

## Classificação Multiestágio Aplicada Ao Monitoramento Não-Intrusivo De *String* Fotovoltaica

Caroline T. S. Passos\*, Wanderley C. Celeste\*, Leonardo J. Silvestre\*, Hélder R. O. Rocha\*\*

\*Centro Universitário do Norte do Espírito Santo, Rodovia Governador Mário Covas, Km 60, 29932-540, São Mateus, ES, Brasil (e-mails: caroltedescosantos@gmail.com, wanderley.celeste@ufes.br, leonardo.silvestre@ufes.br, daniel.coura@ufes.br, luis.rigo@ufes.br).

\*\* Universidade Federal do Espírito Santo, Av. Fernando Ferrari, 514, 29075-910, Vitória, ES, Brasil (e-mail: helder.rocha@ufes.br)

---

**Abstract:** The great current demand for photovoltaic electric energy generation has required the search for solutions to problems of operation of such systems. Among these, the so-called atypical photovoltaic *string* operating conditions stand out, leading the system to a loss of efficiency in its generation capacity or even to severe failures. Regardless of the cause of the atypical *string* condition, its effect is somehow registered in the electrical generation itself. Each causative factor results in an electrical signature, which allows it to be identified. Thus, the objective of this article is to identify the operating condition of a photovoltaic *string* among twenty possible states, one is the normal condition, and the others are atypical ones. It is used a non-intrusive monitoring method based on the use of electrical voltage and current samples generated by the PV *string* itself. A multistage classification methodology is adopted to divide a complex problem into smaller and less complicated sub-problems. Therefore, they are considered two classification stages and, in each one, the K-nearest Neighbors, Support Vector Machine, and Multi-Layer Perceptron classification technics are used. The results achieved led to an average accuracy of 93.9% when using the classifier with the best performance in each subproblem treated.

**Resumo:** A grande demanda atual por geração de energia elétrica fotovoltaica tem exigido a busca de soluções para problemas de operação de tais sistemas. Dentre estas, destacam-se as chamadas condições de operação atípicas de *string* fotovoltaica (PV), levando o sistema a uma perda de eficiência em sua capacidade de geração ou mesmo a falhas graves. Independentemente da causa da condição atípica da *string*, seu efeito é de alguma forma registrado na própria geração elétrica. Cada fator causador resulta em uma assinatura elétrica, que permite identificá-lo. Assim, o objetivo deste artigo é identificar a condição de operação de uma cadeia fotovoltaica dentre vinte estados possíveis, um é a condição normal e os demais são atípicos. É utilizado um método de monitoramento não intrusivo baseado no uso de tensão elétrica e amostras de corrente geradas pela própria cadeia fotovoltaica. Uma metodologia de classificação multiestágio é adotada para dividir um problema complexo em subproblemas menores e menos complicados. Para tanto, consideram-se duas etapas de classificação e são utilizadas, em cada uma delas, as técnicas de classificação K-neest Neighbors (kNN), Support Vector Machine (SVM) e Multi-Layer Perceptron (MLP). Os resultados alcançados levaram a uma acurácia média de 93,9% ao utilizar o classificador com melhor desempenho em cada subproblema tratado.

**Keywords:** Photovoltaic string; Photovoltaic panels; Artificial Intelligence; Machine Learning; Multistage classification.

**Palavras-chaves:** *String* fotovoltaica; Painéis fotovoltaicos; Inteligência artificial; Aprendizado de máquina; Classificação multiestágio.

---

### 1. INTRODUÇÃO

De acordo com a Agência Internacional de Energia (IEA), a geração solar fotovoltaica (PV) se manterá como a principal forma de geração de energia elétrica renovável no mundo pelo menos até o ano de 2026, com um crescimento anual estimado em 17%, alcançando a marca de cerca de 200 GW de capacidade instalada naquele ano, conforme pode ser visto na Fig. 1 (IEA, 2021).

A produção de energia a partir de fonte solar tem se tornado uma opção cada vez mais utilizada, tanto em pequena, quanto em larga escala. O motivo desse sucesso é o fato de ser uma fonte renovável e gratuita. Além disso, o processo final de geração de energia elétrica a partir de fonte solar é limpo, ou seja, não emite qualquer tipo de poluição ou emissão de gases de efeito estufa. Soma-se, ainda, o fato de que os equipamentos responsáveis pela captação de energia solar serem silenciosos, não tendo interferência acústica no meio-

ambiente, permitindo a instalação inclusive no meio urbano (Micheletti, *et al.* 2021).

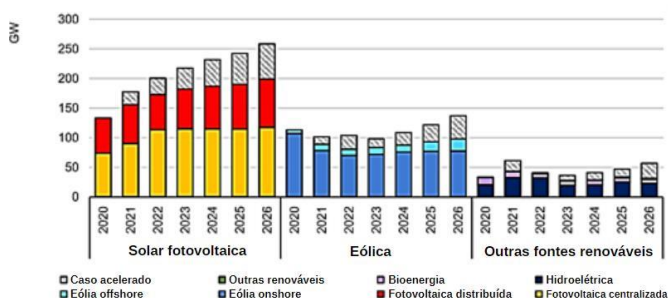


Fig 1. Evolução da energia solar PV no mundo. Fonte: IEA, 2021.

De acordo com Monteiro (2018), sistemas PVs são compostos por uma cadeia de módulos (ou painéis) PVs conectados eletricamente em série, formando um subsistema de geração de energia elétrica chamado de *string* PV. Assim, toda e qualquer falha associada a uma *string* PV a pode afetar o desempenho global do sistema.

