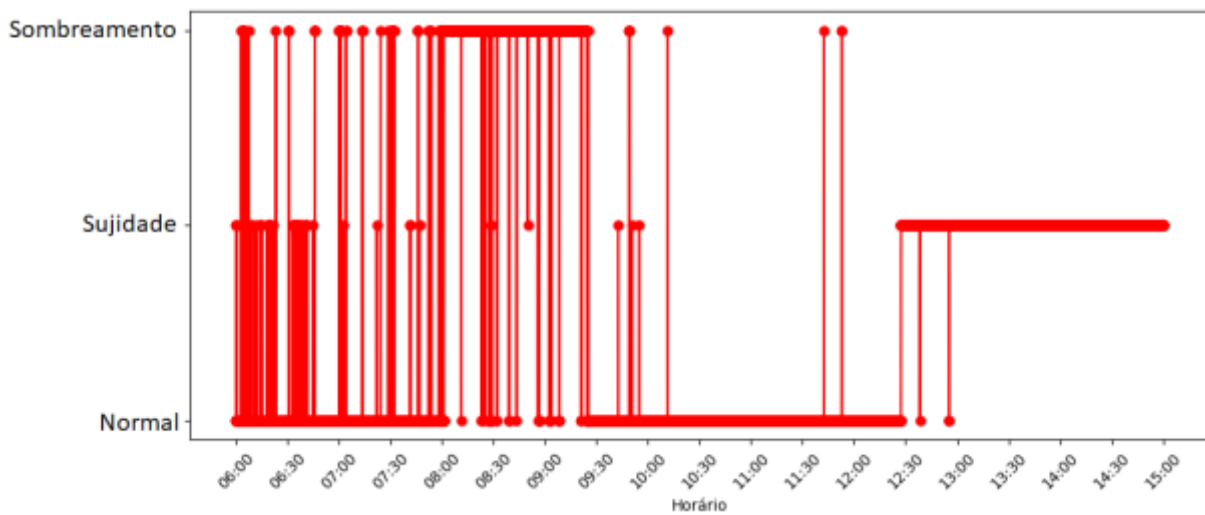


(c)

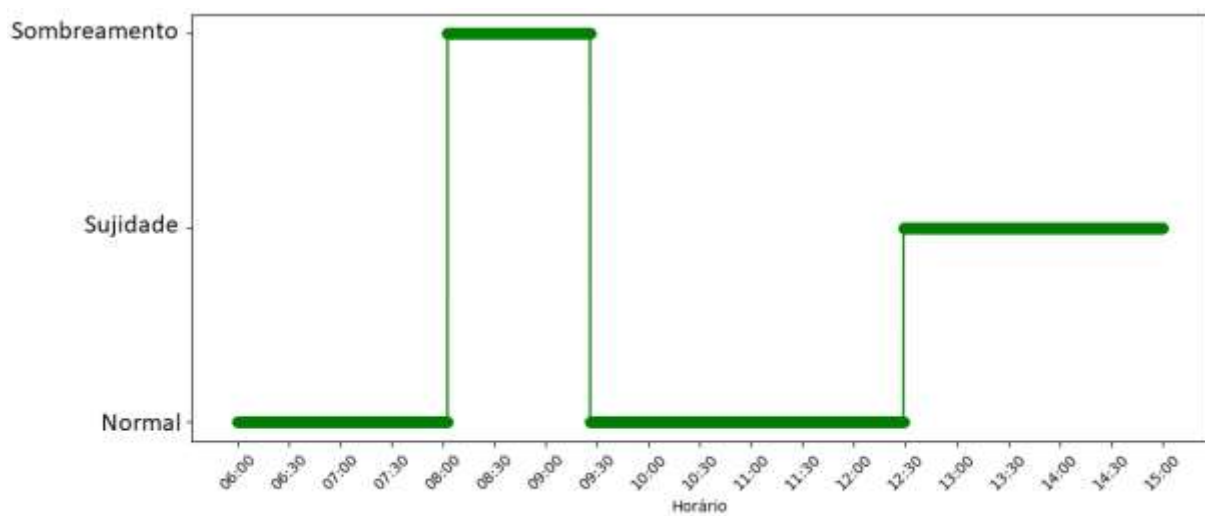
Figura 42 – Avaliação sequencial de $k = 180$ amostras. (a) Conjunto de 31535 amostras sequenciais; (b) Classificação instantânea; (c) Classificação informada.



(a)

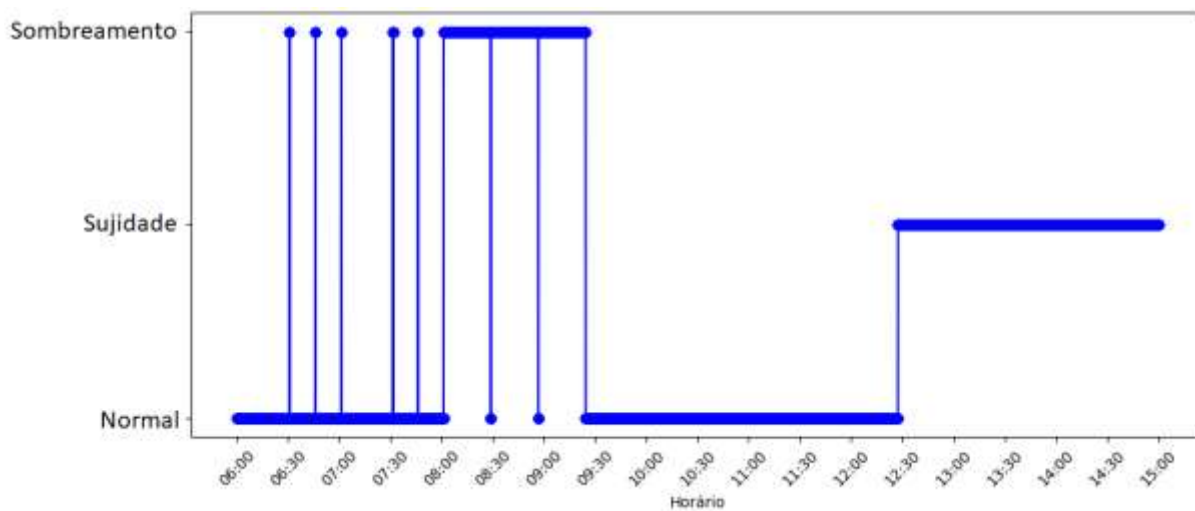


(b)

