



UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO TECNOLÓGICO
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

Bruno Castro Valle

**Um Método para a Estimativa de Strings Ausentes em Usinas Fotovoltaicas
de Grande Escala**

Florianópolis
2024

Bruno Castro Valle

**Um Método para a Estimativa de Strings Ausentes em Usinas Fotovoltaicas
de Grande Escala**

Dissertação submetida ao Programa de Pós-graduação em Engenharia Elétrica da Universidade Federal de Santa Catarina para a obtenção do título de Mestre em Processamento de Informação

Orientador: Prof. Richard Demo Souza, Dr.

Coorientador: Prof. Arliones Hoeller Jr, Dr.

Florianópolis

2024

Ficha catalográfica gerada por meio de sistema automatizado gerenciado pela BU/UFSC.
Dados inseridos pelo próprio autor.

Castro Valle, Bruno

Um Método para a Estimativa de Strings Ausentes em Usinas Fotovoltaicas de Grande Escala / Bruno Castro Valle ; orientador, Richard Demo Souza, coorientador, Arliones Hoeller Jr, 2024.

42 p.

Dissertação (mestrado) - Universidade Federal de Santa Catarina, Centro Tecnológico, Programa de Pós-Graduação em Engenharia Elétrica, Florianópolis, 2024.

Inclui referências.

1. Engenharia Elétrica. 2. Usina Fotovoltaica. 3. Operação e Manutenção. 4. Desempenho e Falhas. 5. Grande Escala. I. Demo Souza, Richard. II. Hoeller Jr, Arliones. III. Universidade Federal de Santa Catarina. Programa de Pós-Graduação em Engenharia Elétrica. IV. Título.

Bruno Castro Valle

**Um Método para a Estimativa de Strings Ausentes em Usinas Fotovoltaicas
de Grande Escala**

O presente trabalho em nível de mestrado foi avaliado e aprovado, em 24 de setembro de 2024, pela banca examinadora composta pelos seguintes membros:

Prof, Guilherme Luiz Moritz, Dr.
Universidade Tecnológica Federal do Paraná

Prof, Roberto Alexandre Dias, Dr.
Instituto Federal de Santa Catarina

Certificamos que esta é **versão original e final** do trabalho de conclusão que foi julgado adequado para obtenção do título de Mestre em Processamento de Informação.

Coordenador do Programa de
Pós-Graduação

Prof. Richard Demo Souza, Dr.
Orientador

Florianópolis, 24 de setembro de 2024

Este trabalho é dedicado a minha querida família e namorada.

AGRADECIMENTOS

Gostaria de agradecer primeiramente aos meus pais por todos os ensinamentos, apoio, incentivos e principalmente pelo amor que sempre me deram. Sou e serei eternamente grato.

A minha irmã pelo apoio e amor durante toda a vida e apesar das discussões e brigas sempre que necessitei estava lá pra mim.

A minha namorada Camila por ser a melhor companheira que eu poderia ter, por ter me suportado nesse período de estresse e sempre me incentivar a crescer e melhorar, tanto como pessoa, quanto profissionalmente.

A minha família, que mesmo dos mais diversos sempre estiveram lá para me apoiar e fornecer energias positivas para o meu sucesso e futuro.

Ao professor orientador Richard Demo, que apesar das condições adversas me acolheu e me ajudou a crescer e desenvolver tanto quanto pessoa quanto como profissional, além de ter me auxiliado em todas as etapas desde o meu trabalho de conclusão de curso até o final do meu mestrado.

Ao professor Arliones que apesar de ser de outra instituição me auxiliou, apoiou e me fez ir além para conseguir completar essa etapa da melhor maneira possível.

Gostaria de expressar meus sinceros agradecimentos à CAPES e à ENGIE Brasil pelo suporte que tornou possível o desenvolvimento deste projeto de pesquisa.

E finalmente a UFSC por ter me fornecido oportunidades de aprendizado e crescimento incríveis, tanto na graduação quanto no mestrado.

*“Viva como se você fosse morrer amanhã.
Aprenda como se você fosse viver para sempre.”
(Mahatma Gandhi)*

RESUMO

Esta dissertação apresenta uma nova metodologia para detectar strings ausentes em sistemas fotovoltaicos de grande escala (VLSPV) utilizando apenas dados adquiridos no nível das stringboxes. Ao aproveitar a análise de dados e técnicas de aprendizado de máquina não supervisionadas, o método proposto estima o número de strings ausentes por stringbox, comparando a corrente contínua de cada stringbox com stringboxes dentro da mesma região. Esta abordagem preenche uma lacuna na literatura existente, oferecendo uma solução adaptada ao nível típico de instrumentação de plantas VLSPV. O trabalho inclui uma análise das perdas de energia causadas por strings ausentes, quantificando seu impacto no desempenho geral do sistema. A avaliação com dados reais demonstrou que o método proposto atinge uma precisão de aproximadamente 90% na detecção de strings ausentes. Essas descobertas oferecem *insights* valiosos para equipes de operações e manutenção, permitindo a identificação e mitigação de strings problemáticas em plantas VLSPV.

Palavras-chave: 1.Fotovoltaica. 2.Grande Escala. 3.Desempenho. 4.Falhas. 5.Manutenção.

ABSTRACT

This dissertation presents a novel methodology for detecting missing strings in Very Large-Scale Photovoltaic (VLSPV) systems, utilizing only data acquired at the stringbox level. By leveraging data analysis and unsupervised machine learning techniques, the proposed method estimates the number of missing strings per stringbox by comparing the direct current from each stringbox with neighboring stringboxes within the same region. This approach addresses a gap in the existing literature by providing a solution tailored to the typical instrumentation level of VLSPV plants. The work includes an analysis of the energy losses caused by missing strings, quantifying their impact on the overall system performance. Evaluation against real-world data demonstrated that the proposed method achieves a precision of approximately 90% in detecting missing strings. These findings offer valuable insights for operations and maintenance teams, enabling the identification and mitigation of problematic strings in VLSPV plants.

Keywords: 1.Photovoltaic. 2.Large-Scale. 3.Performance. 4.Failures. 5.Maintenance.

LISTA DE FIGURAS

Figura 1	– <i>Layout</i> de uma usina Fotovoltaico em Muito Grande Escala (VLSPV).	20
Figura 2	– Ilustração da irradiância medida no Plano do Array (POA) durante um dia, considerando um <i>tracker</i> regular e um <i>tracker</i> em repouso, no dia 1º de janeiro, para diferentes horas.	23
Figura 3	– Ilustração mostrando o agrupamento geográfico com $K = 4$ realizado usando <i>k-means</i> , com cada ponto representando um Grupo de Inversores (ITS).	24
Figura 4	– Um exemplo de agrupamento de correntes com $L = 4$ agrupamentos, conduzido usando <i>k-means</i> , onde cada ponto representa a corrente média da <i>string</i> em uma determinada <i>stringbox</i> . O agrupamento 1 (marcador quadrado) é o agrupamento saudável.	25
Figura 5	– Diagrama ilustrando a metodologia proposta, delineando a aquisição de dados, pré-processamento, estimativa da <i>string</i> saudável e, finalmente, o cálculo das <i>strings</i> faltantes.	26
Figura 6	– Número médio estimado de strings faltantes por dia, para diferentes números de agrupamentos geográficos K para um número fixo de agrupamentos de corrente $L = 2$, ao longo de um período de 15 dias.	28
Figura 7	– Número estimado de strings faltantes por dia, para diferentes números de agrupamentos de corrente L para um número fixo de $K = 4$ agrupamentos geográficos, ao longo de um período de 15 dias.	30
Figura 8	– Soma diária da corrente contínua medida com perdas de corrente estimadas (%) usando parâmetros $K = 4$ e $L = 2$, erro médio de 2,68%.	33
Figura 9	– Soma diária da corrente contínua medida com perdas de corrente estimadas (%) usando parâmetros $K = 36$ e $L = 2$, erro médio de 2,11%.	33

LISTA DE ABREVIATURAS E SIGLAS

ANEEL	Agência Nacional de Energia Elétrica
ANN	Rede Neural Artificial
BJT	Transistor de Junção Bipolar
CA	Corrente Alternada
CC	Corrente Contínua
CNN	Rede Neural Convolucional
DT	Árvore de Decisão
FN	Falso Negativo
FP	Falso Positivo
ITS	Grupo de Inversores
kNN	K-Vizinhos Mais Próximos
LG	Linha para Terra
LL	Linha para Linha
LSPV	Fotovoltaico em Grande Escala
LSTM	Memória de Longo e Curto Prazo
MAE	Erro Médio Absoluto
ML	Aprendizado de Máquina
MLP	Perceptron Multicamadas
MPPT	Rastreamento do Ponto de Máxima Potência
P	Precisão
POA	Plano do Array
PV	Fotovoltaico
R	Revoação
SCADA	Supervisão, Controle e Aquisição de Dados
SSL	Aprendizado Semi-Supervisionado
SVM	Máquina de Vetores de Suporte
VLSPV	Fotovoltaico em Muito Grande Escala

SUMÁRIO

1	INTRODUÇÃO	12
1.1	OBJETIVOS	14
1.1.1	Objetivos Gerais	14
1.1.2	Objetivos Específicos	14
1.2	OUTRAS INFORMAÇÕES	15
2	REVISÃO DA LITERATURA	16
3	MODELO DE UMA USINA DE GRANDE PORTE	19
4	METODOLOGIA	21
4.1	PRÉ-PROCESSAMENTO	21
4.2	AGRUPAMENTO GEOGRÁFICO	22
4.3	AGRUPAMENTO DE CORRENTES	23
4.4	CÁLCULO DAS <i>STRINGS</i> FALTANTES	24
4.5	ESTIMATIVA DAS PERDAS	25
5	RESULTADOS	27
5.1	USINA	27
5.2	AGRUPAMENTO GEOGRÁFICO	27
5.3	AGRUPAMENTO DE CORRENTE	29
5.4	VALIDAÇÃO COM DADOS EM CAMPO	29
5.5	PERDAS ESTIMADAS	32
5.6	ASPECTOS PRÁTICOS	32
6	CONCLUSÕES	35
6.1	TRABALHOS FUTUROS	35
	REFERÊNCIAS	37

1 INTRODUÇÃO

As fontes renováveis têm experimentado um notável aumento na capacidade instalada global, com um crescimento previsto de aproximadamente 2400 GW (quase 75%) entre 2022 e 2027, de acordo com a previsão principal da IEA (IEA, 2022). Destacando-se neste aumento está a energia fotovoltaica, impulsionada pela redução dos custos dos painéis fotovoltaicos e pela maior previsibilidade da sua geração em comparação com outras fontes renováveis, como a energia eólica. Conforme discutido em (COELHO; SCHMITZ; MARTINS, 2022), o avanço da energia solar fotovoltaica reflete uma tendência global impulsionada tanto pela inovação tecnológica quanto por políticas de incentivo, o que se reflete nos números: em 2010, a participação da energia solar fotovoltaica na capacidade instalada mundial era de apenas 0,8%, escalando significativamente para 12,8% em 2022. Com base nas projeções, prevê-se que, até 2027, a energia fotovoltaica superará outras categorias (IEA, 2022), contribuindo com 22,2% da capacidade instalada global.

No Brasil, o ano de 2022 foi considerado o melhor ano de energia solar segundo o site O Setor Elétrico (O SETOR ELÉTRICO, 2022). O país saltou de 7,9 GW, ao final de 2020, para 13 GW ao final de 2021, crescimento de 65%, mesmo em meio a um ano desafiador de pandemia global. Já para o ano de 2022 a mesma fonte afirma que haverá um crescimento de 105% na geração própria e 67.8% nas grandes usinas solares, reafirmando assim a grande importância dessa energia para o futuro do Brasil.