O grau de dificuldade para detectar falhas em sistemas PVs é maior se comparado a sistemas de geração elétrica convencionais (hidráulicos, térmicos e eólicos). Isso devido à variação da potência gerada (tensão e corrente elétricas) e da não linearidade da curva corrente versus tensão, também chamada de curva I-V (Trindade, 2020). Dessa maneira, a detecção de falhas por dispositivos convencionais de proteção, como fusíveis e disjuntores, torna-se difícil, pois eles se baseiam em grandezas elétricas individuais, como a corrente elétrica (Phinikarides, *et al.* 2014).

Assim, sobre os dados de entrada para detecção de falhas, neste trabalho, foram utilizados apenas dados de tensão e corrente como entrada no sistema classificador. Esses dados são suficientes para resolução desse problema e podem-se conseguir com equipamentos específicos. Em trabalhos que buscam resolver problemas parecidos com os problemas aqui propostos, são comumente utilizadas tensão, corrente, irradiância e temperatura, para que sejam analisadas as falhas em painéis. Em Ji *et al.* (2017), por exemplo, foi proposto um método de reconfiguração de arranjo PV dinâmico para detectar automaticamente, classificar e localizar falhas, analisando a curva I-V para classificar as falhas e usou-se uma abordagem sistemática de cima para baixo para detectar e localizar falhas que podem ocorrer simultaneamente. Os mesmos tipos de dados de entrada também são usados em Chen *et al.* (2018), Belaout *et al.* (2018), Garoudja *et al.* (2017), Khoshnami & Sadeghkhan (2018), Navid *et al.* (2020), Aziz *et al.* (2020), Li *et al.* (2019), Wang *et al.* (2016), Dhimish *et al.* (2017), Qureshi *et al.* (2020) e Hussain *et al.* (2020).

Em relação ao uso de Inteligência Artificial (em inglês, *Artificial Intelligence*, AI) para realizar a detecção das falhas, diversos modelos são propostos, variando de redes neurais a classificadores de menor complexidade. O sistema aqui proposto buscou classificar os dados através da comparação

do resultado de acurácia entre três classificadores: K-nearest Neighbors (kNN), o Support Vector Machine (SVM) e o Multi-Layer Perceptron (MLP). Nos trabalhos encontrados na literatura para resolver basicamente os mesmos problemas, o trabalho de Chen *et al.* (2018) propõem um modelo de vetor autoregressivo (AR), em Belaout *et al.* (2018) foi usado Classificador Neuro-Fuzzy Adaptável Multiclasse (MC-NFC), os quais têm a capacidade de lidar com informações incertas e imprecisas, mas não podem atualizar e ajustar seus parâmetros automaticamente, assim como em Dhimish *et al.* (2017). Para superar essa desvantagem, alguns algoritmos de aprendizagem supervisionada foram aplicados com base no treinamento do conjunto de dados. Os autores em Garoudja *et al.* (2017) propuseram o algoritmo ABC (Artificial Bee Colony) para detectar falhas, o qual é usado para prever as coordenadas MPP (Máximo Ponto de Potência) de corrente, tensão e potência de todo o painel.

Khoshnami & Sadeghkhan (2018) foram os autores que criaram o sistema mais parecido com o sistema multiestágio desse trabalho dentre todos os trabalhos aqui listados, que foi um algoritmo de detecção de falha de dois estágios, porém; os estágios consistiram em fins diferentes dos estágios aqui apresentados. O Estágio 1 de Khoshnami & Sadeghkhan (2018) baseia-se em identificar os ruídos de tensão e corrente a fim de serem mitigados usando os filtros passa-baixo (LPFs). Seguido da identificação, os dados são normalizados para aumentar a escalabilidade do algoritmo proposto. O Estágio 2 é um processo de tomada de decisão do algoritmo entre se o painel opera com falha ou se está sofrendo um sombreamento parcial. Para distinguir entre a condição de falha ou sombreamento parcial, o estudo propõe um recurso baseado em forma de onda. A base desse recurso é a taxa de variação do *array* com os dados de potência. Esse estudo não identifica qual o painel que pode estar causando a baixa potência por estar operando com falha, como não identifica a falha a qual ele está submetido. A identificação acaba sendo superficial e incompleta devido à falta de informações sobre como resolver o problema de operação sob falha.

Navid *et al.* (2020) desenvolveram um algoritmo que tem a técnica validada por meio de imagens do comportamento térmico da superfície frontal do módulo solar. Por meio da câmera de imagem térmica, o ponto de acesso no módulo solar é detectado e essas medições são comparadas com a Rede Neural Artificial (ANN) desenvolvida em duas camadas. Já em Aziz *et al.* (2020), foram utilizadas redes neurais convolucionais (CNN) profundas bidimensionais (2-D) para extrair recursos 2-D de escalogramas gerados a partir de dados do sistema PV, a fim de detectar e classificar com eficácia as falhas. Em Li *et al.* (2019), foi proposto um Modelo de Simulação de Falha Rápida Baseado em Código (CFFSM) que é estabelecido para simular curvas I-V do painel PV sob várias condições de falhas e posteriormente compará-las à condição normal do painel, para, assim, detectar falhas. Hussain *et al.* (2020) desenvolveram, implantaram e validaram um sistema PV de detecção de falhas usando a função de base radial (RBF), exigindo apenas dois parâmetros como entrada para a RNA (irradiância solar e potência de saída). Wang *et al.* (2016) desenvolveram um algoritmo de Otimização de Enxame de Partículas

Modificado (MPSO). É aplicável para resolver de problemas de extração de parâmetros intrínsecos, ou pouco conhecidos. As grandezas elétricas mensuráveis nesta aplicação incluem apenas tensão e corrente do painel. Em Qureshi et al. (2020), uma técnica de análise de componente independente (ICA)-based wireless para diagnóstico de falhas é proposta.