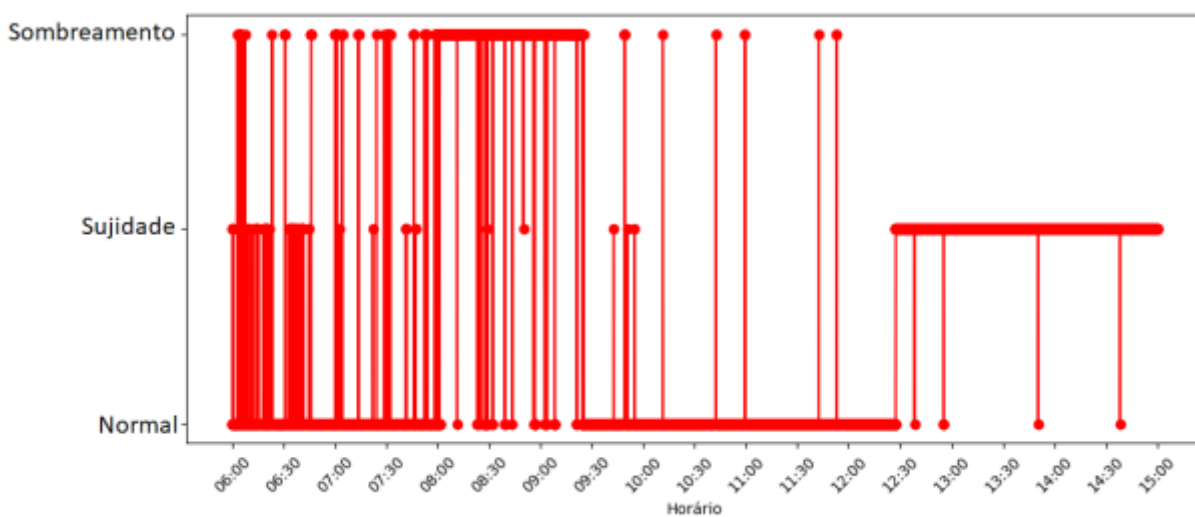


(c)

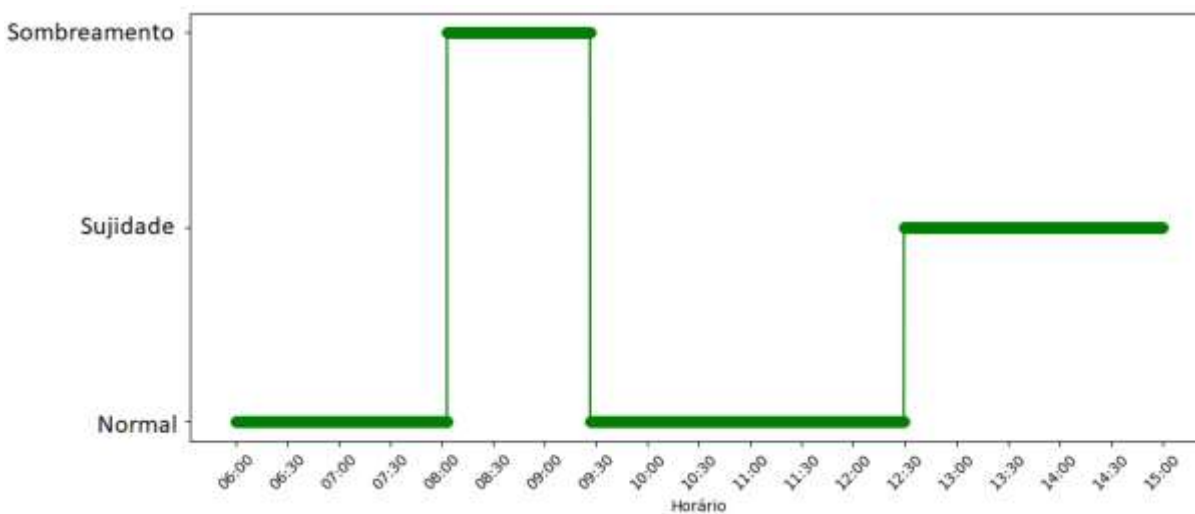
Figura 43 – Avaliação sequencial de $k = 240$ amostras. (a) Conjunto de 31535 amostras sequenciais; (b) Classificação instantânea; (c) Classificação informada.



