

Universidade Federal de São Carlos – UFSCar  
Programa de Pós-Graduação em Ciência da Computação  
Leonardo Aparecido Lira Diniz

**Transferência de aprendizado profundo para  
previsão da irradiação solar: uma análise para  
diferentes bioclimas e escassez de dados**

São Carlos - SP

Março/2026



Leonardo Aparecido Lira Diniz

**Transferência de aprendizado profundo para previsão da  
irradiação solar: uma análise para diferentes bioclimas e  
escassez de dados**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de São Carlos, como parte dos requisitos para a obtenção do título de Mestre em Ciência da Computação. Área de Concentração: Inteligência Artificial

Orientador: Prof. Dr. Ricardo Augusto Souza Fernandes

São Carlos - SP

Março/2026



**Folha de Aprovação**

Defesa de dissertação de mestrado do(a) candidato(a) Leonardo Aparecido Lira Diniz, realizada em 02/03/2026

**Comissão Julgadora**

Prof(a) Dr(a) Ricardo Augusto Souza Fernandes (USP)

Prof(a) Dr(a) Vinicius Ponte Machado (UFPI)

Prof(a) Dr(a) Thamyres Tamulla Cavalcante Palito (UFPB)

O relatório de defesa assinado pelos membros da comissão Julgadora encontra-se arquivado junto ao Programa de Pós-Graduação em Ciência da Computação

# Agradecimentos

À minha esposa Leide que, incansavelmente, deu todo o suporte emocional que um ser humano é capaz de oferecer.

Ao meu orientador Prof. Dr. Ricardo A. S. Fernandes, pela imensa paciência frente aos meus equívocos incipientes.



# Resumo

A natureza intermitente da irradiação solar tem sido um grande desafio no planejamento e na operação de sistemas de energia elétrica. No Brasil, devido às distintas zonas bioclimáticas, cada uma com suas peculiaridades e comportamentos climáticos, o cenário torna-se complexo. Por este motivo, uma metodologia baseada em transferência de aprendizado aliada ao estado da arte em previsão de irradiação solar pode trazer vantagens, principalmente quando se tem uma baixa quantidade de dados de determinadas estações climatológicas. Nesse sentido, a transferência de aprendizado poderá ser usada tanto entre dados de estações pertencentes a uma mesma zona bioclimática (visando inclusive resolver o problema de dados faltantes) quanto entre dados provenientes de estações alocadas em zonas bioclimáticas distintas (buscando resolver o problema da disparidade climática). O trabalho propõe, a partir de um *baseline* construído com o modelo de média móvel integrada auto-regressiva sazonal, avaliar o desempenho de redes neurais do tipo LSTM (*long-short term memory*) e sua hibridização com redes neurais convolucionais (CNN, do inglês *convolutional neural network*), com o intuito de realizar a previsão horária de irradiação solar. Os resultados demonstram que, em relação ao teste de modelos construídos unicamente por LSTMs, em comparação aos modelos híbridos CNN-LSTM, os primeiros se mostraram ligeiramente melhores, em aproximadamente 1%, em média. Em relação aos testes de transferência de aprendizado, as abordagens de extração de características e ajuste fino tiveram razoável desempenho quando comparadas à adaptação *zero-shot*, ficando equiparados aos resultados dos modelos LSTM de cada estação climatológica, principalmente na configuração de estações por distância. Em relação à configuração de estações com baixa quantidade de dados, a abordagem teve bom desempenho, principalmente dentro da métrica MAE.

**Palavras-chaves:** Aprendizado de máquina; Irradiação Solar; Modelos estatísticos; Transferência de aprendizado.



# Abstract

The intermittent nature of solar irradiation has been a major challenge in planning and operating electrical power systems. In Brazil, due to the distinct bioclimatic zones, each with its own peculiarities and climatic behaviors, the scenario becomes complex. For this reason, a methodology based on transfer learning combined with state-of-the-art solar irradiation forecasting can bring advantages, especially when there is a low amount of data from certain climatological stations. In this sense, transfer learning can be used both between data from stations belonging to the same bioclimatic zone (aiming to solve the problem of missing data) and between data from stations located in different bioclimatic zones (seeking to solve the problem of climatic disparity). This work proposes, based on a baseline constructed with the seasonal autoregressive integrated moving average model, to evaluate the performance of LSTM (long-short term memory) neural networks and their hybridization with convolutional neural networks (CNNs) in order to perform hourly solar irradiance forecasting. The results demonstrate that, in relation to the test of models constructed solely by LSTMs, compared to the CNN-LSTM hybrid models, the former showed slightly better performance, by approximately 1% on average. Regarding the learning transfer tests, the feature extraction and fine-tuning approaches performed reasonably well when compared to zero-shot adaptation, being comparable to the results of the LSTM models for each climatological station, especially in the configuration of stations by distance. Regarding the configuration of stations with a low amount of data, the approach performed well, mainly within the MAE metric.

**Keywords:** Machine learning; Solar Irradiance; Statistical models; Transfer learning.



# Lista de ilustrações

Figura 1 – Esquema simplificado do funcionamento do Transfer Learning. Adaptado de (GAO et al., 2024). . . . .	30
Figura 2 – Representação gráfica da metodologia. . . . .	43
Figura 3 – Posição no mapa de todas as Zonas Bioclimáticas. . . . .	45
Figura 4 – Média dos Valores de Irradiação Solar por Hora para as Distintas Zonas Bioclimáticas. . . . .	46
Figura 5 – Porcentagens dos valores nulos por estação. . . . .	46
Figura 6 – Diagrama do modelo LSTM parametrizado para previsão de GHI. . . . .	49
Figura 7 – Diagrama do modelo BiLSTM parametrizado para previsão de GHI. . . . .	50
Figura 8 – Diagrama dos modelos híbridos parametrizados para previsão de GHI. . . . .	51
Figura 9 – Importâncias de cada característica climática dos dados de Teresina. . . . .	52
Figura 10 – Representação gráfica da previsão de Teresina. . . . .	55
Figura 11 – Representação gráfica da previsão de Brasília. . . . .	55
Figura 12 – Representação gráfica da previsão em Teresina utilizando o modelo LSTM. . . . .	56
Figura 13 – Cenário #2: MSE dos modelos baseados em LSTM para a própria estação comparados às abordagens de transferência de aprendizado Zero-Shot, Extração de Características e Ajuste Fino. . . . .	62
Figura 14 – Cenário #2: MAE dos modelos baseados em LSTM para a própria estação comparados às abordagens de transferência de aprendizado Zero-Shot, Extração de Características e Ajuste Fino. . . . .	63
Figura 15 – Cenário #2: $R^2$ dos modelos baseados em LSTM para a própria estação comparados às abordagens de transferência de aprendizado Zero-Shot, Extração de Características e Ajuste Fino. . . . .	63
Figura 16 – Cenário #3: MSE dos modelos singulares comparados às abordagem Zero-Shot, Extração de Características e Ajuste Fino . . . . .	65
Figura 17 – Cenário #3: MAE dos modelos singulares comparados às abordagens Zero-Shot, Extração de Características e Ajuste Fino . . . . .	66
Figura 18 – Cenário #3: $R^2$ dos modelos singulares comparados às abordagem Zero-Shot, Extração de Características e Ajuste Fino . . . . .	66
Figura 19 – Análise de Correlação das Variáveis Explicativas do Conjunto de Dados de Teresina. . . . .	84
Figura 20 – <i>Resample</i> dos dados de Brasília do ano todo, de horas para dias. . . . .	84
Figura 21 – Teste <i>AdFuller</i> para a Estação Climatológica de Brasília. . . . .	85

Figura 22 – Decomposição da série temporal de Brasília em Janeiro/Março de 2024.	85
Figura 23 – Gráficos de Autocorrelação e Autocorrelação Parcial de Brasília . . . . .	88

# Lista de tabelas

Tabela 1 – Conjunto de Estações Seleccionadas para o Cenário #1. . . . .	44
Tabela 2 – Conjunto de Estações Seleccionadas para o Cenário #3. . . . .	47
Tabela 3 – Parametrizações do Modelo SARIMA. . . . .	48
Tabela 4 – Organização das análises realizadas para transferência de aprendizado. . . . .	53
Tabela 5 – Resultados das melhores parametrizações do modelo SARIMA para cada estação climatológica. . . . .	55
Tabela 6 – Resultados médios do LSTM e do BiLSTM para as execuções de cada modelo. . . . .	56
Tabela 7 – Resultados médios do CNN-LSTM e do CNN-BiLSTM para as execuções de cada modelo. . . . .	57
Tabela 8 – Melhor modelo por estação considerando as métricas MAE e $R^2$ . . . . .	58
Tabela 9 – Resultados da Transferência de Aprendizado obtidos para o cenário #2 em Curitiba. . . . .	59
Tabela 10 – Resultados da Transferência de Aprendizado obtidos para o cenário #2 em Santa Maria. . . . .	59
Tabela 11 – Resultados da Transferência de Aprendizado obtidos para o cenário #2 em Juiz de Fora. . . . .	60
Tabela 12 – Resultados da Transferência de Aprendizado obtidos para o cenário #2 em Brasília. . . . .	60
Tabela 13 – Resultados da Transferência de Aprendizado obtidos para o cenário #2 em Campos dos Goytacazes. . . . .	61
Tabela 14 – Resultados da Transferência de Aprendizado obtidos para o cenário #2 em Campo Grande. . . . .	61
Tabela 15 – Resultados da Transferência de Aprendizado obtidos para o cenário #2 em Manaus. . . . .	62
Tabela 16 – Resultados da Transferência de Aprendizado obtidos para o cenário #3 em Alvorada do Gurgueia. . . . .	67
Tabela 17 – Resultados da Transferência de Aprendizado obtidos para o cenário #3 em Angical do Piauí . . . . .	68
Tabela 18 – Resultados da Transferência de Aprendizado obtidos para o cenário #3 em Baixa Grande do Ribeiro. . . . .	68
Tabela 19 – Resultados da Transferência de Aprendizado obtidos para o cenário #3 em Bom Jesus do Piauí. . . . .	69
Tabela 20 – GHI Médio e Desvio Padrão para cada Estação Climatológica. . . . .	83



# Sumário

<b>1</b>	<b>Introdução</b>	<b>15</b>
1.1	Motivação e Justificativa	16
1.2	Objetivos e Contribuições	17
1.3	Organização da Dissertação	18
<b>2</b>	<b>Fundamentação teórica</b>	<b>20</b>
2.1	Séries temporais	20
2.2	Horizonte de previsão	21
2.3	Componentes da Irradiação Solar	22
2.3.1	Irradiância Normal Direta	22
2.3.2	Irradiância Horizontal Difusa	22
2.3.3	Irradiância Horizontal Global	22
2.4	Modelos de Previsão	23
2.4.1	Modelo de Céu Claro	23
2.4.2	Modelo Físico	23
2.4.3	Modelos Estatísticos	24
2.4.4	Modelos Baseados em Aprendizado de Máquina	25
2.4.4.1	Aprendizado Raso	25
2.4.4.2	Aprendizado Profundo	26
2.5	Transferência de Aprendizado	29
2.5.1	Estratégias de Transferência de Aprendizado	31
2.6	Métricas de Avaliação de Desempenho	32
<b>3</b>	<b>Levantamento Bibliográfico</b>	<b>34</b>
3.1	Abordagens Baseadas em Modelos de Aprendizado Profundo	34
3.2	Abordagens Baseadas em Modelos Híbridos	37
3.3	Abordagens Baseadas em Transferência de Aprendizado	39
<b>4</b>	<b>Metodologia Proposta</b>	<b>43</b>
4.1	Coleta dos Dados e Pré-processamento	44
4.1.1	Cenário #1	45
4.1.2	Cenário #2	46
4.1.3	Cenário #3	47
4.2	Análises das Séries Temporais	47
4.3	Modelo de Referência	48
4.4	Modelos de Aprendizado Profundo	48

4.4.1	Estratégia de modelagem . . . . .	49
4.4.1.1	Construção dos modelos . . . . .	49
4.4.1.2	Treinamento e Validação dos Modelos . . . . .	51
4.5	Transferência de Aprendizado . . . . .	52
<b>5</b>	<b>Resultados . . . . .</b>	<b>54</b>
5.1	Cenário #1 . . . . .	54
5.1.1	Modelo SARIMA . . . . .	54
5.1.2	Modelos Baseados em Aprendizado Profundo . . . . .	54
5.2	Cenários #2 – Transferência de Aprendizado entre Estações de Diferentes Zonas Bioclimáticas . . . . .	58
5.2.1	Curitiba . . . . .	59
5.2.2	Santa Maria . . . . .	61
5.2.3	Juiz de Fora . . . . .	62
5.2.4	Brasilia . . . . .	63
5.2.5	Campos dos Goytacazes . . . . .	64
5.2.6	Campo Grande . . . . .	64
5.2.7	Manaus . . . . .	64
5.3	Cenário #3 – Transferência de Aprendizado entre Estações Circunvizinhas . . . . .	65
5.3.1	Alvorada do Gurguéia . . . . .	65
5.3.2	Angical do Piauí . . . . .	66
5.3.3	Baixa Grande do Ribeiro . . . . .	67
5.3.4	Bom Jesus . . . . .	68
<b>6</b>	<b>Considerações finais . . . . .</b>	<b>70</b>
	<b>Referências . . . . .</b>	<b>73</b>
	<b>APÊNDICE A Análises Estatísticas . . . . .</b>	<b>82</b>
A.1	Média e Desvio Padrão . . . . .	82
A.2	Correlação . . . . .	83
A.3	Estacionariedade . . . . .	84
A.4	Tendência e Sazonalidade . . . . .	85
	<b>APÊNDICE B Parametrização do SARIMA . . . . .</b>	<b>87</b>
B.1	ACF e PACF . . . . .	87
	<b>APÊNDICE C xLSTM . . . . .</b>	<b>89</b>

# 1 Introdução

Nas últimas décadas, questões ambientais têm sido um dos principais pontos em que governos e corporações têm empregado esforços para o desenvolvimento sustentável. Relatórios globais mostraram que, na década de 90, os níveis de emissões globais de CO<sub>2</sub> foram de aproximadamente 22,7 milhões de quilotons. No entanto, segundo [Energy Institute \(2024\)](#), em 2024, os níveis de emissões de CO<sub>2</sub> atingiram 40,8 bilhões de toneladas. Ainda é válido mencionar que, o mesmo relatório, apontou os Estados Unidos como responsável por 1/5 dessas emissões. Ademais, segundo a [International Energy Agency \(2024\)](#), a China é responsável por cerca de 75% do consumo de combustíveis fósseis, enquanto a Índia demonstra forte dependência de gás natural, responsável por cerca de 50% do consumo global.

Nesse sentido, nota-se que o cenário de deterioração ambiental tem sido fortemente impactado pelo crescimento das emissões de CO<sub>2</sub>, motivado principalmente pelo aumento da demanda por energia ([YI et al., 2023](#)). Conforme evidenciado pela [International Energy Agency \(2024\)](#), o impacto ambiental está bastante atrelado à produção de energia por meio de combustíveis fósseis e, conseqüentemente, o efeito estufa está diretamente relacionado à emissão de CO<sub>2</sub>. Além disso, o relatório demonstra que os subsídios aos combustíveis fósseis aumentam a procura por energia dessas fontes, causando alterações significativas no clima e dificultando o cumprimento das metas climáticas. Conforme mostrado por [Taghvae et al. \(2023\)](#), o Irã foi o país que mais subsidiou o consumo de combustíveis fósseis entre os anos de 2010 e 2019. Por conseqüência, foi considerado um dos maiores responsáveis pelas alterações climáticas no Oriente Médio.

Nações de todo o planeta têm feito grandes esforços para reduzir o impacto ambiental decorrente da produção e do uso de combustíveis fósseis, como o petróleo e o carvão. De acordo com [Sehrawat, Vashisht e Singh \(2023\)](#), a Índia demonstrou uma queda de 16% no uso de combustíveis poluentes na produção de energia no país entre 2014 e 2022 (com 74% e 58% de uso desse tipo de fonte, respectivamente). Apesar de previsões realizadas pelo [International Energy Agency \(2024\)](#) indicarem que as emissões de CO<sub>2</sub> atingirão um pico em uma tendência crescente antes de 2030, a China, segundo o mesmo relatório, impulsionará um declínio anual de 3% nas emissões.

Ainda assim, a utilização de combustíveis fósseis e os processos convencionais de geração de energia são predominantes. A economia global, de forma geral, ainda depende fortemente de combustíveis fósseis, como o carvão e o petróleo, gerando impactos negativos e agressivos ao meio ambiente, como o efeito estufa e a poluição dos rios e do ar ([ZANG et al., 2020](#)). Portanto, é evidente que, para diminuir a deterioração ambiental, torna-se

cada vez mais importante o uso de fontes de energia seguras e ambientalmente sustentáveis (RABAIA *et al.*, 2021).

Com a grande demanda por fontes renováveis de energia na última década, esforços cada vez maiores têm sido empregados na produção de energia eólica e solar (LIU *et al.*, 2022), sendo esta última o foco do presente trabalho. Além disso, dada a abundância do potencial de geração de energia solar em determinados países, há grande esforço para viabilizar sua produção, apesar de sua característica intermitente (ZANG *et al.*, 2020). Além da intermitência, pode-se destacar desafios relativos à sazonalidade dos dados, bem como questões relacionadas às medições de irradiação solar realizadas por estações climatológicas como, por exemplo, as condições de sombreamento dos sensores (AGUILAR-LÓPEZ *et al.*, 2022) e a existência de dados faltantes (ZHANG *et al.*, 2024). Portanto, tais questões aumentam a complexidade na previsão de séries temporais atreladas à energia solar e, conseqüentemente, aumenta a dificuldade quanto ao planejamento e à operação das fazendas solares e à sua integração com os sistemas de energia elétrica (KUMARI; TOSHNIWAL, 2021c). Nesse sentido, a previsão horária da irradiação solar incidente no solo, embora seja um processo complexo, é fundamental para garantir uma melhor operação dos sistemas elétricos (KRISHNAN; KUMAR; INDA, 2023). Com isso, metodologias baseadas em aprendizado de máquina têm se destacado e podem ser consideradas o estado da arte atual, conforme mostrado na revisão sistemática proposta por Chu *et al.* (2024).

## 1.1 Motivação e Justificativa

O cenário mundial mostra que, nas próximas décadas, o consumo total de energia mais do que dobrará. E, ao contrário do passado, o crescimento da procura está relacionado principalmente às energias renováveis, particularmente ao setor elétrico (KOBBER *et al.*, 2020).

Em 2022, 83% das novas fontes de geração de energia eram oriundas de energias renováveis. Esta expansão robusta representa um progresso considerável no cumprimento dos objetivos do Acordo de Paris de limitar o aumento da temperatura global da superfície a 1,5 °C acima dos níveis pré-industriais e reduzir substancialmente as emissões de CO<sub>2</sub> relacionadas à produção de energia até 2030. Para cumprir esse objetivo, a capacidade global de produção de energia renovável deve aumentar consideravelmente até 2030 (IRENA; WMO, 2023). Em termos de energia solar, a natureza estocástica das variáveis (movimento das nuvens e outras características meteorológicas) faz com que a precisão da previsão seja impactada. Ao propor soluções que envolvam a produção de energia elétrica oriunda de fontes renováveis, deve-se levar em conta a integração de tais soluções aos sistemas elétricos (HUSSAIN *et al.*, 2024), sendo um dos principais problemas a natureza intermitente da energia solar, que adiciona desafios à implementação e à confiabilidade.

Portanto, o impacto da intermitência de recursos renováveis na rede elétrica é uma questão crítica que deve ser considerada (PAZIKADIN et al., 2020) e, assim, faz-se necessário o desenvolvimento de modelos que possam fazer previsões dentro deste cenário de complexidade inerente a esses recursos. Ao incorporar mecanismos de previsão em suas operações, redes elétricas *Smart Grids* podem gerenciar a demanda e o fornecimento de energia com mais eficiência, aumentando a segurança e a confiabilidade da rede ao integrar-se a malha elétrica tradicional (BHUTTA et al., 2024).

Como estado da arte, os modelos de aprendizado profundo, ou suas hibridizações, têm sido amplamente empregados para lidar com os desafios da previsão da irradiação solar horária (RAHIMI et al., 2023). Todavia, por vezes não há disponibilidade de dados suficientes, seja por problemas técnicos nos sensores ou pela ausência de estações em determinadas localizações (NIE et al., 2024). Além disso, outro grande desafio na previsão de energia solar é considerar a diversidade climática de um país, especialmente em países de dimensões continentais, como o Brasil. Nesse contexto, estratégias de transferência de aprendizado podem ser empregadas na resolução de problemas de generalização de modelos em distintas localidades ou em cenários de escassez de dados (BAK et al., 2025). Tais estratégias necessitam de um retreinamento e de ajustes mínimos, de forma a reduzir a complexidade de treinar diversos modelos especialistas para cada localidade e manter a precisão da previsão (LIAO et al., 2024).

## 1.2 Objetivos e Contribuições

Considerando as motivações e justificativas previamente apresentadas, essa dissertação busca avaliar o cenário brasileiro, ou seja, a previsão de irradiação solar para estações climatológicas localizadas em cada uma das oito zonas bioclimáticas distintas. Para tanto, considerou-se a utilização de modelos híbridos de aprendizado profundo, bem como de modelos estatísticos como *baseline*. Tais modelos foram ajustados tanto sem qualquer conhecimento *a priori* quanto por meio de uma estratégia de transferência de aprendizado por ajuste fino. Vale mencionar que a transferência de aprendizado foi utilizada tanto para suprir a possível escassez de dados de determinadas estações climatológicas quanto para evitar múltiplas configurações de hiperparâmetros dos modelos de cada localidade.

Nesse sentido, o objetivo principal dessa pesquisa foi analisar a capacidade de generalização do modelo híbrido ao empregar a estratégia de transferência de aprendizado entre estações das distintas zonas bioclimáticas brasileiras, utilizando modelos de aprendizado profundo. Além disso, buscou-se também realizar a transferência de aprendizado para estações com escassez de dados, ou seja, aquelas que apresentaram maior quantidade de dados faltantes. Para tanto, objetivos secundários podem ser elencados para que fosse possível atingir esse objetivo primário, a saber:

- Coletar dados de estações climatológicas brasileiras que sejam representativas das zonas bioclimáticas que compõem o território nacional, obedecendo a critérios de diversidade entre as zonas bioclimáticas e de quantidade de registros nulos (dados faltantes). Assim, foram selecionadas estações climatológicas com dados de diferentes taxas de integridade;
- Propor um estágio de engenharia de características responsável por selecionar as séries temporais de interesse, principalmente em função da correlação com a irradiação solar e da ocorrência de dados faltantes. Portanto, ainda que a análise de dados faltantes não seja o alvo desta pesquisa, a integridade das séries temporais tem ênfase em sua seleção;
- Implementar um modelo autorregressivo integrado de média móvel com sazonalidade (SARIMA, do inglês *seasonal autoregressive integrated moving average*) que foi usado como caso base para a previsão horária de irradiação solar;
- Implementar e comparar o desempenho de modelos baseados em aprendizado profundo para a previsão da irradiação solar. Os modelos utilizados foram a LSTM (*long-short term memory*) e sua variação bidirecional, bem como suas hibridizações com a CNN (*convolutional neural network*);
- Implementar uma estratégia de transferência de aprendizado para distintas estações climatológicas e para estações com menor integridade de dados, considerando apenas o modelo de melhor desempenho;
- Avaliar o desempenho dos modelos e as abordagens da estratégia de transferência de aprendizado, utilizando as métricas MSE (*mean square error*), RMSE (*root mean square error*), MAE (*mean absolute error*) e  $R^2$  (*r-square*).

### 1.3 Organização da Dissertação

Essa dissertação de mestrado está organizada em capítulos, os quais são descritos na sequência:

- Capítulo 1 – apresentou uma introdução ao problema de estimação de séries temporais de irradiação solar, destacou as motivações e justificativas que pautaram a pesquisa, bem como os objetivos e contribuições da dissertação para avançar o estado da arte;
- Capítulo 2 – apresenta conceitos fundamentais sobre séries temporais, irradiação solar, modelos de previsão, transferência de aprendizado e as métricas de avaliação de desempenho de modelos preditivos comumente encontradas na literatura;

- 
- Capítulo 3 – traz um levantamento bibliográfico focado em modelos preditivos baseados em aprendizado de máquina (rasos, profundos e hibridizações), bem como aqueles que fazem uso de estratégias de transferência de aprendizado;
  - Capítulo 4 – detalha a metodologia proposta, apresentando as estações selecionadas e análises estatísticas dos dados, bem como análises próprias de séries temporais. Ademais, o capítulo destaca o modelo de referência e os modelos de aprendizado profundo empregados à previsão da irradiação solar horária;
  - Capítulo 5 – apresenta os resultados obtidos tanto pelo modelo de referência quanto pelos modelos de aprendizado profundo. Destacam-se os resultados obtidos a partir de transferência de aprendizado, visto que essa é a maior contribuição da dissertação;
  - Capítulo 6 – traz as conclusões da dissertação e as possíveis pesquisas futuras a serem realizadas, a fim de continuar avançando o estado da arte.

## 2 Fundamentação teórica

Este capítulo traz conceitos fundamentais relacionados à previsão da irradiação solar, detalhando desde definições básicas sobre previsão de séries temporais até o funcionamento dos modelos de aprendizado raso e profundo, bem como as técnicas de transferência de aprendizado. Além disso, traz as métricas de avaliação de desempenho dos modelos que são comumente usadas na literatura.

### 2.1 Séries temporais

As séries temporais podem ser definidas como conjuntos de observações relativas a um fenômeno, em que cada observação corresponde ao valor obtido em um determinado instante (WANG et al., 2025). Em aplicações práticas, o tempo é discreto, devido à aquisição de sinais dos *data loggers* das estações climatológicas (BROCKWELL; DAVIS, 1991). A importância desses valores obtidos não está apenas nas observações individuais em si, mas também na maneira como essas observações se relacionam entre si (WEISS, 2023), em que o comportamento sequencial das séries temporais revela uma dependência temporal e uma relação complexa entre os valores da série, permitindo a modelagem por meio de técnicas estatísticas (WANG et al., 2025). Fundamentalmente, séries temporais são compostas por alguns conceitos, a saber:

1. Estacionariedade – representa a condição estática das medidas estatísticas da série temporal, ou seja, por mais que esteja sujeita às leis da probabilidade que governam o comportamento do processo, as medidas estatísticas dos valores da série (como média e variância) não mudam ao longo do tempo (CRYER; CHAN, 2008);
2. Tendência – obtida a partir da decomposição de uma série temporal, a qual representa uma mudança sistemática dos valores de média e variância ao longo do tempo, podendo ser de alta ou baixa (MUDELSEE, 2019);
3. Sazonalidade – é caracterizada na série temporal pela presença de um padrão periódico que ocorre sistematicamente, com um intervalo que compreende o período em que um determinado comportamento nos dados está se repetindo (ATIQUÉ et al., 2019).

Dispondo destes recursos para análise, é possível, portanto, capturar as relações e dependências complexas entre valores, revelando uma dinâmica subjacente e dando suporte a um amplo número de tarefas subsequentes, incluindo classificação, imputação,

detecção de anomalias e previsão de valores futuros da série (WANG et al., 2025), por meio de modelagem estatística ou de técnicas de aprendizado de máquina, sendo estas últimas o foco do presente trabalho. Ao utilizar uma determinada série temporal  $ST = (x_1, x_2, \dots, x_t)$  com o intuito de predizê-la, onde  $T$  é o tamanho da janela de observação, deve-se definir um horizonte de previsão de comprimento  $H$ , tal que  $\hat{Y} = (\hat{y}_t + 1, \hat{y}_t + 2, \dots, \hat{y}_t + H)$  (LIU; WANG, 2024).

Além das previsões univariadas, nas quais as séries temporais contêm uma única variável, como, por exemplo, a temperatura de um local ao longo do dia, também podem ser multivariadas, adicionando ao conjunto de observações os valores de umidade relativa do ar para essa mesma discretização temporal (JIN et al., 2023). Devido a essa dissertação tratar de previsão univariada, em termos matemáticos, uma série temporal pode ser definida como  $\{x_1, x_2, \dots, x_t\}$ , ou seja, uma sequência de  $t$  amostras de dados ordenadas no tempo, sendo  $x_t \in \mathbb{R}$  o valor da série no tempo  $t$  (JIN et al., 2023).

## 2.2 Horizonte de previsão

O horizonte de previsão define o intervalo de tempo que separa a última observação utilizada no ajuste do modelo de previsão e o valor futuro a ser previsto, ou seja, determina o período futuro para o qual a previsão deve ser feita. Além disso, também pode ser descrito como o intervalo de tempo entre o tempo real e o tempo para o qual uma previsão é feita (KUMARI; TOSHNIWAL, 2021a).