Embora os sistemas fotovoltaicos geralmente exijam manutenção relativamente reduzida em comparação com outros sistemas (BADR; HAMAD *et al.*, 2021), eles são suscetíveis a falhas ocasionais e perdas de energia, especialmente em sistemas Fotovoltaico em Grande Escala (LSPV). Tipicamente, esses sistemas LSPV são caracterizados por uma capacidade instalada superior a 1 MW (ELMALLAH *et al.*, 2023), enquanto os sistemas VLSPV possuem uma capacidade instalada superior a 100 MW (ITO *et al.*, 2003). Esta distinção é crucial ao considerar os desafios operacionais únicos, os custos e as necessidades de manutenção associadas a essas escalas de instalações fotovoltaicas. Essas usinas, compostas por configurações série-paralelo de módulos fotovoltaicos (também conhecidos como arranjos fotovoltaicos), também apresentam maior tensão e corrente. Portanto, falhas não detectadas podem causar grandes perdas de energia e monetárias (BADR; HAMAD *et al.*, 2021), sendo, portanto, uma causa relevante de preocupação.

Conforme relatado em (MAPS, 2023), o subdesempenho devido a anomalias quase dobrou, passando de 1,61% em 2019 para 3,13% em 2022. Anomalias em sistemas fotovoltaicos referem-se a qualquer desvio que afeta o funcionamento ideal dos componentes, podendo incluir falhas elétricas, problemas de instalação, degradação de módulos, defeitos em inversores, entre outros. Em sistemas VLSPV, o aumento foi ainda mais pronunciado, subindo de 1,10% em 2019 para 4,04% em 2022. Extrapolando esses dados para a capacidade fotovoltaica instalada globalmente, conforme relatado por (MAPS, 2023) (até o

final de 2021, excluindo instalações residenciais), resulta em uma perda de receita anual de 2,5 bilhões de dólares para a indústria fotovoltaica, com uma perda média de 4,320 dólares por MW para sistemas VLSPV (MAPS, 2023). Se a perda de energia causada por anomalias continuar a crescer na taxa observada, espera-se que atinja quase 6% até 2025. A análise conduzida em (MAPS, 2023) examinou as principais perdas de receita categorizadas por anomalias. Em resumo, anomalias em strings, inversores, combinadores de strings e rastreadores emergiram como os quatro principais fatores de perda de receita devido a problemas de hardware em ativos fotovoltaicos inspecionados em 2022. O estudo em (MAPS, 2023) encontrou que anomalias em strings representaram 34% da receita perdida, anomalias em inversores representaram 22% da perda de receita, anomalias em combinadores resultaram em uma perda de 21%, enquanto anomalias em rastreadores, apesar de serem as mais comuns, impactaram 26,86% de todos os módulos com anomalias, contribuindo com apenas 7,94% para a perda total de receita.

A queda no desempenho devido a anomalias levou a uma expansão notável na busca por métodos de detecção e classificação de falhas em sistemas fotovoltaicos, uma área que, até recentemente, não recebia tanta atenção em instalações de grande escala, conforme indicado por (ZHAO *et al.*, 2019). Nesse sentido, um dos desafios identificados (ANDERSON *et al.*, 2022; BADR; HAMAD *et al.*, 2021) em usinas de grande escala está relacionado à escassez de instrumentação detalhada para cada módulo fotovoltaico, geralmente adquirida por meio de um sistema de Supervisão, Controle e Aquisição de Dados (SCADA). Nesse nível, o sistema SCADA frequentemente carece de granularidade para monitorar módulos ou strings individuais (ANDERSON *et al.*, 2022), pois a instrumentação necessária para a aquisição de dados na usina muitas vezes implica em despesas consideráveis. Portanto, não é geralmente viável obter dados no nível de strings, muito menos em cada painel fotovoltaico, em sistemas VLSPV. Nessa escala, os dados são tipicamente coletados no nível de uma *stringbox* (também conhecida como caixa de combinação) (ZHAO *et al.*, 2019), que representa um grupo de strings interconectadas, cada uma composta por múltiplos módulos fotovoltaicos dispostos em série.

Considerando os aspectos mencionados, este trabalho foca na detecção do número de strings ausentes (desconectadas¹) por *stringbox* em usinas VLSPV. Para alcançar esse objetivo, foram utilizadas técnicas de análise de dados e aprendizado de máquina não supervisionado para estimar a quantidade de strings ausentes por *stringbox*. Este trabalho foca no nível de instrumentação típico encontrado em usinas VLSPV. Portanto, a metodologia se baseia exclusivamente nos dados de corrente contínua coletados de cada *stringbox*, eliminando assim a necessidade de instrumentação por string, módulo ou dados mais granulares. Esta abordagem preenche uma lacuna na literatura existente, pois metodologias similares adequadas para usinas VLSPV ainda não foram identificadas.

Assim, aproveitando os dados de corrente da saída das *stringboxes*, o método

¹ Uma string desconectada geralmente está relacionada a um fusível queimado na *stringbox*.

proposto os compara com as *stringboxes* vizinhas para estabelecer uma média estimada de corrente saudável por string, que seria uma corrente de uma string sem anomalias operando em condições normais. A partir dessa comparação, determina-se o número de strings operacionais e, conseqüentemente, a quantidade de strings ausentes, auxiliando a equipe de manutenção da usina a identificar strings problemáticas. A avaliação do método não supervisionado proposto contra dados reais mostrou boa precisão, sendo capaz de detectar cerca de 90% das strings ausentes, um grande benefício para as equipes de operações e manutenção de usinas VLSPV. Adicionalmente, como subproduto do método, é possível estimar as perdas de energia geradas por essas strings ausentes. Finalmente, como os dados no nível das *stringboxes* já estão disponíveis no sistema SCADA em usinas VLSPV, os benefícios do método proposto vêm sem investimento adicional, além do esforço de análise de dados.

Esta dissertação está organizada na seguinte forma: O Capítulo 2 apresenta uma revisão da literatura. O Capítulo 4 descreve a metodologia proposta para estimar o número de strings ausentes. O Capítulo 5 apresenta os resultados obtidos. Por fim, o Capítulo 6 discute as conclusões e trabalhos futuros.

1.1 OBJETIVOS

1.1.1 Objetivos Gerais

O objetivo do trabalho é desenvolver e testar através de dados reais, técnicas e estratégias a serem aplicadas no cálculo de strings ausentes em uma *stringbox*, de modo que seja possível identificar a *stringbox* com problema e quantificar a perda energética resultante.

Foi necessário verificar a funcionalidade da solução para diferentes tipos de configurações, diferentes condições meteorológicas como um dia ensolarado, nublado ou chuvoso, diferentes usinas solares, entre outros. Além disso, foi feito um *dashboard* para a integração dos resultados de forma interativa com o site da empresa Engie, de forma que ficou fácil localizar e identificar as falhas ocorrendo de maneira diária.

1.1.2 Objetivos Específicos

- Pesquisa sobre o estado da arte de artigos relacionados a detecção e classificação de falhas em sistemas fotovoltaicos, principalmente em sistemas de grande porte;
- Desenvolvimento de um modelo de análise de dados utilizando técnicas de aprendizado de máquina não supervisionado para a detecção de strings ausentes;
- Realização de testes e validações do modelo com dados reais provenientes de uma usina solar de grande porte, sob diversas condições meteorológicas;

- Análise quantitativa das perdas energéticas decorrentes de strings ausentes, avaliando o impacto no desempenho geral do sistema;

1.2 OUTRAS INFORMAÇÕES

Este estudo foi realizado como parte de um projeto no âmbito do Programa de Pesquisa e Desenvolvimento da empresa ENGIE Brasil Energia S.A., regulado pela Agência Nacional de Energia Elétrica (ANEEL) (PD-00403-0056/2023). O projeto em questão tem título “Ferramenta Computacional para Detecção, Identificação e Quantificação de Perdas Energéticas em Usinas Fotovoltaicas”, e visa avançar no campo da detecção e classificação de eventos de falha em usinas fotovoltaicas utilizando técnicas de aprendizado de máquina.

No decorrer do projeto, foi conduzida uma extensa pesquisa sobre o estado da arte em detecção de falhas e técnicas de aprendizado de máquina aplicáveis ao contexto específico de plantas fotovoltaicas. Um dos frutos desse esforço é o artigo publicado na IEEE Journal of Photovoltaics, intitulado “A Method for the Estimation of Missing Strings in Very-Large-Scale Photovoltaic Power Plants” (SIMÃO *et al.*, 2024). Este trabalho, além de contribuir para o avanço acadêmico, reforça a aplicabilidade prática das técnicas desenvolvidas no âmbito do projeto, oferecendo uma solução robusta para a estimativa de strings ausentes em usinas fotovoltaicas de grande porte. Um segundo artigo, intitulado “Método de Estimação de Strings Ausentes em Usinas Fotovoltaicas de Grande Escala usando Aprendizado de Máquina não Supervisionado”, foi submetido ao XVIII EDAO - Reunião de Debate sobre Questões Operacionais. As atividades também incluíram a análise detalhada dos dados históricos disponíveis na empresa, seguida pela definição de estratégias para preparação, segmentação e rotulagem desses dados, essenciais para o treinamento, validação e/ou teste dos modelos de aprendizado de máquina.

As opiniões, resultados e conclusões expressas neste trabalho são de responsabilidade exclusiva dos autores e não refletem necessariamente as visões das instituições financiadoras.

2 REVISÃO DA LITERATURA

Embora não tenha sido identificado nenhum artigo específico abordando a questão de strings ausentes em *stringboxes*, vários artigos relacionados à detecção e classificação de falhas em diversos sistemas fotovoltaicos foram encontrados. Esses estudos apresentam abordagens semelhantes, proporcionando técnicas que podem ser aplicados neste projeto. Alguns desses artigos inclusive realizam quantificações das perdas energéticas, contudo, nenhum deles aborda o cálculo específico do número de strings ausentes. Essa lacuna na literatura destaca a relevância e originalidade do presente projeto de mestrado.

O trabalho descrito em (HASHEMI *et al.*, 2021) utilizou modelagem matemática do sistema fotovoltaico para prever a potência de saída ideal do sistema e classificar nove tipos de falhas com base na perda de potência. Em contraste, o trabalho em (LAURINO; PILIOUGINE; SPAGNUOLO, 2022) realizou simulações de curvas I-V defeituosas a partir de parâmetros elétricos do módulo solar, classificando com Aprendizado de Máquina (ML) quatro tipos de falhas e validando os resultados com dados reais registrados de uma usina na Itália. Em (PILIOUGINE; SPAGNUOLO, 2022) também foi empregada a simulação de curvas I-V, classificando as falhas com 15 modelos de Rede Neural Artificial (ANN) e realizando validação através de um experimento próprio para o teste. Já em (ESKANDARI; MILIMONFARED; AGHAEI, 2021) os autores se concentraram na simulação de curvas I-V, usando Rastreamento do Ponto de Máxima Potência (MPPT), identificando com ML apenas falhas de curto-circuito, separando-as em Linha para Linha (LL) e Linha para Terra (LG).

O trabalho em (VOUTSINAS *et al.*, 2022) empregou curvas I-V e ANN para classificação das falhas e regressão da potência de saída. Em contraste, o artigo (GUO *et al.*, 2022) utilizou um modelo de Memória de Longo e Curto Prazo (LSTM) com dados temporais de uma usina na Bélgica para classificar diferentes tipos de sombreamento no sistema, diferenciando entre sombreamento permanente e temporário. Por sua vez, (MEMON *et al.*) optou por um modelo de Rede Neural Convolucional (CNN), classificando cinco tipos de falhas com base em dados de corrente e tensão de saída, bem como medidas meteorológicas. O artigo (COSTA *et al.*, 2019) concentrou-se na comparação da acurácia de vários modelos de ML, como K-Vizinhos Mais Próximos (kNN), Máquina de Vetores de Suporte (SVM), Árvore de Decisão (DT) e ANN para a classificação de seis tipos de falhas, onde o modelo de ANN obteve o melhor resultado com 99,65% de acurácia. Os autores utilizaram os dados de corrente e tensão de saída, potência gerada e previsões meteorológicas¹. Além disso, os autores de (ZHOU *et al.*, 2020) aplicaram uma SVM para quantificar a perda de potência de saída causada por diferentes falhas de sombreamento. Já o trabalho em (UTAMA *et al.*, 2023) baseou-se em um *dataset* previamente etiquetado

¹ Note que este é o único trabalho publicado em um evento que foi selecionado nesta revisão bibliográfica. A razão para tal é que o trabalho descreve resultados de um projeto executado no Brasil, com financiamento da ANEEL, o que estabelece uma ligação interessante com o projeto em questão.

no artigo (LAZZARETTI *et al.*, 2020) na cidade de Curitiba, para classificar a partir de uma Perceptron Multicamadas (MLP) cinco tipos de falhas.