(a)

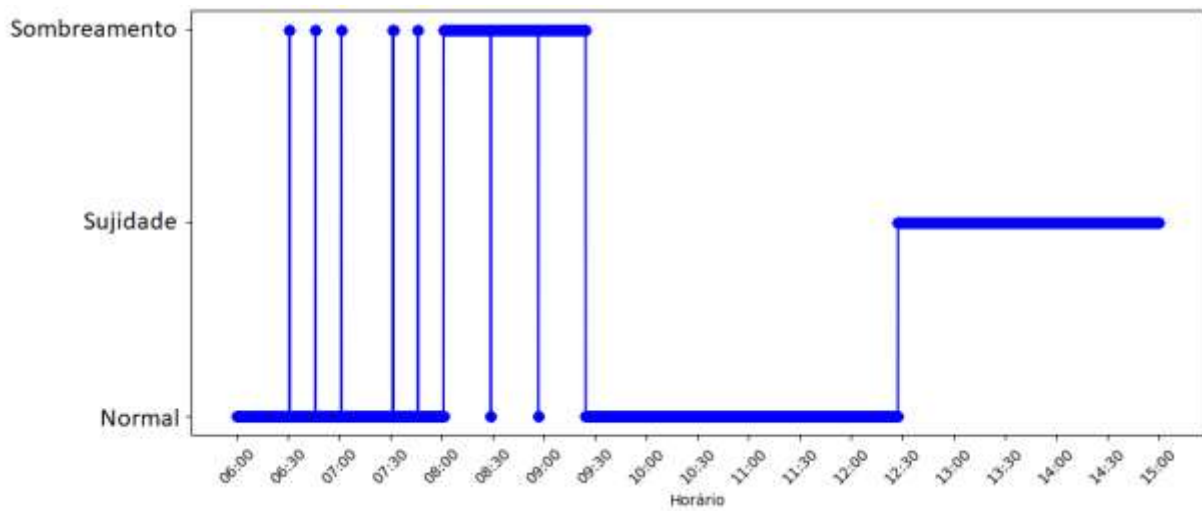


(b)

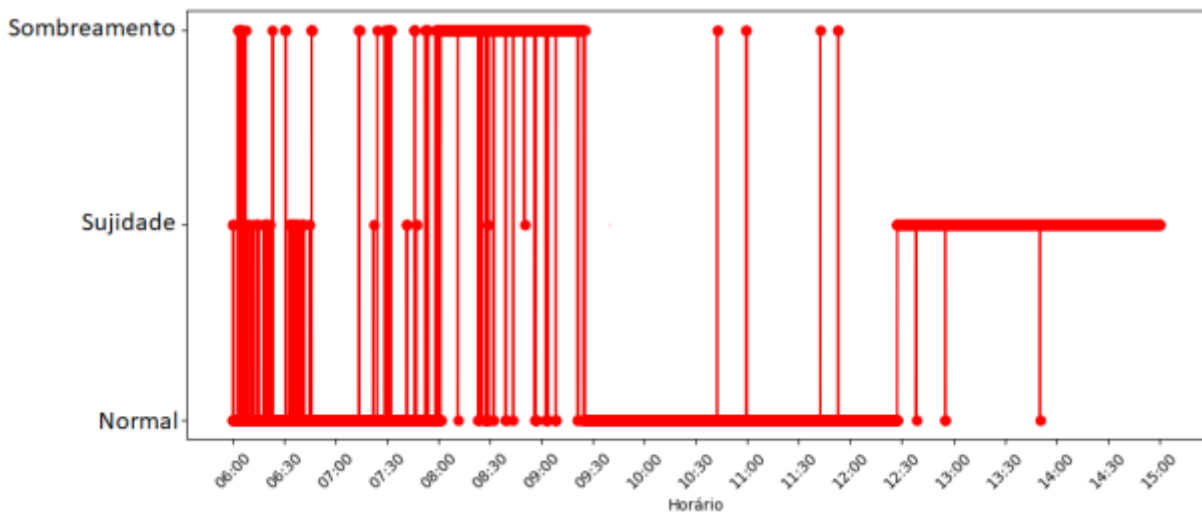


(c)

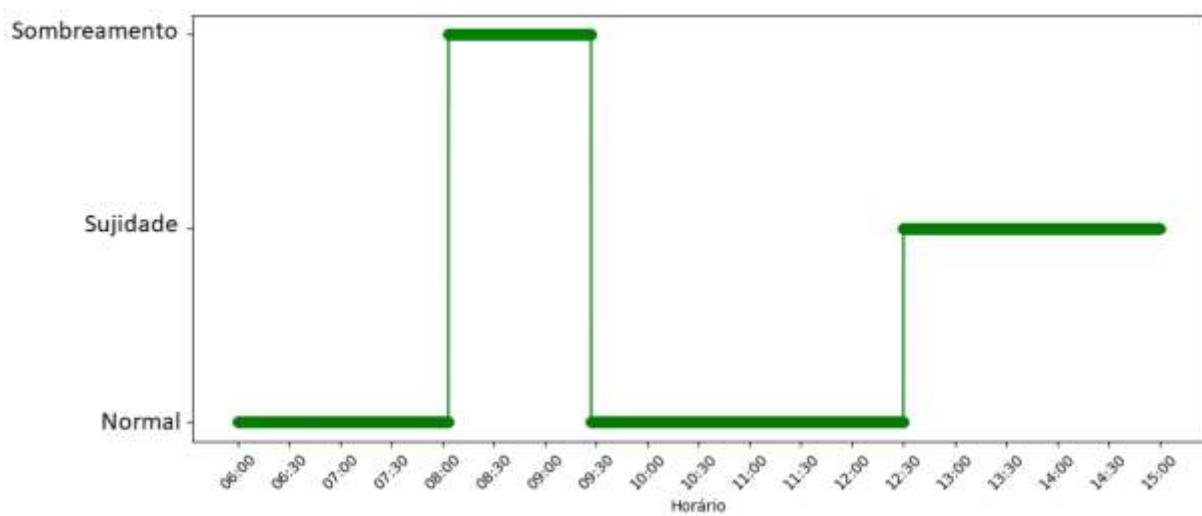
Figura 44 – Avaliação sequencial de $k = 300$ amostras. (a) Conjunto de 31535 amostras sequenciais; (b) Classificação instantânea; (c) Classificação informada.