Contudo, não existe um consenso na literatura sobre uma classificação universal de horizontes de previsão. Porém, segundo Kumari e Toshniwal (2021a), de forma geral, os horizontes podem ser definidos como:

1. Curtíssimo prazo – de alguns segundos, minutos ou horas à frente, sendo aplicado geralmente para o monitoramento em tempo real;
2. Curto prazo – contempla horizontes de mais de uma hora a dias à frente;
3. Médio prazo – engloba previsões para mais de uma semana à frente;
4. Longo Prazo – pode abranger de meses a anos, sendo uma janela de tempo menos precisa do que as demais, dada a incerteza natural associada à variável de interesse.

É importante salientar que o horizonte de previsão é um fator fundamental a ser considerado no processo de previsão, pois influencia diretamente a escolha dos modelos preditivos. Em face disso, do ponto de vista da previsão de irradiação solar, os modelos numéricos de previsão (NWP, do inglês *Numerical Weather Prediction*) apresentam

melhores resultados em horizontes maiores, enquanto os modelos de persistência e os modelos orientados a dados têm melhor desempenho em horizontes menores (KUMAR et al., 2020). Dessa forma, o horizonte de previsão utilizado no presente trabalho foi o de curtíssimo prazo, visto que os experimentos computacionais realizados utilizaram modelos estatísticos e modelos de aprendizado profundo para previsões horárias.

## 2.3 Componentes da Irradiação Solar

A irradiação solar possui componentes que são medidos por sensores em estações climatológicas, sendo esses valores registrados em janelas temporais (registro horário, diário, dentre outros) e utilizados nesse formato. Na sequência, esses componentes são apresentados com mais detalhes.

### 2.3.1 Irradiância Normal Direta

Irradiação Normal Direta (DNI, do inglês *Direct Normal Irradiance*) representa a quantidade de luz que incide perpendicularmente à superfície (LAGUARDA; ALONSO-SUÁREZ; ABAL, 2023). Em suma, essa irradiação é devida aos raios que incidem diretamente do Sol na superfície terrestre (PAZIKADIN et al., 2020).

### 2.3.2 Irradiância Horizontal Difusa

Irradiância Horizontal Difusa (DHI, do inglês *Diffuse Horizontal Irradiance*) é a quantidade de irradiação recebida (por unidade de área) em uma superfície. No entanto, ela se difunde a partir de partículas na atmosfera terrestre (AGBO et al., 2023).

### 2.3.3 Irradiância Horizontal Global

Irradiância Horizontal Global (GHI, do inglês *Global Horizontal Irradiance*) representa a quantidade total de irradiação de ondas curtas incidente em uma superfície horizontal (paralela ao solo), sendo, portanto, o parâmetro mais importante para o cálculo da produção de energia solar. Tanto o DHI quanto o DNI são, na verdade, componentes do GHI e, matematicamente, podem ser representados por:

$$GHI = DNI \cdot \cos(\theta) + DHI, \quad (2.1)$$

na qual o ângulo zenital solar ( $\theta$ ) refere-se ao fator de iluminação na situação hipotética da superfície terrestre (PAZIKADIN et al., 2020). Os dados utilizados nos experimentos desta dissertação apresentam os valores de  $GHI$  em uma distribuição horária.

## 2.4 Modelos de Previsão

Modelos de previsão são algoritmos que, utilizando um conjunto de dados históricos, buscam valores únicos que representam a melhor estimativa de um resultado futuro (TYRALIS; PAPACHARALAMPOUS, 2024). Ao passo que modelos de classificação aprendem padrões e regras a partir dos dados, modelos de previsão ou regressão tipicamente consideram a não estacionariedade da série temporal, as correlações entre os valores ao longo da série e/ou variáveis e as distribuições de probabilidade dos dados para inferir valores futuros (KONG et al., 2025).

### 2.4.1 Modelo de Céu Claro

Modelos de Céu Claro (CSM, do inglês *Clear-Sky Models*) estimam valores de DNI, DHI e GHI que atingem o solo na ausência de nuvens, geralmente considerando parâmetros como as coordenadas geográficas do local, o ângulo zenital solar, parâmetros atmosféricos, entre outros (ANTONANZAS-TORRES et al., 2019). Esse tipo de abordagem baseia-se em imagens de satélite para capturar propriedades das nuvens e calculá-las, geralmente com medidas importantes, como os índices solares citados, para, posteriormente, combinar ou utilizar esses dados como entrada de outros modelos (SENGUPTA et al., 2021). Um exemplo clássico de Modelo de Céu Claro é o modelo *McClear*, que simula algumas situações relacionadas aos fatores atmosféricos e à posição solar. Cada simulação produz valores de irradiação, sendo esses valores armazenados em tabelas para, em seguida, variáveis como vapor d'água e pressão ao nível do mar serem utilizadas como entradas do modelo, além do ângulo zenital solar, obtido através do algoritmo *SG2* (LEFÈVRE et al., 2013). Por fim, o modelo *McClear* localiza os parâmetros atmosféricos mais próximos dentro das tabelas previamente calculadas e faz uma interpolação para obter valores de DNI e DHI, a fim de calcular o GHI (LEFÈVRE et al., 2013)

### 2.4.2 Modelo Físico

O modelo físico ou empírico é aquele que utiliza informações geoespaciais e meteorológicas, imagens obtidas de satélites e modelos numéricos de previsão. Ou seja, trata-se de um modelo matemático que simula a dinâmica atmosférica de acordo com princípios físicos e mecânicos. Como este modelo depende de simulações numéricas, demanda abundância de recursos computacionais e, portanto, recomenda-se sua utilização em previsões de longo prazo (ALKHAYAT; MEHMOOD, 2021). Além disso, esses modelos exigem um alto custo computacional e tempo para lidar com grandes conjuntos de dados, tornando-se ineficaz prever janelas menores utilizando-os (ASSAF et al., 2023).

### 2.4.3 Modelos Estatísticos

Os modelos estatísticos buscam estabelecer uma relação entre as variáveis explicativas (entradas) e a variável de interesse (saída) (ALKHAYAT; MEHMOOD, 2021). Ao contrário das abordagens físicas, que se baseiam apenas nas condições meteorológicas e utilizam modelos matemáticos sem a coleta de dados históricos, as técnicas de previsão estatística se baseiam em dados experimentais e históricos. Portanto, modelos estatísticos estão atrelados à análise de séries temporais, centrando-se no pressuposto de que os dados do passado ocorrerão no futuro (BELMAHDI; LOUZAZNI; BOUARDI, 2020).

Dentre os modelos estatísticos, tem-se o modelo de persistência como um dos mais utilizados na previsão de irradiação solar (KUMAR et al., 2020), em que este modelo assume que os valores futuros de uma série temporal são obtidos sob a hipótese de que as condições climáticas permanecerão as mesmas entre um tempo atual  $t$  e um tempo futuro  $t + T_H$  (KLEISSL, 2013). Um exemplo de implementação de um modelo de persistência é dado por  $\hat{y}_t(t + T_H) = y(t)$ , onde  $\hat{y}_t$  é o valor observado da variável no instante atual  $t$ ,  $T_H$  é o horizonte de previsão e  $y_t(t + T_H)$  é o valor previsto para o instante futuro  $t + T_H$  (KLEISSL, 2013). Uma característica do modelo de persistência em relação ao seu desempenho é que este possui uma relação indireta com o horizonte de previsão, ou seja, quanto maior o horizonte, menor será sua precisão (PAZIKADIN et al., 2020).

Outro exemplo dentro dos modelos estatísticos é a classe de modelos chamada Modelos Autorregressivos e de Média Móvel (ARMA, inglês *Autoregressive Moving Average*), que também inclui o modelo Autorregressivo Integrado de Média Móvel (ARIMA, *Autoregressive Integrated Moving Average*) (BELMAHDI; LOUZAZNI; BOUARDI, 2020), sendo utilizados para a previsão dos recursos já mencionados, como estacionariedade, tendência, sazonalidade e correlações entre valores ao longo de uma série temporal (SZOSTEK et al., 2024).

Especificamente, o modelo *ARIMA* possui um parâmetro adicional que representa a diferenciação da série temporal. Assim, o parâmetro de integração refere-se ao número de vezes necessárias para diferenciar uma série com o intuito de que esta se torne estacionária (BELMAHDI; LOUZAZNI; BOUARDI, 2020). Esse modelo pode ser expresso por:

$$\hat{Y}_t = C + \sum_{i=1}^p \varphi_i y_{t-i} + a_t + \sum_{j=0}^q \theta_j a_{t-j}, \quad (2.2)$$

sendo  $\hat{Y}_t$ , o valor previsto no tempo  $t$ ;  $Y_t$ , o valor observado no tempo  $t$ ;  $C$  é uma constante de flexibilização;  $a_t$  é o termo de erro aleatório no tempo  $t$ ;  $\varphi_i$  e  $\theta_i$  são os parâmetros autoregressivo e de média móvel, respectivamente. Os parâmetros  $p$  e  $q$  correspondem à ordem da autorregressão e à da média móvel, respectivamente. Finalmente,  $d$  é o grau de diferenciação das séries (BELMAHDI; LOUZAZNI; BOUARDI, 2020).

Nos casos em que os dados da série temporal apresentam sazonalidade, há uma variação do modelo ARIMA chamada modelo de Média Móvel Integrada Autorregressiva Sazonal (SARIMA, *Seasonal Autoregressive Integrated Moving Average*), que contém um componente sazonal e adiciona quatro novos parâmetros. Os três primeiros parâmetros são os mesmos do modelo ARIMA, porém de ordem sazonal, e um parâmetro da defasagem sazonal dos dados (BOUZERDOUM; MELLIT; Massi Pavan, 2013). O modelo SARIMA é representado matematicamente por:

$$\Phi(B^s) \phi(B) (1 - B)^d (1 - B^s)^D X_t = \Theta(B^s) \theta(B) e_t, \quad (2.3)$$

na qual,  $B$  é a propriedade de defasagem, ou seja,  $BX_t = X_{t-1}$ ;  $s$  é o período da sazonalidade;  $\phi(B)$  e  $\theta(B)$  são os operadores autorregressivo e de média móvel não sazonais, respectivamente;  $\Phi(B^s)$  e  $\Theta(B^s)$  são os operadores autorregressivo e de média móvel sazonais;  $d$  e  $D$  são as ordens de diferenciação não sazonal e sazonal;  $X_t$  é o valor da série no instante  $t$  e  $e_t$  é o resíduo ou erro no instante  $t$ .

De acordo com Bouzerdoum, Mellit e Massi Pavan (2013), a modelagem do SARIMA ocorre em quatro etapas:

1. Identificação das variáveis para detectar uma possível estacionariedade da série temporal;
2. Busca pela combinação mais relevante de autorregressão e média móvel;
3. Validação para verificar a precisão do modelo escolhido e buscar possíveis melhorias;
4. Previsão da série temporal com base nas parametrizações estabelecidas, utilizando um intervalo de confiança.

#### 2.4.4 Modelos Baseados em Aprendizado de Máquina

Os modelos baseados em aprendizado de máquina são aqueles que podem extrair padrões dos dados, lidando bem com relações não lineares (BZDOK; KRZYWINSKI; ALTMAN, 2017). Dentro dessa definição, esses modelos são divididos principalmente em aprendizado raso e aprendizado profundo. No entanto, vale mencionar que a presente dissertação considerou algoritmos de aprendizado profundo supervisionados, ou seja, eles modelam uma função com pares de entrada e saída (dados rotulados), inferindo a saída da função modelada para novos dados (MAHESH, 2019).

##### 2.4.4.1 Aprendizado Raso

O aprendizado raso é uma classe de modelos orientados a dados e, em geral, possui duas fases: a fase de treinamento, na qual as características de entrada e os alvos são

inseridos nos modelos, e os parâmetros internos dos modelos são ajustados de acordo com algumas regras específicas; e a fase de teste, onde os alvos são previstos com base nos modelos treinados durante a fase de treinamento (XU; SONG; HAO, 2022). Tais modelos têm maior capacidade de lidar com dados mais complexos do que as abordagens baseadas em modelos estatísticos, mas menor do que os modelos de aprendizado profundo (XU; SONG; HAO, 2022), detalhados na seção 2.4.4.2.

Nos últimos anos, muitos algoritmos de Aprendizado de Máquina foram amplamente aplicados na previsão de irradiação solar (KUMARI; TOSHNIWAL, 2021a). Um deles é o algoritmo de Máquina de Vetores de Suporte (SVM, do inglês *Support Vector Machine*), que busca maximizar a distância entre as classes em um hiperplano. O algoritmo recebe como entrada um conjunto de  $N$  amostras, e cada amostra é representada por um vetor  $x_i$  de características, ou seja,  $x_i = (x_1, x_2, x_3, \dots, x_p)$ , que são os respectivos atributos. Cada amostra, por sua vez, também tem um rótulo correspondente, representado por  $\in (-1, +1)$ , que é utilizado para a classificação da amostra, na qual o algoritmo tenta separar as classes no hiperplano (CHEN; YIN; TIAN, 2024).

Outro algoritmo popular na literatura é o chamado Árvore de Decisão, conhecido por seu alto grau de interpretabilidade (frequentemente classificado como modelo *White Box*) (MATZAVELA; ALEPIS, 2021). A Árvore de Decisão opera dividindo os dados de entrada em partes menores, com base em perguntas simples (regras). O resultado é uma estrutura parecida com uma árvore, com nós e ramos, em que cada nó é um ponto de tomada de decisão, podendo ser utilizada para classificação (variáveis discretas e contínuas) e regressão (somente variáveis contínuas) (XU, 2023).

#### 2.4.4.2 Aprendizado Profundo

O aprendizado profundo refere-se a um conjunto de técnicas e modelos/algoritmos que geralmente são construídos com uma arquitetura em camadas (SAMEK et al., 2021). Ademais, os modelos profundos têm se mostrado mais adequados para a previsão de séries temporais, visto que, em grande parte das aplicações, obtiveram melhores resultados (SAMEK et al., 2021).

No escopo da previsão de séries temporais, as Redes Neurais Recorrentes (RNN, do inglês *Recurrent Neural Network*) são amplamente utilizadas, pois sua arquitetura é voltada ao tratamento de dados sequenciais (KUMARI; TOSHNIWAL, 2021a). A RNN possui uma estrutura semelhante à de redes denominadas por *Feed Forward*, com uma camada de entrada, uma camada neural oculta e uma camada neural de saída (SALEHINEJAD et al., 2018). Entretanto, diferente das redes *Feed Forward*, as RNNs possuem conexões recorrentes, em que normalmente as saídas realimentam as entradas. Dessa forma, há a atualização de um componente fundamental da arquitetura denominado estado oculto (do inglês *hidden state*), que representa a memória da rede (MIENYE; SWART; OBAIDO,

2024). Assim, a cada tempo  $t$ , a rede recebe um vetor de entrada  $x_t$  e atualiza o estado oculto  $h_t$ . A representação da atualização desse estado oculto é expressa por:

$$h_t = \sigma_h (W_{xh}x_t + W_{hh}h_{t-1} + b_h), \quad (2.4)$$

em que  $W_{xh}$  são os pesos sinápticos das camadas de entrada e neural oculta;  $W_{hh}$  são os pesos sinápticos da conexão recorrente;  $b_h$  é o vetor com valores de viés (*bias*); e  $\sigma_h$  é a função de ativação.

Após essa etapa, a saída é calculada no tempo  $t$ , utilizando o resultado da etapa anterior, conforme segue:

$$y_t = \sigma_y (W_{hy}h_t + b_y), \quad (2.5)$$

sendo  $W_{hy}$  os pesos das camadas neural oculta e de saída;  $b_y$  é o vetor com valores de viés; e  $\sigma_y$  é a função de ativação da camada de saída (MIENYE; SWART; OBAIDO, 2024).

Vários estudos têm dado atenção à eficácia de soluções de ponta a ponta que utilizam algoritmos de aprendizado profundo, como, por exemplo, as CNNs (PALETTA et al., 2023). Embora existam atualmente diversas variações de CNN, a arquitetura fundamental permanece a mesma, contendo uma camada de entrada, camadas que alternam entre processos de convolução e de agrupamento (*pooling*), uma ou mais camadas totalmente conectadas, funções de ativação e uma camada de saída (CHEN et al., 2023).

Como uma evolução das já citadas RNNs, atualmente, a LSTM tem sido uma das arquiteturas de redes neurais mais utilizadas na previsão de séries temporais (YU; CAO; ZHU, 2019; VANCSURA; TATAY; BAREITH, 2025). As redes LSTM têm o mesmo fluxo de controle que as RNNs, processando os dados sequencialmente e transmitindo informações à medida que se propagam. Contudo, ao contrário das RNNs, as operações da rede LSTM permitem que ela esqueça ou retenha informações por um longo período, estabelecendo relações de longo prazo (KHAREL; ZAREAN; KAUR, 2024). Mais especificamente, essa arquitetura foi desenvolvida para resolver dois problemas das RNNs, conhecidos como Desvanecimento do Gradiente (do inglês, *Vanishing Gradient*) e Explosão do Gradiente (do inglês, *Exploding Gradient*), que ocorriam ao propagar o resultado de uma interação e ao tentar atualizar os pesos, multiplicando-o várias vezes por valores pequenos ou grandes, o que faz esse número tender a zero ou a infinito (ZARZYCKI; ŁAWRYŃCZUK, 2023). Cada neurônio da rede LSTM é interpretado como uma célula e funciona com os seguintes componentes (ZARZYCKI; ŁAWRYŃCZUK, 2023):

- O portão de esquecimento decide qual informação proveniente da célula anterior deve ser retida ou descartada, podendo ser representado por:

$$f(k) = \sigma(W_f X_{LSTM} + R_f h(k-1) + b_f), \quad (2.6)$$

em que  $\sigma$  é a função de ativação sigmoide,  $W_f$  é a matriz de pesos sinápticos relacionada à entrada atual no instante  $k$ ,  $X_{LSTM}$  é a entrada,  $R_f$  é a matriz de pesos sinápticos relacionada ao estado oculto anterior (memória de curto prazo),  $h(k-1)$  é o estado oculto anterior ao momento  $k$  e  $b_f$  representa o vetor com valores de viés;

- O portão de entrada controla quais novas informações devem ser adicionadas à célula no instante atual. Esta etapa seleciona valores do estado oculto anterior e da entrada atual, para fins de atualização, atuando como um filtro e decidindo, a partir dos dados de entrada e do histórico, o que é relevante o suficiente para ser incorporado à memória de longo prazo da rede. A seguinte equação define essa etapa:

$$i(k) = \sigma(W_i X_{LSTM} + R_i h(k-1) + b_i), \quad (2.7)$$

sendo  $\sigma$  a função de ativação sigmoide,  $W_i$  é a matriz de pesos sinápticos relacionada à entrada atual no instante  $k$ ,  $R_i$  é a matriz de pesos sinápticos relacionada ao estado oculto anterior (memória de curto prazo) e  $b_i$  representa o vetor com valores de viés;

- O candidato ao estado da célula prepara informações com potencial de entrar na memória principal (estado da célula); ou seja, o resultado contém os novos valores que poderão ser adicionados à memória, dependendo do que o portão de entrada permitir. A representação matemática dessa etapa pode ser descrita como:

$$g(k) = \tanh(W_g X_{LSTM} + R_g h(k-1) + b_g), \quad (2.8)$$

em que  $\tanh$  é a função de ativação tangente hiperbólica, limitada ao intervalo  $[-1, 1]$ ,  $W_g$  é a matriz de pesos sinápticos relacionada à entrada atual no instante  $k$ ,  $R_g$  é a matriz de pesos sinápticos relacionada ao estado oculto anterior (memória de curto prazo) e  $b_g$  representa o vetor com valores de viés.

Como alternativa aos experimentos computacionais realizados nessa dissertação, a xLSTM, recentemente desenvolvida, foi empregada. Ela representa um aprimoramento do modelo original em alguns aspectos, como, por exemplo, a melhoria em relação à memória e ao seu paralelismo (BECK et al., 2024). Dentro dessa nova arquitetura, existem duas sub-redes (sLSTM e mLSTM) que são combinadas em blocos, formando as células da xLSTM e empilhando essas células para formar redes profundas voltadas principalmente à modelagem de dados sequenciais. A sub-rede *sLSTM* traz as chamadas portas exponenciais que, em vez das portas tradicionais (*input/forget/output*), que multiplicam

linearmente, contêm agora operações exponenciais, proporcionando ao modelo maior dinâmica de escala e mais flexibilidade para controlar o fluxo de informação ao abrir ou fechar as portas.

Outro recurso da sub-rede *sLSTM* é o uso da memória escalar por célula, com uma combinação de memórias entre células por meio de conexões recorrentes. No LSTM clássico, o estado da célula  $C_t$  é atualizado misturando a memória antiga  $C_{t-1}$ , controlada pela porta de esquecimento, e a nova informação  $\tilde{C}_t$  candidata, controlada pela porta de entrada. No *sLSTM*, cada célula tem uma memória escalar. Porém, essas memórias não ficam isoladas e podem ser misturadas entre diferentes células (BECK et al., 2024). Em relação à sub-rede *mLSTM*, a nova abordagem traz uma memória matricial, e não mais em valor escalar ou vetorial, permitindo a combinação de memórias em blocos (matrizes), além de aumentar a capacidade da memória em si. No modelo tradicional, a memória é representada por um vetor  $c_t \in \mathbb{R}^d$  e, na nova abordagem, é representada por  $M_t \in \mathbb{R}^{d \times d}$ . A recuperação da memória é realizada por meio de uma multiplicação de matrizes. No tempo  $t$ , a rede irá armazenar um par de vetores, com chave e valor, pelo mesmo mecanismo dos *transformers*, no formato  $k_t \in \mathbb{R}^d$  para a chave,  $v_t \in \mathbb{R}^d$  para o valor, e para a recuperação posterior dessa informação, o vetor de consulta terá o formato  $q_{t+\tau} \in \mathbb{R}^d$  (BECK et al., 2024).

Na literatura, os modelos CNN e LSTM geralmente tem desempenhos superiores aos modelos de aprendizado raso, tanto em um individualmente quanto em uma hibridização dos modelos (HASSAN; KUMAR, 2025). Portanto, no presente trabalho, o modelo LSTM foi utilizado em ambas as abordagens (individual e híbrida), juntamente com o modelo CNN. Além disso, a xLSTM também foi considerada nos testes, por se tratar de uma melhoria recente da LSTM original.

## 2.5 Transferência de Aprendizado

O conceito de transferência de aprendizado refere-se à capacidade de reutilizar modelos de aprendizado de máquina pré-treinados para adaptar-se (generalizar) a tarefas de contextos distintos ou para realizar ajuste fino em um mesmo contexto. A aprendizagem por transferência visa treinar modelos de forma robusta para um domínio de destino, aproveitando o conhecimento de um domínio de origem (WANG et al., 2020). Como os domínios de origem e de destino são geralmente de distribuições diferentes, os métodos existentes concentram-se principalmente na adaptação das distribuições marginais ou condicionais entre domínios (WANG et al., 2020). A Figura 1 ilustra o funcionamento básico da estratégia de Transferência de Aprendizado, onde um Modelo A é treinado utilizando dados de um Domínio A (origem), que são dados extraídos de uma Localidade A. Em seguida, o Modelo A é utilizado para previsão em uma Localidade B, com pequenos ajustes

em relação aos dados de um Domínio B (destino), dependendo da abordagem.

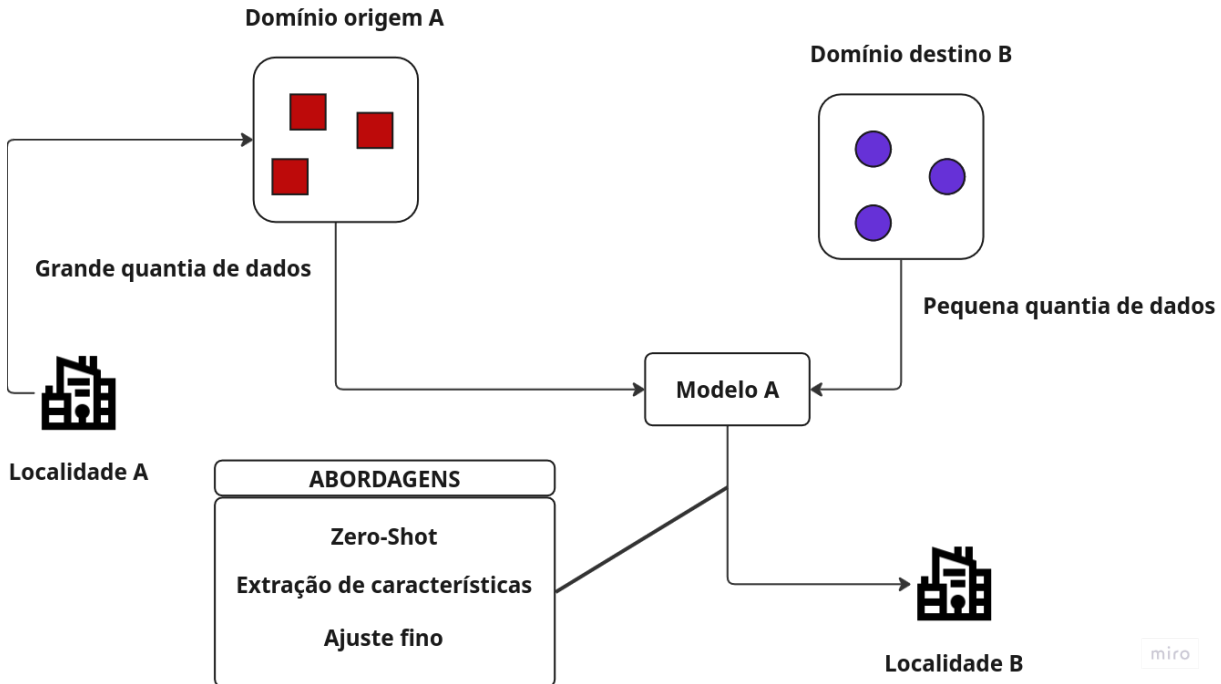


Figura 1 – Esquema simplificado do funcionamento do Transfer Learning. Adaptado de (GAO et al., 2024).

A definição de domínio representa um conjunto de dados  $X$  de  $N$  dimensões, com sua própria distribuição  $P$ . A representação matemática pode ser descrita por  $\mathcal{D} = \{X, P(X)\}$ , na qual  $X$  representa o espaço que contém as características analisadas, ou seja, os dados em si; e  $P(X)$  representa a distribuição de probabilidade desses dados (GAO et al., 2024).

No caso do presente trabalho, cada domínio representa uma estação climatológica diferente utilizada no treinamento do modelo para uso posterior em outro domínio (de uma zona bioclimática diferente, por exemplo).

Outro conceito fundamental é o de tarefa, representado por  $\tau = Y, f(\cdot)$ , na qual  $Y$  são os valores-alvo da predição e  $f(\cdot)$  é a função objetivo que visa estabelecer uma correlação entre as características de entrada e os valores-alvo (GAO et al., 2024). Neste estudo, a tarefa (na origem e no destino) é prever a irradiação solar.

De forma geral, o processo de transferência de aprendizado pode ser descrito formalmente como um conjunto de domínios de origem  $D_{s_i}$ , que possui, cada um, um conjunto de observações  $o^s$  e tarefas de aprendizagem correspondentes  $\tau_{s_i}$  (ZHAO et al., 2024), ou seja:

$$(D_{s_i}, \tau_{s_i}) \mid i = 1, 2, \dots, o^s. \quad (2.9)$$

Em relação às observações  $o^t$  e às tarefas de aprendizado  $\tau_t$  no domínio de des-

tino, a técnica de transferência de aprendizagem visa melhorar uma função preditiva  $f_t$  no domínio  $D_t$ , com base no aprendizado anterior do domínio  $D_{s_i}$  em uma tarefa de aprendizagem específica  $\tau_{s_i}$ , de forma que  $D_{s_i} \neq D_t$  ou  $\tau_{s_i} \neq \tau_t$  (ZHAO et al., 2024).

No Brasil, diferentes regiões podem ter um mesmo padrão para o GHI diário, com algumas variações e especificidades locais. Diante disso, a técnica de Transferência de Aprendizado pode ser uma ferramenta útil, pois os modelos de previsão solar dependem de inúmeros fatores, como, por exemplo, as condições climáticas locais, a sazonalidade dos dados e as demais condições regionais (GOSWAMI et al., 2022). Dessa forma, pode ser necessário um ou mais modelos para uma única estação solar. Entretanto, não é somente a construção do modelo que exigirá esforços, mas também os ajustes, os testes, o *deploy* (implantação) e a própria manutenção (GOSWAMI et al., 2022).