Em relação às falhas identificadas nas *strings*, Chen et al. (2018) identificou em seu algoritmo sombreamento, falha de aterramento, falha de linha-linha e falha de arco em corrente contínua (CC). Ji et al. (2017) detectou falhas de circuito aberto, incompatibilidade e curto-circuito. Em Belaout et al. (2018), foram verificadas as falhas de sombreamento parcial, resistência em série aumentada, diodo de by-pass curto-circuitado, impedância do diodo de by-pass, módulo PV em curto-circuito. As falhas que foram identificadas em Garoudja et al. (2017) foram: curto-circuito, circuito aberto, sombreamento parcial, que foram as mesmas identificadas por Navid et al. (2020), porém, além dessas, também foi identificada a falha de aterramento. Khoshnami & Sadeghkhan (2018) buscaram detectar se o painel estava com falha (não especificada) ou estava em sombreamento parcial. Aziz et al. (2020) buscaram por falhas de sombreamento parcial, alta impedância, baixa incompatibilidade de localização e a presença de Ponto de Potência Máxima (MPPT). Li et al. (2019) estabeleceram parâmetros para padronizar os valores que os painéis tinham que atender para que não fosse considerado operação com falha.

É muito comum utilizar AI para o reconhecimento de atuação sob falha em painéis PV; porém, os autores, em sua maioria, não se preocuparam com a complexidade dos algoritmos desenvolvidos. O que acontece na proposta apresentada neste artigo, primeiramente, é buscar diminuir a complexidade de um sistema desenvolvido para resolver o mesmo problema de detecção de falhas. Por isso, utilizou-se mais de uma etapa com dados apenas de tensão e corrente como dados de entrada.

Assim, o objetivo neste trabalho é desenvolver um sistema baseado em inteligência artificial capaz de identificar falhas em *strings* fotovoltaicas e que seja passível de ser implementado em plantas fotovoltaicas diversas, isto é, desde sistemas distribuídos de baixa potência a grandes centrais de geração de média e alta potência.

Neste sentido, o presente artigo está organizado da seguinte forma: a Seção 2 apresenta a metodologia desenvolvida através da utilização dos classificadores, bem como todo o pré-processamento realizado no banco de dados. Na Seção 3 são apresentados os resultados obtidos e na Seção 4 apresenta-se uma conclusão dos resultados obtidos e propostas de trabalhos futuros.

## 2. METODOLOGIA

### 2.1 String PV e equipamentos de medição

O sistema de identificação proposto usa os dados de resposta de uma planta fotovoltaica composta de uma *string* com seis

módulos PVs, conforme visto na Fig. 2. A *string* gera tensão em CC, a qual é convertida para corrente alternada (CA) através do inversor, para que possa ser injetada em uma rede elétrica de CA.



Fig 2. String fotovoltaica.

Os painéis utilizados na *string* são de silício policristalino com potência de 260 Wp na condição padrão de teste (irradiância de 1000W/m<sup>2</sup>, temperatura de 25°C e espectro AM1.5). Nessa potência, a tensão e a corrente fornecidas, respectivamente, são 30,4 V e 8,56 A. Já a tensão de circuito aberto é de 37,5 V e a corrente de curto-circuito é de 9,12 A. As dimensões de cada módulo são de 1638 x 982 x 40 mm.

A corrente nominal da *string* é a mesma de um único módulo, isto é 8,56A, visto que os painéis estão conectados em série. Sabe-se que em uma *string*, a tensão e a potência nominais são a multiplicação da tensão e da potência de um único painel pelo número de painéis da *string*. Então, a tensão nominal da *string* é de 182,4 VCC e sua potência nominal é de 1560 W.

A planta conta, ainda, com um módulo com sensores de tensão e corrente, alocado entre a *string* e o inversor, que permite monitorar a potência instantânea gerada. Um módulo de aquisição de dados permite a amostragem e digitalização de tal potência na forma de amostras de tensão e corrente de geração elétrica. Por fim, tais amostras podem ser armazenadas para fins de desenvolvimentos *offline* ou em sistemas rodando *online*. A Fig. 3 apresenta o diagrama de blocos do sistema descrito. Os sensores estão identificados como S1, S2, S3, S4, S5, S6 e S7.

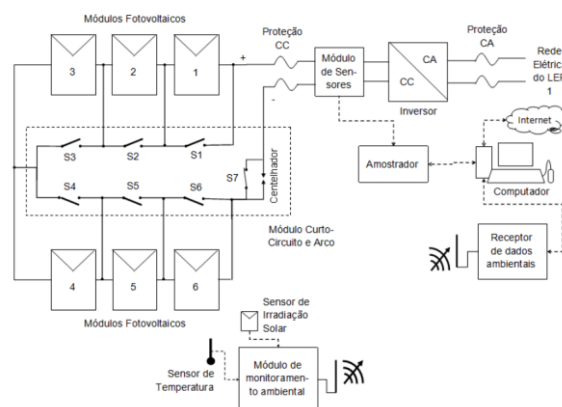


Fig 3. Estrutura experimental da planta fotovoltaica.