(a)



(b)



(c)

Na Figura 40c verifica-se que para uma janela de $k=60$ amostras, ou 60 segundos, pois segundo Pereira (2018) as amostras dessa base de dados foram coletadas a cada um segundo, a indicação do conjunto de máquinas de comitê não atinge 100% de acertos, mesmo com picos emulados de condições com apenas 15 segundos de duração. Portanto, evidencia-se que esse modelo computacional possui certa dificuldade em classificar amostras de sombreamento no início da manhã de 06:00h às 09:30h. Isto inclusive pode ser verificando nas Figuras 40b, 41b, 42b, 43b e 44b, pois apresentam muita instabilidade nesse período do dia. No entanto, a partir de $k=120$ amostras, conforme observado nas Figuras 41c, 42c, 43c e 44c, o sistema de identificação de condições em SFV se comporta bem, e passa acertar todas as classificações esperadas. Em um sistema identificação de condições real, com rede MLP embarcada em plataforma microcontrolada ou em nuvem, treinada com essas amostras com um segundo de intervalo, esse período de “espera” de no mínimo 120 segundos é tolerável, visto que alguns instrumentos de medição de irradiação solar emitem respostas com um intervalo de tempo até maior, no entanto, somente com o sistema em operação é possível confirmar essa condição.

4.5 Conclusões parciais

Neste capítulo verificou-se que a aplicação do conjunto de máquinas de comitê proporciona uma generalização eficaz na classificação de amostras de dados submetidos aos testes. Apesar das máquinas de comitê ocasionalmente confundir amostras de “sujidade” com as do tipo “normal”, o modelo demonstra robustez na classificação dos dados de forma geral.

Também foi desenvolvida uma nova estratégia para verificação dos dados, na qual é realizada classificação sequencial das amostras no decorrer do tempo. Neste contexto a classificação ocorre de maneira contínua para um conjunto de dados, garantindo assim uma identificação precisa e confiável para as condições do SFV. Esta abordagem foi validada através da emulação de amostras do tipo “normal”, “sujidade” e “sombreamento” durante o período de 06:00h às 15:00h, horário no qual haviam amostras disponíveis para análise.

São realizados testes para k igual a 60, 120, 180, 240 e 300 amostras, valores esses que também representam tempo em segundos, pois essas amostras foram coletadas a cada segundo. Assim para k igual a 120 amostras, o gráfico da Figura 41c demonstrou que o sistema de identificação em SFV apresenta excelente precisão nas classificações realizadas, mesmo com um pequeno retardo de dois minutos, o que é aceitável para aplicações práticas.

5 CONCLUSÕES

Neste trabalho foi desenvolvido um sistema para identificação de condições de sombreamento e sujidade em um SFV conectado à rede elétrica. Durante a pesquisa é investigada a utilização de RNA do tipo perceptron e MLP para classificar o estado de funcionamento da planta fotovoltaica utilizando a intensidade de irradiação solar sobre os módulos e também da corrente e tensão elétrica no lado de corrente contínua do SFV. São implementadas máquinas de comitê através da combinação de redes MLP que realizam classificação binária entre as condições de normalidade e sujidade, normalidade e sombreamento e entre sujidade e sombreamento, formando um conjunto de classificadores neurais para realizar a classificação das amostras de dados dessas condições. Por fim, é desenvolvido um algoritmo para a classificação sequencial de amostras de dados e utiliza uma janela deslizante de amostras no tempo para informar a real situação do SFV.

Durante a pesquisa verifica-se que a utilização das redes perceptron não permitem alcançar altas taxas de acertos na identificação das condições no SFV pesquisado. A explicação desse fato decorre das características dos dados disponíveis para treinamento dessas RNA, pois provavelmente não são linearmente separáveis. Devido essa questão recorreu-se a utilização das redes MLP, que são classificadores neurais para realizar classificação de dados complexos e não lineares. Porém, é observado nos testes realizados que uma única MLP não seria suficiente para realizar a identificação de condições em SFV com acertos próximos a totalidade, no entanto, a combinação de três máquinas de comitês realizaram com êxito a tarefa de realizar essas classificações, alcançando acurácia de 98,59% durante a validação de dados com todas as amostras disponíveis e 100% de acertos com a metodologia de classificação sequencial de amostras, utilizando um algoritmo tipo de janela deslizante no decorrer do tempo para um grupo de 120 ou mais amostras, ou seja, o sistema só emite uma resposta sobre a classificação das condições verificadas, após definir a classe majoritária dos dados analisados, sendo essa verificação realizada constantemente a cada avaliação de uma nova amostra de informações.

Apesar desse sistema para identificação de condições de sombreamento e sujidade no SFV estudado mostrar-se satisfatório, verificou-se que ele possui algumas limitações. Foi observado o fato dele ter uma leve tendência em identificar melhor as condições de sombreamento em relação às de sujidade. Acredita-se que esse fato deriva da natureza dos dados de sujidade possuírem características muito próximas com as amostras de rótulo normal. Constatou-se também que o sistema apresenta dificuldades em classificar corretamente

amostras de sombreamento no início da manhã, especificamente entre o horário de 06:00h às 09:30h, período este em que intensidade de corrente elétrica gerada pelos módulos fotovoltaicos está abaixo de 80% do nominal do sistema, mas no decorrer do dia as inconsistências verificadas são mínimas e praticamente inexistentes com a identificação de amostras sequenciais utilizando uma janela deslizante para classificações no tempo.

Em um contexto mais amplo, destaca-se ser fundamental o desenvolvimento de sistemas para identificação de condições em SFV, pois eles desempenham um papel fundamental na maximização da eficiência, segurança e confiabilidade desses sistemas, ganhando relevância ainda maior diante da expansão contínua das instalações FV, as quais devido ao seu elevado custo para grande parte da população.