Nesse sentido, a transferência de aprendizado pode se aplicar ao caso da previsão de irradiação solar para estações diferentes em que os domínios são semelhantes, ao passo que, para um resultado satisfatório utilizando-se a técnica, são necessários alguns critérios: 1) deve existir semelhança entre as tarefas de aprendizagem nos dois domínios; 2) da mesma forma, deve haver uma razoável semelhança entre as distribuições nos dados dos domínios de origem e destino; 3) o modelo utilizado deve estar adequadamente parametrizado (DESPOTOVIC et al., 2024). No presente trabalho, as tarefas são as mesmas entre origem e destino; e as medidas estatísticas entre as estações utilizadas, de distintas zonas bioclimáticas, apesar de diferentes (o que é esperado), não possuem um desvio elevado (média de aproximadamente 1295.0 e desvio padrão 211.19 para dados do período de verão). Maiores detalhes podem ser consultados no Apêndice A.

### 2.5.1 Estratégias de Transferência de Aprendizado

As estratégias existentes em transferência de aprendizado são variadas. Existem categorizações, por exemplo, por disponibilidade dos dados dentro de um domínio, como as abordagens Indutiva, Transdutiva e Não Supervisionada, e por similaridade na representação de características, como as abordagens Homogênea e Heterogênea (GHOLIZADE et al., 2025). Dentro da categoria Indutiva, na qual os dados utilizados nesta pesquisa são representados, destacam-se principalmente as seguintes abordagens:

- *Zero-shot* – é uma abordagem que consiste em utilizar o modelo pré-treinado para a generalização de um domínio de destino sem nunca ter visto nenhum dado rotulado desse domínio. Nesse sentido, o modelo é submetido apenas à validação com dados do domínio de destino. Em resumo, o modelo pré-treinado é utilizado de forma direta, sem extração de características e/ou ajuste fino (MISHRA et al., 2025). Tal estratégia apresenta algumas vantagens, como a ausência da necessidade de dados do domínio de destino para continuar o treinamento do modelo pré-treinado.

Entretanto, se o *zero-shot* não obtiver boa generalização para o domínio de destino, há a necessidade de considerar estratégias de extração de características ou de ajuste fino;

- Extração de características – o modelo extrai representações comuns do domínio de origem que sejam úteis para o domínio de destino, preservando as camadas mais fundamentais do modelo pré-treinado e substituindo somente as camadas totalmente conectadas. É uma abordagem para situações nas quais o domínio de destino possui uma distribuição semelhante, porém diferente do domínio de origem (TANG et al., 2022). Um ponto negativo em relação a essa estratégia é que ela pode apresentar baixo desempenho em termos de generalização, pois ao manter as primeiras camadas do modelo intactas, os pesos permanecem fixos e não se adaptam bem aos requisitos do domínio de destino. Por outro lado, a estratégia de extração de características funciona bem em domínios com distribuições semelhantes e, ao manter os pesos do modelo, estes são impedidos de se ajustar excessivamente à nova tarefa (evitando o *overfitting*), preservando as características aprendidas no treinamento original (RAJ; QIAN; IBRAHIM, 2024).
- Ajuste fino – é geralmente utilizado em situações em que o domínio de destino apresenta uma distribuição razoavelmente diferente daquela do domínio de origem. O modelo pré-treinado é transferido para o domínio de destino, sem o congelamento de nenhuma camada, e a abordagem de ajuste fino aproveita esse conhecimento do domínio de origem e ajusta marginalmente os parâmetros para se adequar ao novo domínio, com uma taxa de aprendizagem menor durante o treinamento (TANG et al., 2022). Como vantagem, tal estratégia contribui para o refinamento das características aprendidas nas tarefas iniciais, a fim de alinhá-las aos objetivos específicos do classificador subsequente, alcançando, assim, um desempenho melhor. Apesar disso, o modelo pode apresentar um problema comum em modelos de aprendizado de máquina que é o *gradient vanishing*, em que o conhecimento adquirido previamente se perde durante o aprendizado da nova tarefa (RAJ; QIAN; IBRAHIM, 2024).

As abordagens supracitadas foram avaliadas durante os testes desenvolvidos neste trabalho, e seus resultados foram comparados, conforme apresentado no Capítulo 5.

## 2.6 Métricas de Avaliação de Desempenho

Como métricas de avaliação dos resultados, a literatura destaca algumas das mais utilizadas, como o MSE e sua extensão, o RMSE. Ambas métricas calculam a distância do valor previsto em relação ao seu valor real e a dispersão dos valores previstos entre

si, sendo ambas sempre positivas e quanto menores, melhor o desempenho do modelo avaliado. Uma característica do RMSE é que seu valor está na mesma unidade que o valor real, tornando-o mais compreensível em comparação com o MSE. Ambas as métricas também penalizam erros maiores em relação às demais métricas (TEMPORAL-NETO et al., 2025). Matematicamente, as métricas MSE e RMSE podem ser representadas por:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2, \quad (2.10)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2}, \quad (2.11)$$

sendo  $y_i$  o valor real; e  $\hat{y}_i$  o valor da previsto.

Outra métrica muito utilizada na literatura é o MAE, que representa o valor médio da distância absoluta entre o valor previsto e o valor real (ACIKGOZ, 2022). A métrica é descrita, conforme segue:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i|, \quad (2.12)$$

Uma quarta métrica comumente utilizada na literatura é o  $R^2$ , ou *R-square*, que representa a proporção de variância do atributo-alvo (variável dependente) predita a partir das variáveis independentes (CHICCO; WARRENS; JURMAN, 2021). A métrica avalia se o modelo preditivo está ajustado e indica o nível de desvio da previsão, que varia de 0 a 1 (YU; CAO; ZHU, 2019). Portanto, se  $R^2 = 0$ , diz-se que o modelo não está bem ajustado. Caso contrário, se  $R^2 = 1$ , o modelo não tem erros, conforme apresentado na sequência:

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2}, \quad (2.13)$$

## 3 Levantamento Bibliográfico

Este capítulo destina-se a apresentar o estado da arte das técnicas de previsão de irradiação solar, destacando os principais modelos preditivos de aprendizado de máquina empregados na literatura correlata.

### 3.1 Abordagens Baseadas em Modelos de Aprendizado Profundo

O estudo de [Mishra e Palanisamy \(2018\)](#) trouxe um processo de ponta a ponta utilizando uma RNN para prever o Índice de Céu Claro ( $K_t$ ), parâmetro utilizado para quantificar e prever as mudanças na irradiação solar devido às condições atmosféricas, como as nuvens. Os dados são coletados entre 2010 e 2012, provenientes de 7 locais distintos dos Estados Unidos (Colorado, Nevada, Montana, Dakota do Sul, Illinois, Mississippi e Pensilvânia). O trabalho apresenta duas abordagens: uma delas focada na previsão de horizonte único (1 a 4 horas) e a outra baseada na previsão de múltiplos horizontes (que considera os mesmos quatro horizontes), mas desenvolve um único modelo responsável por prever todos eles ao mesmo tempo. Os resultados foram adequados em relação à métrica de RMSE, que apresentou valores menores em comparação ao estudo de [Florita et al. \(2017\)](#), por exemplo. Em relação a abordagem com um único horizonte para o modelo RNN proposto, o valor médio de RMSE entre os 4 horizontes e as 7 localidades foi de 26.31; enquanto na segunda abordagem, os resultados demonstraram que o modelo, para múltiplos horizontes, obteve menores valores de RMSE quando comparados à literatura, mesmo para modelos baseados em previsão de horizonte único, com o valor médio de 16.85, ambos os valores sendo em unidade do Índice de Céu Claro. Ainda dentro do estudo, os autores reportaram *insights* significativos sobre o modelo, como a já demonstrada característica das conexões das RNNs atuarem como módulos de memória, permitindo boa adesão a dados em formato de série temporal ([MISHRA; PALANISAMY, 2018](#)). Entretanto, um ponto importante que os autores não abordaram é o problema do *vanishing gradient* e do *exploding gradient*, que ocorrem quando o gradiente de atualização dos pesos da rede se torna muito pequeno ou muito grande, respectivamente.

O trabalho de [Yu, Cao e Zhu \(2019\)](#) apresentou um estudo empregando RNN e LSTM para a previsão de irradiação solar sob condições climáticas adversas, como mudanças climáticas abruptas e características sazonais distintas. Nos resultados de previsão, o coeficiente  $R^2$  da LSTM em dias nublados e dias mistos foi superior a 0,90, enquanto o coeficiente  $R^2$  da RNN foi variou entre 0,70 e 0,79. A partir dos resultados da previsão diária, todos os coeficientes  $R^2$  em dias nublados foram de aproximadamente 0,85. No entanto, a LSTM ainda mostrou-se mais eficaz do que a RNN, em geral, e mais precisa

do que outros modelos. O trabalho em questão utilizou diversas variáveis, além da irradiação solar, para a previsão, o que torna os modelos mais robustos em dias nublados e sob condições adversas. Vale destacar que a metodologia proposta, previamente à previsão, realiza uma classificação do dia céu limpo ou não para aumentar a precisão dos modelos de previsão.

Haider et al. (2022) avaliaram distintos modelos preditivos baseados em aprendizado de máquina, como *Multilayer Perceptron*, CNN e LSTM. Para comparar seus desempenhos com modelos estatísticos, consideraram-se o SARIMAX e o Prophet. Para horizontes de previsão menores, os modelos de aprendizado de máquina apresentaram melhores desempenhos, enquanto, para um prazo maior, os modelos estatísticos mostraram-se mais adequados. Os dados compreendem o período de novembro/2018 a agosto/2019 para Atlanta, New York e Hawaii. Para previsões horárias, o LSTM apresentou os melhores resultados, com RMSE variando entre  $41,37 W/m^2$  e  $66,69 W/m^2$ , MAE entre  $30,19 W/m^2$  e  $46,04 W/m^2$  e  $R^2$  entre 0,92 e 0,99.

Singla, Duhan e Saroha (2022) propuseram a utilização de redes BiLSTM e da Transformada *Wavelet* (WT, do inglês *Wavelet Transform*), em que a WT decompõe a série temporal, gerando sub-séries (denominadas por detalhes e aproximações), cada uma representando padrões de baixa e alta frequência. Em seguida, para cada sub-série gerada, foi utilizado um modelo BiLSTM, e cada saída foi somada para se obter o resultado final da previsão. O modelo proposto foi comparado a outros modelos de aprendizado profundo, tais como LSTM e GRU, bem como foram consideradas outras famílias de Wavelets em conjunto com o BiLSTM. Os resultados mostram que o modelo proposto supera os demais com RMSE médio anual mínimo de  $45,61 W/m^2$  e  $R^2$  de 0,94. O estudo, apesar de conter cenários de teste diversos e de utilizar modelos diferentes, deixa de explorar abordagens híbridas, um cenário já consolidado à época da publicação do artigo.

Demir e Citakoglu (2023) trouxeram um experimento de estimativa mensal da irradiação solar usando cinco algoritmos baseados em aprendizado de máquina, a saber: regressão por vetor de suporte (SVR, do inglês *Support Vector Regression*), LSTM, regressão por processo gaussiano (GPR, do inglês *Gaussian Process Regression*), máquinas de aprendizado extremo (ELM, do inglês *Extreme Learning Machine*) e K vizinhos mais próximos (KNN, do inglês *k-nearest neighbors*), sendo que cada um desses algoritmos foi treinado em todas as 163 estações abordadas no estudo. Na primeira etapa de modelagem, foi aplicada a análise do fator de inflação da variância, utilizada para identificar multicolinearidade e, conseqüentemente, remover variáveis que poderiam reduzir o desempenho do modelo. Em seguida, as variáveis restantes foram utilizadas na modelagem por meio da correlação com os valores de irradiação solar. O algoritmo LSTM foi o que apresentou os melhores desempenhos, com valores médios de RMSE entre 2,297 e 4,422, e os valores médios de  $R^2$  de 0,895. Os pesquisadores utilizaram 163 estações distintas para a coleta

de dados na Turquia, abrangendo todo o território. Sua variedade climática contribui para uma maior diversidade de dados, o que pode promover uma capacidade de generalização ainda maior do modelo utilizado no momento do treinamento.

A WT também foi utilizada por [Alizamir et al. \(2023\)](#), os quais propuseram o uso combinado com redes neurais profundas. Dentre os modelos preditivos, a LSTM combinada à WT obteve o melhor desempenho, com um  $R^2$  que variou de 0,91 a 0,97 para cidades do estado de Illinois, nos Estados Unidos.

[Balal et al. \(2023\)](#) apresentam uma análise de vários algoritmos de aprendizado de máquina, incluindo o *Random Forest* para regressão e o LSTM, que obtiveram os melhores resultados na previsão da geração de energia fotovoltaica na cidade de Lubbock, no Texas, para horizontes de 1 dia, 4 dias, 10 dias e 1 mês. Os resultados, em termos de  $R^2$ , obtidos pela *Random Forest* e pelo LSTM foram de 0,977 e 0,975, respectivamente.

O estudo de [Bektaş e Altaş \(2024\)](#) traz uma abordagem com dois modelos para previsão de 24 horas, cada um utilizando uma entrada diferente. O primeiro usa apenas o GHI e dados climáticos, enquanto o segundo adiciona a velocidade do vento, a temperatura e a umidade relativa do ar para avaliar o impacto dessas variáveis nos resultados. A comparação também se aplica à utilização de BiLSTM, LSTM e RNN. Os dados analisados foram decompostos com a WT para serem utilizados como entrada em todos os modelos. Os resultados mostraram que, apesar dos modelos apresentarem desempenhos muito próximos, o tempo de treinamento da RNN é maior devido à parametrização adicional. Os resultados também demonstraram que o modelo BiLSTM apresentou desempenho superior ao dos demais, além de obter resultados melhores quando comparado a outros estudos.

No estudo de [Guen e Thome \(2020\)](#), os autores propuseram a combinação de cálculos físicos e redes neurais artificiais para previsão em um horizonte de cinco minutos. A rede neural em questão foi a PhyDNet, uma arquitetura profunda que utiliza equações diferenciais parciais para predição de vídeo. Nesse sentido, a rede neural aprendeu o comportamento das equações diferenciais que descrevem as dinâmicas das nuvens (presentes em modelos NWP), introduzindo conhecimento *a priori* para melhorar o desempenho do modelo preditivo.

Ainda, é cabível destacar a revisão da literatura apresentada por [Chandel et al. \(2023\)](#), a qual analisou e comparou técnicas de aprendizado profundo no âmbito da previsão de geração de energia fotovoltaica. A revisão destacou algumas lacunas de pesquisa, como a questão do baixo potencial de generalização de alguns modelos, especialmente ao buscar essa generalização para distintas regiões geográficas. Outro ponto levantado na revisão foi a necessidade de métodos de classificação climática e previsão probabilística do movimento das nuvens. Além disso, o trabalho ressaltou a necessidade de extrair características espaciais e temporais dos dados para prever a irradiação solar. Por fim, o

trabalho apontou a arquitetura LSTM como o modelo que obteve os melhores resultados entre as arquiteturas de aprendizado profundo testadas, independentemente do horizonte de previsão. No entanto, destacou também a GRU (*Gated Recurrent Unit*), pois ela possui menos parâmetros ajustáveis que a LSTM, além de, normalmente, apresentar melhores resultados em conjuntos de dados limitados (menores).

## 3.2 Abordagens Baseadas em Modelos Híbridos

Na revisão de [Rajagukguk, Ramadhan e Lee \(2020\)](#), os autores propuseram um experimento comparando abordagens baseadas em técnicas de aprendizado de máquina com e sem hibridização, utilizando os seguintes modelos: RNN, LSTM, GRU e o modelo híbrido de CNN–LSTM. Os parâmetros utilizados para comparação foram a precisão dos modelos, os dados de entrada, o horizonte de previsão, as diferenças climáticas e o tempo de treinamento. Utilizando a métrica RMSE, os resultados demonstraram que o modelo híbrido em questão supera os três modelos abordados, embora exija um maior tempo de treinamento. Além disso, existem pesquisas anteriores que abordaram modelos baseados em BiLSTM (variação bidirecional do LSTM), os quais obtiveram bons resultados ([WANG et al., 2020](#); [XU et al., 2019](#)).

Na pesquisa de [Zang et al. \(2020\)](#), os autores propuseram um modelo híbrido utilizando uma CNN-LSTM em paralelo, ou seja, as entradas com características espaciais e temporais são aplicadas separadamente às redes CNN e LSTM. A CNN é utilizada para extrair características espaciais de uma matriz bidimensional composta por parâmetros meteorológicos associados a um local-alvo e seus locais vizinhos, criando correlações entre as mesmas características de lugares diferentes e características diferentes de um mesmo local. Em seguida, aplica-se uma LSTM para extrair características temporais dos dados históricos das séries temporais de irradiância solar associadas ao local-alvo. Finalmente, as correlações espaço-temporais são mescladas para prever a irradiação solar para o horizonte de hora a frente. Os resultados demonstraram que a abordagem híbrida obteve vantagens em relação a outros modelos, inclusive para previsões anuais ao considerar cinco locais-alvo (Dallas, San Jacinto, Zapata, Moore e Lamb). Os MAEs obtidos variaram de  $37,20 W/m^2$  a  $52,00 W/m^2$ . A metodologia proposta utilizou dados de três zonas bioclimáticas, o que fornece um potencial de generalização das redes ao serem validadas em locais diferentes, uma vez que, ao reconstruir as entradas com as características espaciais, são utilizados os dados de todas as regiões mencionadas no estudo.

O trabalho de [Kumari e Toshniwal \(2021a\)](#) trouxe modelos híbridos de redes neurais profundas (LSTM e CNN) para a previsão de GHI. A entrada das redes LSTM foi composta pela série temporal do GHI, enquanto a entrada da CNN é a saída da LSTM adicionada de características climáticas. A rede CNN extrai as características espaciais

da entrada e modela a relação entre os dados de entrada e o GHI. Por fim, duas camadas totalmente conectadas são adicionadas ao final do modelo LSTM–CNN. O modelo híbrido baseado em LSTM–CNN apresentou melhorias de 3,62% a 37,37% em relação a modelos profundos como RNN, LSTM e CNN.

A revisão de [Assaf et al. \(2023\)](#) apresentou uma visão geral sobre o uso de modelos de previsão de GHI baseados em aprendizado profundo. Dentro deste conjunto de modelos, incluem-se LSTM, GRU, RNN, CNN e GAN, sendo os modelos híbridos, como o modelo CNN-LSTM, aqueles que os autores consideram como sendo o atual estado da arte, demonstrando  $R^2 > 0,90$  e melhorias de mais de 10% para períodos diários, mensais e sazonais quando comparados a outros modelos.

A pesquisa de [Kumari e Toshniwal \(2021b\)](#) trouxe uma abordagem utilizando o modelo de Aumento Extremo do Gradiente (XGBoost, do inglês *Extreme Gradient Boosting*) e redes neurais profundas para a previsão de GHI horária, empregando um subconjunto de características que inclui temperatura, índice de céu claro, umidade relativa e a hora do dia. O modelo XGBoost foi utilizado para minimizar o ruído, a variância e o viés dos dados, enquanto a rede neural profunda foi utilizada para lidar com o comportamento não linear. Como resultado, os modelos combinados apresentaram um desempenho consistente para todos os locais considerados, independentemente de suas condições climáticas e variações sazonais. O modelo proposto superou os métodos de regressão clássicos e reduziu o erro de previsão em 40%, em comparação a métodos de persistência, SVM e *Random Forest*. Como limitações, o modelo demanda um tempo de treinamento maior em relação aos demais, além de apresentar uma maior complexidade.

O trabalho de [Kumari e Toshniwal \(2021c\)](#) propôs um modelo híbrido baseado em LSTM–CNN para previsão de GHI utilizando dados climatológicos de 23 localidades dos Estados Unidos. O desempenho da abordagem foi avaliado para diferentes estações do ano e apresentou um desempenho superior a outros modelos analisados. A LSTM faz uma primeira predição do GHI e sua saída é usada como entrada da CNN que, por sua vez, utiliza esses dados em conjunto com os demais dados climatológicos para formar uma matriz de correlação entre localidades vizinhas e a localidade alvo. O modelo proposto apresentou melhorias entre 37% e 45% em relação a modelos baseados em SVM, LSTM e CNN. Durante a etapa de validação, foi empregada uma clusterização do índice de céu claro por meio do *K-Means*, garantindo maior assertividade do modelo LSTM–CNN para cenários de céu claro, parcialmente nublado e nublado.

[Rai, Shrivastava e Jana \(2020\)](#) apresentaram um processo de previsão de GHI em um horizonte de dois meses, utilizando o modelo híbrido CNN–BiLSTM. A CNN foi utilizada para extrair as características das séries temporais, enquanto a BiLSTM explorou as dependências não lineares das séries temporais. O modelo híbrido teve melhor desempenho do que os modelos LSTM, GRU, CNN–LSTM e GRU–LSTM. Essa abordagem utilizou

dados de zonas climáticas diferentes, com as mesmas latitudes, mas longitudes diferentes, contribuindo para a generalização dos modelos propostos. Além disso, trouxe uma arquitetura pouco explorada, que é a utilização de redes CNN para extrair características não climáticas, mas sim dos próprios valores de GHI, o que é incomum na literatura. Entretanto, o trabalho explorou pouco a relação com variáveis como temperatura, vento e umidade.

Rajaprasad e Mukkamala (2023) propuseram uma abordagem baseada em um modelo híbrido de CNN-BiLSTM para prever GHI. Ambas as redes foram conectadas de forma direta, sendo a CNN utilizada para extrair características não lineares dos dados, enquanto a BiLSTM foi empregada na previsão de GHI um minuto à frente. O modelo proposto foi comparado com outras redes profundas, como a LSTM e sua variação bidirecional, a BiLSTM, e obteve os melhores MAEs e  $R^2$ , de 46,84 (GHI como unidade) e 0,95, respectivamente. O estudo apresentou uma métrica de  $R^2$  muito precisa para o modelo híbrido, o que, por vezes, pode indicar sobreajuste. Ao mesmo tempo, as camadas utilizadas no modelo híbrido não utilizam nenhum tipo de recurso para evitar esse cenário.

### 3.3 Abordagens Baseadas em Transferência de Aprendizado

Em Mylonas (2023), o autor utilizou a transferência de aprendizado em conjunto com CNNs para a previsão da energia elétrica gerada 24 horas à frente. A princípio, as CNNs foram treinadas em um grande conjunto de dados de séries temporais de vários países da Europa. Em seguida, as previsões foram realizadas para os países do primeiro conjunto de dados e, por fim, para países com poucos dados. Para o segundo conjunto de dados, houve transferência de aprendizado por meio da abordagem de ajuste fino. Os resultados mostraram que as CNNs, quando treinadas em um grande conjunto de dados de qualidade e com uma variedade de exemplos de vários países, são capazes de generalizar e transferir seu conhecimento para melhorar a precisão das previsões, tanto para os países com muitos dados quanto para os países com poucos dados disponíveis. O experimento gerou três CNNs pré-treinadas, que superaram consideravelmente os *baselines* do estudo ao prever os resultados para os países nos quais foram treinadas, com o erro sendo reduzido em até três vezes em alguns casos. Além disso, as CNNs treinadas em diversos países tiveram um desempenho melhor do que aquelas com exatamente as mesmas arquiteturas e que foram treinadas individualmente em cada país.

O estudo de Gao et al. (2024) apresentou o uso de uma estratégia de transferência de aprendizado para dados de GHI das 168 horas anteriores de duas cidades no Japão: Tóquio (origem) e Okinawa (destino). Assim, realizou-se a previsão hora a frente, com o intuito de realizar um experimento com três estratégias de transferência de aprendizado: (i) utilizando a adaptação entre os domínios de origem e destino, feita por meio de uma

GAN; (ii) utilizando o ajuste fino do modelo ao aplicá-lo no domínio de destino; e (iii) utilizando ambas as abordagens previamente mencionadas. Também houve um modelo de *baseline* que, assim como nas abordagens citadas, utilizou o modelo GRU para fazer a previsão ao domínio de destino, porém sem adaptação ou ajuste fino. Os resultados do experimento indicaram que o uso da estratégia baseadas em adaptação de domínios e ajuste fino dos modelos superou o desempenho de qualquer um dos modelos na maioria dos cenários propostos, alcançando uma precisão de previsão de até 98,89%. O trabalho abordou diversas estratégias de transferência de aprendizado, que envolveram o uso direto do modelo na localidade de destino, treinado previamente na localidade de origem, ajuste fino, adaptação de domínio e a combinação desses cenários.

O trabalho de [Goswami et al. \(2022\)](#) trouxe uma abordagem de transferência de aprendizado para evitar a construção de múltiplos modelos para vários locais e suas variações climáticas. O estudo foi conduzido com dados de seis estações, em três zonas climáticas e duas estações do ano na Índia. O modelo de Rede Neural de Unidade Recorrente com Porta Bidirecional (BGRUNN, do inglês *Bidirectional Gated Recurrent Unit Neural Network*) foi utilizado, aplicando-se a técnica de transferência de aprendizado ao treiná-lo em uma das estações para realizar previsão de curto prazo (1 hora e 40 minutos) nos demais locais. O modelo em questão obteve melhor precisão de previsão em comparação com modelos específicos para esses locais e apresentou uma variância menor. O modelo também foi configurado com 39,6% menos parâmetros e precisou de 76,1% menos tempo de treinamento.

Em [Miraftabzadeh et al. \(2023\)](#), os autores utilizaram uma estratégia de transferência de aprendizado para que modelos já treinados em antigas usinas fotovoltaicas pudessem ser treinados com dados de novas usinas, dados estes em quantidade reduzida. A previsão em um horizonte de 24 horas foi feita por nove modelos distintos pré-selecionados, com o melhor resultado obtido por uma LSTM, a qual obteve MSE de 0,55 e um valor de Erro Percentual Absoluto Médio Ponderado (WMAPE, do inglês *Weighted Mean Absolute Percentage Error*) de 47,07%, enquanto o modelo LSTM com transferência de aprendizado melhora a precisão da previsão para MSE de 0,168 e WMAPE de 32,04%.

A pesquisa de [Sarmas et al. \(2022\)](#) utiliza um modelo LSTM com três estratégias de transferência de aprendizado para fornecer previsões precisas da produção de usinas solares. A transferência de aprendizado nesse trabalho foi utilizada tanto para a inicialização dos pesos do modelo LSTM quanto para a extração de características, usando diferentes abordagens de congelamento. Comparando-se com modelos convencionais, os resultados indicam que as estratégias de transferência de aprendizado os superam significativamente, com uma melhoria de 12,6% em termos de RMSE. Apesar de apresentar a metodologia geral de forma clara, os autores não abordam com detalhes, por exemplo, as configurações da rede LSTM utilizada, como número de camadas, número de neurônios

em cada camada, dentre outras.

Em [Tang et al. \(2022\)](#), os autores trouxeram uma abordagem em três etapas para a previsão de energia fotovoltaica: (i) mecanismo de atenção utilizando LSTM reconstrói os dados de entrada, evidenciando características mais importantes; (ii) estrutura paralela composta por uma CNN densa e uma BiLSTM é responsável por receber os dados reconstruídos da etapa anterior e extrair as características espaciais (dados climáticos) e as características temporais (valores de GHI) de forma independente; e (iii) saída da etapa anterior é utilizada como entrada em uma rede neural que mapeia e estabelece relações não lineares entre as características espaciais e temporais. Neste estudo, empregou-se uma transferência de aprendizado por ajuste fino dos parâmetros do modelo, no entanto, ela foi aplicada em cada uma das etapas do processo para verificar o comportamento e a importância de cada uma. Os resultados demonstraram que, ao congelar os parâmetros da etapa de reconstrução da entrada e realizar o ajuste fino, obteve-se os melhores resultados.

A pesquisa de [Nie et al. \(2024\)](#) realizou uma comparação de desempenho entre três modelos utilizando imagens do céu: (i) modelos locais treinados individualmente em um único conjunto de dados local; (ii) modelos globais treinados em conjunto, fazendo a fusão de vários conjuntos de dados; e (iii) modelos ajustados localmente e treinados por meio da transferência de aprendizado com base em um modelo pré-treinado. Os resultados sugeriram que, com a estratégia de treinamento local, os modelos funcionam bem quando aplicados localmente, mas apresentam erros significativos quando aplicados fora do local. Os modelos globais, com uma normalização adequada dos alvos de previsão, podem se sair bem em locais individuais, caso esforços adicionais de treinamento sejam empregados. Por fim, os modelos pré-treinados em um conjunto de dados de origem grande e diversificado e transferidos para um local de destino tiveram desempenho superior em relação às outras duas estratégias.

O trabalho de [Shanker et al. \(2025\)](#) propõe uma comparação entre o modelo BGRU e o modelo de Aprendizagem por Transferência com Unidade Recorrente com Porta Bidirecional (T-BGRU, do inglês *Transfer Learning-Bidirectional Gated Recurrent Unit*) para a previsão do GHI de curto prazo em Mumbai, na Índia. Os autores demonstraram que houve uma melhora significativa em relação às métricas  $nRMSE$  e  $MAE$  do modelo T-BGRU em comparação com o modelo BGRU padrão. Desta forma, valores de 0,001444 e de 0,0155 para o MAE e o  $nRMSE$  foram alcançados usando o modelo T-BGRU, respectivamente. Isso comprova que o T-BGRU oferece mais precisão e previsibilidade. Assim, esses resultados demonstram a viabilidade de aplicar a transferência de aprendizagem para o aprimoramento do desempenho de modelos, especialmente quando os dados de treinamento são limitados a determinadas localidades ou condições climáticas.