## 2.2 Conjunto de Dados Experimentais

O conjunto de dados utilizado é composto por amostras de tensão e de corrente oriundas da geração de energia elétrica da *string* fotovoltaica quando operando em uma dentre as 20 condições de funcionamento consideradas. A Tabela 1 apresenta tais condições de funcionamento, com as respectivas codificações adotadas.

**Tabela 1. Condições de funcionamento da *string* PV**

Código	Condição de funcionamento
0	Normal
1	Painel #1 inteiro sombreado
2	Painel #2 inteiro sombreado
3	Painel #3 inteiro sombreado
4	Painel #4 inteiro sombreado
5	Painel #5 inteiro sombreado
6	Painel #6 inteiro sombreado
7	Painel #1 com setor sombreado
8	Painel #2 com setor sombreado
9	Painel #3 com setor sombreado
10	Painel #4 com setor sombreado
11	Painel #5 com setor sombreado
12	Painel #6 com setor sombreado
13	Painel #1 em curto-circuito
14	Painel #2 em curto-circuito
15	Painel #3 em curto-circuito
16	Painel #4 em curto-circuito
17	Painel #5 em curto-circuito
18	Painel #6 em curto-circuito
19	Arco-elétrico

Para cada uma das condições de funcionamento apresentadas na Tabela 1, foram adquiridas amostras da tensão e da corrente em diversos horários do dia, entre os dias 30/12/2019 e 28/01/2020, a fim de obter dados em diferentes níveis de irradiação solar e temperatura ambiente.

As aquisições são organizadas em ciclos de coleta, onde cada ciclo deve ocorrer no menor intervalo de tempo possível, a fim de garantir que a *string* PV seja submetida à menor variação possível de condição climática (irradiação solar e temperatura ambiente). Cada ciclo de coleta subdivide-se em 20 etapas de aquisição, onde cada etapa correspondente a uma condição de funcionamento diferente. Em cada etapa de aquisição são coletadas amostras de tensão e de corrente.

## 2.3 Classificação Multiestágio

A classificação multiestágio consiste em escalonar o processo de classificação em duas ou mais etapas até chegar ao resultado. O objetivo disso é quebrar um sistema complexo em múltiplos subsistemas de menor complexidade, reduzindo

assim tanto o esforço necessário para o treinamento dos sistemas, quanto para a classificação propriamente dita. A estrutura de classificação multiestágio pode ser vista na Fig. 4.

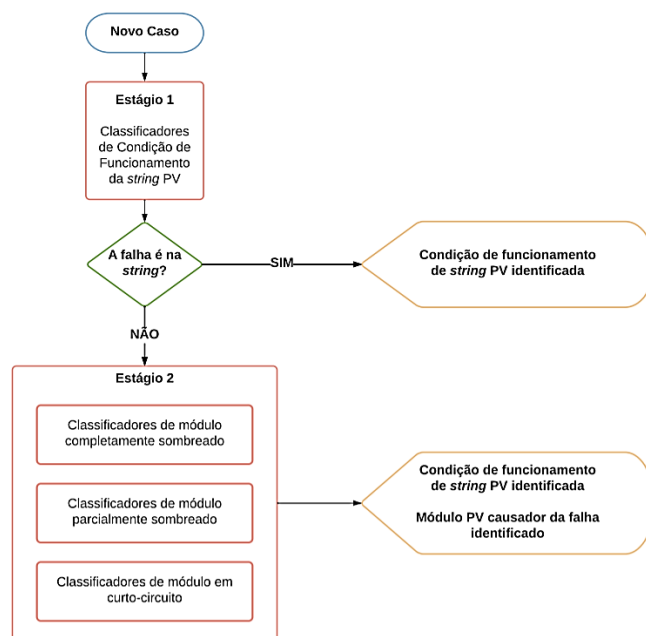


Fig 4. Fluxograma do sistema de classificação multiestágio.

Observa-se no Fluxograma da Fig. 4 que, ao surgir um novo caso para classificação, ele passa pelo primeiro estágio, o qual consiste em identificar a condição de funcionamento da *string* PV. Dependendo da condição de operação de *string* fotovoltaica identificada, o sistema define o problema como resolvido. É o caso, por exemplo, de quando a *string* PV está operando em condição normal, ou quando é identificada uma falha na linha devido ao rompimento de um cabo condutor.

Por outro lado, quando a condição de funcionamento da *string* PV identificada é motivada pela condição de funcionamento de um determinado painel PV que compõe aquela *string*, faz-se necessário um segundo estágio de classificação, a fim de identificar o módulo supostamente defeituoso, resolvendo por completo o problema de identificação.

De acordo com a Fig. 5, o desenvolvimento do primeiro estágio de classificação, que consiste em treinamento e teste dos sistemas, baseia-se em um conjunto de dados reservado para tal propósito. Tais dados são agrupados de acordo com as condições de funcionamento de *string* PV, conforme definido na Tabela I.

O segundo estágio de classificação consiste em, dada uma condição de funcionamento de *string* fotovoltaica provocada por um determinado painel da *string*, identifica-se o painel causador da anormalidade.

Como pode ser visto na Fig. 6, o segundo estágio de classificação subdivide-se em *i* problemas de classificação, onde cada qual conta com o seu próprio subconjunto de

amostras de desenvolvimento. Os detalhes de desenvolvimento são tratados nas seções posteriores.