Em (LAZZARETTI *et al.*, 2020) foi desenvolvido um sistema embarcado de baixo custo para a coleta de dados de cada painel Fotovoltaico (PV) e a classificação com ML de cinco falhas. O artigo (MADETI; SINGH, 2018) destacou-se por sua classificação com kNN e separação de agrupamentos de diferentes falhas, além de uma pesquisa aprofundada nas falhas de sombreamento. Em contraste, o trabalho em (CHO *et al.*, 2020) focou exclusivamente na detecção de falhas, utilizando uma SVM para identificar falhas na string ou no inversor. O artigo (BADR; HAMAD *et al.*, 2021) priorizou a detecção instantânea de falhas, podendo alguns instantes após a coleta dos dados identificar a falha e classificando-as com ML em um de seis tipos. A abordagem de curvas I-V e CNN foi aplicada novamente, mas pelo trabalho descrito em (AZIZ *et al.*, 2020), classificando entre quatro diferentes tipos de falha.

O trabalho de (DYAMOND; RIX) apresentou uma pesquisa sobre a detecção de falhas e a visualização de baixo desempenho em sistemas PV, utilizando modelagem matemática de todos os elementos do sistema fotovoltaico. O trabalho em (HOJABRI *et al.*, 2022) concentrou-se na implementação de um sistema embarcado de baixo custo para o monitoramento de parâmetros elétricos e meteorológicos no local, combinando simulações de falhas com validação em um sistema real. Os autores de (IBRAHIM *et al.*, 2022) concentraram-se na detecção de falhas no inversor, utilizando dados reais de sistemas PV na Índia e testando três modelos de ML para detecção.

O trabalho descrito em (HARROU *et al.*, 2021) realizou modelagem do sistema real, prevendo a potência de saída e implementando um algoritmo de ML para classificar entre três falhas. O autor de (LEVA; MUSSETTA; OGLIARI, 2019) utilizou um *forecasting* de potência gerada para apenas detectar possíveis falhas e não classificá-las. Em contraparte, o artigo (ØGAARD *et al.*, 2021) concentrou-se na detecção via formulação matemática de pequenas perdas de potência a nível de string, sendo capaz de detectar falhas de até 1%. Os autores de (YURTSEVEN; KARATEPE; DENIZ, 2021) utilizaram uma estratégia baseada em diferenças entre medidas próximas de potência para detectar a ocorrência de falhas.

Além disso, alguns artigos exploraram métodos alternativos, como a utilização de satélites para medição de dados meteorológicos em vez de dados registrados no local do sistema feita por (GOMPEL; SPINA; DEVELDER, 2022). Outros trabalhos desenvolveram sistemas embarcados para o monitoramento de curvas I-V de cada painel para classificação de falhas, como (MELLIT *et al.*, 2023). A regressão logística foi o foco do algoritmo de classificação desenvolvido em (WANG; XU; YAN, 2023), que classificou com a mesma quatro diferentes falhas, enquanto o artigo (HE *et al.*, 2023) inovou ao classificar diferentes falhas simultaneamente, identificando se o sistema estava com duas, três ou quatro falhas ao mesmo tempo. O artigo (BADR; ABDEL-KHALIK *et al.*, 2023) aplicou

um complexo algoritmo de Aprendizado Semi-Supervisionado (SSL) para classificar cinco tipos de falhas, provendo também um estudo aprofundado sobre diferentes algoritmos de SSL. O trabalho publicado em (SAKTHIVEL; ARUNACHALAM; JAGATHEESAN, 2023) focou na implementação de um sensor de baixo custo para a classificação de falhas a partir de técnicas de processamento de sinal, projetado para obter dados de tensão em cada célula do sistema PV.

O artigo (KUMAR; MISHRA; DASH, 2023) utilizou uma placa de sistema embarcado com um Transistor de Junção Bipolar (BJT) e um diodo Zener para classificar a partir da análise do espectro de tensão e corrente e localizar as falhas. O algoritmo AdaBoost foi otimizado para classificar três falhas de forma rápida e com alta acurácia em (LODHI *et al.*, 2022). O artigo (JIAO *et al.*, 2023) é um trabalho focado em grandes usinas, que com curvas I-V simuladas consegue classificar via ML em quatro tipos de falhas. O trabalho descrito em (MADICHETTY *et al.*, 2022) não apenas detectou analisando as perdas de potência, mas também quantificou a potência perdida, utilizando as correntes que passam pelos diodos de bypass para detectar módulos defeituosos.

No contexto de usinas fotovoltaicas de grande porte VLSPV, o artigo (ZHAO *et al.*, 2019) ofereceu duas metodologias de classificação de falhas, um método hierárquico de detecção de anomalias com reconhecimento de contexto usando aprendizado não supervisionado e um método de classificação de anomalias multimodal, ambos dependendo do sistema SCADA para a aquisição de informações e abordando seis tipos de falhas diferentes. Em (MUSTAFA *et al.*, 2023) os autores focaram em organizar uma base de dados desbalanceada, utilizando técnicas de *under* e *over sampling* para equilibrar o número de falhas e melhorar a acurácia do modelo de LSTM. O artigo (BAGHERZADEH *et al.*, 2022) concentrou-se na detecção usando técnicas de processamento e análise de sinais e também na localização de falhas a partir de dados de corrente e tensão de saída de cada módulo. Por fim, o trabalho em (ANDERSON *et al.*, 2022) se dedicou exclusivamente às falhas nos *trackers*, realizando testes com diferentes algoritmos de detecção para determinar onde o *tracker* está parado e calcular a perda energética gerada por cada falha no sistema. Esses estudos reforçam a importância do uso de técnicas de aprendizado de máquina e processamento de sinais para melhorar a detecção de anomalias em usinas de grande escala, evidenciando a necessidade de abordagens eficientes e adaptadas às características específicas de instalações VLSPV.

3 MODELO DE UMA USINA DE GRANDE PORTE

Antes da apresentação da metodologia, é importante compreender o modelo genérico das usinas fotovoltaicas contemporâneas em grande escala. Esta seção discute o modelo genérico dessas usinas, fornecendo uma visão geral da operação da planta e apresentando uma perspectiva macro do sistema explorado neste trabalho. Esta abordagem visa oferecer uma compreensão geral do tamanho e da complexidade envolvida em usinas fotovoltaicas de grande porte.

Nas usinas fotovoltaicas, a estrutura é hierarquicamente definida para facilitar a compreensão, com uma progressão dos menores aos maiores componentes. Começando pelas células fotovoltaicas individuais, estas são agrupadas em módulos ou painéis. Esses módulos são então agrupados para formar as strings, cujas conexões são feitas normalmente em série ou em paralelo (BOXWELL, 2010). É importante notar que os rastreadores (*trackers*), que são comumente utilizados em tais instalações, podem acomodar um certo número de painéis e conter uma ou mais strings dentro do mesmo rastreador (ANDERSON *et al.*, 2022). Em instalações de grande escala, o número de strings pode alcançar dezenas de milhares, dependendo da configuração e do tamanho específico da usina.

Partindo das strings, o sistema incorpora as *stringboxes* (ZHAO *et al.*, 2019), assim chamadas por sua função de agregar conjuntos de strings e proteger o segmento de Corrente Contínua (CC) do sistema fotovoltaico. As *stringboxes* são elementos intermediários fundamentais entre as strings e os inversores, oferecendo proteção contra sobretensão e sobrecorrente, além de facilitar a desconexão dos circuitos quando necessário. Os inversores, considerados o “cérebro” do sistema fotovoltaico, vêm a seguir na hierarquia. Esses componentes críticos são responsáveis por converter a CC em Corrente Alternada (CA), enquanto fornecem isolamento e proteção tanto para o sistema quanto para a rede elétrica (ZHAO *et al.*, 2019).

Além disso, em certos sistemas de grande ou muito grande escala, a definição de um ITS pode ser utilizada para melhorar a organização e a gestão. Um ITS representa um grupo de inversores, contribuindo para uma operação e manutenção mais eficientes.

Uma ilustração do *layout* de uma usina fotovoltaica de grande escala, conforme descrito acima, é fornecida na Figura 1. Nessa figura, denota-se o número de *stringboxes* no sistema por N , enquanto o número de strings por *stringbox* é M_n , onde $n \in \{1, \dots, N\}$. Observa-se também, a corrente de uma *string*, denotada por $I_{n,m}$, e a corrente de saída de uma *stringbox*, denotada por I_n , destacadas em vermelho. Considerando essa estrutura, o número total de painéis fotovoltaicos pode chegar às centenas de milhares. Devido ao custo significativo associado à obtenção de dados de toda a planta, é razoável assumir que, na maioria dos casos práticos (ZHAO *et al.*, 2019; ANDERSON *et al.*, 2022), as informações estão disponíveis apenas a partir do nível das *stringboxes* e além.

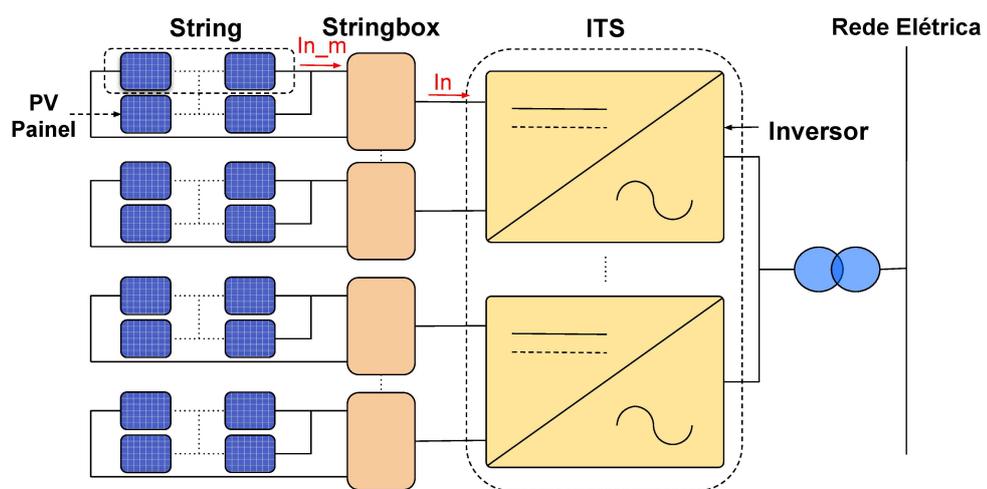


Figura 1 – *Layout* de uma usina VLSPV.

4 METODOLOGIA

Esta seção descreve o método proposto para calcular *strings* faltantes em *stringboxes* dentro de usinas VLSPV. A metodologia baseia-se em análise de dados e aprendizado de máquina, e está estruturada em quatro fases. A abordagem proposta incorpora parâmetros ajustáveis para acomodar a variabilidade entre diferentes plantas e a disponibilidade de dados fornecidos pelo sistema SCADA. Dessa forma, garante uma estratégia flexível e eficiente que pode ser adaptada às características específicas de cada cenário, possibilitando sua aplicabilidade não apenas a uma planta específica, mas também a uma ampla gama de instalações fotovoltaicas.