Em [Mishra et al. \(2025\)](#), os autores propõem uma nova técnica chamada Previsão de Curto Prazo da Irradiação Solar em Transferência de Aprendizado com Disparo Zero

(SPIRIT, do inglês *Short-term Prediction of solar IRradIance for zero-shot Transfer Learning*). Utilizando imagens de câmeras no céu em combinação com características físicas relacionadas à previsão de irradiação solar e variáveis geográficas auxiliares, sem a necessidade de dados históricos no domínio de destino. Os dados de imagem são transformados em um vetor que, posteriormente, é concatenado com as características físicas e as variáveis auxiliares. Em média, a técnica superou os modelos mais modernos em 70%, dentro de cenários de *zero-shot*, ou seja, sem a utilização de dados históricos.

Em [Zhang et al. \(2025\)](#), os autores utilizam a técnica de Transferência de Aprendizado para revisão de curtíssimo prazo, dentro de um contexto de completa disparidade entre estações climáticas: a estação de origem, na Califórnia, com um clima seco e predominantemente claro, e a estação alvo em Nottingham, no Reino Unido, onde a condição climática é, em sua maioria, úmida e chuvosa. Utilizando um modelo genérico e combinando imagens do céu com dados climáticos. Como resultado, apenas duas semanas de dados no modelo alvo foram necessárias para superar o modelo base, utilizando 80% menos dados para atingir desempenho superior ao modelo base treinado do zero. Além disso, a técnica de transferência de aprendizado reduziu o tempo de treinamento em até 60%.

## 4 Metodologia Proposta

O objetivo deste capítulo é detalhar a metodologia utilizada na aplicação do modelo de *baseline* e dos modelos baseados em aprendizado profundo. Além disso, os dados das séries temporais são explorados e analisados por meio do processo de Análise Exploratória dos Dados (EDA, do inglês *Exploratory Data Analysis*), que evidenciam alguns aspectos importantes à configuração dos modelos. Com o intuito de melhor ilustrar a metodologia proposta, a Figura 2 apresenta um diagrama de blocos da mesma.

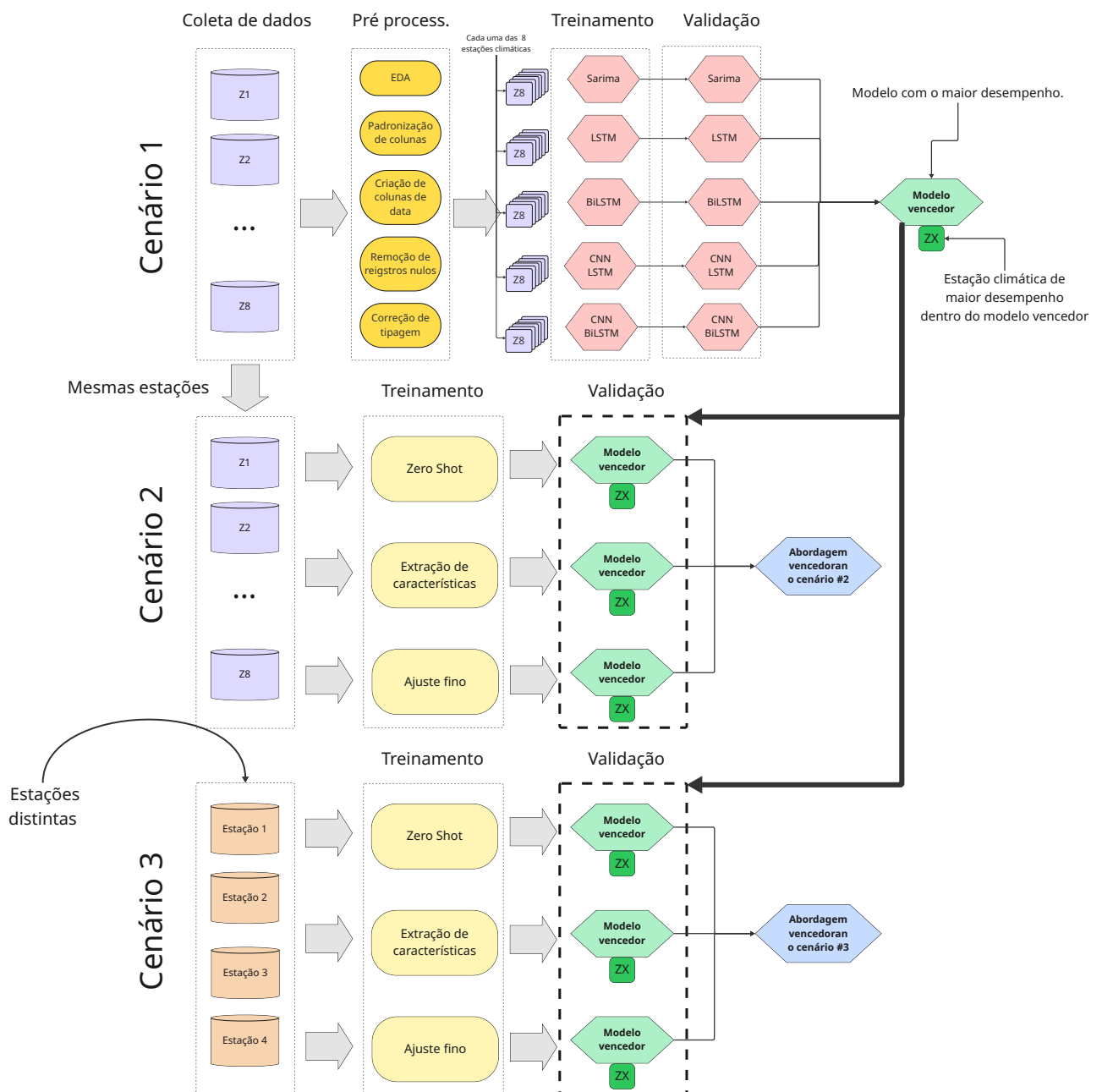


Figura 2 – Representação gráfica da metodologia.

## 4.1 Coleta dos Dados e Pré-processamento

Todos os conjuntos de dados foram coletados no portal do INMET (Instituto Nacional de Meteorologia) (INMET, 2025), de janeiro de 2023 a dezembro de 2024, de 8 estações, uma por zona bioclimática, conforme a Tabela 1 e a Figura 3, que apresenta a posição de cada cidade no mapa. A proposta foi utilizar dados de uma estação (AXXX) de cada zona bioclimática diferente (ZX) para averiguar suas respectivas diferenças, com uma análise separada para cada uma. O processo de seleção obedeceu ao critério da estação com o maior número de dados disponíveis, dentro de cada uma das 8 zonas bioclimáticas, cada zona com as 4 maiores cidades em termos de população. No caso da ausência da cidade no site do INMET, uma quinta cidade foi selecionada, e assim por diante. Os dados segmentados por zona bioclimática podem ser encontrados na Norma Brasileira ABNT NBR 15220-3 (ABNT, 2005), com os respectivos números populacionais obtidos no Censo 2022 do IBGE (IBGE, 2023).

Tabela 1 – Conjunto de Estações Selecionadas para o Cenário #1.

Cidade	ID	Zona Bioclimática
Curitiba	A807	Z1
Santa Maria	A803	Z2
Juiz de Fora	A518	Z3
Brasília	A001	Z4
Campos dos Goytacazes	A607	Z5
Campo Grande	A702	Z6
Teresina	A312	Z7
Manaus	A101	Z8

A estrutura dos conjuntos de dados segue uma granularidade horária, ou seja, cada instância do conjunto de dados representa os valores coletados naquela hora do dia. No entanto, por se tratar de dados de irradiação solar, utilizou-se apenas a faixa horária de 07h00 a 18h00 (pelo horário BRT), visto que, nesse período, os sensores de GHI coletaram os sinais mais significativos, conforme mostrado no gráfico da Figura 4, que considera todas as zonas bioclimáticas. Além disso, outras transformações foram necessárias, como a atribuição correta de tipagens, ao passo que todas as colunas são originalmente do tipo texto, e a reconstrução das colunas de data, que não se encontravam em um formato padronizado no conjunto original.

A proposta dessa dissertação foi dividida em três cenários, a saber: (i) cenário #1 – são criados modelos especialistas em cada uma das estações climatológicas selecionadas, sendo uma para cada zona bioclimática, ou seja, não há aplicação de qualquer estratégia de transferência de aprendizado; (ii) cenário #2 – utiliza dados de estações localizadas em

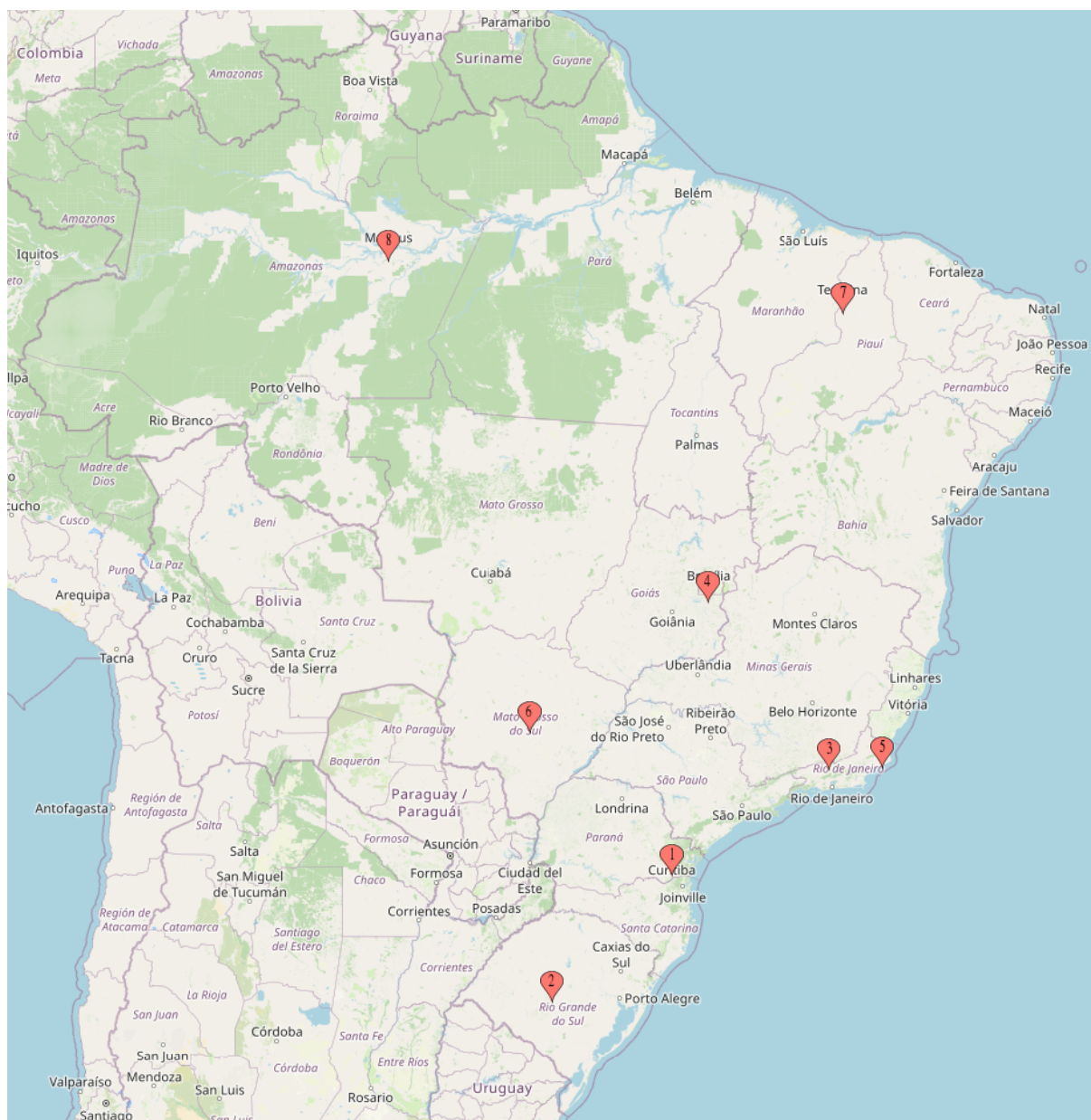


Figura 3 – Posição no mapa de todas as Zonas Bioclimáticas.

distintas zonas bioclimáticas para realizar o processo de transferência de aprendizado entre elas; e (iii) cenário #3 – utiliza dados de estações circunvizinhas e realiza a transferência de aprendizado entre elas.

#### 4.1.1 Cenário #1

Considerando o cenário #1, foram empregados conjuntos de dados das estações e das respectivas zonas bioclimáticas apresentadas na Tabela 1.

Para o cenário #1, apesar de terem sido escolhidas estações climatológicas com a maior integridade de dados (menor quantidade de dados faltantes), ainda assim houve a necessidade de realizar uma limpeza dos conjuntos de dados de cada uma das estações

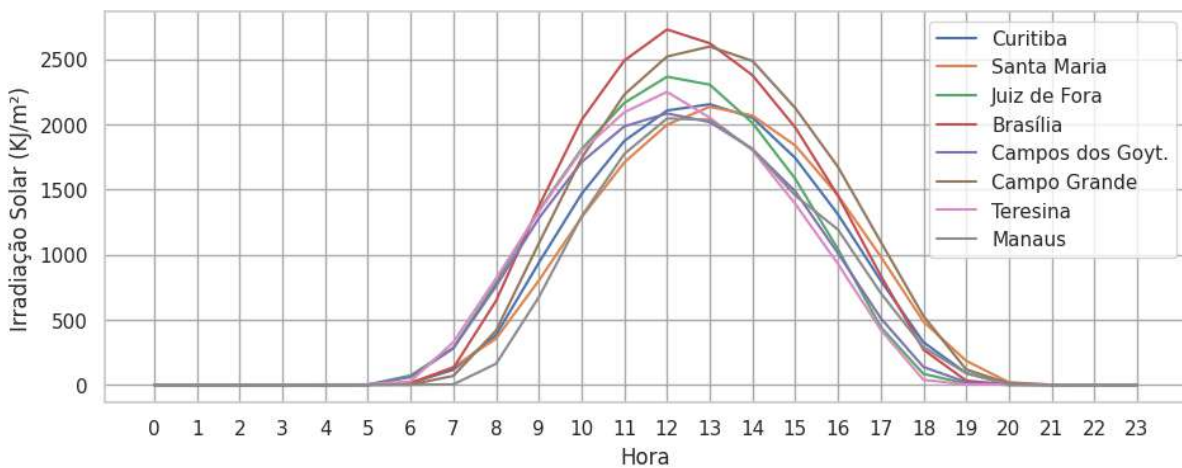


Figura 4 – Média dos Valores de Irradiação Solar por Hora para as Distintas Zonas Bioclimáticas.

previamente mencionadas. Nesse sentido, os dados faltantes foram apenas removidos. Com o objetivo de ilustrar a integridade dos dados, a Figura 5 apresenta o percentual de valores nulos de cada estação climatológica.

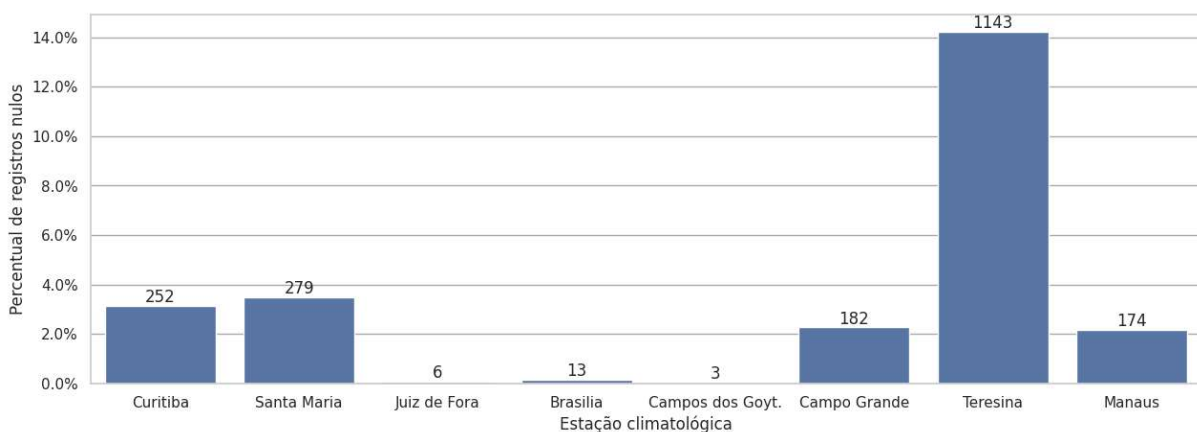


Figura 5 – Percentagens dos valores nulos por estação.

A preparação dos dados também envolveu a modelagem das entradas, que resultou em duas entradas (uma para a rede CNN e outra para a rede LSTM) e em um vetor alvo de saída ( $Y$ ). No caso da rede CNN, utilizaram-se dados climáticos, como velocidade do vento, pressão atmosférica e temperatura, além do próprio GHI, produzindo uma matriz 2D, enquanto a entrada da rede LSTM/BiLSTM é o mesmo GHI distribuído no tempo, como série temporal. Ambos os inputs foram construídos com janelas de 11 horas (tópico mais detalhado na seção B.1) e um horizonte de previsão de 1 hora de diferença entre as duas entradas e o vetor alvo.

#### 4.1.2 Cenário #2

O cenário #2 foi destinado a avaliar a transferência de aprendizado entre estações localizadas em zonas bioclimáticas distintas. Nesse sentido, considerou-se como domínio

de origem a estação climatológica que apresentou a maior precisão preditiva no cenário #1 e, portanto, seus dados foram considerados o domínio de origem.

Com o intuito de verificar o desempenho das estratégias de transferência de aprendizado consideradas nessa dissertação, foram mantidas as mesmas estações climatológicas do cenário #1, visto que cada uma já se encontra em uma zona climatológica distinta.

### 4.1.3 Cenário #3

Como o cenário #3 focou na transferência de aprendizado entre estações climatológicas circunvizinhas, optou-se por manter o modelo preditivo da estação com maior precisão, definido no cenário #1. Entretanto, novas estações climatológicas foram selecionadas em relação às suas distâncias à estação considerada como domínio de origem e às suas respectivas quantidades de dados, priorizando quantidades pequenas de dados utilizáveis, para averiguar os efeitos da Transferência de Aprendizagem em uma situação de escassez de dados. Tais estações e suas respectivas proporções de dados nulos são apresentadas na Tabela 2.

Tabela 2 – Conjunto de Estações Selecionadas para o Cenário #3.

Cidade	ID	% Nulos
Alvorada do Gurgueia	A336	79.59%
Angical do Piauí	A377	65.49%
Baixa Grande do Ribeiro	A375	15.17%
Bom Jesus	A326	9.96%

## 4.2 Análises das Séries Temporais

Cabe mencionar que todas as séries temporais tiveram seus dados avaliados quanto à média, desvio padrão, correlação, estacionariedade, tendência e sazonalidade. Essas análises encontram-se detalhadas no Apêndice A, as quais demonstraram uma elevada correlação positiva entre a temperatura e a velocidade do vento. Ainda foi observada uma elevada correlação negativa para a umidade.

Notou-se também, pelo teste de AdFuller (NAZARI et al., 2025), que os dados das estações climatológicas são estacionários. Ainda, ao realizar a decomposição das séries temporais, foi possível observar que elas apresentam um comportamento sazonal, porém sem uma tendência expressiva que pudesse indicar variação nas estatísticas dos dados.

### 4.3 Modelo de Referência

Para validar os resultados obtidos pelos modelos baseados em aprendizado profundo, foi considerada a resposta de um modelo de *baseline* (referência). Para tanto, empregou-se o SARIMA, pois este incorpora, em seu modelo matemático, a sazonalidade das séries temporais. Portanto, esse modelo serve como ponto de partida para se avaliar o desempenho, de forma comparativa, em relação aos modelos de aprendizado profundo. Os testes foram realizados utilizando três configurações, conforme a Tabela 3. Os parâmetros não sazonais são representados por  $\mathbf{p}$ ,  $\mathbf{d}$  e  $\mathbf{q}$ , enquanto os parâmetros sazonais são representados por  $\mathbf{P}$ ,  $\mathbf{D}$ ,  $\mathbf{Q}$  e  $\mathbf{S}$ . A primeira configuração foi definida de acordo com a análise de Autocorrelação (ACF, do inglês *AutoCorrelation Function*) e Autocorrelação Parcial (PACF, do inglês *Partial AutoCorrelation Function*), ambas apresentadas no Apêndice B. As demais configurações são variações arbitrárias para fins de comparação com o resultado obtido pelo ajuste via ACF e PACF. Embora os testes de ACF e PACF tenham sinalizado uma sazonalidade de 12 horas, os melhores resultados foram obtidos com a configuração do parâmetro  $\mathbf{S}$  do modelo para 11 horas.

Tabela 3 – Parametrizações do Modelo SARIMA.

Configuração	$\mathbf{p}$	$\mathbf{d}$	$\mathbf{q}$	$\mathbf{P}$	$\mathbf{D}$	$\mathbf{Q}$	$\mathbf{S}$
1	1	0	0	1	0	1	11
2	0	0	0	1	0	1	11
3	1	0	1	1	0	1	11

Cada uma das 8 estações foi testada em todas as configurações, o que gerou 24 resultados.

### 4.4 Modelos de Aprendizado Profundo

Como abordagem baseada em inteligência artificial, modelos de Aprendizado Profundo foram utilizados na previsão de irradiação solar, servindo como fundação para a técnica de Transferência de Aprendizado aplicada posteriormente. Os modelos LSTM e sua variação Bidirecional (BiLSTM), juntamente com o modelo híbrido CNN-LSTM, foram utilizados conforme o estado da arte em termos de modelos preditivos aplicados para previsão de séries temporais. Após o pré-processamento, os dados foram segmentados em GHI e outras variáveis climáticas, os quais foram então usados como entradas separadas nas redes LSTM e CNN, respectivamente, utilizando o horizonte de previsão de 1 hora.

A metodologia do teste consistiu em comparar os seguintes modelos: LSTM, BiLSTM, CNN-LSTM e CNN-BiLSTM. Assim, para esse teste, os modelos foram executados (treinamento e teste) cinco vezes para cada estação, gerando 160 resultados (8 estações x 4

modelos x 5 execuções). Ao final, o modelo com o melhor desempenho foi utilizado nos cenários 2 e 3, considerando três distintas estratégias de Transferência de Aprendizado (*zero-shot*, extração de características e ajuste fino).

#### 4.4.1 Estratégia de modelagem

##### 4.4.1.1 Construção dos modelos

O modelo LSTM foi configurado com duas camadas, sendo a primeira com 64 neurônios e a segunda com 32, ou seja, apenas com valores múltiplos de 8 ou potências de 2 que, de acordo com a documentação oficial da NVIDIA (NVIDIA Corporation, 2023), são padrões que otimizam o uso de CPUs e GPUs. Em seguida, há duas camadas densas (*fully-connected*), sendo a primeira com 32 neurônios e ativação não linear (*ReLu*) e a segunda (camada de saída) com 1 neurônio e ativação linear. Para o modelo BiLSTM, empregaram-se os mesmos hiperparâmetros. Vale ressaltar que a presente parametrização foi a que obteve os melhores resultados em testes realizados com diferentes números de camadas e neurônios.

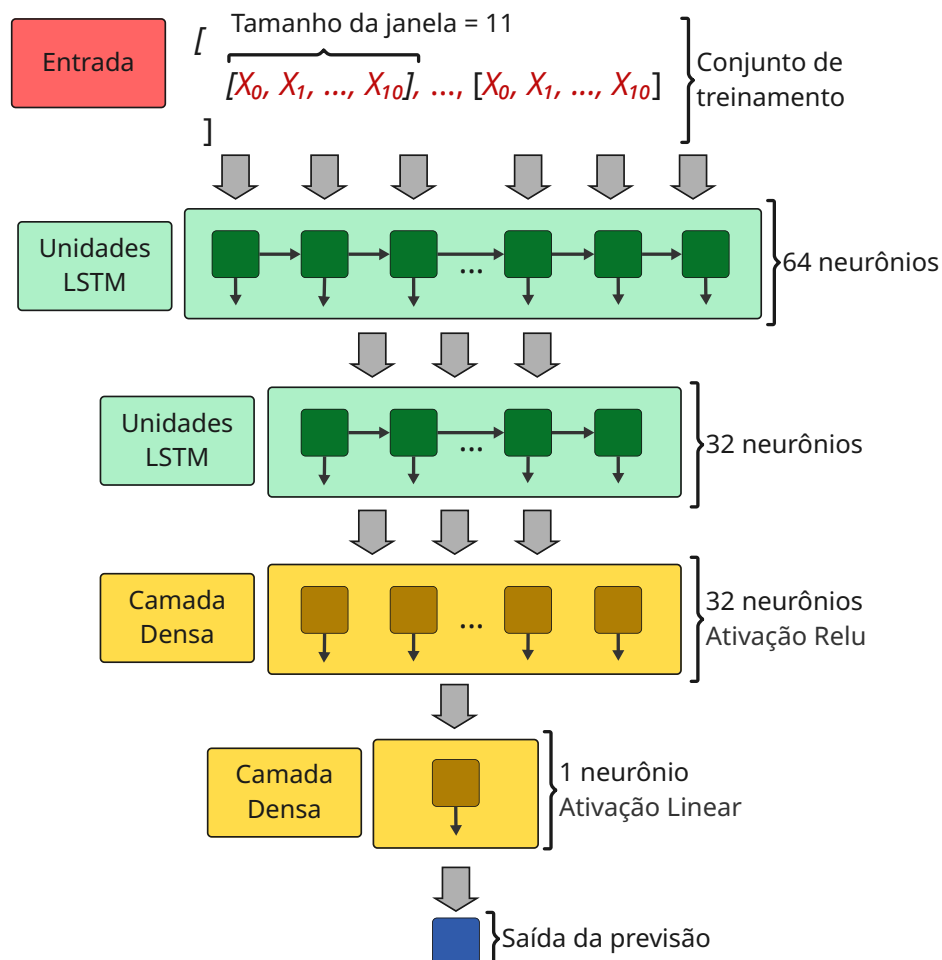


Figura 6 – Diagrama do modelo LSTM parametrizado para previsão de GHI.

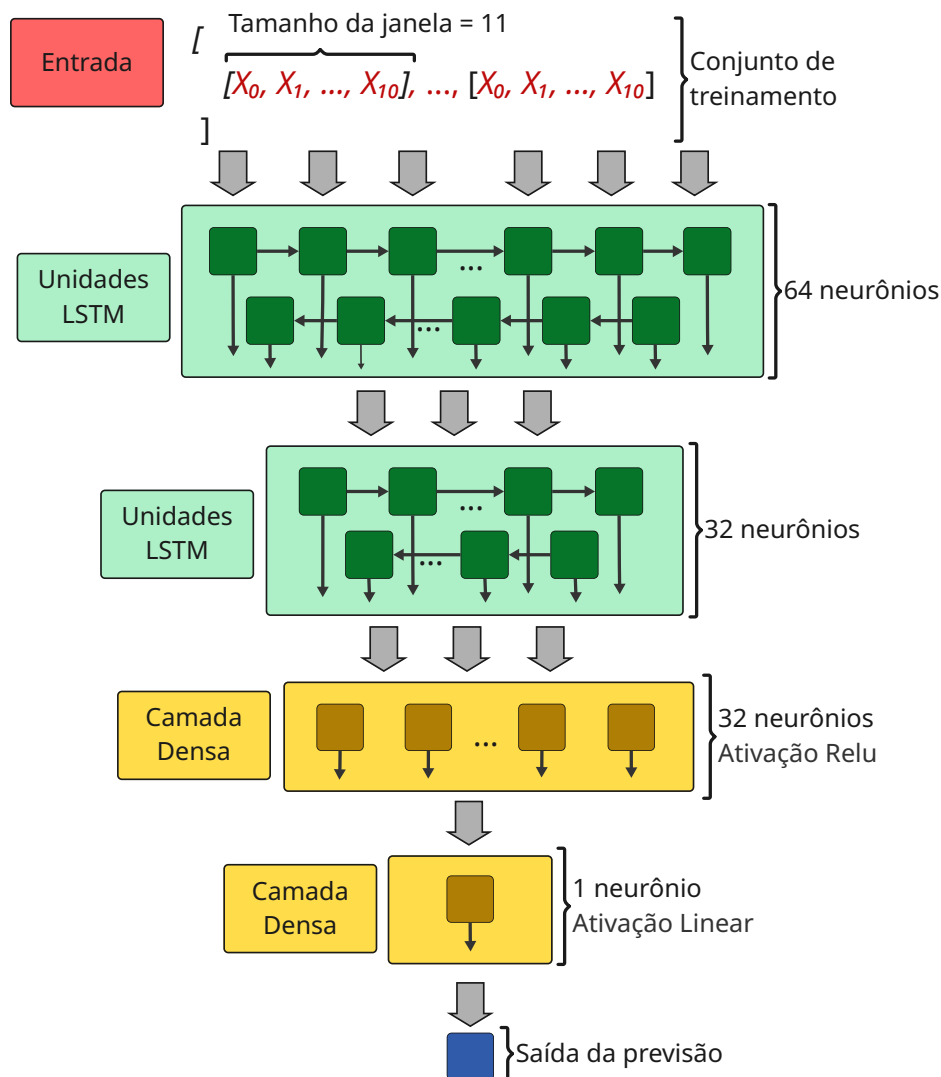


Figura 7 – Diagrama do modelo BiLSTM parametrizado para previsão de GHI.