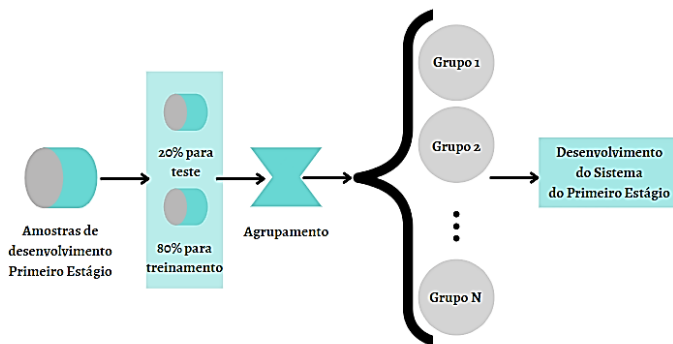


Fig 5. Diagrama de desenvolvimento do primeiro estágio de classificação.

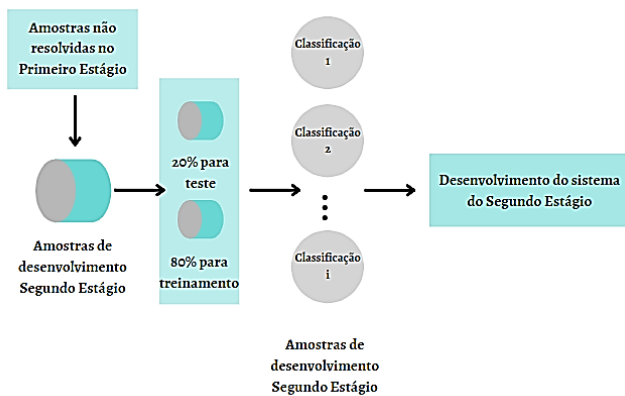


Fig 6. Diagrama de desenvolvimento do segundo estágio de classificação.

#### 2.4 Definição de casos

O conjunto de dados encontram-se na chamada “forma bruta”, ou seja, representam amostras de tensão e de corrente elétricas geradas pela *string* fotovoltaica quando operando nas diversas condições descritas na Subseção 2.2. Logo, para o desenvolvimento dos sistemas aqui propostos, faz-se necessária uma série de manipulações para divisões, adequações e pré-processamentos necessários. Na sequência, são descritas cada etapa de manipulação realizada sobre o conjunto de dados.

Ambos os estágios de classificação passam por duas etapas de desenvolvimento chamadas de treinamento e teste. Além disso, é aqui considerada uma etapa após o desenvolvimento, a qual é chamada de teste final. As métricas do teste final servem de comparação entre diferentes soluções para um mesmo problema tratado. A Fig.7 ilustra como é feita a subdivisão do conjunto de amostras a ser empregado em cada uma das três etapas, isto é, os dois estágios de classificação e o teste final.

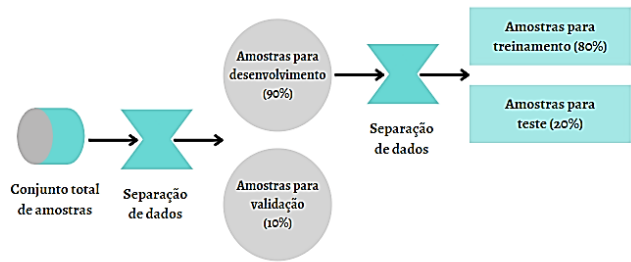


Fig.7. Separação de amostras do conjunto de dados para os dois estágios.

Embora se possa associar que cada ciclo de coleta corresponde a um caso observado, é possível que tal ciclo seja representado por dois ou mais casos. Isso implica em uma redução no número de amostras por caso. Deste modo, a fim de viabilizar um estudo para analisar o efeito em se utilizar uma menor quantidade de amostras para representar um caso, é definido que:

$$n\_amostras\_por\_caso = \frac{N\_AMOSTRAS\_POR\_CICLO}{AJUSTE\_JANELA}, \quad (1)$$

onde  $n\_amostras\_por\_caso$  é um valor inteiro que representa a quantidade de amostras consecutivas que constituem um caso,  $N\_AMOSTRAS\_POR\_CICLO$  é igual a 10.000 e  $AJUSTE\_JANELA$  é uma constante que pode ser definida como 1, 2, 4, 5, 8, 10 ou 20. Ressalta-se que  $AJUSTE\_JANELA$  igual a 1 consiste em um número de amostras por caso igual ao número de amostras por ciclo.

#### 2.5 Extração de Características

A extração de características é uma técnica aplicada sobre o conjunto de dados, a fim de reduzir o espaço dimensional representado por ele. Trata-se de uma etapa de pré-processamento fundamental para o bom funcionamento de certos sistemas de aprendizagem. No entanto, é necessário extrair um conjunto de características que de fato permita a separação de sujeitos distintos.

Como o conjunto de dados usados neste trabalho representa grandezas elétricas, isto é, tensão e corrente elétricas, são consideradas aqui características elétricas determinadas a partir de tais grandezas. Essas características dependem da quantidade de amostras por caso, conforme descrito na Subseção 2.4.

São propostas como características as seguintes grandezas elétricas, as quais são normalmente usadas para caracterizar sistemas elétricos: o valor médio quadrático (rms) da tensão e da corrente, potências ativa, aparente e reativa (Equações 4 a 6) e fator de potência, como apresentado nas equações (2) a (7), respectivamente. Observe que  $i(t)$  e  $v(t)$  correspondem à  $t$ -ésima amostra de corrente e tensão elétrica, respectivamente, de um caso composto por  $T$  amostras.

$$V_{rms} = \frac{1}{T} \sum_{t=0}^{T-1} v^2[t] \quad (2)$$

$$I_{rms} = \frac{1}{T} \sum_{t=0}^{T-1} i^2[t] \quad (3)$$

$$P = \frac{1}{T} \sum_{t=0}^{T-1} v[t] \cdot i[t] \quad (4)$$

$$S = V_{rms} \cdot I_{rms} \quad (5)$$

$$Q = \sqrt{S^2 - P^2} \quad (6)$$

$$FP = \frac{P}{S} \quad (7)$$

Tendo em vista que as características possuem unidades e limites de variação distintos, faz-se necessário normalizá-las, a fim de melhorar o desempenho dos sistemas a serem desenvolvidos e diminuir o tempo de treinamento de tais sistemas (Faceli, *et al.* 2011). Além disso, a não normalização pode causar a saturação de parâmetros do sistema sob treinamento. Essa saturação ocorre devido à existência de valores de entrada muito grandes, o que causa a não convergência do processo de treinamento (Faceli, *et al.* 2011).