Especificamente, o método proposto emprega técnicas de análise de dados e aprendizado de máquina para estimar o número de *strings* faltantes em cada *stringbox*. Dado o contexto de uma planta VLSPV, o procedimento começa segmentando a planta em micro-regiões, os chamados agrupamentos geográficos, a fim de isolar efeitos microclimáticos, envelhecimento, sujidade ou outros tipos de variações locais. Subsequentemente, estabelece-se uma estimativa da corrente média de uma *string* saudável dentro de cada agrupamento geográfico, servindo como referência para comparação com a corrente média observada por *string* em cada *stringbox*. Em seguida, é possível estimar a quantidade de *strings* operacionais e, conseqüentemente, a de *strings* faltantes em cada *stringbox*. O método proposto utiliza exclusivamente dados obtidos da saída dos *stringboxes* através do sistema SCADA, sendo executado diariamente.

A seguir, cada subseção aborda uma fase específica do método proposto.

4.1 PRÉ-PROCESSAMENTO

Inicialmente, o foco é no processamento dos valores de corrente associados à saída de cada *stringbox* na planta. Além dos dados de corrente, cada registro inclui carimbos de tempo e a *stringbox* à qual os dados pertencem. Na etapa inicial, realiza-se a limpeza de valores ausentes e inválidos nos dados, permitindo um início mais direcionado do pré-processamento específico deste projeto. É importante notar que um dos principais desafios na geração de energia está relacionado ao mau funcionamento dos *trackers* (ANDERSON *et al.*, 2022). Os *trackers* são equipamentos mecânicos responsáveis por ajustar a orientação dos painéis solares, acompanhando o movimento do sol ao longo do dia para maximizar a captação de luz solar e, conseqüentemente, a eficiência da geração de energia. Portanto, é necessário evitar trabalhar com dados medidos afetados por falhas nos *trackers* que são bastante comuns.

Isso pode ser feito ao desconsiderar as *stringboxes* que possuem *strings* sobre *trackers* que apresentam falha. No entanto, isso levaria a dois problemas: i) Primeiro detectar falhas nos *trackers* para selecionar apenas as *stringboxes* cujas *strings* estão conectadas a *trackers* operacionais; ii) *Strings* ausentes sobre *trackers* com mau funcionamento não

seriam detectadas. Para contornar tais problemas, primeiro observamos que, por padrão, *trackers* com mau funcionamento são automaticamente configurados para a posição de repouso (horizontal). Em seguida, é determinada uma janela de tempo durante o dia correspondente a períodos em que os *trackers* operacionais devem estar aproximadamente alinhados horizontalmente com o sol. Durante essa janela de tempo, tanto os *trackers* operacionais quanto os em repouso estariam aproximadamente na mesma posição. Essa estratégia nos permite isolar eficientemente o problema de *trackers* com mau funcionamento daquele de *strings* ausentes.

As janelas de tempo específicas são determinadas através de uma simulação de rastreamento utilizando a biblioteca PVlib (HOLMGREN *et al.*, 2018) em Python, levando em consideração a localização da usina, o tipo de instalação de *trackers* e a época do ano. Com essas informações, obtemos a posição dos *trackers* solares usando o modelo de céu claro, que estima a irradiância solar em condições ideais de céu sem nuvens a partir de dados como localização, data e hora, e aplicamos uma função de rastreamento de eixo, ambos da PVlib. Isso nos fornece uma estimativa da posição de rastreamento, que então filtramos para obter uma janela de tempo compreendida entre -5 e 5 graus em relação ao horizonte. Tipicamente, essa janela de tempo ocorre ao redor do meio-dia. A Figura 2 ilustra a irradiância medida no POA, em W/m^2 , durante um dia, considerando um *tracker* em condições normais de operação, seguindo a posição do sol no céu, e um *tracker* em repouso, na posição horizontal. Ao redor do meio-dia, ambas as irradiâncias são muito próximas, pois o *tracker* regular e o *tracker* em repouso estão em torno de zero graus. Portanto, nesse momento do dia, é razoável adquirir dados de *stringbox* que possuem *trackers* em repouso, pois a diferença na corrente medida é mínima (se houver).

Subsequentemente, em um determinado dia, calcula-se a mediana dos valores de corrente CC de cada um dos N *stringbox* durante essa janela de tempo¹, resultando em um único valor diário para cada *stringbox*, denotado aqui como I_n , onde $n \in 1, \dots, N$. Finalmente, assumindo que todas as *strings*, em média, contribuem igualmente para a corrente total de saída do *stringbox*, a corrente estimada na m -ésima *string* conectada ao n -ésimo *stringbox*, onde $m \in 1, \dots, M_n$, é obtida como

$$I_{n,m} = \frac{I_n}{M_n}. \quad (1)$$

4.2 AGRUPAMENTO GEOGRÁFICO

O objetivo desta fase é estabelecer microrregiões dentro da usina fotovoltaica que exibam perfis de irradiação solar ou desempenho esperado semelhantes, permitindo resultados mais precisos, uma vez que as *strings* nessas regiões devem ter desempenho similar. Esta abordagem é relevante devido à vasta área física ocupada por usinas VLSPV,

¹ A mediana é uma alternativa melhor do que a média em termos de evitar o impacto de outliers.

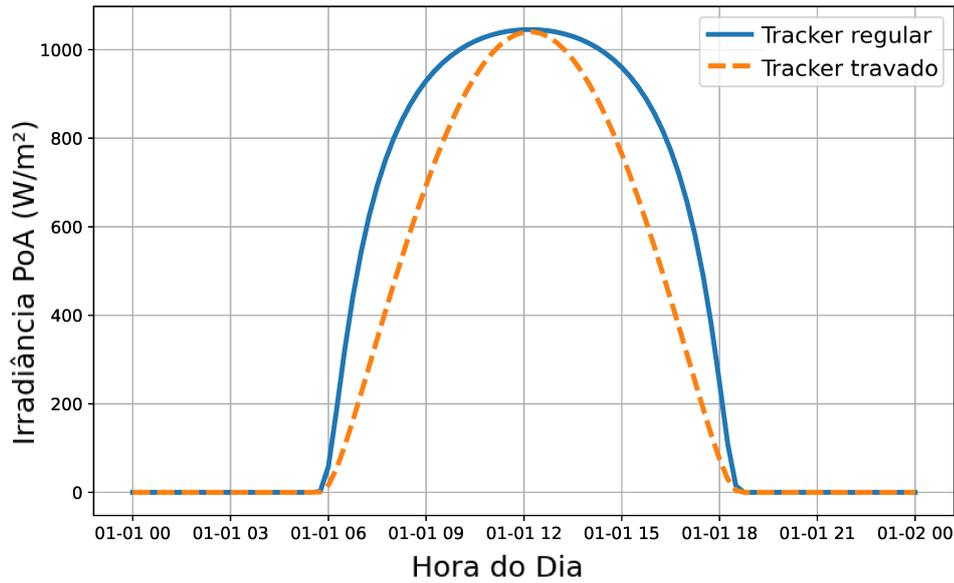


Figura 2 – Ilustração da irradiância medida no POA durante um dia, considerando um *tracker* regular e um *tracker* em repouso, no dia 1º de janeiro, para diferentes horas.

que cobrem áreas substanciais, podendo chegar a milhões de metros quadrados. Consequentemente, essas usinas estão sujeitas a variações nas condições de irradiação solar em diferentes seções, conforme observado em (ZHAO *et al.*, 2019), que descreve uma instalação com mais de 400.000 painéis fotovoltaicos distribuídos por uma vasta área terrestre. O agrupamento geográfico visa mitigar essas variações, proporcionando uma análise mais precisa.

Para alcançar este objetivo, o método requer coordenadas de latitude e longitude para cada *stringbox*², permitindo o agrupamento automatizado. Utilizando o bem conhecido algoritmo *k-means* (MACQUEEN, 1967), empregamos os dados geográficos para fins de agrupamento. Especificando o número desejado de agrupamentos geográficos, denotado aqui por K , esta abordagem assegura um método direto e adaptável. A partir de agora, cada agrupamento é tratado de forma independente nas fases subsequentes. Podemos observar um exemplo de agrupamento geográfico na Figura 3, onde uma área é separada em $K = 4$ agrupamentos. Em nosso contexto, os pontos representam a posição de cada ITS.

4.3 AGRUPAMENTO DE CORRENTES

O principal objetivo desta fase é estabelecer um valor estimado para a corrente de uma *string* saudável. Além disso, nesta e nas próximas fases cada agrupamento geográfico

² Na ausência de dados de localização de cada *stringbox*, pode-se adotar uma abordagem manual. Isso envolve inspecionar visualmente o mapa da usina e delinear os agrupamentos de acordo, definindo assim o número de regiões.

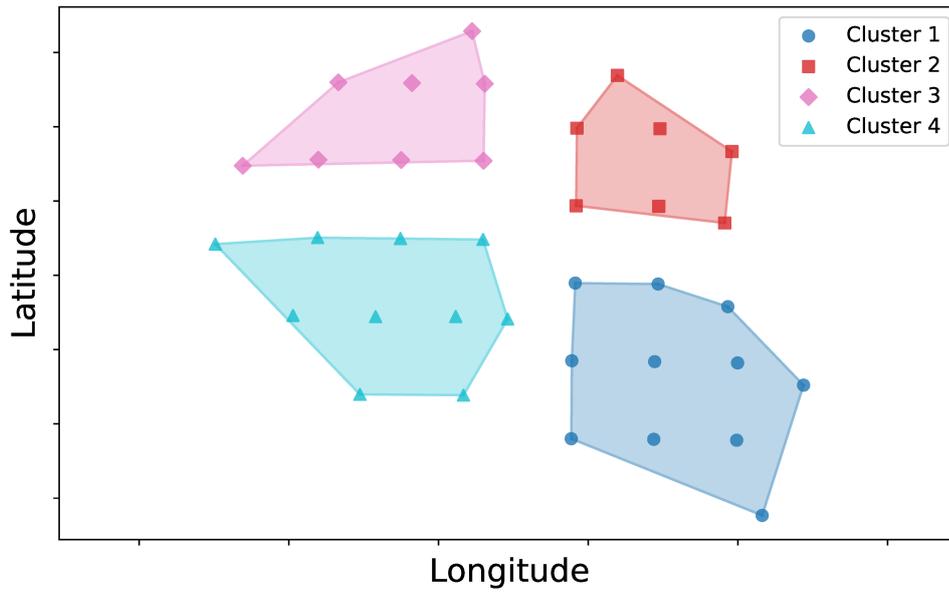


Figura 3 – Ilustração mostrando o agrupamento geográfico com $K = 4$ realizado usando k -means, com cada ponto representando um ITS.

é tratado separadamente. Primeiramente, é preciso denotar a corrente estimada de uma *string* saudável no k -ésimo agrupamento geográfico, $k \in 1, \dots, K$, como H_k . Para obter tal estimativa, procedemos com outra operação de agrupamento, mas desta vez considerando o conjunto de correntes de *string* obtidas em (1) que pertencem apenas ao k -ésimo agrupamento geográfico. Para isso, empregamos novamente o algoritmo k -means dentro deste conjunto de correntes, separando-os em L agrupamentos. Em seguida, o agrupamento com a maior corrente mediana entre os L agrupamentos é selecionado como o agrupamento saudável para esse agrupamento geográfico, levando a H_k . Podemos observar um exemplo de agrupamento de correntes na Figura 4, com $L = 4$ agrupamentos de corrente, e onde cada ponto representa a corrente média da *string* em uma determinada *stringbox*. O eixo x é ordenado pelo número do ITS, onde cada ITS contém várias *stringboxes*. Os pontos em vermelho pertencem ao agrupamento das *stringboxes* com a maior corrente média por *string*. Portanto, a corrente mediana do agrupamento em azul escuro se tornaria a *string* saudável estimada para este agrupamento geográfico.