Nos modelos híbridos CNN-LSTM e CNN-BiLSTM, são empregadas CNNs para estabelecer relações entre as demais variáveis climáticas e a variável alvo (GHI), propagando tais relações para a LSTM ou para a BiLSTM. A CNN utilizou uma camada de convolução unidimensional (*Conv1D*), com tamanho do filtro de 64 pesos sinápticos e *kernel* de 3. Em seguida, uma segunda camada de normalização (*Batch Normalization*) foi adicionada para evitar o problema de sobreajuste. A terceira camada foi composta por configurações semelhantes às da primeira camada, porém com filtro de 32 pesos sinápticos e o mesmo valor de *kernel*. A quarta camada foi usada para agregação (*MaxPooling*), sendo essa unidimensional. Na sequência, tem-se uma camada *flatten*, que remodela os dados de entrada, empilhando-os em um vetor unidimensional. Por fim, duas camadas completam o modelo CNN, sendo uma delas densa e totalmente conectada (com 32 neurônios) e a outra uma camada de normalização, também usada para evitar o problema de sobreajuste. A parte LSTM e BiLSTM do modelo híbrido foi construída de forma idêntica ao modelo individualizado e previamente apresentado. Em um primeiro momento, tanto

a CNN quanto a LSTM/BiLSTM processam os dados de forma individual e, em seguida, os concatenam. Assim, camadas densas de 128, 64, 32 e 1 neurônios são empilhadas, intercalando cada uma com uma camada de normalização, para gerar a saída do modelo.

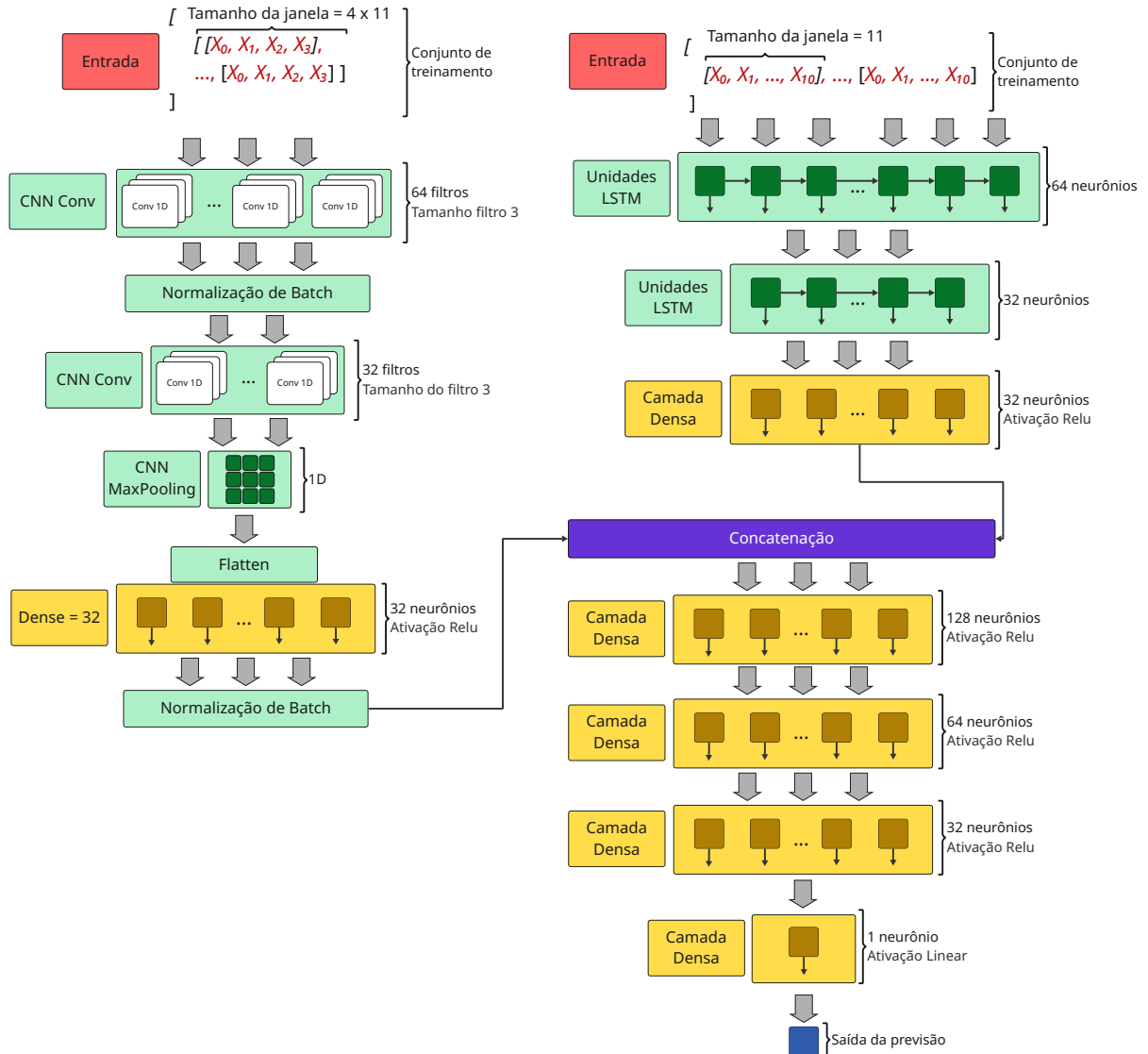


Figura 8 – Diagrama dos modelos híbridos parametrizados para previsão de GHI.

#### 4.4.1.2 Treinamento e Validação dos Modelos

Para o treinamento de LSTM e BiLSTM, os dados foram estruturados de modo que suas entradas recebessem uma janela temporal de 11 horas (*lag*) para prever a hora seguinte (horizonte de previsão de 1 hora). Essa mesma estruturação de dados foi empregada no modelo SARIMA. Portanto,  $X_t = (x_{t-11}, x_{t-10}, \dots, x_{t-1})$ , na qual  $X_t$  é o valor a ser previsto e  $x_{t-n}$  são os valores anteriores ao atual. A validação dos modelos de previsão foi realizada para os 9 meses de dados finais, correspondendo a 30% dos conjuntos de dados.

No entanto, para a CNN houve a necessidade de estruturar os dados de entrada no formato matricial. Nesse sentido, executou-se um processo de seleção de *features* por meio da técnica de PFI (do inglês, *Permutation Feature Importance*), que mede a importância das *features* ao treinar um modelo de Regressor de Floresta Randômica, avaliando o impacto da permutação aleatória de cada característica sobre o resultado. Todas as *features* utilizadas podem ser visualizadas na Figura 9, bem como o *ranking* de cada uma delas.

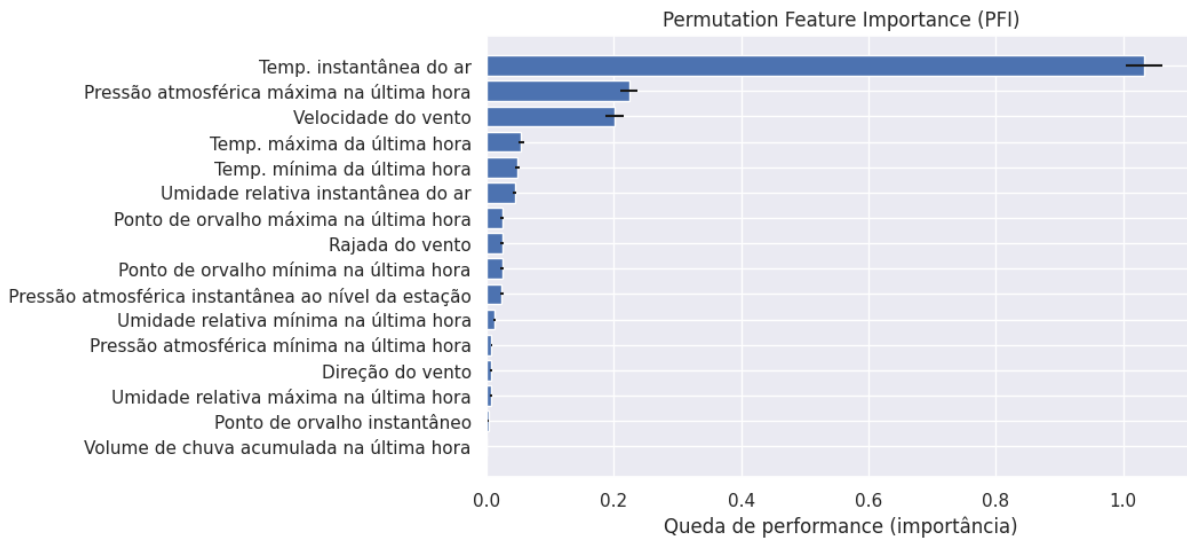


Figura 9 – Importâncias de cada característica climática dos dados de Teresina.

A partir dessa análise, as *features* de maior importância foram: temperatura do ar, pressão atmosférica máxima e velocidade do vento.

## 4.5 Transferência de Aprendizado

A técnica de transferência de aprendizado foi aplicada com base em três estratégias distintas, utilizadas nos cenários #2 e #3. Para tanto, considerou-se apenas o modelo de aprendizado profundo, que demonstrou o melhor desempenho, como alvo das estratégias de transferência de aprendizado, tanto para distintas zonas bioclimáticas quanto para estações climatológicas circunvizinhas. Portanto, empregaram-se como dados de base (ou seja, domínio de origem) aqueles usados no treinamento da estação climatológica de melhor desempenho. As três estratégias de transferência de aprendizado são apresentadas na sequência:

- *Zero-shot* – os dados do domínio de origem são utilizados para treinar o modelo preditivo. Em seguida, o mesmo modelo, sem nenhum ajuste ou treinamento adicional, é utilizado para prever dados de outra estação climatológica;
- Extração de características – as camadas do modelo preditivo pré-treinado são carregadas e os pesos são fixados. Em seguida, tais camadas (com os mesmos hiperparâ-

metros do modelo pré-treinado) são empilhadas com novas camadas densas. Por fim, os dados do domínio de destino são apresentados ao modelo parcialmente ajustado para que o mesmo seja treinado e a previsão da série temporal seja realizada;

- Ajuste fino – o modelo pré-treinado é carregado e treinado com os dados do domínio de destino, utilizando uma taxa de aprendizado menor ( $1e-5$ ) do que a utilizada no modelo pré-treinado ( $1e-3$ ).

Em resumo, as análises realizadas nesta dissertação podem ser organizadas conforme as descrições da Tabela 4.

Tabela 4 – Organização das análises realizadas para transferência de aprendizado.

Cenário	Transferência de Aprendizado	Descrição
1	–	Modelos treinados e validados para apenas uma estação climatológica
2	Zero-shot Extração de características Ajuste Fino	Modelo pré-treinado é usado na previsão direta para estações em distintas zonas bioclimáticas Modelo pré-treinado é re-treinado, somente nas camadas posteriores, com dados de estações em distintas zonas bioclimáticas Modelo pré-treinado é re-treinado, em todas as camadas e com menor taxa de aprendizado, com dados de estações em distintas zonas bioclimáticas
3	Zero-shot Extração de características Ajuste Fino	Modelo pré-treinado é usado na previsão direta para estações circunvizinhas Modelo pré-treinado é re-treinado, somente nas camadas posteriores, com dados de estações circunvizinhas Modelo pré-treinado é re-treinado, em todas as camadas e com menor taxa de aprendizado, com dados de estações circunvizinhas

# 5 Resultados

Esse capítulo apresenta os resultados obtidos a partir da metodologia proposta, que foram, a princípio, analisados para a previsão horária da irradiação solar no cenário #1, ou seja, no contexto da própria estação climatológica e em comparação com o modelo base (SARIMA). Na sequência, as respostas dos modelos preditivos foram analisadas para a transferência de aprendizado, de forma a atender aos cenários #2 (treinamento realizado para dados da estação localizada na zona bioclimática do modelo vencedor do cenário #1 com ajuste fino para estações localizadas em outras zonas bioclimáticas) e #3 (treinamento realizado para dados da estação localizada na zona bioclimática do modelo vencedor do cenário #1, e ajuste fino para estações circunvizinhas, isto é, contidas na mesma zona bioclimática).

## 5.1 Cenário #1

Conforme previamente mencionado, o cenário #1 contempla a análise geral dos modelos preditivos em comparação ao modelo de referência (SARIMA). Nesse sentido, a princípio, ajustou-se o modelo SARIMA para previsões horárias.

### 5.1.1 Modelo SARIMA

Os testes do modelo SARIMA foram realizados a partir dos 30% finais dos dados. Os resultados são apresentados na Tabela 5 para cada uma das estações utilizadas (destacando-se as cidades em que estão localizadas).

Vale destacar que os melhores resultados, em termos de RMSE, foram obtidos com ordens diferentes do SARIMA, sendo que, na maior parte dos casos, a ordem 2 foi adotada. Vale mencionar, ainda, que a ordem 2 considerou o valor 0 para os parâmetros não sazonais  $p$  (autorregressão) e  $q$  (média móvel). Dentre as estações, o melhor desempenho em termos de MAE foi obtido na localizada em Teresina, com 1779,94, enquanto a pior previsão foi realizada para a estação localizada em Brasília, com 2607,09.

Para ilustrar o comportamento do modelo SARIMA nas estações de Teresina e Brasília, as previsões do modelo são apresentadas nas Figuras 10 e 11, respectivamente.

### 5.1.2 Modelos Baseados em Aprendizado Profundo

Em relação aos modelos de aprendizado profundo, a princípio, são apresentados em suas formas individualizadas, ou seja, LSTM e BiLSTM. Assim, seus resultados podem ser visualizados na Tabela 6.

Tabela 5 – Resultados das melhores parametrizações do modelo SARIMA para cada estação climatológica.

Cidade	Ordem	MSE	RMSE	MAE	$R^2$
Curitiba	2	2152,43	46,39	2110,50	-0,008
Santa Maria	2	2250,29	47,43	2149,68	0,067
Juiz de Fora	2	2115,49	45,99	2074,83	0,259
Brasília	1	2895,78	53,81	2607,09	-0,310
Campos dos Goytacazes	1	2187,27	46,76	2120,66	0,172
Campo Grande	1	2596,82	50,95	2482,81	-0,004
Teresina	2	1845,24	42,95	1779,94	0,272
Manaus	1	2048,20	45,25	1927,72	0,025

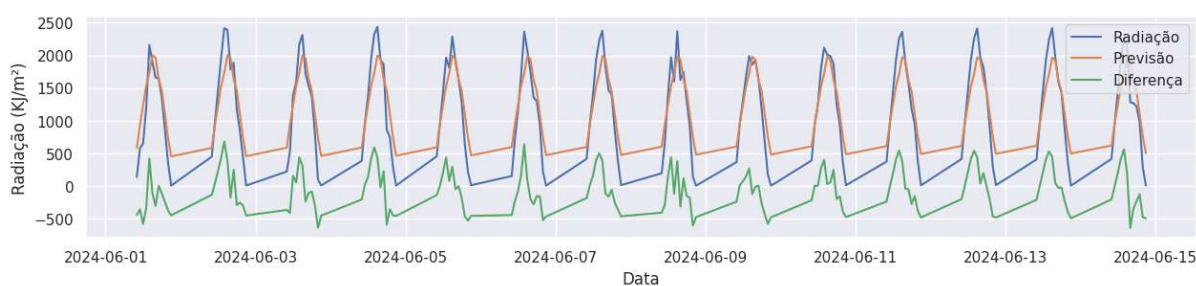


Figura 10 – Representação gráfica da previsão de Teresina.

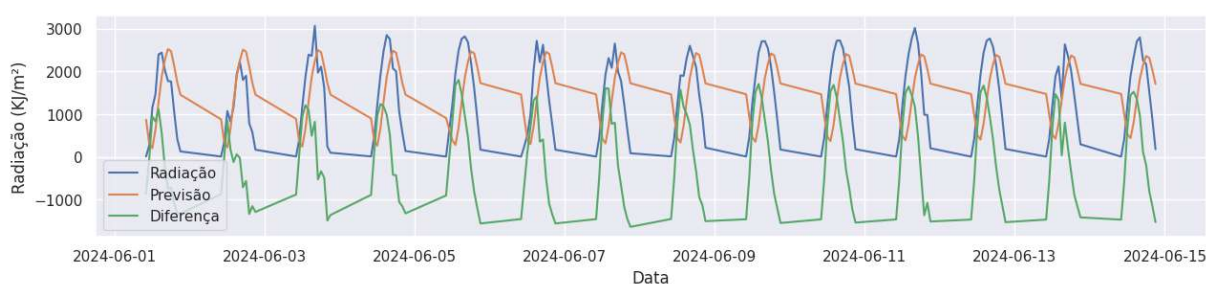


Figura 11 – Representação gráfica da previsão de Brasília.

Para cada modelo, foram realizadas cinco execuções por estação climatológica, visto que são estocásticos. Em termos de MAE, nota-se que o modelo LSTM apresentou melhor desempenho na maioria das estações. Entretanto, a diferença dos valores da métrica MAE variou 29%, entre 1434,51 e 1850,84, correspondendo a Teresina e Brasília, respectivamente. Uma representação gráfica da previsão em Teresina pode ser vista na Figuras 12

Vale ressaltar que, apesar da estação climatológica de Manaus obter, ainda que com relativa proximidade, valores de MSE e RMSE até maiores que os de Teresina, também apresentou valores com menor desempenho, principalmente em relação ao  $R^2$ , evidenciando uma maior capacidade de explicação da variabilidade da série temporal de Teresina. Além disso, Campos

Tabela 6 – Resultados médios do LSTM e do BiLSTM para as execuções de cada modelo.

Cidade	Tipo de Rede	MSE	RMSE	MAE	$R^2$
Brasília	BiLSTM	1795,77	42,37	1876,86	0,778
	LSTM	1759,88	41,95	1850,84	0,811
Campo Grande	BiLSTM	1770,68	42,07	1874,28	0,820
	LSTM	1732,85	41,62	1842,50	0,857
Campos dos Goytacazes	BiLSTM	1486,77	38,55	1597,25	0,867
	LSTM	1457,66	38,17	1576,55	0,892
Curitiba	BiLSTM	1518,81	38,97	1622,56	0,792
	LSTM	1495,48	38,67	1603,74	0,810
Juiz de Fora	BiLSTM	1529,49	39,10	1643,22	0,812
	LSTM	1500,86	38,74	1612,18	0,846
Manaus	BiLSTM	1372,79	37,05	1455,75	0,746
	LSTM	1345,70	36,68	1438,52	0,754
Santa Maria	BiLSTM	1457,37	38,17	1576,51	0,872
	LSTM	1443,59	37,99	1560,66	0,880
Teresina	BiLSTM	1364,91	36,94	1447,68	0,873
	LSTM	1352,17	36,77	1434,51	0,888

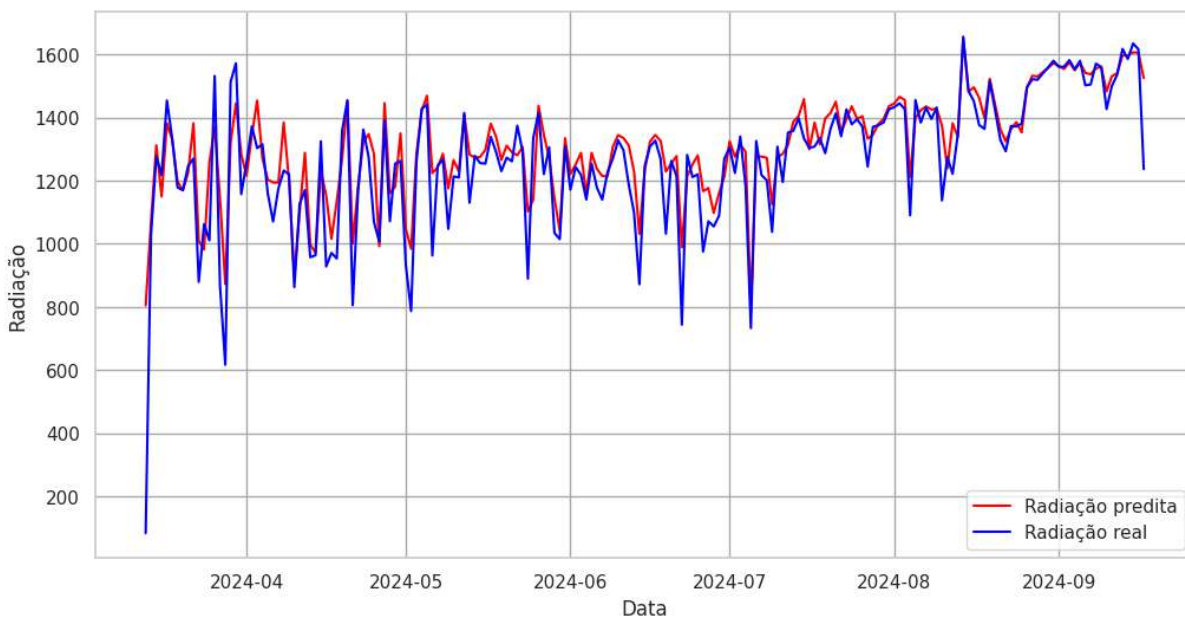


Figura 12 – Representação gráfica da previsão em Teresina utilizando o modelo LSTM.

dos Goytacazes apresentou um  $R^2$  ligeiramente melhor do que Teresina, com 0,89. Mas, em contrapartida, seus resultados nas demais métricas não foram satisfatórios. Sendo assim, Teresina foi a estação que apresentou os resultados com maior estabilidade.

Em seguida, foram considerados os mesmos modelos (LSTM e BiLSTM), porém com arquiteturas híbridizadas com uma CNN. Os resultados dos modelos híbridos são apresentados na Tabela 7. Tais modelos apresentaram MAEs médios em que o CNN-BiLSTM mostrou-se

ligeiramente melhor do que o CNN-LSTM. Além disso, seus resultados, apesar de semelhantes aos obtidos com LSTM e BiLSTM, mostraram um decréscimo dos MAEs médios em relação a Teresina, mas um acréscimo em relação a Brasília, com valores respectivos de 1419,51 e 1887,86, para a abordagem CNN-BiLSTM. Em relação a variação de resultados, a estação de Manaus com a CNN-BiLSTM obteve a melhor métrica de MAE, com 1409,23, apesar de resultados não tão satisfatórios em outras métricas, seguida por Teresina, com valores da mesma métrica mais equilibrados entre CNN-LSTM e CNN-BiLSTM, em termos de MSE, RMSE, MAE e  $R^2$ . Vale ressaltar que Campos dos Goytacazes com a abordagem CNN-BiLSTM também se destacou, obtendo o maior valor de  $R^2$  do teste, de 0,90.

Tabela 7 – Resultados médios do CNN-LSTM e do CNN-BiLSTM para as execuções de cada modelo.

Estação	Tipo de Rede	MSE	RMSE	MAE	$R^2$
Brasília	BiLSTM	1806,73	42,50	1887,86	0,781
	LSTM	1811,88	42,56	1887,71	0,778
Campo Grande	BiLSTM	1763,43	41,99	1872,12	0,833
	LSTM	1761,50	41,97	1875,22	0,831
Campos dos Goytacazes	BiLSTM	1480,65	38,47	1581,14	0,903
	LSTM	1481,60	38,49	1581,34	0,895
Curitiba	BiLSTM	1502,68	38,76	1611,13	0,795
	LSTM	1500,50	38,73	1609,98	0,800
Juiz de Fora	BiLSTM	1530,90	39,12	1642,59	0,829
	LSTM	1539,26	39,23	1649,53	0,818
Manaus	BiLSTM	1312,39	36,22	1409,23	0,778
	LSTM	1324,14	36,38	1425,00	0,756
Santa Maria	BiLSTM	1465,59	38,28	1586,12	0,876
	LSTM	1469,61	38,33	1592,33	0,870
Teresina	BiLSTM	1337,01	36,56	1419,51	0,873
	LSTM	1338,10	36,58	1423,53	0,861

Os modelos não hibridizados tiveram maior desempenho em mais estações, apresentando resultados globalmente mais estáveis e demonstrando maior explicabilidade. A consistência global superior do modelo baseado apenas em LSTM pode ser atribuída à menor complexidade da sua arquitetura, que reduz a variância e melhora a estabilidade entre as estações. Por outro lado, o modelo híbrido CNN-LSTM se beneficia da extração de características correlacionadas no processo convolucional da CNN, que permite um melhor desempenho em cenários mais desafiadores, caracterizados por maior ruído e menor previsibilidade. Esse comportamento dos modelos reflete uma compensação clássica entre viés e variância, em que a arquitetura híbrida se destaca em casos complexos, mas, de forma geral, pode perder robustez.

Uma comparação entre a média dos MAEs dos melhores modelos para cada estação é mostrada na Tabela 8 em relação aos modelos híbridos. Apesar do valor de  $R^2$  de Teresina ser menor do que o de Campos dos Goytacazes, a distância entre os valores de MAE, que representam

o erro absoluto, é razoavelmente grande. Portanto, Teresina foi a estação com os números mais estáveis.

Tabela 8 – Melhor modelo por estação considerando as métricas MAE e  $R^2$ .

Estação	Modelo	MAE	Var	STD	$R^2$	Var	STD
Brasília	LSTM	1850,84	90,07	9,49	0,8119	0,0001	0,01
Campo Grande	LSTM	1842,50	24,63	4,96	0,8576	0,00	0,006
Campos dos Goyt.	CNN BiLSTM	1581,14	225,59	15,01	0,9035	0,00	0,008
Curitiba	LSTM	1603,74	143,58	11,98	0,8107	0,0002	0,014
Juiz de Fora	LSTM	1612,18	45,30	6,73	0,8467	0,0001	0,012
Manaus	CNN BiLSTM	1409,23	18,45	4,29	0,7789	0,00	0,004
Santa Maria	LSTM	1560,67	21,89	4,67	0,8805	0,00	0,008
Teresina	LSTM	1432,90	15,83	3,97	0,8911	0,00	0,003

Em relação ao MAE, notou-se que Teresina e o modelo LSTM apresentaram os menores valores de variância e desvio padrão, com aproximadamente ~8% a menos em relação ao segundo colocado, Manaus. Em relação ao  $R^2$ , os valores se mantiveram mais estáveis, novamente com Teresina apresentando o menor valor de 0,003 e Manaus com o valor de 0,004.

Apesar dos modelos não hibridizados se sobressaírem em termos gerais com o LSTM, os modelos bidirecionais apresentaram resultados ligeiramente melhores do que os modelos não-bidirecionais no cenário híbrido, o oposto do que ocorreu no cenário anterior. Isso pode indicar que o suporte correlacional da CNN contribui para o melhor desempenho do BiLSTM. Em relação ao modelo de referência e às suas configurações, os modelos de aprendizado profundo são superiores, mesmo considerando as mesmas entradas de dados.

## 5.2 Cenários #2 – Transferência de Aprendizado entre Estações de Diferentes Zonas Bioclimáticas

Devido ao melhor desempenho apresentado pelo modelo LSTM no cenário #1, este modelo foi considerado para a etapa de transferência de aprendizado nos cenários #2 e #3. Assim, em ambos os cenários, foi realizado o treinamento deste modelo para uma estação de origem e, em seguida, as três estratégias já mencionadas foram aplicadas em cinco momentos diferentes, com conjuntos distintos em termos de quantidade de dados, a saber:

- *Conjunto 1 (padrão)* – 70% dos dados para treinamento;
- *Conjunto 2* – 50% dos dados para treinamento;
- *Conjunto 3* – 35% dos dados para treinamento;
- *Conjunto 4* – 25% dos dados para treinamento;
- *Conjunto 5* – 17% dos dados para treinamento.

Todos os resultados, por estação, obtidos para este cenário são apresentados nas Tabelas 9, 10, 11, 12, 13, 14 e 15. As comparações visuais podem ser observadas nas Figuras 13, 14 e 15, respectivamente, para MSE, MAE e  $R^2$ .