Existem diversas ferramentas de normalização, sendo as principais chamadas de transformações lineares. É o caso, por exemplo, da função MinMax (disponível na biblioteca Sklearn (2021) do Python), a qual é utilizada neste trabalho), que utiliza métodos estatísticos, como chi quadrado, correlação e ANOVA (para medir a relação existente entre as variáveis preditoras e a variável target). Com os resultados obtidos, é possível filtrar apenas os atributos que conduzirão ao desempenho mínimo requerido.

A ferramenta SelectKBest da biblioteca Sklearn é aplicada sobre o conjunto formado pelas seis características elétricas aqui empregadas, a fim de analisar o grau de relevância da característica através de suas aplicações estatísticas na solução de cada parte do problema, isto é, no primeiro e no segundo estágios de classificação.

## 2.6 Ferramentas de Classificação

Os classificadores desenvolvidos neste trabalho, tanto para o primeiro quanto para o segundo estágio de classificação, são baseados em três técnicas distintas de classificação. São elas, o kNN (K-Nearest Neighbors), o SVM (Support Vector Machine) e o MLP (Multi-Layer Perceptron). A linguagem de implementação dos classificadores foi a linguagem Python. Trata-se de ferramentas comumente adotadas na literatura para resolver problemas variados a partir de busca de padrões em um conjunto de dados. Além disso, são ferramentas bem consolidadas e amplamente disponibilizadas em bibliotecas dedicadas ao desenvolvimento de sistemas baseados em IA, como é o caso da biblioteca Sklearn, a qual é amplamente usada neste trabalho.

## 2.6 GridSearchCV e Ensemble

GridSearchCV e Ensemble são ferramentas também disponíveis na biblioteca Sklearn e utilizadas com sucesso neste trabalho. Enquanto a primeira permite automatizar o processo de ajuste dos hiperparâmetros de uma dada ferramenta de classificação, a Ensemble permite combinar um conjunto diversificado de modelos (ou soluções baseadas

em aprendizado de máquina) treinados para resolver um mesmo problema, a fim de melhorar o desempenho de um sistema, associando os algoritmos de forma a minimizar suas desvantagens individuais no modelo final. Isso pode ser feito a partir de uma combinação entre as soluções apresentadas por cada modelo, ou por um esquema de eleição da melhor solução encontrada. Modelos baseados em Ensemble podem ser incorporados tanto em cenários com grande volume de dados ou até mesmo quando os dados são insuficientes (Rokach, 2010).

Assim, para encontrar a acurácia global, é realizado o seguinte desenvolvimento:  $Ac1$  é definida como uma métrica que representa a acurácia ao tomar a melhor solução dentre as apresentadas pelo conjunto de classificadores de primeiro estágio, enquanto  $Acfs$ ,  $Acps$  e  $Acsc$  são definidas como métricas que representam as acurácias ao tomar as melhores soluções dentre as apresentadas pelos conjuntos de classificadores de segundo estágio dedicados aos problemas de painel PV em sombreamento completo, sombreamento parcial e curto-circuito, respectivamente.

Para efeito de análise de desempenho, propõe-se a obtenção da acurácia média de classificação no segundo estágio, dada pela Equação (8):

$$Ac2 = \frac{Acfs + Acps + Acsc}{3} \quad (8)$$

Finalmente, determina-se a acurácia do sistema multistágio a partir da Equação (9).

$$Ac_{global} = \frac{Ac1 + Ac2}{2} \quad (9)$$

## 3. RESULTADOS E DISCUSSÃO

A Tabela 2 mostra as métricas de desempenho do primeiro estágio referentes a cada classificador após o processo de desenvolvimento, que consistiu em otimização de arquiteturas, treinamento e teste. Observa-se que todos os classificadores exibiram uma mesma acurácia de 98,45%, o que mostra que o sistema baseado em kNN, o mais simples entre os três, já é suficiente para lidar com o problema do primeiro estágio.

**Tabela 2. Desempenho de cada Classificador no Primeiro Estágio**

Classificador	Acurácia [%]	Tempo treinamento (ms)	Tempo identificação (us)
kNN	98,45	1,43	57,07
SVM	98,45	13,86	46,25
MLP	98,45	573,97	10,61

Tal simplicidade do sistema kNN reflete no menor tempo de treinamento (que no caso desta técnica em específico, consiste no tempo necessário para organizar os casos nos

conjuntos ou classes do problema). Por outro lado, o processo de identificação de um novo caso (tempo de identificação) é o maior dentre os três classificadores. Porém, dura apenas cerca 50  $\mu$ s, o que é um tempo pequeno para o problema em questão.

Na sequência, são mostrados os resultados obtidos com os sistemas desenvolvidos para o primeiro estágio quando aplicado o chamado ajuste de janela, conforme descrito na Subseção 2.4. Os resultados referentes à acurácia de cada classificador podem ser vistos na Fig.8.