Nesse sentido, é inegável a significância deste estudo, pois contribui substancialmente para a ampliação do conhecimento no campo da identificação de condições em SFV e o impacto positivo dessa pesquisa não se limita apenas a aprimorar a eficiência e a confiabilidade da instalação FV pesquisada, mas também abre novas perspectivas para novos estudos a fim de tornar a energia solar uma alternativa mais acessível e eficaz, beneficiando toda a sociedade.

5.1 Sugestões de trabalhos futuros

Com o objetivo de melhorar o desempenho do trabalho desenvolvido e aprofundar o conhecimento no tema, algumas sugestões de desenvolvimentos futuros são listadas abaixo:

- Implementar a identificação de outros tipos de condições em SFV, tais como desconexão dos módulos e problemas no inversor FV;
- Realizar uma nova redução na dimensionalidade dos dados, excluindo a necessidade de utilizar a informação da irradiação solar, pois na maioria das instalações FV, principalmente na microgeração, não é utilizado piranômetro para medir essa grandeza, já que é um instrumento caro e geralmente usado apenas em grandes centrais de geração solar FV. Outra opção, é a utilização de dados de satélites;
- Embarcar o sistema desenvolvido em plataforma microcontrolada tipo Raspberry Pi, Rock Pi, BeagleBone ou Jetson AGX Xavier, pois possuem sistema operacional aberto e permitem a execução de códigos em linguagem Shell, C ou até mesmo Python.
- Integrar o sistema de identificação de condições em SFV desenvolvido, com aplicativo *web*, Android ou iOS, para monitoramento remoto via *internet*.

REFERÊNCIAS

- ABDOLRASOL, Maher GM et al. **Artificial neural networks based optimization techniques: A review**. Electronics, v. 10, n. 21, p. 2689, 2021.
- ABSOLAR - Associação Brasileira de Energia Solar Fotovoltaica. **Infográfico**. Disponível em: < <https://www.absolar.org.br/mercado/infografico/> > Acesso em 5 fev. 2024.
- ALVES, F. R. R. **Estudo do efeito da sujidade na eficiência de módulos fotovoltaicos**. 2018. 99 f. Dissertação de Mestrado, Engenharia de Produção e Sistemas, Pontifícia Universidade Católica de Goiás (PUC-Goiás).
- ALVES, R. H. F.; JÚNIOR, G. H. D.; MARRA, E. G.; LEMOS, R. P.; **Automatic fault classification in photovoltaic modules using Convolutional Neural Networks**. Renewable Energy, v. 179, p. 502-516, 2021.
- AL-KATHERI, A. A.; AL-AMMAR, E. A.; ALOTAIBI, M.; GHAZI, G. A. "Artificial Neural Network Application for Faults Detection in PV Systems," 2022 IEEE Delhi Section Conference (DELCON), New Delhi, India, 2022, pp. 1-6, doi: 10.1109/DELCON54057.2022.9752837.
- AHMADI, Sajjad et al. **Open-switch and open-clamping diode fault diagnosis for single-phase five-level neutral-point-clamped inverters**. IEEE journal of emerging and selected topics in power electronics, v. 9, n. 4, p. 4676-4686, 2020.
- ARAÚJO, D. N.; CARVALHO, P. C. M; DUPONT, I. M., 2019. **Efeitos da acumulação de sujeira sobre o desempenho de módulos fotovoltaicos**. Revista Tecnologia, vol. 40, n. 2, pp. 1-23. DOI: 10.5020/23180730.2019.9414.
- AZIZ, Farkhanda et al. **A novel convolutional neural network-based approach for fault classification in photovoltaic arrays**. IEEE Access, v. 8, p. 41889-41904, 2020.
- BASNET, Barun; CHUN, Hyunjun; BANG, Junho. **An intelligent fault detection model for fault detection in photovoltaic systems**. Journal of Sensors, v. 2020, p. 1-11, 2020.
- BARBOSA, B. H. G. **Computação evolucionária e máquinas de comitê na identificação de sistemas não-lineares**. Tese de Doutorado. Universidade Federal de Minas Gerais, 2009.
- BRASIL. Congresso nacional. **Lei nº 14.300, de 06 de janeiro de 2022**. Brasília, DF. Disponível em: <https://in.gov.br/en/web/dou/-/lei-n14.300-de-6-de-janeiro-de-2022-372467821>. Acesso em: 10 fev. 2023.
- CIN - Centro de Informática. **Conceito: Protótipos**. Disponível em: https://www.cin.ufpe.br/~gta/rup-vc/core.base_rup/guidances/concepts/prototypes_9D1E67A.html#:~:text=Um%20prot%C3%B3tipo%20%C3%A9%20um%20vers%C3%A3o,para%20resolver%20um%20problema%20espec%C3%ADfico. Acesso em: 10 julho 2023.