Tabela 9 – Resultados da Transferência de Aprendizado obtidos para o cenário #2 em Curitiba.

<b>Estratégia</b>	<b>MSE</b>	<b>RMSE</b>	<b>MAE</b>	<b><math>R^2</math></b>
Cenário 1	1495,48	38,67	1603,74	0,81
Zero-Shot	1572,70	39,66	1702,70	0,735
Extração Caract. 70%	1482,24	38,49	1607,28	0,813
Extração Caract. 50%	1516,55	38,94	1646,03	0,803
Extração Caract. 35%	1546,49	39,32	1678,51	0,801
Extração Caract. 25%	1715,04	41,41	1859,67	0,784
Extração Caract. 15%	1957,69	44,24	2104,75	0,620
Ajuste Fino 70%	1481,68	38,49	1598,38	0,826
Ajuste Fino 50%	1515,82	38,93	1643,46	0,808
Ajuste Fino 35%	1557,92	39,47	1687,63	0,778
Ajuste Fino 25%	1756,02	41,90	1892,08	0,699
Ajuste Fino 15%	1968,67	44,36	2108,84	0,636

Tabela 10 – Resultados da Transferência de Aprendizado obtidos para o cenário #2 em Santa Maria.

<b>Estratégia</b>	<b>MSE</b>	<b>RMSE</b>	<b>MAE</b>	<b><math>R^2</math></b>
Cenário 1	1443,59	37,99	1560,66	0,880
Zero-Shot	1620,73	40,26	1760,22	0,664
Extração Caract. 70%	1447,80	38,05	1577,61	0,871
Extração Caract. 50%	1441,55	37,96	1573,39	0,886
Extração Caract. 35%	1409,04	37,53	1533,58	0,874
Extração Caract. 25%	1539,21	39,23	1679,62	0,852
Extração Caract. 15%	1962,73	44,30	2121,72	0,767
Ajuste Fino 70%	1437,58	37,91	1557,89	0,888
Ajuste Fino 50%	1434,73	37,87	1562,52	0,883
Ajuste Fino 35%	1412,55	37,58	1542,46	0,862
Ajuste Fino 25%	1551,36	39,38	1686,95	0,841
Ajuste Fino 15%	1955,08	44,21	2104,58	0,736

### 5.2.1 Curitiba

Em Curitiba, de modo geral, o modelo do cenário #1 obteve melhor desempenho em relação a todas as estratégias, com exceção dos primeiros cenários das estratégias de Extração de características e de Ajuste Fino. Retreinando o modelo com a totalidade dos dados de

Tabela 11 – Resultados da Transferência de Aprendizado obtidos para o cenário #2 em Juiz de Fora.

<b>Estratégia</b>	<b>MSE</b>	<b>RMSE</b>	<b>MAE</b>	<b><math>R^2</math></b>
Cenário 1	1500,86	38,74	1612,18	0,846
Zero-Shot	1530,20	39,12	1663,40	0,811
Extração Caract. 70%	1494,20	38,65	1610,70	0,849
Extração Caract. 50%	1533,90	39,16	1662,50	0,791
Extração Caract. 35%	1462,46	38,24	1586,70	0,795
Extração Caract. 25%	1531,61	39,13	1659,11	0,760
Extração Caract. 15%	1695,07	41,17	1823,44	0,748
Ajuste Fino 70%	1499,15	38,72	1612,76	0,846
Ajuste Fino 50%	1528,92	39,10	1648,92	0,811
Ajuste Fino 35%	1469,35	38,33	1590,89	0,797
Ajuste Fino 25%	1541,07	39,25	1669,48	0,765
Ajuste Fino 15%	1716,65	41,43	1847,61	0,711

Tabela 12 – Resultados da Transferência de Aprendizado obtidos para o cenário #2 em Brasília.

<b>Estratégia</b>	<b>MSE</b>	<b>RMSE</b>	<b>MAE</b>	<b><math>R^2</math></b>
Cenário 1	1759,88	41,95	1850,84	0,811
Zero-Shot	1753,49	41,88	1859,52	0,815
Extração Caract. 70%	1741,34	41,74	1845,05	0,819
Extração Caract. 50%	1805,61	42,49	1916,63	0,761
Extração Caract. 35%	1763,55	41,99	1871,74	0,720
Extração Caract. 25%	1765,94	42,02	1869,36	0,770
Extração Caract. 15%	1833,92	42,82	1953,49	0,773
Ajuste Fino 70%	1741,26	41,74	1841,87	0,828
Ajuste Fino 50%	1823,09	42,69	1917,77	0,770
Ajuste Fino 35%	1790,41	42,31	1879,24	0,726
Ajuste Fino 25%	1763,05	41,98	1860,40	0,769
Ajuste Fino 15%	1875,79	43,31	1984,48	0,718

treinamento (70%), tanto o MSE quanto  $R^2$  superaram o resultado do cenário #1, com uma pequena margem de diferença de 0,9% e 1,1%, respectivamente. No caso do MAE, somente o Ajuste Fino superou o modelo do cenário #1, com um valor de 1598,38. Observou-se uma queda no desempenho geral à medida que os dados diminuíam no treinamento, com pouca diferença entre Extração de Características e Ajuste Fino. Para efeitos de comparação, até o uso de 35% dos dados, as abordagens de Extração de Características e de Ajuste Fino obtiveram melhores resultados do que a abordagem *Zero-Shot*.

Tabela 13 – Resultados da Transferência de Aprendizado obtidos para o cenário #2 em Campos dos Goytacazes.

<b>Estratégia</b>	<b>MSE</b>	<b>RMSE</b>	<b>MAE</b>	<b><math>R^2</math></b>
Cenário 1	1457,66	38,17	1576,55	0,892
Zero-Shot	1866,49	43,21	2021,68	0,168
Extração Caract. 70%	1516,31	38,94	1712,37	0,770
Extração Caract. 50%	1754,17	41,88	1920,26	0,644
Extração Caract. 35%	1750,40	41,83	1914,26	0,569
Extração Caract. 25%	1812,91	42,57	1979,64	0,581
Extração Caract. 15%	2065,94	45,45	2228,77	0,685
Ajuste Fino 70%	1531,59	39,14	1724,87	0,758
Ajuste Fino 50%	1756,54	41,91	1953,65	0,661
Ajuste Fino 35%	1971,44	44,40	2051,28	0,154
Ajuste Fino 25%	1865,40	43,19	1995,03	0,625
Ajuste Fino 15%	2182,45	46,71	2296,39	0,668

Tabela 14 – Resultados da Transferência de Aprendizado obtidos para o cenário #2 em Campo Grande.

<b>Estratégia</b>	<b>MSE</b>	<b>RMSE</b>	<b>MAE</b>	<b><math>R^2</math></b>
Cenário 1	1732,85	41,62	1842,50	0,857
Zero-Shot	1762,34	41,98	1892,76	0,827
Extração Caract. 70%	1733,02	41,63	1841,98	0,856
Extração Caract. 50%	1744,34	41,76	1860,01	0,828
Extração Caract. 35%	1656,70	40,70	1779,41	0,785
Extração Caract. 25%	1690,34	41,11	1812,22	0,784
Extração Caract. 15%	1821,60	42,68	1962,24	0,695
Ajuste Fino 70%	1734,33	41,65	1841,42	0,850
Ajuste Fino 50%	1746,48	41,79	1862,42	0,815
Ajuste Fino 35%	1659,65	40,73	1778,13	0,794
Ajuste Fino 25%	1696,16	41,18	1816,39	0,784
Ajuste Fino 15%	1835,09	42,83	1968,79	0,717

### 5.2.2 Santa Maria

A estação de Santa Maria obteve um cenário semelhante ao anterior, em que os números da abordagem *Zero-Shot* ficaram acima do resultado do modelo do cenário #1, no caso das métricas MSE/RMSE e MAE, e inferiores no caso da métrica  $R^2$ . O desempenho em MSE/RMSE e MAE melhorou em relação às demais abordagens, até o uso de 35% dos dados. A partir de 25%, houve um decréscimo no desempenho, indicando um determinado limite de dados faltantes. Em relação a  $R^2$ , o decréscimo foi linear, ou seja, quanto menos dados, menos performance. Extração de Características com 35% dos dados foi a menor utilização dos dados que foi melhor do que o

Tabela 15 – Resultados da Transferência de Aprendizado obtidos para o cenário #2 em Manaus.

Estratégia	MSE	RMSE	MAE	$R^2$
Cenário 1	1345,70	36,68	1438,52	0,754
Zero-Shot	1412,86	37,59	1529,67	0,699
Extração Caract. 70%	1325,99	36,41	1428,35	0,781
Extração Caract. 50%	1365,56	36,95	1471,99	0,718
Extração Caract. 35%	1348,74	36,72	1453,54	0,716
Extração Caract. 25%	1285,05	35,84	1370,02	0,703
Extração Caract. 15%	1362,96	36,91	1439,17	0,590
Ajuste Fino 70%	1319,49	36,33	1424,02	0,775
Ajuste Fino 50%	1363,46	36,92	1463,14	0,724
Ajuste Fino 35%	1345,13	36,67	1440,01	0,706
Ajuste Fino 25%	1265,47	35,57	1357,45	0,684
Ajuste Fino 15%	1378,31	37,12	1443,27	0,582

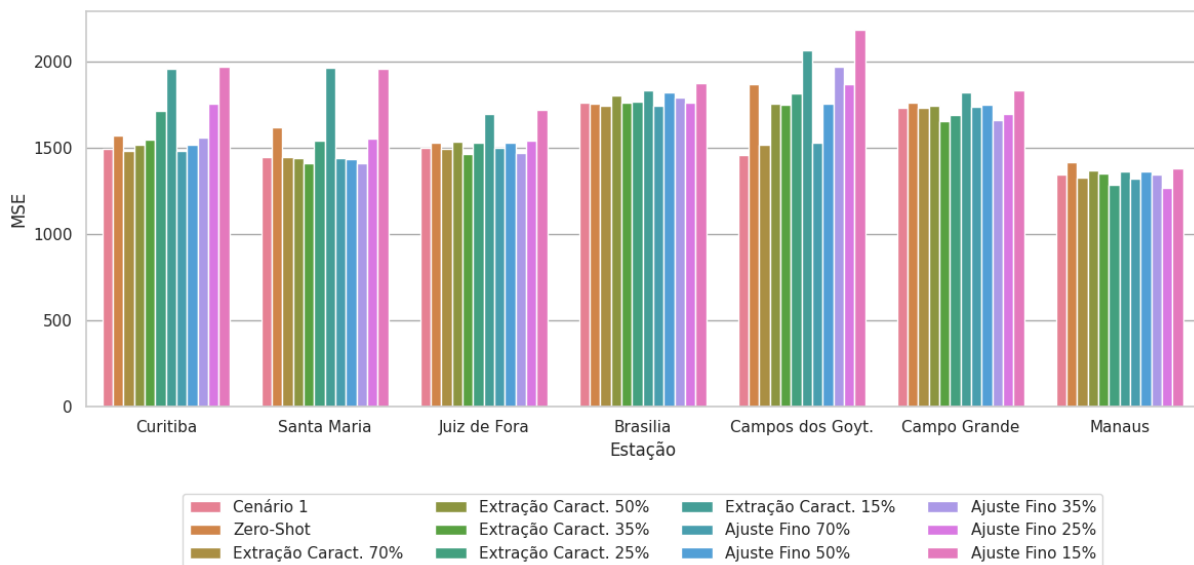


Figura 13 – Cenário #2: MSE dos modelos baseados em LSTM para a própria estação comparados às abordagens de transferência de aprendizado Zero-Shot, Extração de Características e Ajuste Fino.

modelo do cenário #1, com uma melhoria de 2,4% em MSE e 1,7% em relação a MAE, mas com um  $R^2$  de 0,68% menor, não sendo, entretanto, o melhor neste caso. Em Ajuste Fino, ocorre o mesmo cenário: a utilização de 35% dos dados superou o modelo do cenário #1 em 2,1%, 1,1% para MSE e MAE, respectivamente, mas ficou 2% abaixo em relação a  $R^2$ .

### 5.2.3 Juiz de Fora

Nos resultados de Juiz de Fora, os valores de MSE e MAE ficaram bem próximos ao comparar as abordagens com os valores do modelo do cenário #1, sendo as únicas diferenças relevantes as execuções com 15% dos dados em Extração de Características e Ajuste Fino,

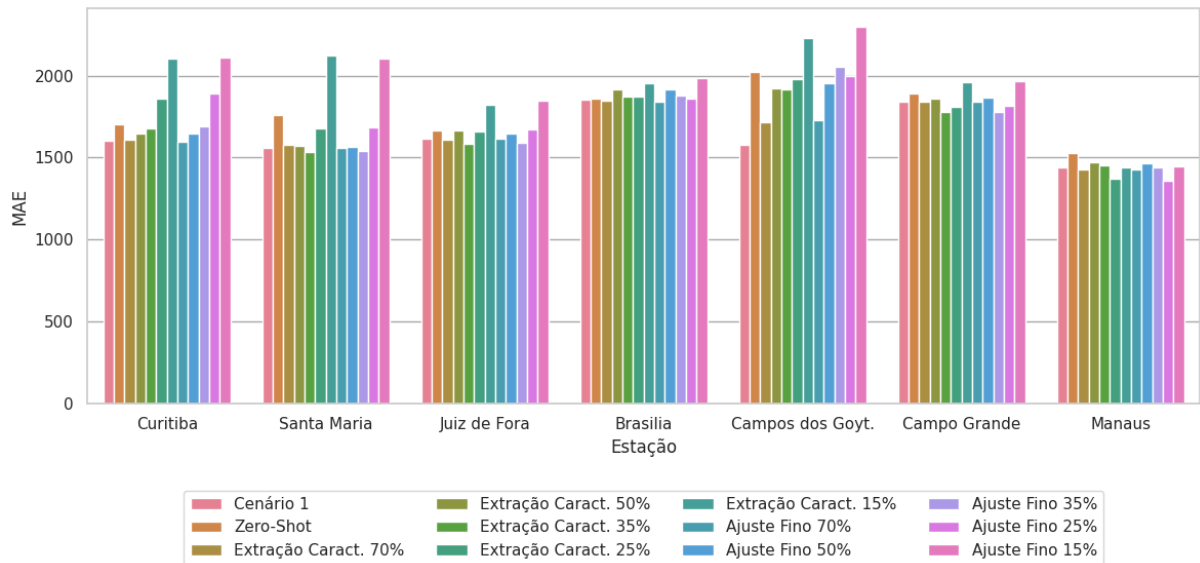


Figura 14 – Cenário #2: MAE dos modelos baseados em LSTM para a própria estação comparados às abordagens de transferência de aprendizado Zero-Shot, Extração de Características e Ajuste Fino.

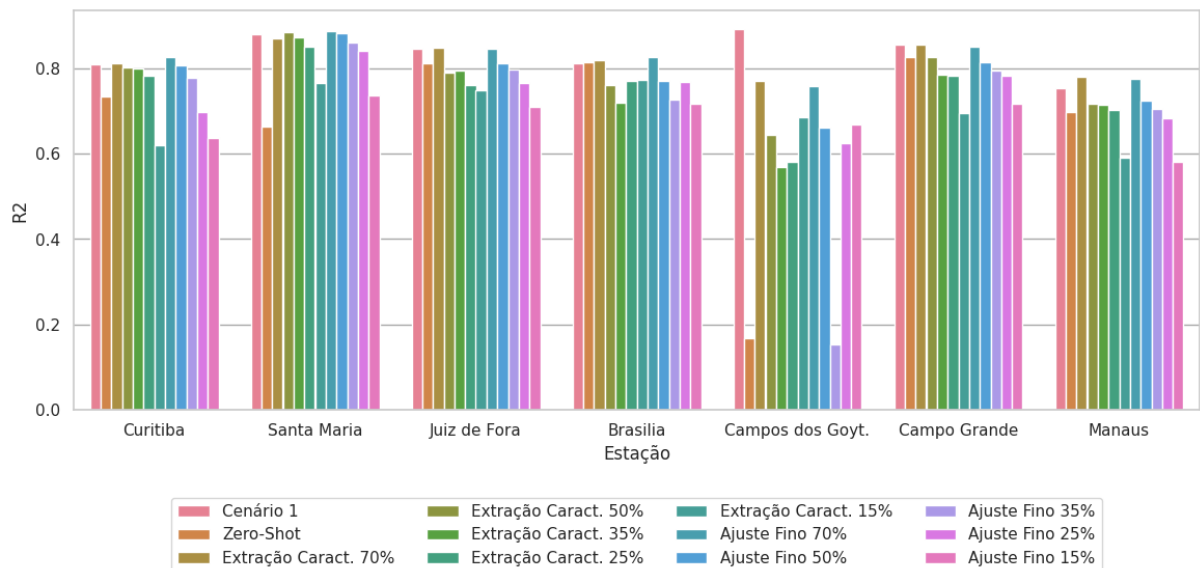


Figura 15 – Cenário #2:  $R^2$  dos modelos baseados em LSTM para a própria estação comparados às abordagens de transferência de aprendizado Zero-Shot, Extração de Características e Ajuste Fino.

cujo desempenho apresentou uma razoável inferioridade. Em relação a métrica  $R^2$ , somente a execução com 70% em treinamento supera o modelo base, no caso da Extração de Características, com um valor de 0,849, ou empata, no caso do Ajuste Fino, com 0,846. Novamente, observa-se uma tendência de decréscimo no desempenho ao utilizar menos dados na Transferência de Aprendizado, sendo mais evidente em  $R^2$ .

## 5.2.4 Brasília

Brasília apresenta números mais estáveis do que Juiz de Fora, com os resultados das abordagens de Transferência de Aprendizado (em ambas as abordagens) e os resultados do modelo do cenário #1 equiparados em termos de MSE e MAE. No caso da utilização de 70% dos

dados, tais valores até superam o modelo base em todas as métricas, com melhoria na Extração de Características de 1%, 0,3% e ~1%, e para o Ajuste Fino de de 1%, 0,4% e 2%, respectivamente para MSE, MAE e  $R^2$ . Apesar de Teresina e Brasília estarem em zonas bioclimáticas diferentes e evidentemente possuírem características climáticas distintas, a Transferência de Aprendizagem conseguiu prever com razoável precisão, mesmo para o treinamento com 15% dos dados.

### 5.2.5 Campos dos Goytacazes

Em Campos dos Goytacazes, a performance das abordagens de Transferência de Aprendizado ficou razoavelmente reduzida em comparação com estações anteriores, quanto ao  $R^2$ . Houve um decréscimo substancial na métrica, com um valor de 0,892 para o cenário #1, enquanto na estratégia *Zero-Shot*, o valor obtido foi de 0.168; utilizando a Extração de Características, os valores diminuem conforme a quantidade de dados de treinamento diminui, com o melhor valor sendo 0,77 na utilização de 70% em treinamento; em Ajuste Fino, o mesmo ocorre, com o melhor valor novamente em 70%, desta vez de 0,758. Em relação às métricas MSE e MAE, ocorre o mesmo cenário, em que nenhuma estratégia supera a execução do modelo do cenário #1, com os melhores resultados em Extração de Características ficando 4% e 8,6% menores, e Ajuste Fino ficando 5% e 9,4% abaixo, respectivamente, para MSE e MAE.

### 5.2.6 Campo Grande

Os resultados de Campo Grande foram semelhantes a estação de Juíz de Fora, em termos de comparação entre as abordagens: os valores de MSE e MAE ficaram bem próximos do resultado do cenário #1, apresentando uma leve piora entre o *Zero-Shot* e a utilização de 50%, mas superando o modelo do cenário #1 na execução com 35% dos dados, sugerindo novamente um possível limite no *trade-off* entre a quantidade de dados e o desempenho da Transferência de Aprendizado. Vale destacar que a estratégia *Zero-Shot*, apesar de apresentar desempenho inferior, ficou razoavelmente equiparada ao cenário #1.

### 5.2.7 Manaus

Os resultados de Manaus, assim como em outras estações, apresentaram resultados semelhantes ao cenário #1 nas estratégias de Transferência de Aprendizado, com os resultados de MSE e MAE para a execução de 25% superando o modelo base, em média, 5,2%. Comparando as estratégias de Extração de Características e Ajuste Fino, a última obteve um desempenho de aproximadamente ~21% maior em relação a primeira. Já em  $R^2$ , o padrão se manteve em relação a outras estações, em que a execução com 70% dos dados apresentou melhores resultados do que o modelo do cenário #1; porém, o desempenho cai à medida que os dados de treinamento diminuem, até se igualar ao *Zero-Shot* ou até ficar abaixo, no caso da execução com apenas 15% dos dados no treinamento.

## 5.3 Cenário #3 – Transferência de Aprendizado entre Estações Circunvizinhas

O cenário 3 inicia os testes da configuração com dados faltantes. As estações foram selecionadas com o critério de que contivessem, de forma crescente (uma estação com o menor índice de dados, a segunda com mais, a terceira com mais do que a segunda e assim por diante), uma razoável falta de dados, todas pertencentes à mesma zona bioclimática da estação, em um raio de aproximadamente ~500 km. As estações selecionadas foram:

- Alvorada do Gurguéia, com um percentual de ~17% de dados utilizáveis;
- Angical do Piauí, com um percentual de ~18% de dados utilizáveis;
- Baixa Grande do Ribeiro, com um percentual de ~39% de dados utilizáveis;
- Bom Jesus, com um percentual de ~47% de dados utilizáveis;

Os resultados são descritos nas Tabelas 16, 17, 18 e 19; as comparações visuais podem ser observadas nas figuras 16, 17 e 18, respectivamente, para MSE, MAE e  $R^2$ .

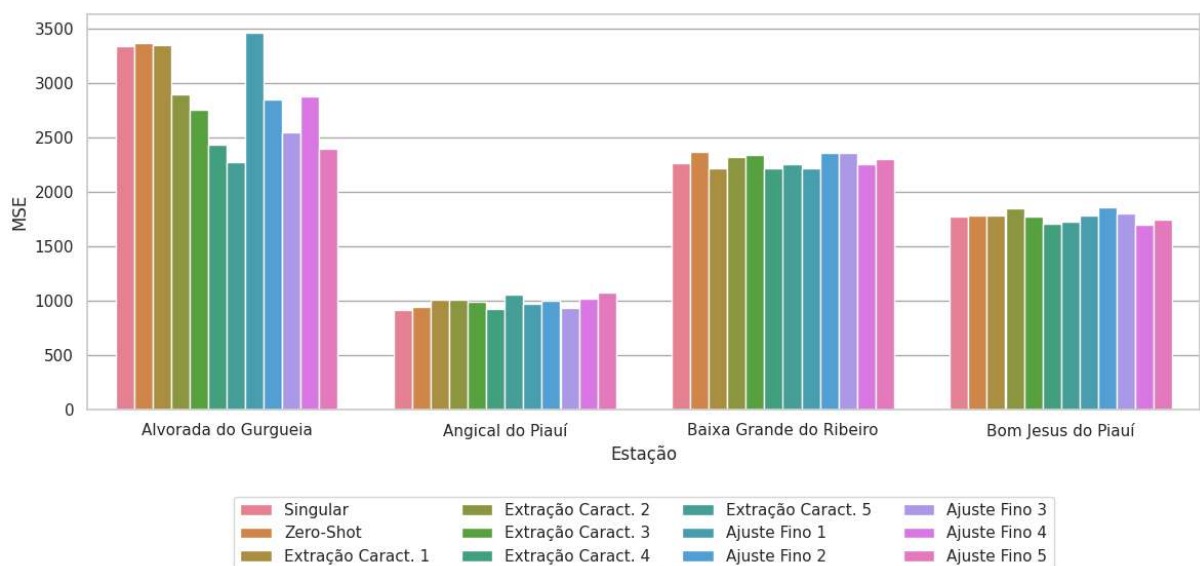


Figura 16 – Cenário #3: MSE dos modelos singulares comparados às abordagens Zero-Shot, Extração de Características e Ajuste Fino

### 5.3.1 Alvorada do Gurguéia

Em Alvorada do Gurguéia, as execuções do modelo singular, *Zero-Shot* e as execuções dos modelos utilizando a Extração de Características e Ajuste Fino com 70% dos dados ficaram semelhantes em termos de MSE e MAE. Nesta estação, o comportamento dos modelos ficou diferente dos demais: conforme os dados diminuem, os valores dessas métricas melhoram, sendo os resultados com maior performance as execuções com 15% dos dados, indicando uma possível

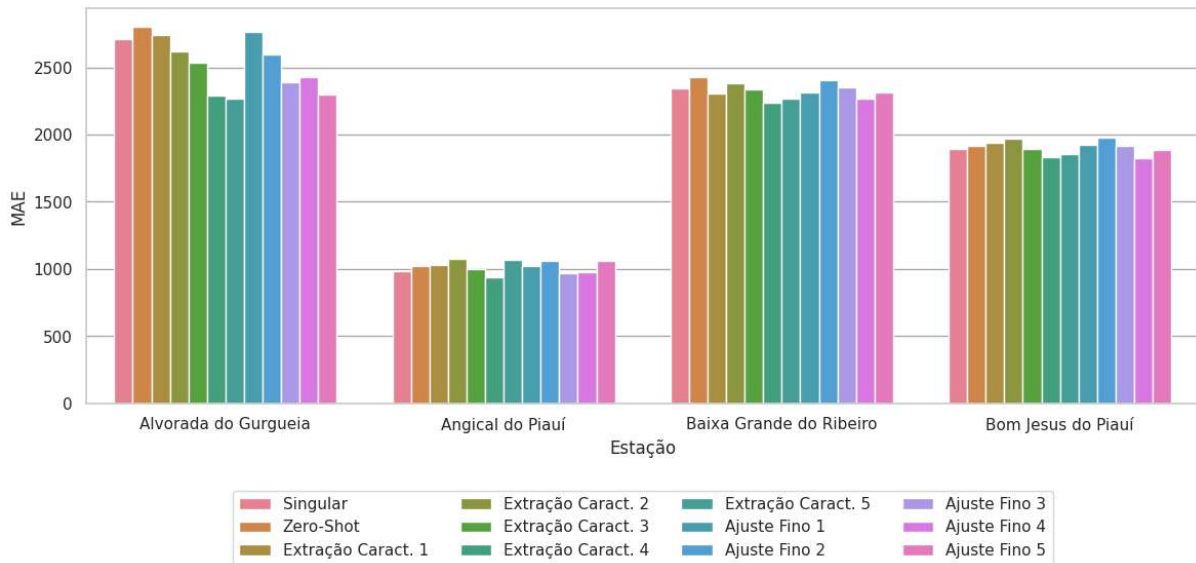


Figura 17 – Cenário #3: MAE dos modelos singulares comparados às abordagens Zero-Shot, Extração de Características e Ajuste Fino

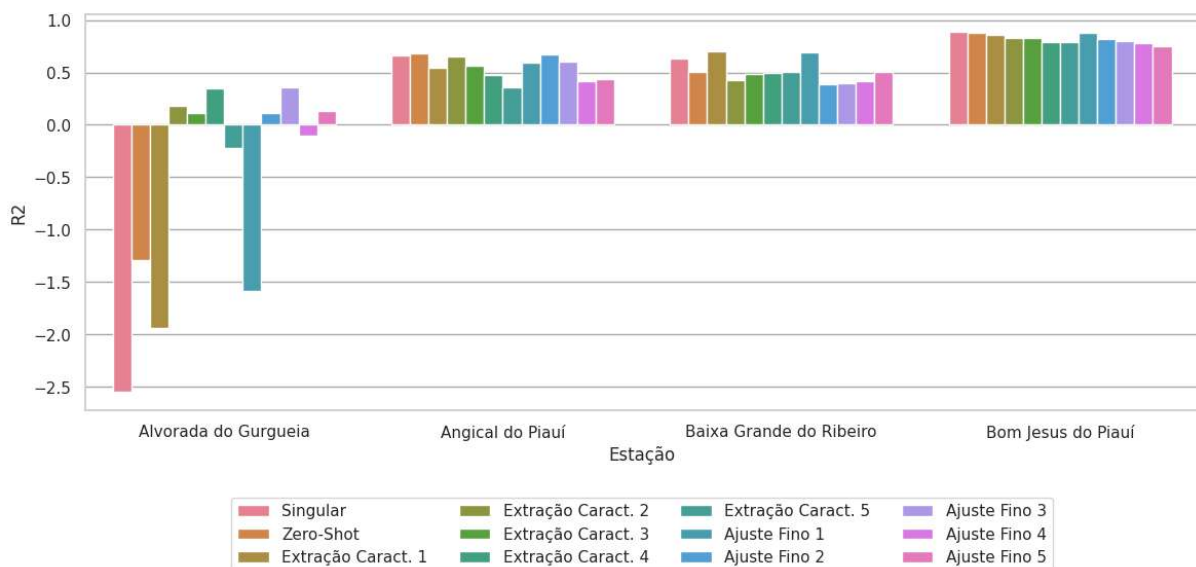


Figura 18 – Cenário #3:  $R^2$  dos modelos singulares comparados às abordagem Zero-Shot, Extração de Características e Ajuste Fino

incapacidade do modelo em lidar com as diferentes distribuições entre a estação utilizada para treinar o modelo singular e a estação em questão. Houve uma melhoria de 31,74% e 16,28% para Extração de Características e para Ajuste Fino de 28% e 15,17%, respectivamente, para MSE e MAE. Em  $R^2$ , os números ficaram bem abaixo dos demais resultados, inclusive gerando valores negativos, e as execuções com poucos dados novamente superaram as demais.