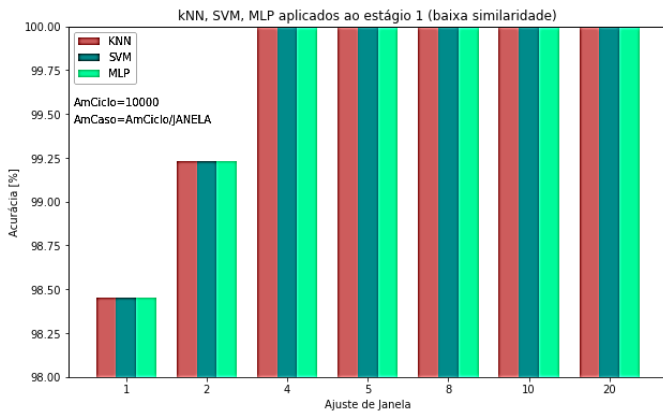


Fig 8. Desempenho de sistemas do primeiro estágio de classificação em função do ajuste de janela.

Percebe-se que, para o primeiro estágio, os três tipos de classificadores exibem comportamentos idênticos para qualquer ajuste de janela. Logo, de acordo com os resultados da Fig 8, um ajuste de janela para o mínimo de 4 (o que corresponde a 776 casos totais por classe, contendo 2500 amostras por caso) é suficiente para proporcionar uma acurácia de 100%.

A Tabela 3 traz as métricas encontradas no Estágio 2 para o melhor desempenho de cada classificador após o processo de desenvolvimento de arquiteturas otimizadas, isto é, treinamento e teste. Observa-se que, em termos de acurácia, os classificadores exibiram valores superiores a 90% de precisão para as condições atípicas denominadas sombreamento total e curto-circuito, e próximo a 90% de precisão na condição também atípica chamada de sombreamento parcial. Isso mostra que até mesmo o sistema baseado em kNN, o mais simples entre os três, já é suficiente para lidar com o problema também do segundo estágio. Por fim, são apresentados também os tempos de treinamento e de identificação de um novo caso para cada uma das soluções desenvolvidas.

Nos gráficos das Fig 9 a 11, observa-se, que no segundo estágio, ocorre uma degradação considerável do desempenho dos três tipos de classificadores desenvolvidos para as três condições atípicas de painéis PV quando aplicado o ajuste de janela. Trata-se de um efeito contrário ao observado no primeiro estágio, onde as classes eram de baixa similaridade quando comparadas com as classes sob identificação neste segundo estágio. Logo, fica claro que o número de amostras por caso afeta o desempenho de sistemas dedicados a classificar classes de maior similaridade.

**Tabela 3. Desempenho de cada Classificador no Segundo Estágio**

Classificador	Condição atípica	Acurácia [%]	Tempo treinamento (ms)	Tempo identificação (us)
kNN	Sombreamento total	94,02%	1,14	16,49
	Sombreamento parcial	88,89%	0,93	49,49
	Curto-circuito	97,44%	0,98	81,30
SVM	Sombreamento total	94,87%	8,02	10305,18
	Sombreamento parcial	88,03%	16,39	10,29
	Curto-circuito	97,44%	12,25	18,69
MLP	Sombreamento total	94,44%	681,62	8,43
	Sombreamento parcial	87,18%	670,74	7,42
	Curto-circuito	95,73%	683,87	3,54

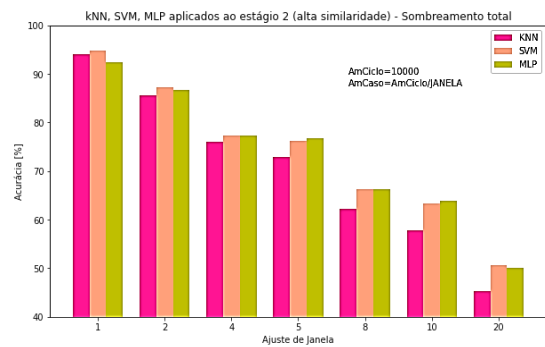


Fig 9. Desempenho de sistemas do segundo estágio de classificação em função do ajuste de janela.

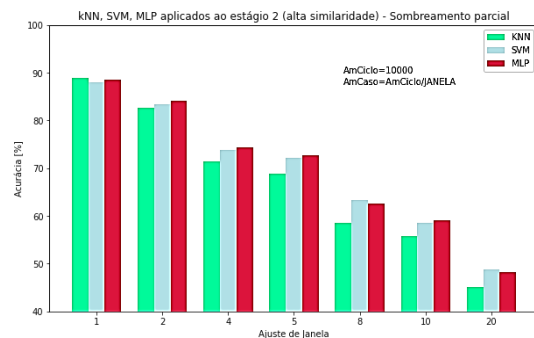


Fig 10. Desempenho de sistemas do segundo estágio de classificação em função do ajuste de janela.

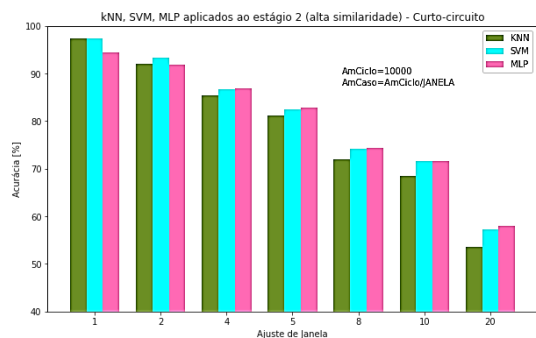


Fig 11. Desempenho de sistemas do segundo estágio de classificação em função do ajuste de janela.