- COMERIO, A.; MUNIZ, P. R.; RAMPINELLI, M.; FARDIN, J. F. **A review on impacts from dust deposition on photovoltaic modules**. 2019. 13th IEEE International Conference on Industry Applications (INDUSCON). p. 502-508, 2019.
- COSTA, C. H. **Classificação de falhas em plantas fotovoltaicas usando aprendizado de máquina**. 2020. 78 p. Dissertação (Mestrado em Engenharia Elétrica e Informática Industrial) – Universidade Tecnológica Federal do Paraná (UTFPR), 2020.
- COSTA, C. H.; MORITZ, G.L.; LAZZARETTI, A. E.; MULINARI, B. M.; ANCELMO, H. C.; RODRIGUES, M. P.; OROSKI, E.; GOES, R. E. **A Comparison of Machine Learning-Based Methods for Fault Classification in Photovoltaic Systems**. 2019 IEEE PES Innovative Smart Grid Technologies Conference - Latin America (ISGT Latin America), Gramado, Brazil, 2019, pp. 1-6, doi: 10.1109/ISGT-LA.2019.8895279.
- CRESESB - Centro de Referência para Energia Solar e Eólica. **Componentes de um sistema fotovoltaico**. Disponível em: <http://www.cresesb.cepel.br/index.php?section=com_content&lang=pt&cid=321>. Acesso em: 15 jun. 2023.
- DA SILVA, Rogério Oliveira; SILVA, Igor Rodrigues Sousa. **Linguagem de Programação Python**. Tecnologias em Projeção, v. 10, n. 1, p. 55-71, 2019.
- DHIMISH, Mahmoud; TYRRELL, Andy M. **Photovoltaic Bypass Diode Fault Detection Using Artificial Neural Networks**. IEEE Transactions on Instrumentation and Measurement, v. 72, p. 1-10, 2023.
- DIVINA, F.; GILSON, A.; GOMÉZ-VELA, F.; GARCÍA TORRES, M.; TORRES, J.F. **Stacking Ensemble Learning for Short-Term Electricity Consumption Forecasting**. Energies 2018, 11, 949. <https://doi.org/10.3390/en11040949>
- DUPONT, I. M. **Metodologia para identificação de condições não ideais de operação de plantas fotovoltaicas**. 2021. 168 f. Tese (Doutorado em Engenharia Elétrica) – Universidade Federal do Ceará, 2021.
- ELBASET, A. A., ABDELWAHAB, S. A. M., IBRAHIM, H. A., & Eid, M. A. E. **Performance analysis of photovoltaic systems with energy storage systems**. Springer International Publishing, 2019.
- ESKANDARI, Aref; MILIMONFARED, Jafar; AGHAEI, Mohammadreza. **Line-line fault detection and classification for photovoltaic systems using ensemble learning model based on IV characteristics**. Solar Energy, v. 211, p. 354-365, 2020.
- FACELI, K.; LORENA, A. C.; GAMA, J.; ALMEIDA, T. A., CARVALHO, A. C. P. L. F. **Inteligência artificial: Uma abordagem de aprendizado de máquina**, 2nd ed., Rio de Janeiro, RJ, Brasil: LTC, 2021, pp. 28-52.
- FERNANDES, J. A. F.; GONZÁLEZ, C.; VALDERRAMA, A.; SILVA, L. N. A.; MARQUES, E. R. F. **Energias renováveis: fonte de energia limpa?** XI Congresso Brasileiro de Gestão Ambiental, Vitória, ES, 23 a 26/11/2020. Disponível em: <www.ibeas.org.br/congresso/Trabalhos2020/X-004.pdf> Acesso em: 10 aug. 2023.

FELTRIN, F. **Redes Neurais Artificiais**. Curitiba, PR, Brasil: Editora Uniorg 2020.

FONSECA, A. C.; SOUSA, C. C.; ALVES, V. P. O.; BARBOSA, C. F. O.; PINHO, J. T. PEREIRA, E. J. S. **Bancada didática demonstrativa da operação de sistemas fotovoltaicos isolados**. Anais do VIII Congresso Brasileiro de Energia Solar, Fortaleza, 01 a 05 de junho de 2020.

FRONIUS. **Fronius Primo 3.0-1 - 8.2-1 Manual de instruções**. 2018. Disponível em: <https://manuals.fronius.com/html/4204102165/pt-BR.html#0_t_0000000140>. Acesso em: 12 fev. 2023.

GANAIÉ, Mudasir A. et al. **Ensemble deep learning: A review**. Engineering Applications of Artificial Intelligence, v. 115, p. 105151, 2022.

GAO, Wei. **PV array fault detection based on deep neural network**. In: 2021 IEEE Green Technologies Conference (GreenTech). IEEE, 2021. p. 42-47.

GÉRON, A. **Mãos à Obra Aprendizado de Máquina com SciKit-Learn & Tensor Flow: conceitos, Ferramentas e Técnicas para a Construção de Sistemas Inteligentes**. Rio de Janeiro, RJ, Brasil: Alta Books, 2019, pp. 259-279.

GOOGLE (2021). **O que é o Colaboratory?** Disponível em: <https://colab.research.google.com/?hl=pt_BR#scrollTo=-Rh3-Vt9Nev9> Acesso em 12 mar. 2021.

HAYKIN, S. **Redes neurais - princípios e prática**. 2. ed., Bookman, 2001.

HICKEL, Bernardo Meyer. **O impacto no desempenho de sistemas fotovoltaicos causado pelo acúmulo de sujeira sobre os módulos FV-metodologia e avaliação através de curvas IxV em campo**. 2017. Dissertação de Mestrado, Universidade Federal de Santa Catarina (UFSC).

HONG, Ying-Yi; PULA, Rolando A. **Detection and classification of faults in photovoltaic arrays using a 3D convolutional neural network**. Energy, v. 246, p. 123391, 2022.

HU, Y. H., HWANG, J. N. **Handbook of neural network signal processing**. CRC Press LLC, 2000 N.W. Corporate Blvd., Boca Raton, Florida. 2001. ISBN 0-8493-2359-2.

ILSE, Klemens K. et al. **Fundamentals of soiling processes on photovoltaic modules**. Renewable and Sustainable Energy Reviews, v. 98, p. 239-254, 2018.

INMETRO - Instituto Nacional de Metrologia, Qualidade e Tecnologia. **Portaria n.º 140/2022**. Disponível em: <<https://www.in.gov.br/en/web/dou/-/portaria-n-140-de-21-de-marco-de-2022-389587680>> Acesso em: 05 sep. 2023.

JAURIS, G. C. **Energia solar fotovoltaica: estudo de viabilidade financeira de um sistema para atendimento de uma unidade consumidora residencial**. 2023. 49 f. Trabalho de Conclusão de Curso (Engenharia Elétrica) - Universidade Federal de Santa Maria, RS, 2023.