### 5.3.2 Angical do Piauí

Os valores de MSE e MAE na estação de Angical do Piauí foram os melhores em relação as demais estações do cenário #3. Os valores das estratégias de Transferência de Aprendizado ficaram muito próximos dos valores do modelo singular, por vezes superando-o. Sem nenhum

Tabela 16 – Resultados da Transferência de Aprendizado obtidos para o cenário #3 em Alvorada do Gurgueia.

Estratégia	MSE	RMSE	MAE	$R^2$
Singular	3335,30	57,77	2710,61	-2,549
Zero-Shot	3369,38	58,04	2800,90	-1,291
Extração Caract. 70%	3349,98	57,87	2738,10	-1,939
Extração Caract. 50%	2893,74	53,79	2620,18	0,177
Extração Caract. 35%	2754,79	52,48	2536,67	0,115
Extração Caract. 25%	2434,60	49,34	2288,43	0,344
Extração Caract. 15%	2276,57	47,71	2269,06	-0,220
Ajuste Fino 70%	3459,12	58,82	2762,03	-1,582
Ajuste Fino 50%	2852,27	53,40	2593,78	0,109
Ajuste Fino 35%	2542,88	50,42	2391,63	0,358
Ajuste Fino 25%	2876,75	53,63	2428,57	-0,107
Ajuste Fino 15%	2398,09	48,97	2299,32	0,129

treinamento, a estratégia *Zero-Shot* ficou apenas 3,3%, 3% e 2,7%, respectivamente, em MSE, MAE e  $R^2$ , abaixo em relação ao modelo base. Em Extração de Características, os melhores resultados foram obtidos com execuções que utilizaram menos dados (35% e 25%), diferentemente do  $R^2$ , que apresentou variações no desempenho ao longo das execuções, onde obteve apenas 1,95% de diferença em relação ao modelo singular na execução com 50% dos dados. Em Ajuste Fino, os valores dos resultados das execuções variaram mais em relação a estratégia anterior, onde, em MSE, nenhuma das execuções superou o modelo singular, mas se aproximou, com o melhor resultado sendo a execução com 35% dos dados e diferença de 2,2% em relação ao modelo singular; em MAE, onde 35% e 25% superaram o modelo base em aproximadamente ~1,7%. Em  $R^2$ , além do *Zero-Shot*, somente a execução com 50% dos dados em Ajuste Fino superou o modelo singular, com um valor de 0,669.

### 5.3.3 Baixa Grande do Ribeiro

Em Baixa Grande do Ribeiro, tanto em MSE quanto em MAE, as estratégias apresentaram melhor desempenho com menos dados de treinamento, com o melhor resultado obtido na execução com 25%, indicando um possível limite do *trade-off* entre a quantidade de dados para generalização e a ausência de dados em termos de especialização do modelo. MSE e MAE tiveram uma melhoria média nas abordagens com 25%, de 2% para MSE e 3,8% para MAE. Em relação a  $R^2$ , os melhores resultados foram das execuções padrão, com 70% dos dados, tanto para Extração de Características quanto para Ajuste Fino, com média de ~0,70. Nessa métrica, o comportamento dos resultados é de queda à medida que diminuem os dados de treinamento.

Tabela 17 – Resultados da Transferência de Aprendizado obtidos para o cenário #3 em Angical do Piauí

Estratégia	MSE	RMSE	MAE	$R^2$
Singular	913,37	30,22	987,43	0,665
Zero-Shot	944,75	30,74	1018,82	0,684
Extração Caract. 70%	1008,50	31,75	1029,54	0,541
Extração Caract. 50%	1007,05	31,73	1076,35	0,652
Extração Caract. 35%	993,00	31,51	1003,09	0,563
Extração Caract. 25%	920,38	30,33	937,32	0,475
Extração Caract. 15%	1057,73	32,52	1067,67	0,362
Ajuste Fino 70%	973,81	31,21	1025,27	0,596
Ajuste Fino 50%	1002,07	31,65	1063,17	0,669
Ajuste Fino 35%	934,63	30,57	970,32	0,605
Ajuste Fino 25%	1020,04	31,93	973,83	0,415
Ajuste Fino 15%	1070,61	32,72	1061,07	0,439

Tabela 18 – Resultados da Transferência de Aprendizado obtidos para o cenário #3 em Baixa Grande do Ribeiro.

Estratégia	MSE	RMSE	MAE	$R^2$
Singular	2258,60	47,53	2343,61	0,636
Zero-Shot	2365,70	48,64	2431,14	0,503
Extração Caract. 70%	2213,37	47,06	2304,07	0,706
Extração Caract. 50%	2323,46	48,20	2385,49	0,430
Extração Caract. 35%	2337,18	48,34	2333,35	0,488
Extração Caract. 25%	2212,71	47,03	2235,86	0,50
Extração Caract. 15%	2254,14	47,47	2270,08	0,507
Ajuste Fino 70%	2219,72	47,11	2315,61	0,690
Ajuste Fino 50%	2357,41	48,55	2402,81	0,392
Ajuste Fino 35%	2358,47	48,56	2351,46	0,398
Ajuste Fino 25%	2257,72	47,51	2269,50	0,413
Ajuste Fino 15%	2301,00	47,96	2315,89	0,504

### 5.3.4 Bom Jesus

As execuções com 25% e 15% de dados de treinamento obtiveram os melhores resultados em Bom Jesus, principalmente os 25% de dados de treinamento, em Extração de Características e Ajuste Fino, com destaque para esta última estratégia, que apresentou uma melhoria de 3,8% e 3,5%, respectivamente, para MSE e MAE. Em relação a  $R^2$ , a estratégia *Zero-Shot* e a execução padrão do Ajuste Fino apresentaram valores próximos ao modelo singular, com destaque para a primeira, que, mesmo sem novo treinamento, ficou apenas 0,5% abaixo do modelo base.

Tabela 19 – Resultados da Transferência de Aprendizado obtidos para o cenário #3 em Bom Jesus do Piauí.

<b>Estratégia</b>	<b>MSE</b>	<b>RMSE</b>	<b>MAE</b>	<b><math>R^2</math></b>
Singular	1769,84	42,07	1895,52	0,884
Zero-Shot	1777,31	42,16	1915,26	0,879
Extração Caract. 70%	1785,78	42,27	1940,24	0,855
Extração Caract. 50%	1844,21	42,94	1968,01	0,833
Extração Caract. 35%	1776,19	42,14	1896,16	0,827
Extração Caract. 25%	1710,57	41,35	1833,22	0,785
Extração Caract. 15%	1721,18	41,48	1856,69	0,794
Ajuste Fino 70%	1781,99	42,23	1921,74	0,874
Ajuste Fino 50%	1856,57	43,08	1974,67	0,820
Ajuste Fino 35%	1802,61	42,45	1913,98	0,803
Ajuste Fino 25%	1701,23	41,24	1828,31	0,778
Ajuste Fino 15%	1743,36	41,75	1882,76	0,750

## 6 Considerações finais

O presente trabalho propôs apresentar as técnicas e ferramentas do estado da arte presentes na literatura, abordando especificamente modelos híbridos e a transferência de aprendizado. Visando trazer essas abordagens para o cenário brasileiro, que possui oito zonas bioclimáticas, cada uma com seu próprio ecossistema climático e, portanto, apresentando um desafio de generalização de modelos aplicados a esse cenário, esta dissertação demonstrou o uso de Redes Neurais LSTM e modelos híbridos utilizando CNN e LSTM para previsão em um horizonte de curto prazo, empregando a Transferência de Aprendizado para superar os problemas de diferenças climáticas entre as diversas estações e a falta de dados.

Na literatura consultada, os autores chegam a um consenso: os modelos híbridos superaram os modelos singulares em quase todos os cenários, com destaque para o modelo híbrido que utiliza CNN em conjunto com o modelo LSTM (e sua variação bidirecional). Em relação à técnica de Transferência de Aprendizado, os resultados da literatura indicam que ela traz ganhos expressivos, independentemente do modelo utilizado, em termos de precisão, redução da parametrização e redução do tempo de treinamento.

No presente trabalho, a proposta foi:

- aplicar o estado da arte, utilizando modelos de aprendizado profundo de forma singular e híbrida, comparando seus resultados;
- aplicar a técnica de TL em diferentes configurações, no modelo com melhor desempenho na etapa anterior, e medir o desempenho em relação ao cenário brasileiro.

Em relação ao desempenho dos modelos, o modelo LSTM singular saiu-se melhor tanto em comparação com a variação bidirecional quanto com a variação híbrida, com uma diferença mínima em relação a esses cenários, diferentemente da comparação com o *baseline*, em que a diferença de desempenho é expressivamente maior. Globalmente, o LSTM singular apresentou o melhor desempenho médio, obtendo o menor MAE agregado e o maior  $R^2$  médio, com melhorias de 0,78%, 1,01% e 1,37% em relação ao CNN + BiLSTM (segundo colocado), CNN + LSTM (terceiro colocado) e BiLSTM (último colocado), respectivamente. O que se percebeu foi um desempenho ligeiramente melhor dos modelos Bidirecionais em relação aos unidirecionais dentro do cenário híbrido, indicando um possível suporte adicional da estrutura correlacional fornecida pelo CNN.

Os testes com a técnica de Transferência de Aprendizado foram promissores, atingindo um desempenho comparável (por vezes até superando) ao modelo base da estação alvo na maioria das estações, dentro do cenário de longa distância e do cenário de dados faltantes.

No cenário #2, com a transferência entre estações de zonas bioclimáticas diferentes, a estratégia *Zero-Shot* apresentou um desempenho frequentemente menor, indicando uma dificuldade do modelo base (cenário #1) em generalizar dados de estações com diferenças climáticas,

sem qualquer tipo de ajuste. Em média, o *Zero-Shot* apresentou uma piora na previsão de 8,1% em relação ao modelo do cenário #1, destacando-se que Brasília foi a única estação que conseguiu igualar e até superar ligeiramente (no caso do MSE) o modelo base. Em Extração de Características, nos treinamentos padrão (70%) e com 50% dos dados, obteve-se um desempenho equivalente e, por vezes, superior ao modelo base; porém, houve uma degradação significativa ao realizar o treinamento com quantidades de dados menores. Em  $R^2$ , os resultados foram, em sua maioria, inferiores ao do modelo do cenário #1, porém superiores à estratégia *Zero-Shot*. Nesta estratégia, nota-se uma tendência de queda no desempenho a partir de um determinado limite, sendo necessário um número mínimo de dados locais para que a transferência seja eficiente. Nesse sentido, a estratégia de Extração de Características é funcional, mas está sujeita a um limite na ausência de dados, que em algumas estações e/ou estratégias pode chegar até 35% de dados no treinamento. Em média, a estratégia reduziu aproximadamente 3,7% de MSE e 3% de MAE, além de aumentar em 1,8% o  $R^2$ . Em Ajuste Fino, notou-se um padrão semelhante a Extração de Características, com desempenho muito próximo ou maior do que o modelo do cenário #1, porém com a mesma queda de performance em  $R^2$  ao diminuir os dados de treinamento. Em um cenário com baixa escassez de dados (50%, por exemplo), o Ajuste Fino tem uma ligeira vantagem sobre a extração de características, o que começa a mudar conforme os dados de treinamento diminuem. Refletindo essa superioridade, em média, a estratégia reduziu aproximadamente 6,4% de MSE e 5% de MAE, além de aumentar em 3,3% o  $R^2$ .

Para o cenário #3, com a transferência entre estações circunvizinhas, a técnica de Transferência de Aprendizado demonstrou desempenho significativo, considerando um cenário de escassez de dados. Observou-se que o *Zero-Shot* não obteve ganhos expressivos, embora tenha ficado bem próximo do modelo singular. Por outro lado, mesmo sem ajustes adicionais e refinamento das camadas, conseguiu alcançar um resultado satisfatório ao igualar-se ao modelo singular, apresentando aproximadamente ~5% de piora nos resultados, em média. Em Extração de Características, os modelos tiveram melhorias principalmente nas estações com maior escassez (Alvorada do Gurgueia e Angical do Piauí), e conforme os dados diminuía entre as execuções, melhorias progressivas foram observadas, com casos de redução significativa do MSE e do MAE, e aumento considerável do  $R^2$ . Nessa estratégia, os resultados demonstram que o conhecimento previamente adquirido nos dados de Teresina compensou a falta de dados, com um ganho de 8,5% em relação ao MSE e ao MAE, e 18,7% em relação a  $R^2$ , comparando-se ao modelo singular. Por fim, Ajuste Fino demonstrou ser a estratégia mais consistente entre as apresentadas, obtendo melhores resultados nas estações com maior escassez de dados, evidenciando a relação indireta entre a quantidade de dados e o desempenho dos modelos, com uma melhoria de 10,3% em relação ao MSE e ao MAE, e 22,3% em relação a  $R^2$ , muito influenciado pelos resultados da estação com menor número de dados, Alvorada do Gurgueia, onde o modelo singular obteve valores negativos.

Vale ressaltar que, em alguns casos, conforme os dados diminuem entre as execuções, as demais estratégias ganham mais desempenho, evidenciando um possível cenário de forte viés, em que, quanto menos dados, maior o viés e menor a capacidade de variância do modelo, impactando a generalização dos modelos ao aplicar as estratégias mais refinadas com dados

escassos e privilegiando execuções em um cenário com um volume razoável de dados. O que se notou, de forma geral, é que, quanto maior a escassez de dados locais, maior é o impacto da Transferência de Aprendizado.

Para trabalhos futuros, uma análise necessária é a exploração de outras abordagens na Transferência de Aprendizado, como o uso de Redes Neurais Adversárias, para a adaptação do domínio de origem ao levar o aprendizado obtido para o domínio de destino. Isso permitiria evitar o uso de ajustes finos, uma vez que as diferentes distribuições estariam uniformizadas, o que facilitaria o aprendizado.

# Referências

- ABNT. *NBR 15220-3: Desempenho térmico de edificações Parte 3: Zoneamento bioclimático brasileiro e diretrizes construtivas para habitações unifamiliares de interesse social*. [S.l.], 2005. Citado na página 44.
- ACIKGOZ, H. A novel approach based on integration of convolutional neural networks and deep feature selection for short-term solar radiation forecasting. *Applied Energy*, v. 305, p. 117912, jan. 2022. ISSN 0306-2619. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0306261921012241>>. Citado na página 33.
- AGBO, E. P. et al. Characteristics of various radiative fluxes: global, tilted, direct, and diffused radiation—a case study of nigeria. *Meteorology and Atmospheric Physics*, v. 135, n. 2, p. 14, Feb 2023. ISSN 1436-5065. Disponível em: <<https://doi.org/10.1007/s00703-023-00951-8>>. Citado na página 22.
- AGUILAR-LÓPEZ, J. M. et al. Mobile sensor for clouds shadow detection and direct normal irradiance estimation. *Solar Energy*, v. 237, p. 470–482, 2022. ISSN 0038-092X. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0038092X21010811>>. Citado na página 16.
- ALIZAMIR, M. et al. Improving the accuracy of daily solar radiation prediction by climatic data using an efficient hybrid deep learning model: Long short-term memory (lstm) network coupled with wavelet transform. *Engineering Applications of Artificial Intelligence*, v. 123, p. 106199, 2023. ISSN 0952-1976. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0952197623003834>>. Citado na página 36.
- ALKHAYAT, G.; MEHMOOD, R. A review and taxonomy of wind and solar energy forecasting methods based on deep learning. *Energy and AI*, v. 4, p. 100060, jun. 2021. ISSN 2666-5468. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S2666546821000148>>. Citado 2 vezes nas páginas 23 e 24.
- ANTONANZAS-TORRES, F. et al. Clear sky solar irradiance models: A review of seventy models. *Renewable and Sustainable Energy Reviews*, v. 107, p. 374–387, 2019. ISSN 1364-0321. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1364032119301261>>. Citado na página 23.
- ASSAF, A. M. et al. A review on neural network based models for short term solar irradiance forecasting. *Applied Sciences*, v. 13, n. 14, 2023. ISSN 2076-3417. Citado 2 vezes nas páginas 23 e 38.
- ATIQUÉ, S. et al. Forecasting of total daily solar energy generation using arima: A case study. In: *2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC)*. [S.l.: s.n.], 2019. p. 0114–0119. Citado na página 20.
- BAK, S. et al. Transfer learning for photovoltaic power forecasting across regions using large-scale datasets. *IEEE Access*, v. 13, p. 136175–136190, 2025. Citado na página 17.
- BALAL, A. et al. Forecasting solar power generation utilizing machine learning models in lubbock. *Emerg. Sci. J.*, Ital Publication, v. 7, n. 4, p. 1052–1062, jul. 2023. Citado na página 36.

- BECK, M. et al. *xLSTM: Extended Long Short-Term Memory*. 2024. Disponível em: <<https://arxiv.org/abs/2405.04517>>. Citado 2 vezes nas páginas 28 e 29.
- BEKTAŞ, S. Ç.; ALTAŞ, I. H. Dwt-bilstm-based models for day-ahead hourly global horizontal solar irradiance forecasting. *Neural Computing and Applications*, v. 36, n. 21, p. 13243–13253, Jul 2024. ISSN 1433-3058. Disponível em: <<https://doi.org/10.1007/s00521-024-09701-2>>. Citado na página 36.
- BELMAHDI, B.; LOUZAZNI, M.; BOUARDI, A. E. One month-ahead forecasting of mean daily global solar radiation using time series models. *Optik*, v. 219, p. 165207, out. 2020. ISSN 0030-4026. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0030402620310433>>. Citado na página 24.
- BHUTTA, M. S. et al. Optimizing solar power efficiency in smart grids using hybrid machine learning models for accurate energy generation prediction. *Scientific Reports*, v. 14, n. 1, p. 17101, Jul 2024. ISSN 2045-2322. Disponível em: <<https://doi.org/10.1038/s41598-024-68030-5>>. Citado na página 17.
- BOUZERDOUM, M.; MELLIT, A.; Massi Pavan, A. A hybrid model (sarima–svm) for short-term power forecasting of a small-scale grid-connected photovoltaic plant. *Solar Energy*, v. 98, p. 226–235, 2013. ISSN 0038-092X. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0038092X13004039>>. Citado na página 25.
- BROCKWELL, P. J.; DAVIS, R. A. *ITSM: An interactive time series modelling package for the PC*. New York, NY: Springer, 1991. Citado na página 20.
- BZDOK, D.; KRZYWINSKI, M.; ALTMAN, N. Machine learning: a primer. *Nat. Methods*, Springer Science and Business Media LLC, v. 14, n. 12, p. 1119–1120, dez. 2017. Citado na página 25.
- CHANDEL, S. S. et al. Review of deep learning techniques for power generation prediction of industrial solar photovoltaic plants. *Solar Compass*, v. 8, p. 100061, 2023. ISSN 2772-9400. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S2772940023000292>>. Citado na página 36.
- CHEN, F. et al. Review of lightweight deep convolutional neural networks. *Archives of Computational Methods in Engineering*, Springer Science and Business Media LLC, v. 31, n. 4, p. 1915–1937, nov. 2023. ISSN 1886-1784. Disponível em: <<http://dx.doi.org/10.1007/s11831-023-10032-z>><http://dx.doi.org/10.1007/s11831-023-10032-z>. Citado na página 27.
- CHEN, X.; YIN, Z.; TIAN, H. Support vector machines principles and actually example. *Procedia Computer Science*, v. 243, p. 2–11, 2024. ISSN 1877-0509. The 4th International Conference on Machine Learning and Big Data Analytics for IoT Security and Privacy. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S187705092402009X>>. Citado na página 26.
- CHICCO, D.; WARRENS, M. J.; JURMAN, G. The coefficient of determination r-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Comput Sci*, United States, v. 7, p. e623, jul. 2021. Citado na página 33.
- CHU, Y. et al. A review of distributed solar forecasting with remote sensing and deep learning. *Renewable and Sustainable Energy Reviews*, v. 198, p. 114391, 2024. ISSN 1364-0321. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S136403212400114X>>. Citado na página 16.

CRYER, J. D.; CHAN, K.-S. *Time series analysis*. 2. ed. New York, NY: Springer, 2008. (Springer texts in statistics). Citado na página 20.

DEMIR, V.; CITAKOGLU, H. Forecasting of solar radiation using different machine learning approaches. *Neural Computing and Applications*, v. 35, n. 1, p. 887–906, Jan 2023. ISSN 1433-3058. Disponível em: <<https://doi.org/10.1007/s00521-022-07841-x>>. Citado na página 35.

DESPOTOVIC, M. et al. Solar irradiance time series forecasting using auto-regressive and extreme learning methods: Influence of transfer learning and clustering. *Applied Energy*, v. 365, p. 123215, 2024. ISSN 0306-2619. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0306261924005981>>. Citado na página 31.

Energy Institute. *Statistical Review of World Energy 2024*. London: Energy Institute, 2024. Accessed: 2025-11-06. Disponível em: <<https://www.energyinst.org/statistical-review>>. Citado na página 15.

FLORITA, A. R. et al. Short-term solar forecasting performance of popular machine learning algorithms: Preprint. In: NATIONAL RENEWABLE ENERGY LABORATORY (NREL), GOLDEN, CO (UNITED STATES). 2017. Disponível em: <<https://www.osti.gov/biblio/1406990>>. Citado na página 34.

GAO, Y. et al. Solutions to the insufficiency of label data in renewable energy forecasting: A comparative and integrative analysis of domain adaptation and fine-tuning. *Energy*, v. 302, p. 131863, 2024. ISSN 0360-5442. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0360544224016360>>. Citado 3 vezes nas páginas 9, 30 e 39.

GHOLIZADE, M. et al. A review of recent advances and strategies in transfer learning. *International Journal of System Assurance Engineering and Management*, v. 16, 02 2025. Citado na página 31.

GOSWAMI, S. et al. A novel transfer learning-based short-term solar forecasting approach for india. *Neural Computing and Applications*, v. 34, n. 19, p. 16829–16843, Oct 2022. ISSN 1433-3058. Disponível em: <<https://doi.org/10.1007/s00521-022-07328-9>>. Citado 2 vezes nas páginas 31 e 40.

GUEN, V. L.; THOME, N. A deep physical model for solar irradiance forecasting with fisheye images. In: *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. [S.l.: s.n.], 2020. p. 2685–2688. ISSN 2160-7516. Citado na página 36.

HAIDER, S. A. et al. Deep learning and statistical methods for short- and long-term solar irradiance forecasting for islamabad. *Renewable Energy*, v. 198, p. 51–60, 2022. ISSN 0960-1481. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0960148122011387>>. Citado na página 35.

HASSAN, I. U.; KUMAR, A. Time series prediction of photovoltaic energy using machine learning approach-a review. In: *2025 International Conference on Computing Technologies Data Communication (ICCTDC)*. [S.l.: s.n.], 2025. p. 1–6. Citado na página 29.

HUSSAIN, D. et al. Solar energy integration into smart grids: Challenges and opportunities. *Letters in High Energy Physics*, v. 4, p. 2313–2324, 2024. Citado na página 16.

IBGE. *Censo Demográfico 2022: População e Domicílios*. [S.l.], 2023. Disponível em: <<https://www.ibge.gov.br/estatisticas/sociais/saude/22827-censo-demografico-2022.html>>. Citado na página 44.

- INMET. *Portal do INMET (Instituto Nacional de Meteorologia)*. 2025. Disponível em: <<https://portal.inmet.gov.br/>>. Citado na página 44.
- International Energy Agency. *World Energy Outlook 2024*. Paris: IEA, 2024. Accessed: 2025-11-06. Disponível em: <<https://www.iea.org/reports/world-energy-outlook-2024>>. Citado na página 15.
- IRENA; WMO. *2022 Year in Review: Climate-driven Global Renewable Energy Potential Resources and Energy Demand*. [S.l.], 2023. Citado na página 16.
- JIN, M. et al. *Large Models for Time Series and Spatio-Temporal Data: A Survey and Outlook*. 2023. Disponível em: <<https://arxiv.org/abs/2310.10196>>. Citado na página 21.
- KHAREL, A.; ZAREAN, Z.; KAUR, D. Long short-term memory (lstm) based deep learning models for predicting univariate time series data. 2024. Citado na página 27.
- KLEISSL, J. *Solar Energy Forecasting and Resource Assessment*. Academic Press, 2013. ISBN 9780123977724. Disponível em: <[https://books.google.com.br/books?id=94KI0\\_SPwW8C](https://books.google.com.br/books?id=94KI0_SPwW8C)>. Citado na página 24.
- KOBER, T. et al. Global energy perspectives to 2060 – wec’s world energy scenarios 2019. *Energy Strategy Reviews*, v. 31, p. 100523, 2020. ISSN 2211-467X. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S2211467X20300766>>. Citado na página 16.
- KONG, X. et al. Deep learning for time series forecasting: a survey. *International Journal of Machine Learning and Cybernetics*, v. 16, n. 7, p. 5079–5112, Aug 2025. ISSN 1868-808X. Disponível em: <<https://doi.org/10.1007/s13042-025-02560-w>>. Citado na página 23.
- KRISHNAN, N.; KUMAR, K. R.; INDA, C. S. How solar radiation forecasting impacts the utilization of solar energy: A critical review. *Journal of Cleaner Production*, v. 388, p. 135860, 2023. ISSN 0959-6526. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0959652623000185>>. Citado na página 16.
- KUMAR, D. S. et al. Solar irradiance resource and forecasting: a comprehensive review. *IET Renewable Power Generation*, v. 14, n. 10, p. 1641–1656, 2020. Disponível em: <<https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/iet-rpg.2019.1227>>. Citado 2 vezes nas páginas 22 e 24.
- KUMARI, P.; TOSHNIWAL, D. Deep learning models for solar irradiance forecasting: A comprehensive review. *Journal of Cleaner Production*, v. 318, p. 128566, out. 2021. ISSN 0959-6526. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0959652621027736>>. Citado 3 vezes nas páginas 21, 26 e 37.
- KUMARI, P.; TOSHNIWAL, D. Extreme gradient boosting and deep neural network based ensemble learning approach to forecast hourly solar irradiance. *Journal of Cleaner Production*, v. 279, p. 123285, 2021. ISSN 0959-6526. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0959652620333308>>. Citado na página 38.
- KUMARI, P.; TOSHNIWAL, D. Long short term memory–convolutional neural network based deep hybrid approach for solar irradiance forecasting. *Applied Energy*, v. 295, p. 117061, 2021. ISSN 0306-2619. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0306261921005158>>. Citado 2 vezes nas páginas 16 e 38.

- LAGUARDA, A.; ALONSO-SUÁREZ, R.; ABAL, G. Improved estimation of hourly direct normal solar irradiation (dni) using geostationary satellite visible channel images over moderate albedo areas. *Solar Energy*, v. 259, p. 30–40, 2023. ISSN 0038-092X. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0038092X23002864>>. Citado na página 22.
- LEFÈVRE, M. et al. Mcclear: a new model estimating downwelling solar radiation at ground level in clear-sky conditions. *Atmospheric Measurement Techniques*, v. 6, n. 9, p. 2403–2418, 2013. Disponível em: <<https://amt.copernicus.org/articles/6/2403/2013/>>. Citado na página 23.
- LIAO, R. et al. Enhanced photovoltaic power generation forecasting for newly-built plants via physics-infused transfer learning with domain adversarial neural networks. *Energy Conversion and Management*, v. 322, p. 119114, 2024. ISSN 0196-8904. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0196890424010550>>. Citado na página 17.
- LIU, F. et al. Correct and remap solar radiation and photovoltaic power in china based on machine learning models. *Applied Energy*, v. 312, p. 118775, 2022. ISSN 0306-2619. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0306261922002240>>. Citado na página 16.
- LIU, X.; WANG, W. Deep time series forecasting models: A comprehensive survey. *Mathematics*, v. 12, n. 10, 2024. ISSN 2227-7390. Disponível em: <<https://www.mdpi.com/2227-7390/12/10/1504>>. Citado na página 21.
- MAHESH, B. Machine learning algorithms -a review. *International Journal of Science and Research (IJSR)*, v. 9, 01 2019. Citado na página 25.
- MATZAVELA, V.; ALEPIS, E. Decision tree learning through a predictive model for student academic performance in intelligent m-learning environments. *Computers and Education: Artificial Intelligence*, v. 2, p. 100035, 2021. ISSN 2666-920X. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S2666920X21000291>>. Citado na página 26.
- MIENYE, I. D.; SWART, T. G.; OBAIDO, G. Recurrent neural networks: A comprehensive review of architectures, variants, and applications. 2024. Citado na página 27.
- MIRAF'TABZADEH, S. M. et al. A day-ahead photovoltaic power prediction via transfer learning and deep neural networks. *Forecasting*, v. 5, n. 1, p. 213–228, 2023. ISSN 2571-9394. Disponível em: <<https://www.mdpi.com/2571-9394/5/1/12>>. Citado na página 40.
- MISHRA, A. et al. *SPIRIT: Short-term Prediction of solar IRradIance for zero-shot Transfer learning using Foundation Models*. 2025. Disponível em: <<https://arxiv.org/abs/2502.10307>>. Citado 2 vezes nas páginas 31 e 41.
- MISHRA, S.; PALANISAMY, P. Multi-time-horizon solar forecasting using recurrent neural network. In: *2018 IEEE Energy Conversion Congress and Exposition (ECCE)*. [S.l.: s.n.], 2018. p. 18–24. ISSN 2329-3748. Citado na página 34.
- MUDELSEE, M. Trend analysis of climate time series: A review of methods. *Earth-Science Reviews*, v. 190, p. 310–322, 2019. ISSN 0012-8252. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0012825218303726>>. Citado na página 20.
- MYLONAS, D. Solar generation forecasting with transfer learning. Aristotle University of Thessaloniki, 2023. Disponível em: <<http://ikee.lib.auth.gr/record/349447>><http://ikee.lib.auth.gr/record/349447>. Citado na página 39.