4.4 CÁLCULO DAS STRINGS FALTANTES

Com a corrente estimada de uma *string* saudável em cada agrupamento geográfico, a estimativa do número de *strings* faltantes pode prosseguir. Primeiro, para cada *stringbox* n pertencente ao k -ésimo agrupamento geográfico, o número de *strings* conectadas dentro daquele *stringbox* pode ser determinado como

$$M^{\text{con}}_n = \frac{In}{H_k}. \quad (2)$$

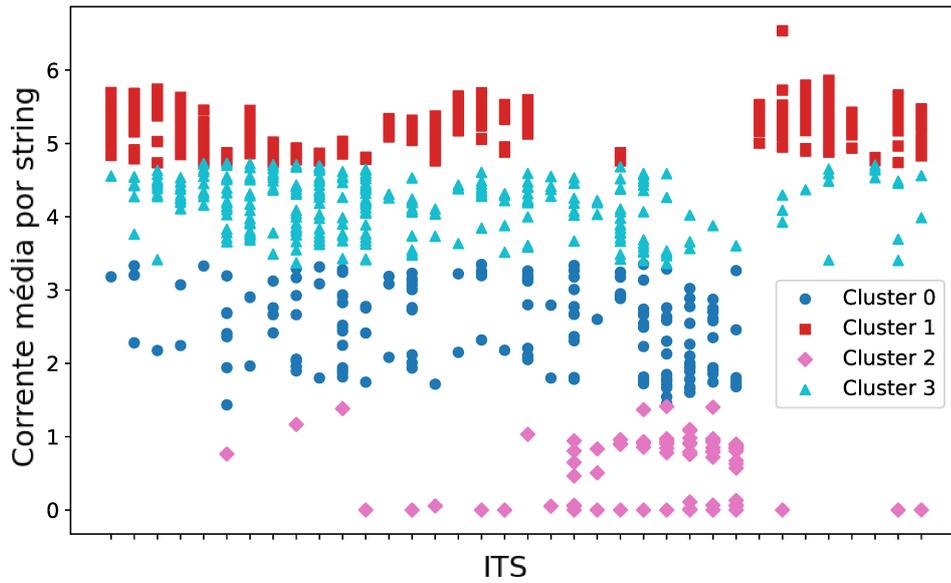


Figura 4 – Um exemplo de agrupamento de correntes com $L = 4$ agrupamentos, conduzido usando k -means, onde cada ponto representa a corrente média da *string* em uma determinada *stringbox*. O agrupamento 1 (marcador quadrado) é o agrupamento saudável.

Em seguida, o número de *strings* faltantes em cada *stringbox* pode ser finalmente estimado como

$$M_n^{\text{mis}} = M_n - M_n^{\text{con}}. \quad (3)$$

O resultado de (3) é um número real. No entanto, na prática, este número só pode ser um inteiro. A aproximação inteira de M_n^{mis} pode ser obtida pela aplicação das operações de arredondamento, piso ou teto.

Finalmente, um diagrama que resume a metodologia proposta é ilustrado na Figura 5. O método possui alguns parâmetros ajustáveis, o tamanho do agrupamento geográfico K e o tamanho do agrupamento de corrente L , que podem ser explorados para melhor acomodar variações de desempenho dentro da usina e em termos de sensibilidade do método. Tais possibilidades são melhor discutidas na Seção 5.

4.5 ESTIMATIVA DAS PERDAS

Com o número de *strings* faltantes, torna-se possível avaliar as perdas de corrente correspondentes em uma determinada *stringbox* como

$$I_n^{\text{loss}} = M_n^{\text{mis}} \times H_k. \quad (4)$$

Com esta equação, podemos determinar a corrente ideal como $I_n^{\text{loss}} + I_n$, possibilitando o cálculo da perda de corrente por *stringbox*. Uma abordagem semelhante pode ser aplicada para a usina inteira, somando as perdas de todas as *stringboxes* e comparando-as com a soma da corrente ideal em todas as *stringboxes*.

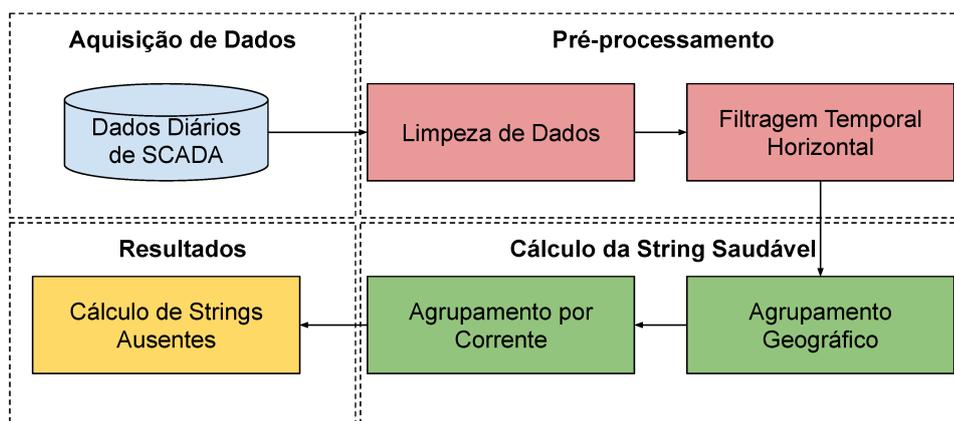


Figura 5 – Diagrama ilustrando a metodologia proposta, delineando a aquisição de dados, pré-processamento, estimativa da *string* saudável e, finalmente, o cálculo das *strings* faltantes.

5 RESULTADOS

No presente Capítulo são apresentados os resultados da aplicação do método proposto a dados reais de uma usina VLSPV, incluindo o efeito dos parâmetros ajustáveis.

Os resultados obtidos são discutidos em termos de eficácia na estimativa de strings faltantes, análise de perdas de corrente e sua aplicabilidade em diferentes condições operacionais da usina. Além disso, são exploradas as variações de desempenho observadas em função dos parâmetros ajustáveis do método, como o tamanho dos agrupamentos geográficos e de corrente.

5.1 USINA

Os resultados utilizam dados de uma usina de energia fotovoltaica no Brasil, que possui uma capacidade de aproximadamente 138.6 MW, pertencendo, portanto, à classe VLSPV. A usina é composta por quatro grandes seções, cada uma equipada com nove ITS, que, por sua vez, são compostos de até quatro inversores cada. No total, a usina possui 36 ITS e 132 inversores. Esta configuração inclui dois tipos de *stringboxes*, em termos de capacidade de strings: o Tipo I possui 18 strings instaladas por *stringbox*, enquanto o Tipo II possui 15 strings. Todos os inversores contêm uma *stringbox* Tipo I e sete *stringboxes* Tipo II. Portanto, a usina possui 132 *stringboxes* Tipo I e 924 *stringbox* Tipo II, totalizando 16.236 strings. Cada string é composta por um conjunto de 30 painéis fotovoltaicos, resultando em um total de 487.080 painéis.

A usina de grande escala descrita destaca um número vasto de painéis fotovoltaicos. Isso reforça a noção de que adquirir dados da saída de cada painel é muitas vezes impraticável, tanto em termos da quantidade de sensores necessários quanto dos custos associados de monitoramento e manutenção.

O método proposto possui dois parâmetros ajustáveis, o número de agrupamentos geográficos K e o número de agrupamentos de corrente L . Esses parâmetros podem ser utilizados para, sob a disponibilidade de dados de referência, ajustar o método às condições de uma usina específica. A seguir, ilustramos o comportamento do método proposto sob diferentes configurações.

5.2 AGRUPAMENTO GEOGRÁFICO

O número de micro-regiões no agrupamento geográfico (K) influencia a quantidade de *stringboxes* dentro de cada agrupamento de corrente, que são então utilizados para estimar valores de corrente saudável por string (corrente de uma string sem falhas). A corrente saudável refere-se ao valor de referência da corrente elétrica gerada por uma *string* operando em condições ideais, sem falhas ou anomalias. Diminuir o número de agrupamentos geográficos resulta em um conjunto de dados maior usado para calcular a

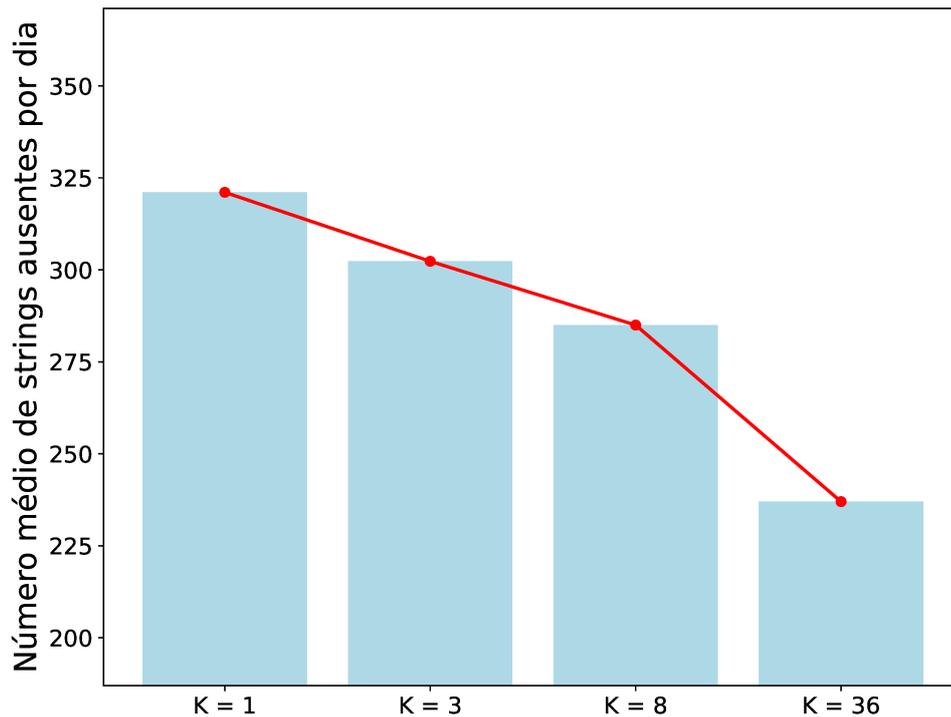


Figura 6 – Número médio estimado de strings faltantes por dia, para diferentes números de agrupamentos geográficos K para um número fixo de agrupamentos de corrente $L = 2$, ao longo de um período de 15 dias.

corrente saudável dentro de cada agrupamento. Isso pode resultar em um maior número de *stringboxes* de alto desempenho dentro da região, aumentando consequentemente a corrente “saudável” estimada. Como resultado, as estimativas de strings operacionais podem diminuir, enquanto as estimativas de strings faltantes podem aumentar. Por outro lado, um maior número de agrupamentos geográficos permite uma análise mais granular, acomodando microclimas ou efeitos de baixo desempenho e potencialmente resultando em um desempenho mais uniforme entre as *stringboxes* dentro de cada agrupamento geográfico. Consequentemente, isso resultaria em uma estimativa menor do que constitui uma corrente saudável e, portanto, em um número estimado reduzido de strings faltantes. Observa-se que um número excessivamente pequeno de agrupamentos geográficos pode gerar resultados não confiáveis devido à insuficiência de dados para o agrupamento de corrente preciso e a estimativa da corrente saudável. Portanto, a escolha de um K adequado, o número de agrupamentos geográficos, é muito relevante na abordagem proposta.

A Figura 6 ilustra o impacto do número de agrupamentos geográficos K , na soma total de strings faltantes, mantendo o número de agrupamentos de corrente constante em $L = 2$. A figura mostra a média do número de strings faltantes por dia, em um período de 15 dias.

Por exemplo, se ao considerar um único agrupamento geográfico ($K = 1$), todas as *stringboxes* são agrupadas em uma única região, aumentando a corrente saudável de

referência e, conseqüentemente, aumentando a estimativa de strings faltantes, possivelmente gerando mais detecções falsas. À medida que o número de agrupamentos geográficos aumenta, a corrente saudável estimada naturalmente diminui, pois as *stringboxes* de alto desempenho podem agora ser divididas em diferentes agrupamentos geográficos, resultando em uma estimativa menor de strings faltantes.