- KURUKURU, VS Bharath; HAQUE, Ahteshmaul; KHAN, Mohammed Ali. **Fault classification for photovoltaic modules using thermography and image processing**. In: 2019 IEEE Industry Applications Society Annual Meeting. IEEE, 2019. p. 1-6.
- LI, B.; DELPHA, C.; DIALLO, D.; MIGAN-DUBOIS, A. **Application of Artificial Neural Networks to photovoltaic fault detection and diagnosis: A review**. Renewable and Sustainable Energy Reviews, Volume 138, 2021, 110512, ISSN 1364-0321, <https://doi.org/10.1016/j.rser.2020.110512>.
- LICHAO, Wang et al. **Preprocessing method for photovoltaic power plant data based on gru neural network**. Acta Energaie Solaris Sinica, v. 43, n. 11, p. 78, 2022.
- LÓPEZ-VARGAS, A.; FUENTES, M.; VIVAR, M.; MUÑOZ-RODRÍGUEZ, F. J. **Low – cost datalogger intended for remote monitoring of solar photovoltaic stand - alone systems based on Arduino**. IEEE Sensors Journal, v. 1748, n. c, 2019.
- LU, Xiaoyang et al. **Fault diagnosis for photovoltaic array based on convolutional neural network and electrical time series graph**. Energy Conversion and Management, v. 196, p. 950-965, 2019.
- MADETI, S. R.; SINGH, S. N. **A comprehensive study on different types of faults and detection techniques for solar photovoltaic system**. Solar Energy, Elsevier, v. 158, n. June, p. 161–185, 2017. ISSN 0038092X. Disponível em: <<https://doi.org/10.1016/j.solener.2017.08.069>>
- MADUREIRA, C. P.; MENEZES, L. L.; SANTOS JUNIOR, R. B.; ASSIS, R. J. S. **A Utilização de energia fotovoltaica para transição energética na américa latina pós-acordo de paris (2015) e o regime internacional de mudanças climáticas**. Revista Conjuntura Global v. 12(2), 2023. p. 1-19.
- MANFREDINI, R. A, CORRÊA, G. N, OLIVEIRA, B. R., PAULI, S. T. Z. **Aplicações de Machine Leaning**. Pantanal Editora. 2021.
- MELLIT, A. et al. **An embedded system for remote monitoring and fault diagnosis of photovoltaic arrays using machine learning and the internet of things**. Renewable Energy, v. 208, p. 399-408, 2023.
- MOREIRA, J. R. S. et al. **Energias Renováveis, Geração Distribuída e Eficiência Energética**. p. 520. 2ª ed. Editora: LTC, 2021.
- MAIA, F. C. T. **Estudo de modelos e técnicas de detecção e diagnóstico de falhas aplicados a sistemas fotovoltaicos**. 2019. 141 p. Dissertação (Mestrado Em Engenharia Elétrica) – Universidade Federal de Sergipe, 2019.
- MEDEIROS, C. M. S. **Uma contribuição ao problema de seleção de modelos neurais usando o princípio de máxima correlação dos erros**. 2008. 162 p. Tese (Doutorado em Engenharia de Teleinformática) – Universidade Federal do Ceará, Fortaleza, 2008.

MEYERS, Bennet E.; APOSTOLAKI-IOSIFIDOU, Elpiniki; SCHELHAS, Laura T. **Solar data tools: Automatic solar data processing pipeline**. In: 2020 47th IEEE Photovoltaic Specialists Conference (PVSC). IEEE, 2020. p. 0655-0656.

MORETTIN, P.A.; SINGER, J.M. **Estatística e Ciência de Dados**, LTC, Rio de Janeiro, Brasil, 1ª ed., 2023.

NEXTRACKER. **TrueCapture**. Disponível em: <<https://nextracker.com/truecapture>>. Acesso em: 10 jul. 2023.

NURJANAH, Siti et al. **Dusting and Soiling Effect on PV Panel Performance: Case Study Open-pit Mining in South Sumatra, Indonesia**. In: 2021 International Conference on Electrical and Information Technology (IEIT). IEEE, 2021. p. 251-256.

OBAIDI, Marwah Qasim; DERBEL, Nabil. **IoT-based monitoring and shading faults detection for a PV water pumping system using deep learning approach**. Bulletin of Electrical Engineering and Informatics, v. 12, n. 5, p. 2673-2681, 2023.

PETERS, Ian Marius et al. **The value of stability in photovoltaics**. Joule, v. 5, n. 12, p. 3137-3153, 2021.

PATIL, Dipti D. et al. **Arc fault detection in dc microgrid using deep neural network**. In: 2021 4th Biennial International Conference on Nascent Technologies in Engineering (ICNTE). IEEE, 2021. p. 1-6.

PELISSON, Angelo Abdallah. **Aprendizado de máquina para previsão de geração de energia fotovoltaica em dados de estações solarimétricas**. 2021. 110 f. Dissertação (mestrado) - Universidade Federal de Santa Catarina, Campus Joinville, Programa de Pós-Graduação em Engenharia de Sistemas Eletrônicos, Joinville, 2021.

PEREIRA, M. L. **Desenvolvimento de um sistema de monitoramento e detecção de falhas para microgeração fotovoltaica**. 2018. 128 f. Dissertação (Mestrado em Energias Renováveis) - Instituto Federal de Educação, Ciência e Tecnologia do Ceará, 2018.

PESSANHA, José et al. **An approach for data treatment of solar photovoltaic generation**. IEEE Latin America Transactions, v. 18, n. 09, p. 1563-1571, 2020.