Por fim, realizou-se o chamado teste final com as amostras reservadas para tal propósito, ou seja, os novos casos. Todos os sistemas desenvolvidos foram usados para identificar tais casos e a ferramenta Ensemble, conforme mencionado na Seção 2, possibilitou a escolha automática da melhor solução encontrada, resultando em uma Acurácia Média Global igual a 93,9%.

#### 4. CONCLUSÃO

Mostrou-se, neste trabalho, que é possível realizar a identificação de condição de funcionamento de *string* PV combinando o resultado apresentado por classificadores diversos, os quais foram desenvolvidos para atuar em dois estágios de classificação. A acurácia média global de 93,9%, obtida a partir de novos casos, pode ser considerada adequada para o problema aqui tratado. No entanto, há espaço para a realização de mais pesquisas nesta linha, considerando, por exemplo, outros tipos de condições atípicas de *strings* PV e *strings* compostas por um número maior de painéis PVs.

#### REFERÊNCIAS

Aziz, F.; Haq, A. U.; Ahmad, S.; Mahmoud, Y.; Jalal, M. & Ali, U. (2020). A Novel Convolutional Neural Network-Based Approach for Fault Classification in Photovoltaic Arrays. *IEEE POWER & ENERGY SOCIETY SECTION*, v. 8, pp 41889-41904.

Belaout, A.; Krim, F.; Mellit, A.; Talbi, B & Arabi, A. (2018). Multiclass adaptive neuro-fuzzy classifier and feature selection techniques for photovoltaic array fault detection and classification. *Renewable Energy*, v. 127, pp. 548-558.

Chen, L.; Li, S. & Wang, X. (2018). Quickest Fault Detection in Photovoltaic Systems. *IEEE TRANSACTIONS ON SMART GRID*, v. 9, n. 3.

Dhimish, M.; Holmes, V.; Mehrdadi, B. & Dales, M. (2017). Multi-layer photovoltaic fault detection algorithm. *High Volt*, v. 2 n. 4, pp. 244-252.

Faceli, K., Lorena, A. C., Gama, J. & Carvalho, A. C. P. L. F. (2011). *Inteligência artificial: Uma abordagem de aprendizado de máquina*. Rio de Janeiro: LTC, 2:192.

Garoudja, E.; Harrou, F.; Sun, Y.; Kara, K. Chouder, A. & Silvestre, S. (2017). Statistical fault detection in photovoltaic systems. *Solar Energy*, v. 150, pp. 485-499.

Hussain, M.; Dhimish, M.; Titarenko, S.; Mather, P. (2020). Artificial neural network based photovoltaic fault detection algorithm integrating two bi-directional input parameters. *Renewable Energy*, v. 155, pp. 1272-1292.

IEA. (2021). *Renewables 2021: Analysis and forecast to 2026*.

Khoshtami, A. & Sadeghkhani, I. (2018). Two-stage power-based fault detection scheme for photovoltaic systems. *Solar Energy*, v. 176, pp. 10-21.

Li, Y.; Ding, K.; Zhang, J. Chen, F.; Chen, X. & Wu, J. (2019). A fault diagnosis method for photovoltaic arrays based on fault parameters identification. *Renewable Energy*, v. 143, pp. 52-63.

Micheletti, I.T.S., *et al.* (2021). Energia solar, extra fiscalidade e políticas públicas na promoção ao desenvolvimento sustentável?. *Brazilian Journal of Development*, v.7, n.6, p. 58659-58680.

Monteiro. M.M.M.M.L. (2018). Desenvolvimento de um método de detecção e análise de falhas em centrais fotovoltaicas por termografia. Universidade de Lisboa: Faculdade de Ciências. 2018.

Navid, Q.; Hassan, A.; Fardoun, A. A. & Ramzan, R. (2020). An Online Novel Two-Layered Photovoltaic Fault Monitoring Technique Based Upon the Thermal Signatures. *Sustainability*, v. 12, pp. 9607.

Phinikarides, A., Kindyni, N., Makrides, G. & Georghiou, G. (2014). Review of photovoltaic degradation rate methodologies. *Renewable and Sustainable Energy Reviews*, v. 40, pp. 143-152.

Qureshi, F. A.; Uddin, Z.; Satti, M. B & Ali, M. (2020). CA-based solar photovoltaic fault diagnosis. *Int Trans Electr Energ Syst*, v. 30, pp. 12456.

Rokach, L. (2010). Pattern Classification Using Ensemble Methods. *World Scientific*.

SCIKIT-LEARN.ORG. (2021). sklearn. neural-network. MLP Classifier - scikitlearn 0.19.2 documentation. 2021. Disponível em: <<http://scikitlearn.org/stable/modules/generated/sklearn.neuralnetwork.MLPClassifier.html>>. Acesso em: 19 de abril de 2022.

Trindade, A.B. (2020). Automated Verification of Stand-alone Solar Photovoltaic Systems Optimal Sizing and Project Validation Automated Verification of Standalone. *Solar Photovoltaic Systems. Tese (Doutorado em Informática) – UFAM. Amazonas*, p.130.

Wang, W.; Liu, A. C.; Chung, H. S.; Lau, R. W.; Zhang, J. & Lo, A. W. (2016). Fault Diagnosis of Photovoltaic Panels Using Dynamic Current-Voltage Characteristics. *IEEE TRANSACTIONS ON POWER ELECTRONICS*, v. 31, n. 2, pp. 1588-1599.