- NAZARI, A. et al. Groundwater level forecasting using empirical mode decomposition and wavelet-based long short-term memory (lstm) neural networks. *Groundwater for Sustainable Development*, v. 28, p. 101397, 2025. ISSN 2352-801X. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S2352801X24003205>>. Citado na página 47.
- NIE, Y. et al. Sky image-based solar forecasting using deep learning with heterogeneous multi-location data: Dataset fusion versus transfer learning. *Applied Energy*, v. 369, p. 123467, 2024. ISSN 0306-2619. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S030626192400850X>>. Citado 2 vezes nas páginas 17 e 41.
- NVIDIA Corporation. *Getting Started With Deep Learning Performance*. 2023. Acesso em: 20 fev. 2026. Disponível em: <<https://docs.nvidia.com/deeplearning/performance/pdf/Getting-Started-With-Deep-Learning-Performance.pdf>>. Citado na página 49.
- PALETTA, Q. et al. Advances in solar forecasting: Computer vision with deep learning. *Advances in Applied Energy*, v. 11, p. 100150, set. 2023. ISSN 2666-7924. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S266679242300029X>>. Citado na página 27.
- PAZIKADIN, A. R. et al. Solar irradiance measurement instrumentation and power solar generation forecasting based on artificial neural networks (ann): A review of five years research trend. *Science of The Total Environment*, v. 715, p. 136848, maio 2020. ISSN 0048-9697. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0048969720303582>>. Citado 3 vezes nas páginas 17, 22 e 24.
- RABAIA, M. K. H. et al. Environmental impacts of solar energy systems: A review. *Science of The Total Environment*, v. 754, p. 141989, 2021. ISSN 0048-9697. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0048969720355182>>. Citado na página 16.
- RAHIMI, N. et al. A comprehensive review on ensemble solar power forecasting algorithms. *Journal of Electrical Engineering & Technology*, v. 18, n. 2, p. 719–733, Mar 2023. ISSN 2093-7423. Disponível em: <<https://doi.org/10.1007/s42835-023-01378-2>>. Citado na página 17.
- RAI, A.; SHRIVASTAVA, A.; JANA, K. A cnn-bilstm based deep learning model for mid-term solar radiation prediction. *International Transactions on Electrical Energy Systems*, v. 31, 10 2020. Citado na página 38.
- RAJ, J.; QIAN, L.; IBRAHIM, Z. Fine-tuning – a transfer learning approach. 11 2024. Citado na página 32.
- RAJAGUKGUK, R. A.; RAMADHAN, R. A. A.; LEE, H.-J. A review on deep learning models for forecasting time series data of solar irradiance and photovoltaic power. *Energies*, v. 13, n. 24, 2020. ISSN 1996-1073. Citado na página 37.
- RAJAPRASAD, S.; MUKKAMALA, R. A hybrid deep learning framework for modeling the short term global horizontal irradiance prediction of a solar power plant in india. *Polityka Energetyczna – Energy Policy Journal*, v. 26, n. 3, p. 101–116, 2023. ISSN 1429-6675. Disponível em: <<https://doi.org/10.33223/epj/168115>>. Citado na página 39.
- SALEHINEJAD, H. et al. *Recent Advances in Recurrent Neural Networks*. 2018. Disponível em: <<https://arxiv.org/abs/1801.01078>>. Citado na página 26.

- SAMEK, W. et al. Explaining deep neural networks and beyond: A review of methods and applications. *Proceedings of the IEEE*, v. 109, n. 3, p. 247–278, 2021. Citado na página 26.
- SARMAS, E. et al. Transfer learning strategies for solar power forecasting under data scarcity. *Scientific Reports*, v. 12, n. 1, p. 14643, Aug 2022. ISSN 2045-2322. Disponível em: <<https://doi.org/10.1038/s41598-022-18516-x>>. Citado na página 40.
- SEHRAWAT, N.; VASHISHT, S.; SINGH, A. Solar irradiance forecasting models using machine learning techniques and digital twin: A case study with comparison. *International Journal of Intelligent Networks*, v. 4, p. 90–102, 2023. ISSN 2666-6030. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S2666603023000064>>. Citado na página 15.
- SENGUPTA, M. et al. *Best Practices Handbook for the Collection and Use of Solar Resource Data for Solar Energy Applications: Third Edition*. [S.l.: s.n.], 2021. Citado na página 23.
- SHANKER, M. et al. Transfer learning for solar energy forecasting: A comparative study of bgru and t-bgru models. In: *2025 International Conference on Sustainable Energy Technologies and Computational Intelligence (SETCOM)*. [S.l.: s.n.], 2025. p. 1–6. Citado na página 41.
- SINGLA, P.; DUHAN, M.; SAROHA, S. An ensemble method to forecast 24-h ahead solar irradiance using wavelet decomposition and bilstm deep learning network. *Earth Science Informatics*, v. 15, n. 1, p. 291–306, Mar 2022. ISSN 1865-0481. Disponível em: <<https://doi.org/10.1007/s12145-021-00723-1>>. Citado na página 35.
- SZOSTEK, K. et al. Analysis of the effectiveness of arima, sarima, and svr models in time series forecasting: A case study of wind farm energy production. *Energies*, v. 17, n. 19, 2024. ISSN 1996-1073. Disponível em: <<https://www.mdpi.com/1996-1073/17/19/4803>>. Citado na página 24.
- TAGHVAEE, V. M. et al. Fossil fuel price policy and sustainability: energy, environment, health and economy. *International Journal of Energy Sector Management*, Emerald Publishing Limited, v. 17, n. 2, p. 371–409, Jan 2023. ISSN 1750-6220. Disponível em: <<https://doi.org/10.1108/IJESM-09-2021-0012>>. Citado na página 15.
- TANG, Y. et al. Photovoltaic power forecasting: A hybrid deep learning model incorporating transfer learning strategy. *Renewable and Sustainable Energy Reviews*, v. 162, p. 112473, 2022. ISSN 1364-0321. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1364032122003781>>. Citado 2 vezes nas páginas 32 e 41.
- TEMPORAL-NETO, A. et al. Pin: A new metric to evaluate solar irradiance forecast models. *Energy Reports*, v. 13, p. 2307–2315, 2025. ISSN 2352-4847. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S2352484725000770>>. Citado na página 33.
- TYRALIS, H.; PAPACHARALAMPOUS, G. A review of predictive uncertainty estimation with machine learning. *Artificial Intelligence Review*, v. 57, n. 4, p. 94, Mar 2024. ISSN 1573-7462. Disponível em: <<https://doi.org/10.1007/s10462-023-10698-8>>. Citado na página 23.
- VANCSURA, L.; TATAY, T.; BAREITH, T. Navigating ai-driven financial forecasting: A systematic review of current status and critical research gaps. *Forecasting*, v. 7, n. 3, 2025. ISSN 2571-9394. Disponível em: <<https://www.mdpi.com/2571-9394/7/3/36>>. Citado na página 27.

WANG, J. et al. Transfer learning with dynamic distribution adaptation. *ACM Trans. Intell. Syst. Technol.*, Association for Computing Machinery, New York, NY, USA, v. 11, n. 1, feb 2020. ISSN 2157-6904. Disponível em: <<https://doi.org/10.1145/3360309>>. Citado na página 29.

WANG, Y. et al. *Deep Time Series Models: A Comprehensive Survey and Benchmark*. 2025. Disponível em: <<https://arxiv.org/abs/2407.13278>>. Citado 2 vezes nas páginas 20 e 21.

WANG, Z. et al. Prediction of solar radiation in qinghai lake area based on bilstm-attention method. In: *Proceedings of the 2019 3rd International Conference on Advances in Image Processing*. New York, NY, USA: Association for Computing Machinery, 2020. (ICAIP '19), p. 149–154. ISBN 9781450376754. Disponível em: <<https://doi.org/10.1145/3373419.3373439>>. Citado na página 37.

WEISS, C. H. Discrete-valued time series. *Entropy (Basel)*, v. 25, n. 12, nov. 2023. Citado na página 20.

XU, J. Systematic analysis and application prospect of decision tree. *Highlights in Science, Engineering and Technology*, v. 71, p. 163–170, Nov. 2023. Disponível em: <<https://drpress.org/ojs/index.php/HSET/article/view/12687>>. Citado na página 26.

XU, J. et al. A deep learning methodology for automatic extraction and discovery of technical intelligence. *Technological Forecasting and Social Change*, v. 146, p. 339–351, 2019. ISSN 0040-1625. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0040162518320389>>. Citado na página 37.

XU, S.; SONG, Y.; HAO, X. A comparative study of shallow machine learning models and deep learning models for landslide susceptibility assessment based on imbalanced data. *Forests*, v. 13, n. 11, 2022. ISSN 1999-4907. Disponível em: <<https://www.mdpi.com/1999-4907/13/11/1908>>. Citado na página 26.

YI, S. et al. Environmental concerns in the united states: Can renewable energy, fossil fuel energy, and natural resources depletion help? *Gondwana Research*, v. 117, p. 41–55, 2023. ISSN 1342-937X. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1342937X23000308>>. Citado na página 15.

YU, Y.; CAO, J.; ZHU, J. An lstm short-term solar irradiance forecasting under complicated weather conditions. *IEEE Access*, v. 7, p. 145651–145666, 2019. ISSN 2169-3536. Citado 3 vezes nas páginas 27, 33 e 34.

ZANG, H. et al. Short-term global horizontal irradiance forecasting based on a hybrid cnn-lstm model with spatiotemporal correlations. *Renewable Energy*, v. 160, p. 26–41, 2020. ISSN 0960-1481. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0960148120308557>>. Citado 3 vezes nas páginas 15, 16 e 37.

ZARZYCKI, K.; ŁAWRYŃCZUK, M. Long short-term memory neural networks for modeling dynamical processes and predictive control: A hybrid physics-informed approach. *Sensors*, v. 23, n. 21, 2023. ISSN 1424-8220. Disponível em: <<https://www.mdpi.com/1424-8220/23/21/8898>>. Citado na página 27.

ZHANG, H. et al. A novel hybrid transformer-based framework for solar irradiance forecasting under incomplete data scenarios. *IEEE Transactions on Industrial Informatics*, v. 20, n. 6, p. 8605–8615, 2024. Citado na página 16.

ZHANG, L. et al. Transfer learning in very-short-term solar forecasting: Bridging single site data to diverse geographical applications. *Applied Energy*, v. 377, p. 124353, 2025. ISSN 0306-2619. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0306261924017367>>. Citado na página 42.

ZHAO, Z. et al. A comparison review of transfer learning and self-supervised learning: Definitions, applications, advantages and limitations. *Expert Systems with Applications*, v. 242, p. 122807, 2024. ISSN 0957-4174. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0957417423033092>>. Citado 2 vezes nas páginas 30 e 31.

# APÊNDICE A – Análises Estatísticas

## A.1 Média e Desvio Padrão

A Tabela 20 apresenta a média e desvio padrão dos conjuntos de dados usados no cenário #1 (transferência de aprendizado entre estações de distintas zonas bioclimáticas).

Tabela 20 – GHI Médio e Desvio Padrão para cada Estação Climatológica.

Cidade	Estação do Ano	GHI Médio (W/m <sup>2</sup> )	Desvio Padrão
Curitiba	Outono	1142.77	1047.79
	Inverno	973.17	897.78
	Primavera	1049.40	1041.65
	Verão	1204.21	1087.43
Santa Maria	Outono	920.28	854.96
	Inverno	847.78	839.55
	Primavera	1189.65	1156.47
	Verão	1532.63	1224.34
Juiz de Fora	Outono	1151.40	931.50
	Inverno	1234.10	1013.13
	Primavera	1202.26	1154.58
	Verão	1209.88	1085.11
Brasília	Outono	1461.88	1028.36
	Inverno	1557.99	1032.69
	Primavera	1462.28	1188.65
	Verão	1352.67	1092.41
Campos dos Goytacazes	Outono	699.55	884.83
	Inverno	947.56	997.20
	Primavera	1457.51	1191.76
	Verão	1355.60	1194.06
Campo Grande	Outono	1258.72	1014.35
	Inverno	1309.75	997.96
	Primavera	1455.35	1182.80
	Verão	1565.54	1214.57
Teresina	Outono	779.64	841.91
	Inverno	1276.21	846.90
	Primavera	1309.71	909.57
	Verão	1020.69	835.33
Manaus	Outono	827.47	852.36
	Inverno	1063.79	984.11
	Primavera	1078.95	997.10
	Verão	949.99	985.37

## A.2 Correlação

Em termos de correlação, algumas variáveis apresentam uma relação direta com a irradiação. Na Figura 19, utilizando a estação de Teresina como exemplo, ao analisar a correlação entre as colunas do conjunto de dados, nota-se que a velocidade do vento e, principalmente, a temperatura, possuem uma correlação forte e positiva com a irradiação, o que, *a priori*, é esperado em relação à temperatura. Em contrapartida, a umidade apresenta uma correlação forte e

negativa com a irradiação, indicando uma variação inversamente proporcional.

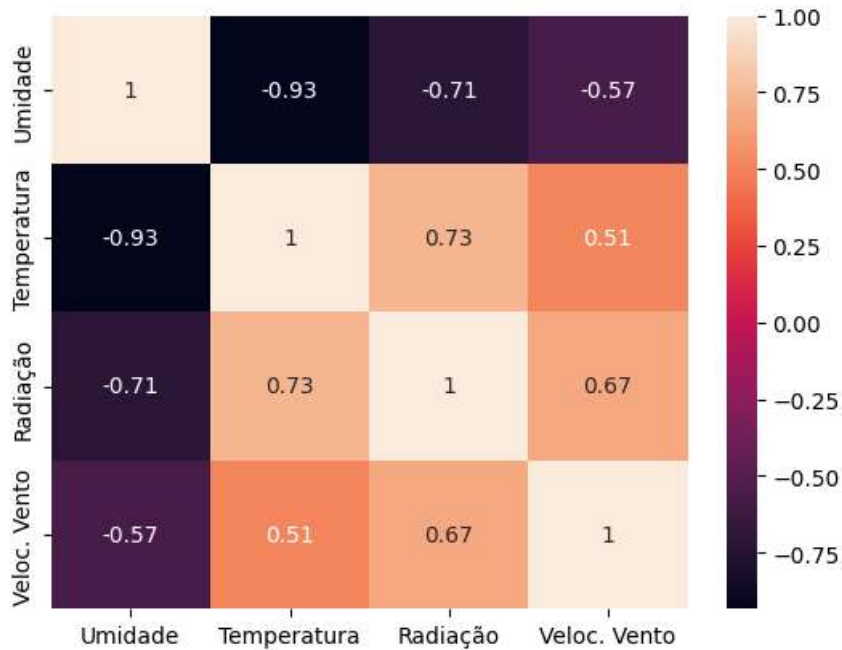


Figura 19 – Análise de Correlação das Variáveis Explicativas do Conjunto de Dados de Teresina.

### A.3 Estacionariedade

Em um primeiro momento, foi proposta uma análise visual das séries em si, com um *resample* dos dados, de horas para dias, para facilitar a visualização. Verificou-se que, visualmente, as séries são estacionárias; porém, existem períodos que demonstram alguma variação nos dados. No exemplo da série de Brasília, na Figura 20, do período de maio a outubro, existe um comportamento diferente do restante da série.

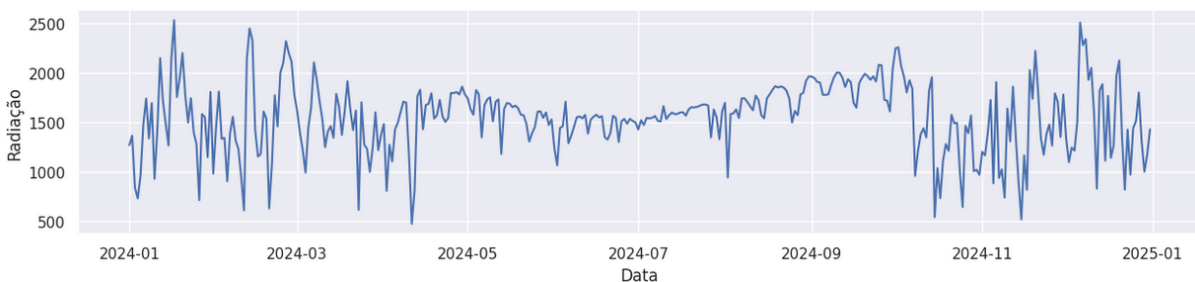


Figura 20 – *Resample* dos dados de Brasília do ano todo, de horas para dias.

Como segunda etapa, a estacionariedade foi medida utilizando o teste *Ad Fuller*, um teste estatístico de hipótese que verifica a existência de uma raiz unitária na série, fazendo com que a série não seja estacionária, sendo essa a hipótese nula. Caso a hipótese nula seja confirmada, então os dados não possuem estacionariedade. Caso contrário, ou seja, se a hipótese nula for rejeitada em detrimento da hipótese alternativa, os dados têm estacionariedade.

Visivelmente, a maioria das séries temporais apresentou algum componente de não-estacionariedade, mas o teste de *Ad Fuller* demonstrou que os dados dessas estações são esta-

cionários, não considerando o *resample* por dia, mas sim a série original (com granularidade horária). O resultado pode ser verificado na Figura 21, em que o  $p$ -valor obtido esteve abaixo de 0,05.

```
adfuller_test(df[["radiacao", "data_hora"]].set_index("data_hora"))
Teste estatístico ADF : -7.260340697890236
p-valor : 1.688009664643017e-10
Num Lags : 31
Numero de observações : 3986
Fortes evidências contra a hipótese nula(Ho), ou seja, pode ser rejeitada. Os dados não possuem uma raiz unitária, por tanto tem estacionariedade.
```

Figura 21 – Teste *AdFuller* para a Estação Climatológica de Brasília.

A série original (horária) não apresentou evidências suficientes para ser classificada como não estacionária. Portanto, não foi necessário aplicar etapas de diferenciação para remover a tendência e a sazonalidade.

## A.4 Tendência e Sazonalidade

Esta etapa visou separar a série temporal em componentes de tendência, sazonalidade e resíduo. Como previamente mencionado, os dados de irradiação solar apresentam um comportamento sazonal, sem tendência de alta ou baixa, com a média e a variância constantes, apesar de conter intervalos temporais com algumas variações. Na Figura 22, esses componentes são destacados em uma faixa arbitrária do primeiro trimestre do ano de 2024.

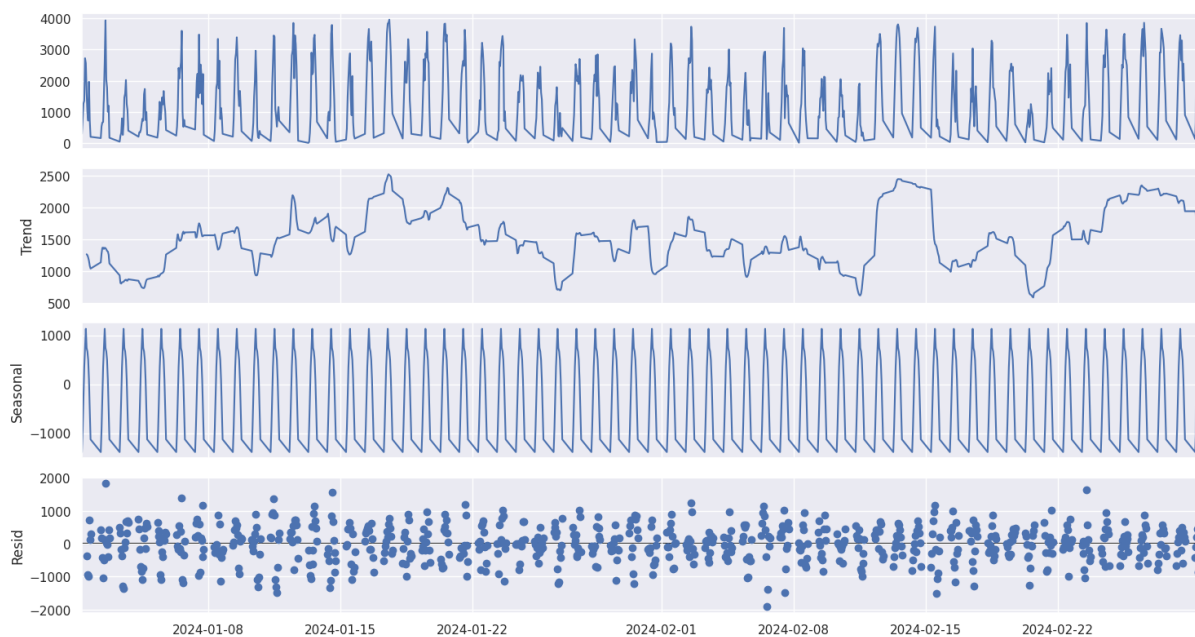


Figura 22 – Decomposição da série temporal de Brasília em Janeiro/Março de 2024.

Conforme mostrado no gráfico de sazonalidade da Figura 22, os dados apresentam uma evidente sazonalidade que se mantém constante, enquanto no gráfico de tendência não foi possível observar uma tendência expressiva que indique uma grande variação nas estatísticas dos dados.

Como a variância não muda ao longo do tempo e a tendência não acompanha a sazonalidade, a escolha do método de decomposição foi o aditivo.

# APÊNDICE B – Parametrização do SARIMA

## B.1 ACF e PACF

O intuito desta etapa é observar os gráficos de Autocorrelação (ACF, do inglês *AutoCorrelation Function*) e Autocorrelação Parcial (PACF, do inglês *Partial AutoCorrelation Function*), disponíveis na Figura 23, obtendo evidências para a configuração dos parâmetros do modelo de *baseline* (SARIMA).

Os parâmetros não sazonais são:  $\mathbf{p}$ , que é o parâmetro de autorregressão;  $\mathbf{q}$ , que é o parâmetro de média móvel; e  $\mathbf{d}$ , que é o número de diferenciações aplicadas aos dados. Em relação a esses parâmetros, o gráfico de ACF mostra uma forte correlação com ciclos de 12 horas, mas não apresenta uma queda expressiva conforme a série avança; em vez disso, apresenta uma queda exponencial, não havendo picos em defasagens específicas. Portanto, um parâmetro de média móvel  $q = 0$ . O gráfico de PACF sugere um parâmetro  $p = 1$ , dado que possui somente um pico relevante dentro dos múltiplos de 12 (janela de horário utilizada). Ou seja, o valor da correlação cai abruptamente após a primeira defasagem. O parâmetro restante,  $d$ , foi configurado como 0, dado que nenhuma etapa de diferenciação foi aplicada aos dados, os quais já eram estacionários de acordo com o teste realizado anteriormente.

Em relação aos parâmetros sazonais, observou-se que, no gráfico ACF, existem picos relevantes na sazonalidade (12 horas) presentes nos dados, indicando um parâmetro  $Q = 1$ , diferentemente do gráfico PACF, que apresenta um pico não tão relevante em relação ao período sazonal, indicando um parâmetro  $P = 1$ . Sobre o parâmetro que representa a diferenciação, como não houve etapas para remover a tendência, o valor mais adequado para o parâmetro é 0. Por fim, o parâmetro indicando a defasagem dos dados tem o valor 12, que é o valor em que a série começa a repetir.

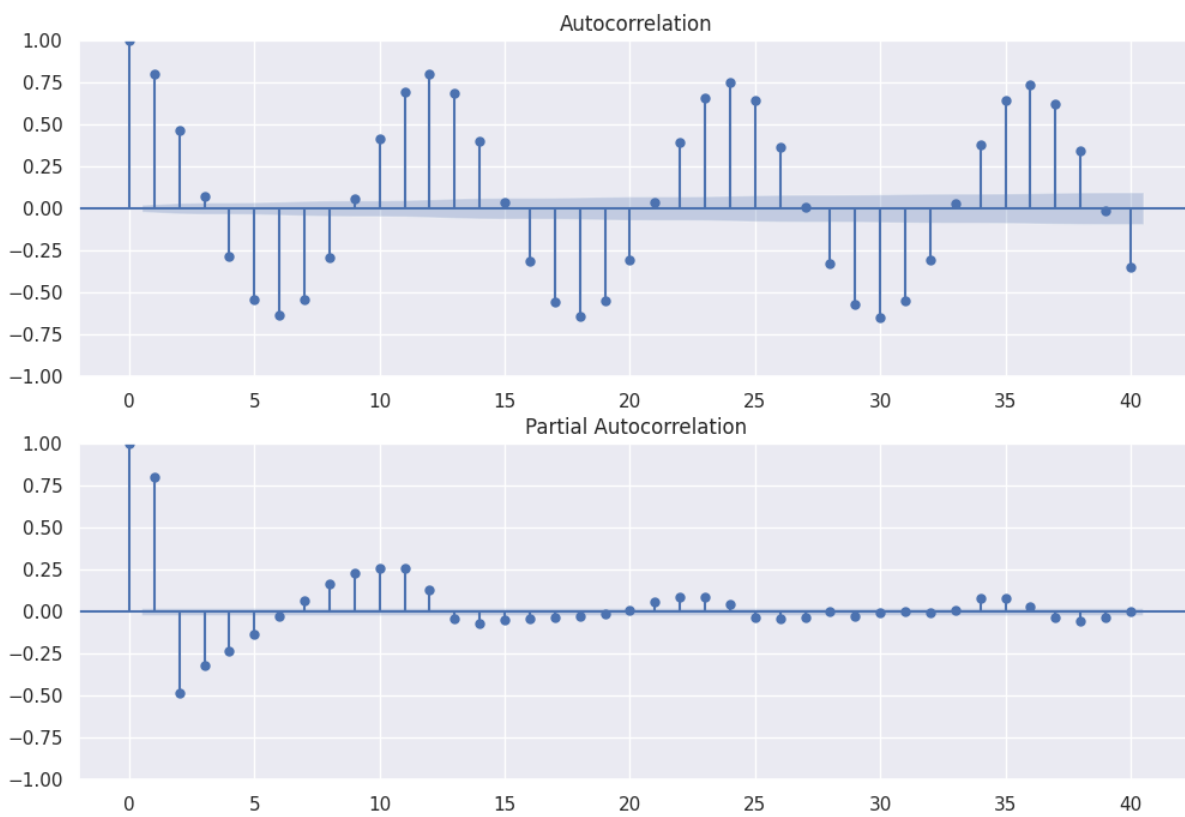


Figura 23 – Gráficos de Autocorrelação e Autocorrelação Parcial de Brasília

## APÊNDICE C – xLSTM

O modelo xLSTM foi testado como uma alternativa adicional de resultados, sendo uma implementação recente e promissora em relação ao aprimoramento do histórico dentro da rede LSTM tradicional. Os dados utilizados foram da estação de Teresina, por ser um dos melhores conjuntos em termos de variabilidade, com o mesmo horizonte de previsão das redes anteriores. Foram realizadas as mesmas cinco execuções, com os seguintes resultados médios: MSE de 0,200, RMSE de 0,447, MAE de 0,381 e  $R^2$  de 0,581. O artigo menciona que os experimentos com o xLSTM foram comparados com métodos existentes que tem foco em modelos de linguagem. Além disso, as técnicas propostas para mitigar os problemas expostos, como o problema da busca do vizinho mais próximo, são da mesma natureza (problemas de linguagem). Portanto, essas abordagens podem não ser adequadas para a previsão de irradiação em séries temporais, o que pode explicar o baixo desempenho quando comparadas a modelos tradicionais.