PEROZA, J.; RAMPINELLI, G. A. **Análise de desempenho e atuação de diodos de bypass em um módulo fotovoltaico comercial**. Anais do VII Congresso Brasileiro de Energia Solar – Gramado/SC, 17 a 20 de abril de 2018.

PIGNATON, E. G. **Avaliação de desempenho de usina de minigeração distribuída da UFES a partir de modelo matemático**. 2022. 100 f. Dissertação (Mestrado Profissional Em Engenharia de Controle e Automação) - Instituto Federal de Educação, Ciência e Tecnologia do Espírito Santo (IFES), 2022.

PILLAI, D. S.; RAJASEKAR, N. **A comprehensive review on protection challenges and fault diagnosis in PV systems**, Renewable and Sustainable Energy Reviews, Volume 91, 2018, Pages 18-40, ISSN 1364-0321, <https://doi.org/10.1016/j.rser.2018.03.082>.

PEREIRA, V. M. S. **Análise das perdas de eficiência em módulos fotovoltaicos**. 2021. Trabalho de Conclusão de Curso (TCC), Universidade Federal Rural do Semi-Árido (UFERSA). Disponível em: <
<https://repositorio.ufersa.edu.br/server/api/core/bitstreams/940eea01-6e83-41d1-a5bf-f9cee0689b23/content>>. Acesso em: 20/03/2023.

PRAKASH, P.; KUMAR, K. **Artificial Neural Network Based Fault Prediction and Detection in Grid Computing**. In: 2022 IEEE 9th Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON). IEEE, 2022. p. 1-5.

REEVES, B. **Solar Power DIY Handbook**. Book Benefits, 2018. Disponível em: <
<https://www.bookbenefits.com/2021/04/solar-power-diy-handbook-so-you-want-to.html>>
 Acesso em: 5 sep, 2023.

REINDERS, Angele et al. **Development of a big data bank for PV monitoring data, analysis and simulation in COST Action ‘PEARL PV’**. In: 2019 IEEE 46th Photovoltaic Specialists Conference (PVSC). IEEE, 2019. p. 1388-1392.

ROMANHOLO, P. V. V. **Análise dos efeitos da sujidade no sistema fotovoltaico de 34 kWp da EMC/UFG**. 2019. 112 f. Dissertação (Mestrado em Engenharia Elétrica e da Computação) - Universidade Federal de Goiás, Goiânia, 2019.

ROSENBLATT, Frank et al. **Principles of neurodynamics: Perceptrons and the theory of brain mechanisms**. Washington, DC: Spartan books, 1962.

ROUANI, L.; HARKAT, M. F.; KOUADRI, A.; MEKHILEF, S. **Shading fault detection in a grid-connected pv system using vertices principal component analysis**. Renewable Energy, Volume 164, 2021, Pages 1527-1539, ISSN 0960-1481,
<https://doi.org/10.1016/j.renene.2020.10.059>. Acesso em: 10 jul. 2023.

SILVA, I. N.; SPATTI, D. H.; FLAUZINO, R. A. **Redes neurais artificiais: para engenharias e ciências aplicadas**, Artliber, São Paulo, Brasil, 2ª ed., 2016.

SILVA, Eduardo Goulart da. **Classificação do tipo de sombreamento e de sujidade em módulos fotovoltaicos utilizando LGBM e curvas I-V**. 2023. Trabalho de Conclusão de Curso (TCC). Universidade Federal de Santa Maria (UFSM), Engenharia de Controle e Automação. Disponível em: <
https://repositorio.ufsm.br/bitstream/handle/1/29986/Silva_Eduardo_Goulart_da_2023_TCC.pdf?sequence=1&isAllowed=y>. Acesso em: 15/06/2023.

SHARIAH, Adnan; AL-IBRAHIM, Ehab Ali. **Impact of Dust and Shade on Solar Panel Efficiency and Development of a Simple Method for Measuring the Impact of Dust**. Journal of Sustainable Development of Energy, Water & Environment Systems (JSDEWES), v. 11, n. 2, 2023.

SOFFIAH, K.; MANOHARAN, P. S.; DEEPAMANGAI, P. **Fault detection in grid connected PV system using artificial neural network**. In: 2021 7th International Conference on Electrical Energy Systems (ICEES). IEEE, 2021. p. 420-424.

SURESH, V.; JANIK, P.; REZMER, J.; LEONOWICZ, Z. **Forecasting Solar PV Output Using Convolutional Neural Networks with a Sliding Window Algorithm**. *Energies* 2020, 13, 723. <https://doi.org/10.3390/en13030723>

SYAFARUDDIN; ZINGER, D. S. **Review on Methods of Fault Diagnosis in Photovoltaic System Applications**. *Journal of Engineering Science and Technology Review*, v. 12, n. 5, 2019. p. 53 – 66.

TI - Texas Instruments. **Analog Front End for Arc Detection in Photovoltaic Applications Reference Design - Revised April 2023**. Disponível em: <https://www.ti.com/lit/ug/tiduez9b/tiduez9b.pdf?ts=1700574340894&ref_url=https%253A%252F%252Fwww.google.com%252F>. Acesso em: Acesso em: 10 jul. 2023.

YAHYAOU, I. **Advances in Renewable Energies and Power Technologies**. v. 1: Solar and Wind Energies. University Carlos III of Madrid, Spain, 2018. Elsevier Inc. 2018.

TWIDELL, J. **Renewable Energy Resources**. Routledge; 4th ed. 2021.

YANG, Nien-Che; ISMAIL, Harun. **Voting-based ensemble learning algorithm for fault detection in photovoltaic systems under different weather conditions**. *Mathematics*, v. 10, n. 2, p. 285, 2022.

ANEXO A

Todos os códigos fonte dos algoritmos desenvolvidos nesta pesquisa podem ser verificados em: <https://abre.ai/dm-isc-codigos>