Utilizar um número muito grande de agrupamentos geográficos pode subestimar a corrente saudável de um agrupamento, reduzindo a estimativa de strings faltantes e, potencialmente, deixando strings com falhas não detectadas. Na avaliação do método apresentada aqui, o valor máximo de K é limitado a 36 porque a configuração atual da usina utilizada como base possui informações de localização de 36 ITSs. Assim, a escolha de K deve ser feita considerando a melhor divisão espacial para manter a homogeneidade dentro dos agrupamentos, sem perder a capacidade de detectar anomalias de desempenho.

5.3 AGRUPAMENTO DE CORRENTE

Ao contrário dos agrupamentos geográficos, um maior número de agrupamentos de corrente eleva o critério de filtragem, levando a uma corrente saudável estimada mais alta. Isso ocorre porque apenas o agrupamento com a maior mediana de corrente é usado para estimar a corrente saudável por string. Portanto, à medida que o número de agrupamentos de corrente aumenta, a mediana do agrupamento “saudável” tende a subir, resultando em um menor número estimado de strings operacionais e, conseqüentemente, uma maior estimativa de strings faltantes.

A Figura 7 ilustra o impacto de L na estimativa de strings faltantes. Consideramos variar o número de agrupamentos de corrente entre $L = 1$ e $L = 5$, enquanto mantemos o número de agrupamentos geográficos fixo em $K = 4$. Novamente, os resultados representam a média do número estimado de strings faltantes por dia ao longo de um período de 15 dias. Os resultados reforçam a intuição de como o método se comporta para diferentes números de agrupamentos de corrente, de modo que o número estimado de strings faltantes aumenta com L . Definir o par L, K de melhor desempenho só é possível com base em dados de referência.

5.4 VALIDAÇÃO COM DADOS EM CAMPO

O método proposto foi validado e analisado usando dados de campo rotulados coletados ao longo de um dia, com o status anotado de um subconjunto de strings da planta. Vale ressaltar que a obtenção de tais dados de campo de um sistema VLSPV é uma tarefa muito cara e demorada, justificando o teste com apenas um dia. Nesse dia, a equipe de campo relatou várias falhas notáveis, incluindo um inversor com defeito que impactou oito *stringboxes*, um fusível aberto no barramento de um inversor e vários pequenos danos

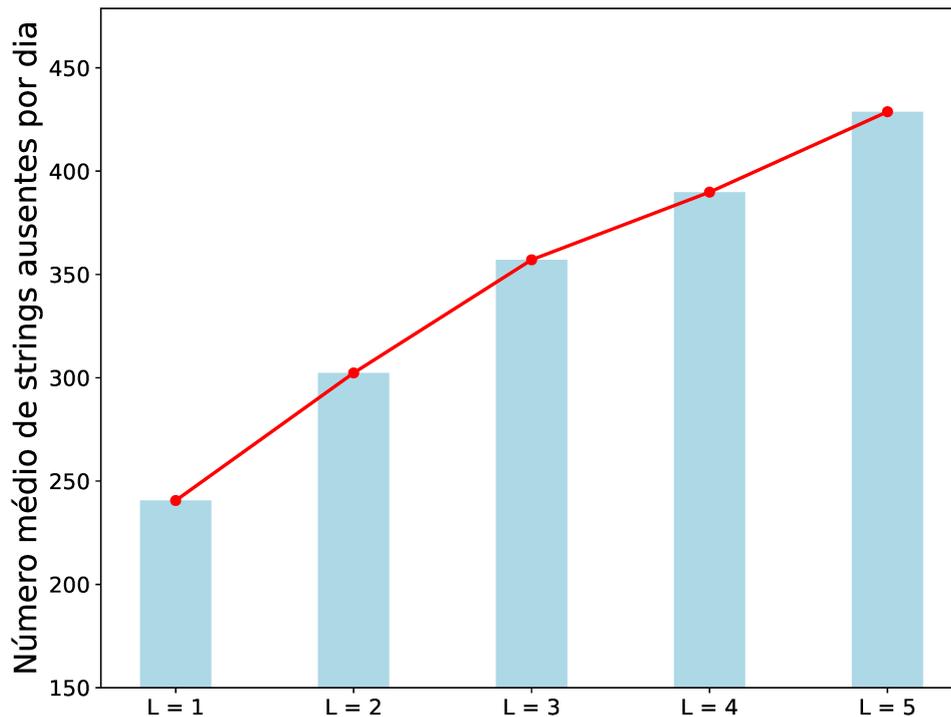


Figura 7 – Número estimado de strings faltantes por dia, para diferentes números de agrupamentos de corrente L para um número fixo de $K = 4$ agrupamentos geográficos, ao longo de um período de 15 dias.

nas strings, como fusíveis queimados e strings abertas. Essas últimas questões afetaram um total de 265 strings.

Antes de continuar com a validação, definimos as métricas usadas. Como o objetivo deste trabalho é detectar strings faltantes, as previsões corretas são strings verdadeiramente faltantes (Verdadeiros Positivos) ou strings verdadeiramente operacionais (Verdadeiros Negativos). Nesse contexto, as métricas de desempenho são:

- Erro Médio Absoluto (MAE): o erro absoluto médio do número de strings faltantes;
- Falso Positivo (FP): o número de strings operacionais marcadas como faltantes;
- Falso Negativo (FN): o número de strings faltantes marcadas como operacionais;
- Precisão (P): a proporção de strings marcadas como faltantes que realmente estão faltando, isto é, $P = TP / (TP + FP)$;
- Revoação (R): a proporção de strings realmente faltantes que foram corretamente marcadas, isto é, $R = TP / (TP + FN)$;
- F-Score(F_1): a média harmônica da precisão e da revocação, isto é, $F_1 = (2 \times P \times R) / (P + R)$, representando, simetricamente, tanto a precisão quanto a revocação em uma única métrica.

Tabela 2 – Validação com Dados de Campo para Diferentes Números de Agrupamentos Geográficos (K) e de Corrente (L).

K	L	MAE	FP	FN	R %	P %	F_1 %
1	2	0.487	23	16	92.0	94.3	93.2
1	3	0.438	31	4	89.5	98.5	93.8
1	4	0.700	54	2	83.1	99.3	90.4
1	5	0.700	54	2	83.1	99.3	90.4
4	2	0.438	22	13	92.3	95.3	93.8
4	3	0.438	31	4	89.5	98.5	93.8
4	4	0.625	48	2	84.7	99.3	91.4
4	5	0.650	50	2	84.1	99.3	91.1
36	2	0.450	25	11	91.4	96.0	93.6
36	3	0.525	33	9	88.9	96.7	92.7
36	4	0.713	48	9	84.7	96.7	90.3
36	5	0.688	51	4	83.9	98.5	90.6

A Tabela 2 mostra os resultados da aplicação da abordagem proposta para estimar strings faltantes, considerando dados rotulados reais. Durante os testes, variamos o número de agrupamentos geográficos e de corrente, K e L , respectivamente. Note que M_n^{mis} é aproximado para o inteiro mais próximo. Primeiramente, observa-se que o desempenho do método, sendo um algoritmo não supervisionado, é bastante robusto, já que todas as configurações analisadas alcançaram um F-Score acima de 90%. Além disso, a variação dos parâmetros ajustáveis oferece a possibilidade de ajustar o modelo com base na métrica mais relevante em cada cenário. O menor MAE foi alcançado, por exemplo, ao separar a planta em $K = 4$ agrupamentos geográficos, com $L = 3$ agrupamentos de corrente, também alcançando o maior F-Score. Um desempenho notável também foi obtido com $K = 4$ e $L = 2$, alcançando o mesmo MAE, mas com tanto R quanto P acima de 90%, um ótimo equilíbrio entre FP e FN.

É importante também analisar as configurações que resultaram na menor precisão. As configurações que levaram ao maior MAE são aquelas com maior L , independentemente do valor de K . Mais agrupamentos de corrente levam a piores estimativas em termos de MAE, pois a corrente saudável estimada aumenta, tornando o método muito mais seletivo. Nesse caso, as strings realmente faltantes são mais facilmente detectadas, reduzindo os falsos negativos, à medida que a corrente da *stringbox* à qual estão associadas se desvia mais da corrente do agrupamento saudável. No entanto, várias strings operacionais, sujeitas a pequenas variações de desempenho devido a outros fatores como sombra, sujeidade, envelhecimento, ou outros, também são mais facilmente detectadas como faltantes, aumentando os falsos positivos.

A escolha dos valores de K e L foi realizada de forma empírica. Foram testadas todas as combinações possíveis de K variando entre 1 e 36, e de L variando entre 1 e 5. No entanto, os resultados apresentados concentram-se nas configurações onde a usina está

inteira, outra na qual a mesma foi dividida em 4 (quatro parques) e a última configuração com K igual a 36, que é o valor máximo de ITS disponível, considerando que a localização está disponível apenas para esses componentes. Já para L , valores acima de 5 resultaram em um número elevado de falsos positivos, tornando a escolha de $L \leq 5$ mais adequada para garantir um equilíbrio satisfatório entre precisão e robustez do modelo.

5.5 PERDAS ESTIMADAS

Tendo um método para estimar se strings individuais estão desconectadas ou com mau funcionamento, e, portanto, impactando a produção geral de energia da planta, investigamos agora as perdas decorrentes das “strings ausentes” no sistema estudado. Ao examinar e quantificar essas perdas, é possível fornecer introspecções valiosas sobre os efeitos das strings ausentes no desempenho e eficiência do sistema fotovoltaico.

A Figura 8 mostra que, com apenas $K = 4$ agrupamentos geográficos, o algoritmo tem um melhor desempenho em dias com alta corrente gerada I_n (por exemplo, dia 13-07), tipicamente dias ensolarados com poucas nuvens, resultando em uma estimativa razoável de perda de energia. No entanto, em dias com baixa I_n (por exemplo, dia 03-07), estimamos perdas incomumente grandes, provavelmente devido à variabilidade da irradiância dentro do mesmo agrupamento geográfico. Para resolver essa questão, em dias de baixa irradiância é aconselhável considerar mais agrupamentos geográficos, utilizando, por exemplo, o máximo de $K = 36$, equivalente ao número total de ITS. Na Figura 9 temos esse resultado, onde alcançamos uma perda estimada de 2,11%, consistente com o desempenho médio reportado por (MAPS, 2023). É também importante notar que alguns dias estão marcados como “SETPOINT”, que são dias em que a planta foi forçada a reduzir sua geração devido a exigências regulatórias, e, portanto, não são incluídos nestas estimativas.

5.6 ASPECTOS PRÁTICOS

Esta seção discute alguns aspectos práticos da implementação e utilização do método proposto. Inicialmente, é necessário alimentar o método com a configuração de cada planta, fornecendo a localização geográfica de todas as *stringboxes* e o número de strings conectados a cada uma. Os resultados na Tabela 2 fornecem introspecções sobre como diferentes valores de K e L impactam a saída do método em termos de falsos positivos, falsos negativos e F-Score. Observe que o método deve ser utilizado para informar as equipes de manutenção sobre strings potencialmente ausentes, sendo natural receber a realimentação dessas equipes e eventualmente alterar parâmetros para melhor atender às necessidades da aplicação.

Ainda assim, boas decisões na escolha de K e L dependem de uma boa compreensão de seus papéis. O número de agrupamentos geográficos (K) pode abordar o tamanho da

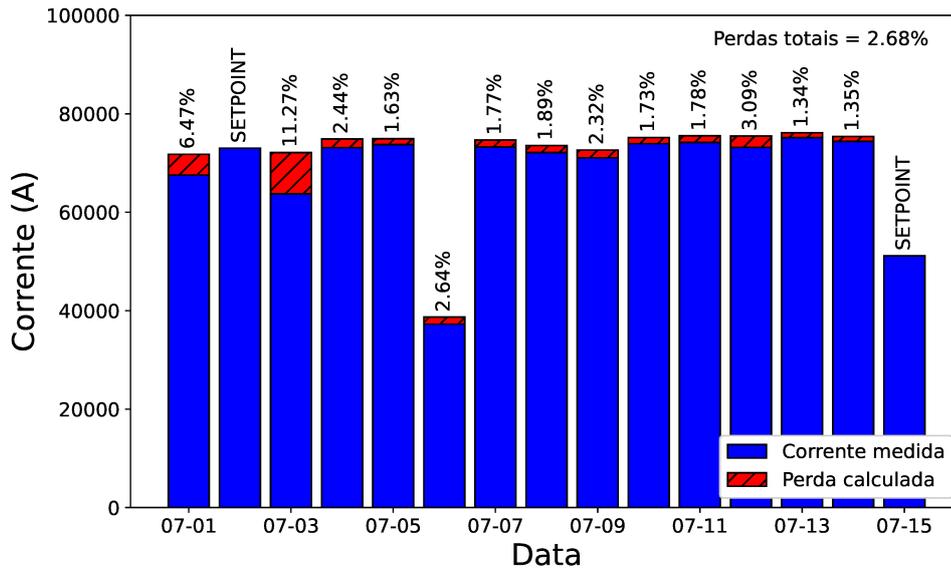


Figura 8 – Soma diária da corrente contínua medida com perdas de corrente estimadas (%) usando parâmetros $K = 4$ e $L = 2$, erro médio de 2,68%.

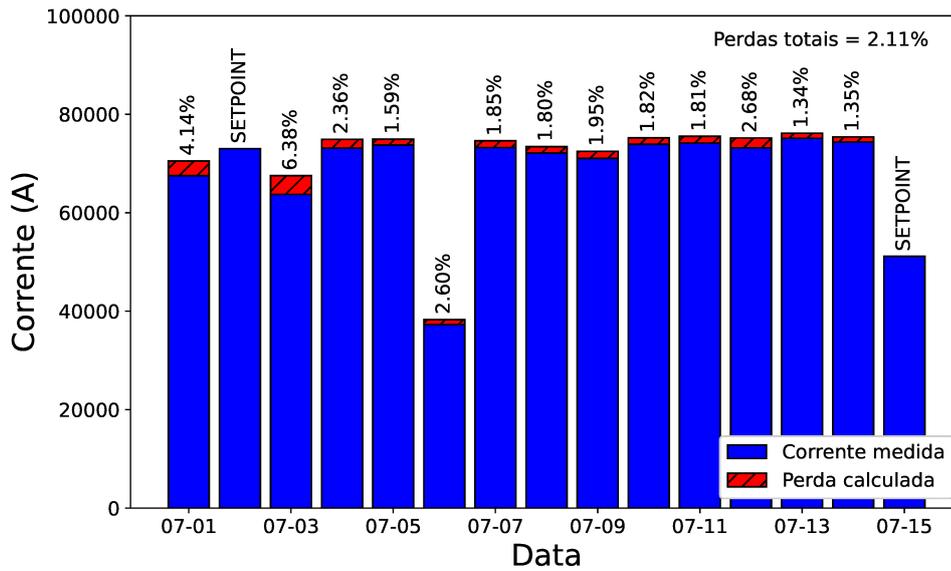


Figura 9 – Soma diária da corrente contínua medida com perdas de corrente estimadas (%) usando parâmetros $K = 36$ e $L = 2$, erro médio de 2,11%.

planta. Plantas muito extensas podem enfrentar condições diferentes em algumas de suas partes (por exemplo, efeitos de microclima) ou estar sujeitas a obstáculos naturais (por exemplo, uma montanha sombreando uma região durante um período do dia), enquanto plantas menores podem não apresentar tais condições. Um valor mais alto de K leva a menos *stringboxes* em cada agrupamento, e conseqüentemente uma maior probabilidade de que as *stringboxes* dentro do agrupamento enfrentem condições semelhantes, além de dar mais peso à corrente de cada *stringbox* ao considerar as correntes medianas dos agrupamentos. Valores mais baixos de K , por outro lado, levam a agrupamentos geográficos com mais

stringboxes, aumentando a diversidade de condições geográficas dentro do agrupamento, mas também filtrando o impacto de uma única *stringbox* com mau funcionamento. A definição de K deve ser revisitada se a planta for expandida ou reformada, mas com base nos resultados, uma recomendação é começar dividindo a planta em vários agrupamentos geográficos.

Valores mais altos de L reduzem os FN e aumentam os FP, pois mais agrupamentos de corrente incluem apenas *stringboxes* de alto desempenho no agrupamento de corrente saudável. Reduzir L tem o efeito oposto. Assim, ajustar L pode ajudar os operadores da planta a ajustar o método às suas necessidades específicas. Por exemplo, enviar equipes de gerenciamento para inspecionar strings que estão funcionando corretamente desperdiça recursos valiosos e limitados. Portanto, se o operador deseja minimizar esse desperdício, o sistema deve ser ajustado para condições com menos FP, ou seja, com maior R . Por outro lado, se houver equipes de manutenção suficientes disponíveis, o sistema pode ser ajustado para uma configuração com menos FN, ou seja, maior P . De qualquer forma, uma recomendação segura para uma configuração inicial do método proposto é usar um pequeno número de agrupamentos de corrente, como $L = 2$, a fim de limitar o número de FP, enquanto L pode aumentar com o tempo de acordo com os objetivos das equipes de operações e manutenção.

Outro aspecto relevante a ser notado é que o número de strings ausentes estimadas não é exato, pois existem outras causas de subdesempenho em uma planta fotovoltaica, como sujeira, envelhecimento e sombreamento. No entanto, a menos que haja diferenças significativas na data de instalação ou limpeza, o efeito do envelhecimento e da sujeira é tipicamente espalhado ao longo da planta, reduzindo a corrente média em todas as *stringboxes* e, portanto, não impactando significativamente a abordagem proposta. O mesmo é verdade para outras anomalias que são uniformemente espalhadas ao longo da planta, tendo impacto limitado em M_n^{mis} . Por outro lado, strings ausentes afetam consideravelmente a corrente de saída da *stringbox*, sendo muito mais facilmente detectadas pelo método proposto. Além disso, o impacto de fontes intermitentes de subdesempenho, como sombras passageiras, pode ser mitigado considerando que uma string está ausente apenas se detectada assim por dois (ou mais) dias consecutivos.

Uma prática recomendada em relação ao método proposto é desconsiderar os casos onde $M_n^{\text{mis}} < 1$, para evitar FP que são apenas devido a outras fontes de subdesempenho. Além disso, também notamos que FP podem ser considerados pelas equipes de operações e manutenção como indicadores de que strings conectados àquela *stringbox* podem estar sob efeito de sujeira excessiva, envelhecimento ou outras fontes de subdesempenho, de modo que devem ser revisados, consistindo assim em um subproduto potencialmente valioso da abordagem proposta.

6 CONCLUSÕES

A abordagem proposta enfrenta o desafio de detectar strings ausentes em plantas VLSPV. Utilizando técnicas de análise de dados e aprendizado de máquina, estimamos a quantidade de strings ausentes por *stringbox*, auxiliando o pessoal no local a identificar strings problemáticas. O método proposto baseia-se exclusivamente em dados de corrente coletados de cada *stringbox*, elimina a necessidade de instrumentação mais granular, que é cara e impraticável em instalações de grande escala.

A validação do método não supervisionado proposto contra dados do mundo real mostrou bons resultados, com uma precisão acima de 90% na detecção de strings ausentes em diversas configurações. Além disso, realizamos uma análise quantitativa das perdas de energia causadas por strings ausentes, oferecendo uma visão sobre as implicações econômicas dessas anomalias. Analisamos dados reais de uma usina por um período de 15 dias, demonstrando as perdas estimadas devido às strings ausentes. Observamos que o método funciona bem em dias com alta irradiância, mas requer ajustes no número de agrupamentos geográficos (K) em dias com baixa irradiância para obter estimativas mais precisas.

Foram discutidos também aspectos práticos da implementação e ajuste do método, destaca-se a importância de escolher corretamente os parâmetros K (número de agrupamentos geográficos) e L (número de agrupamentos de corrente). Nossa abordagem mostrou-se robusta e ajustável às necessidades específicas das operações e equipes de manutenção, proporcionando um equilíbrio entre a detecção de FP e FN.

Em suma, a metodologia proposta não apenas facilita a detecção de strings ausentes, mas também oferece *insights* para a operação e manutenção de plantas VLSPV, melhorando a eficiência e reduzindo as perdas energéticas associadas.

6.1 TRABALHOS FUTUROS

Este estudo abriu várias possibilidades interessantes para pesquisas futuras e melhorias práticas no monitoramento e manutenção de sistemas fotovoltaicos de grande escala. Abaixo são discutidos alguns pontos que merecem atenção adicional:

- **Validação Adicional com Dados Reais:** A validação contínua do método proposto com dados reais é essencial para aprimorar sua robustez e confiabilidade. Com uma extensão do período de coleta de dados ou a obtenção de novos conjuntos de dados, seria possível realizar análises mais detalhadas e precisas, explorando diferentes condições climáticas e sazonais que possam afetar o desempenho do sistema.
- **Geração de Relatórios Automatizados:** Implementar a geração automatizada de relatórios poderia fornecer suporte valioso para as equipes de operação e manutenção das usinas. Relatórios periódicos poderiam destacar anomalias detectadas, como

strings ausentes ou falha operacional, facilitando a identificação rápida e intervenção proativa.

- Relatórios de Perdas Mensais/Anuais: Desenvolver relatórios detalhados sobre as perdas de energia devido a strings ausentes ou com mau funcionamento poderia oferecer uma visão clara e quantitativa dos impactos financeiros dessas anomalias. Gerentes e tomadores de decisão poderiam usar essas informações para ajustar estratégias de manutenção e otimizar o desempenho global da planta.
- Otimização da Rota de Manutenção: Com informações detalhadas sobre a localização de cada ITS ou até mesmo de cada *stringbox*, seria viável estabelecer uma ordem otimizada para a regularização de cada *stringbox*. Estratégias baseadas em algoritmos de otimização poderiam ser desenvolvidas para minimizar o tempo de inatividade e maximizar a eficiência das operações de manutenção.

Essas direções de pesquisa não apenas fortaleceriam a aplicabilidade prática do método proposto, mas também contribuiriam significativamente para a eficiência operacional e econômica de sistemas fotovoltaicos em escala industrial.

REFERÊNCIAS

ANDERSON, Kevin *et al.* A Method for Estimating Time-Series PV Production Loss from Solar Tracking Failures. **IEEE Journal of Photovoltaics**, IEEE Electron Devices Society, v. 12, p. 119–126, 1 jan. 2022. ISSN 21563403. DOI: 10.1109/JPHOTOV.2021.3123872.

AZIZ, Farkhanda *et al.* A Novel Convolutional Neural Network-Based Approach for Fault Classification in Photovoltaic Arrays. **IEEE Access**, Institute of Electrical e Electronics Engineers Inc., v. 8, p. 41889–41904, 2020. ISSN 21693536. DOI: 10.1109/ACCESS.2020.2977116.

BADR, Mohamed M.; ABDEL-KHALIK, Ayman S. *et al.* Intelligent fault identification strategy of photovoltaic array based on ensemble self-training learning. **Solar Energy**, Elsevier Ltd, v. 249, p. 122–138, jan. 2023. ISSN 0038092X. DOI: 10.1016/j.solener.2022.11.017.

BADR, Mohamed M.; HAMAD, Mostafa S. *et al.* Fault Identification of Photovoltaic Array Based on Machine Learning Classifiers. **IEEE Access**, Institute of Electrical e Electronics Engineers Inc., v. 9, p. 159113–159132, 2021. DOI: 10.1109/ACCESS.2021.3130889.

BAGHERZADEH, Hossein *et al.* A novel protection scheme to detect, discriminate, and locate electric faults in photovoltaic arrays using a minimal number of measurements. **International Journal of Electrical Power and Energy Systems**, Elsevier Ltd, v. 141, out. 2022. ISSN 01420615. DOI: 10.1016/j.ijepes.2022.108172.

BOXWELL, Michael. **Solar Electricity Handbook: A Simple, Practical Guide to Solar Energy - Designing and Installing Photovoltaic Solar Electric Systems**. United Kingdom: Greenstream Publishing, 2010.

CHO, Kyeong Hee *et al.* Failure Diagnosis Method of Photovoltaic Generator Using Support Vector Machine. **Journal of Electrical Engineering and Technology**, Korean Institute of Electrical Engineers, v. 15, p. 1669–1680, 4 jul. 2020. ISSN 20937423. DOI: 10.1007/s42835-020-00430-9.

COELHO, Roberto Francisco; SCHMITZ, Lenon; MARTINS, Denizar Cruz. **Energia Solar Fotovoltaica: Geração, Conversão e Aplicações**. Florianópolis, SC: Ed. dos Autores, 2022.

COSTA, Clayton H. Da *et al.* A Comparison of Machine Learning-Based Methods for Fault Classification in Photovoltaic Systems. **2019 IEEE PES Conference on Innovative Smart Grid Technologies, ISGT Latin America 2019**, Institute of Electrical e Electronics Engineers Inc., set. 2019. DOI: 10.1109/ISGT-LA.2019.8895279.

DYAMOND, Wayne Peter; RIX, Arnold J. **Fault Detection and Performance Visualisation for a Grid-Connected Photovoltaic Power Plant Using Sensor Data**. [S.l.]: Stellenbosch University, 2019. Master Thesis.

ELMALLAH, Salma *et al.* Shedding light on large-scale solar impacts: An analysis of property values and proximity to photovoltaics across six U.S. states. **Energy Policy**, Elsevier Ltd, v. 175, abr. 2023. ISSN 03014215. DOI: 10.1016/j.enpo1.2023.113425.

ESKANDARI, Aref; MILIMONFARED, Jafar; AGHAEI, Mohammadreza. Fault Detection and Classification for Photovoltaic Systems Based on Hierarchical Classification and Machine Learning Technique. **IEEE Transactions on Industrial Electronics**, Institute of Electrical e Electronics Engineers Inc., v. 68, p. 12750–12759, 12 dez. 2021. ISSN 15579948. DOI: 10.1109/TIE.2020.3047066.

GOMPPEL, Jonas Van; SPINA, Domenico; DEVELDER, Chris. Satellite based fault diagnosis of photovoltaic systems using recurrent neural networks. **Applied Energy**, Elsevier Ltd, v. 305, jan. 2022. ISSN 03062619. DOI: 10.1016/j.apenergy.2021.117874.

GUO, Hui *et al.* A novel method for quantitative fault diagnosis of photovoltaic systems based on data-driven. **Electric Power Systems Research**, Elsevier Ltd, v. 210, set. 2022. ISSN 03787796. DOI: 10.1016/j.epsr.2022.108121.

HARROU, Fouzi *et al.* Monitoring of photovoltaic systems using improved kernel-based learning schemes. **IEEE Journal of Photovoltaics**, IEEE Electron Devices Society, v. 11, p. 806–818, 3 mai. 2021. ISSN 21563403. DOI: 10.1109/JPHOTOV.2021.3057169.

HASHEMI, Behzad *et al.* Systematic photovoltaic system power losses calculation and modeling using computational intelligence techniques. **Applied Energy**, Elsevier Ltd, v. 284, fev. 2021. ISSN 03062619. DOI: 10.1016/j.apenergy.2020.116396.

HE, Zengxiang *et al.* Compound fault diagnosis for photovoltaic arrays based on multi-label learning considering multiple faults coupling. **Energy Conversion and Management**, Elsevier Ltd, v. 279, mar. 2023. ISSN 01968904. DOI: 10.1016/j.enconman.2023.116742.

HOJABRI, Mojgan *et al.* IoT-Based PV Array Fault Detection and Classification Using Embedded Supervised Learning Methods. **Energies**, MDPI, v. 15, 6 mar. 2022. ISSN 19961073. DOI: 10.3390/en15062097.

HOLMGREN, William *et al.* PVLIB-Python: a python package for modeling solar energy systems. **Journal of Open Source Software**, The Open Journal, v. 3, n. 29, p. 884, 2018.

IBRAHIM, Mariam *et al.* Machine Learning Schemes for Anomaly Detection in Solar Power Plants. **Energies**, MDPI, v. 15, 3 fev. 2022. ISSN 19961073. DOI: 10.3390/en15031082.

IEA. **Renewables 2022**. License: CC BY 4.0. IEA. 2022. Disponível em: <https://www.iea.org/reports/renewables-2022>.

ITO, Masakazu *et al.* **A preliminary study on potential for very large-scale photovoltaic power generation (VLS-PV) system in the Gobi desert from economic and environmental viewpoints**. v. 75. [S.l.: s.n.], 2003. P. 507–517.

JIAO, Xuan *et al.* A Novel Fault Diagnosis Scheme for PV Plants Based on Real-Time System State Identification. **IEEE Journal of Photovoltaics**, IEEE Electron Devices Society, jul. 2023. ISSN 21563403. DOI: 10.1109/JPHOTOV.2023.3262950.

KUMAR, Utkarsh; MISHRA, Sukumar; DASH, Kalyan. An IoT and Semi-Supervised Learning-Based Sensorless Technique for Panel Level Solar Photovoltaic Array Fault Diagnosis. **IEEE Transactions on Instrumentation and Measurement**, v. 72, p. 1–12, 2023. DOI: 10.1109/TIM.2023.3287247.

LAURINO, Marica; PILIOUGINE, Michel; SPAGNUOLO, Giovanni. Artificial neural network based photovoltaic module diagnosis by current–voltage curve classification. **Solar Energy**, Elsevier Ltd, v. 236, p. 383–392, abr. 2022. ISSN 0038092X. DOI: 10.1016/j.solener.2022.02.039.

LAZZARETTI, André Eugênio *et al.* A monitoring system for online fault detection and classification in photovoltaic plants. **Sensors (Switzerland)**, MDPI AG, v. 20, p. 1–30, 17 set. 2020. ISSN 14248220. DOI: 10.3390/s20174688.

LEVA, Sonia; MUSSETTA, Marco; OGLIARI, Emanuele. PV Module Fault Diagnosis Based on Microconverters and Day-Ahead Forecast. **IEEE Transactions on Industrial Electronics**, Institute of Electrical e Electronics Engineers Inc., v. 66, p. 3928–3937, 5 mai. 2019. ISSN 02780046. DOI: 10.1109/TIE.2018.2879284.

LODHI, Ehtisham *et al.* An AdaBoost Ensemble Model for Fault Detection and Classification in Photovoltaic Arrays. **IEEE Journal of Radio Frequency Identification**, Institute of Electrical e Electronics Engineers Inc., v. 6, p. 794–800, 2022. ISSN 24697281. DOI: 10.1109/JRFID.2022.3212310.

MACQUEEN, J. B. Some methods for classification and analysis of multivariate observations. **Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability**, v. 1, p. 281–297, 1967.

MADETI, Siva Ramakrishna; SINGH, S. N. Modeling of PV system based on experimental data for fault detection using kNN method. **Solar Energy**, Elsevier Ltd, v. 173, p. 139–151, out. 2018. ISSN 0038092X. DOI: 10.1016/j.solener.2018.07.038.

MADICHETTY, Sreedhar *et al.* A Novel High-Speed Sensorless Faulty Panel Detection Technique for an SPV String/Array: An accurate and cost-effective approach for SPV industry. **IEEE Power Electronics Magazine**, Institute of Electrical e Electronics Engineers Inc., v. 9, p. 33–39, 1 mar. 2022. ISSN 23299215. DOI: 10.1109/MPEL.2022.3140985.

MAPS, Raptor. **2023 Global Solar Report**. [S.l.], 2023. Disponível em: <https://raptormaps.com/resources/2023-global-solar-report>.

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**, Elsevier Ltd, v. 208, p. 399–408, mai. 2023. ISSN 18790682. DOI: 10.1016/j.renene.2023.03.096.

MEMON, Sufyan Ali *et al.* A Machine-Learning-Based Robust Classification Method for PV Panel Faults. **Sensors**, MDPI, v. 22, 21 nov. 2022. ISSN 14248220. DOI: 10.3390/s22218515.

MUSTAFA, Zain *et al.* Fault identification for photovoltaic systems using a multi-output deep learning approach. **Expert Systems with Applications**, Elsevier Ltd, v. 211, jan. 2023. ISSN 09574174. DOI: 10.1016/j.eswa.2022.118551.

O SETOR ELÉTRICO. **2022: o melhor ano da energia solar no Brasil**. [S.l.]: ABSOLAR, 2022. Acessado em 02 novembro 2022. Disponível em: <https://www.absolar.org.br/noticia/2022-o-melhor-ano-da-energia-solar-no-brasil/>.

ØGAARD, Mari B. *et al.* Robust and fast detection of small power losses in large-scale PV systems. **IEEE Journal of Photovoltaics**, IEEE Electron Devices Society, v. 11, p. 819–826, 3 mai. 2021. ISSN 21563403. DOI: 10.1109/JPHOTOV.2021.3060732.

PILIOUGINE, Michel; SPAGNUOLO, Giovanni. Mismatching and partial shading identification in photovoltaic arrays by an artificial neural network ensemble. **Solar Energy**, Elsevier Ltd, v. 236, p. 712–723, abr. 2022. ISSN 0038092X. DOI: 10.1016/j.solener.2022.03.026.

SAKTHIVEL, Senthamizh Selvan; ARUNACHALAM, Venkadesan; JAGATHEESAN, Karthickraja. Detection, Classification, and Location of Open-Circuit and Short-Circuit Faults in Solar Photovoltaic Array: An Approach Using Single Sensor. **IEEE Journal of Photovoltaics**, v. 13, n. 6, p. 986–990, 2023. DOI: 10.1109/JPHOTOV.2023.3304113.

SIMÃO, Tiago Edmir *et al.* A Method for the Estimation of Missing Strings in Very-Large-Scale Photovoltaic Power Plants. **IEEE Journal of Photovoltaics**, v. 14, n. 5, p. 839–847, 2024. DOI: 10.1109/JPHOTOV.2024.3430977.

UTAMA, Christian *et al.* Explainable artificial intelligence for photovoltaic fault detection: A comparison of instruments. **Solar Energy**, Elsevier Ltd, v. 249, p. 139–151, jan. 2023. ISSN 0038092X. DOI: 10.1016/j.solener.2022.11.018.

VOUTSINAS, Stylianos *et al.* Development of a multi-output feed-forward neural network for fault detection in Photovoltaic Systems. **Energy Reports**, Elsevier Ltd, v. 8, p. 33–42, nov. 2022. ISSN 23524847. DOI: 10.1016/j.egyr.2022.06.107.

WANG, Mengyuan; XU, Xiaoyuan; YAN, Zheng. Online fault diagnosis of PV array considering label errors based on distributionally robust logistic regression. **Renewable Energy**, Elsevier Ltd, v. 203, p. 68–80, fev. 2023. ISSN 18790682. DOI: 10.1016/j.renene.2022.11.126.

YURTSEVEN, Kaan; KARATEPE, Engin; DENIZ, Esref. Sensorless fault detection method for photovoltaic systems through mapping the inherent characteristics of PV plant site: Simple and practical. **Solar Energy**, Elsevier Ltd, v. 216, p. 96–110, mar. 2021. ISSN 0038092X. DOI: 10.1016/j.solener.2021.01.011.

ZHAO, Yingying *et al.* Hierarchical anomaly detection and multimodal classification in large-scale photovoltaic systems. **IEEE Transactions on Sustainable Energy**, Institute of Electrical e Electronics Engineers Inc., v. 10, p. 1351–1361, 3 jul. 2019. ISSN 19493037. DOI: 10.1109/TSTE.2018.2867009.

ZHOU, Siyu *et al.* A Shadow Fault Diagnosis Method Based on the Quantitative Analysis of Photovoltaic Output Prediction Error. **IEEE Journal of Photovoltaics**, Institute of Electrical e Electronics Engineers Inc., v. 10, p. 1158–1165, 4 jul. 2020. ISSN 21563403. DOI: 10.1109/JPHOTOV.2020.2995041.