



UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA MECÂNICA
MESTRADO EM ENGENHARIA MECÂNICA

VICTOR OLIVEIRA SANTOS

**EFEITO DO VIÉS GEOGRÁFICO EM UMA REDE NEURAL PROFUNDA ATRAVÉS
DA TRANSFERÊNCIA DE APRENDIZADO ENTRE AS ESTAÇÕES DE HANFORD
(EUA) E PETROLINA (BRASIL)**

FORTALEZA

2022

VICTOR OLIVEIRA SANTOS

EFEITO DO VIÉS GEOGRÁFICO EM UMA REDE NEURAL PROFUNDA ATRAVÉS DA
TRANSFERÊNCIA DE APRENDIZADO ENTRE AS ESTAÇÕES DE HANFORD (EUA) E
PETROLINA (BRASIL)

Dissertação apresentada ao Curso de Mestrado em Engenharia Mecânica do Programa de Pós-Graduação em Engenharia Mecânica do Centro de Tecnologia da Universidade Federal do Ceará, como requisito parcial à obtenção do título de mestre em Engenharia Mecânica. Área de Concentração: Processos, Equipamentos e Sistemas para Energias Renováveis.

Orientador: Prof. Dr. Paulo Alexandre Costa Rocha

FORTALEZA

2022

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Biblioteca Universitária
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

- S239e Santos, Victor Oliveira.
Efeito do viés geográfico em uma rede neural profunda através da transferência de aprendizado entre as estações de Hanford (EUA) e Petrolina (Brasil) / Victor Oliveira Santos. – 2022.
72 f. : il. color.
- Dissertação (mestrado) – Universidade Federal do Ceará, Centro de Tecnologia, Programa de Pós-Graduação em Engenharia Mecânica, Fortaleza, 2022.
Orientação: Prof. Dr. Paulo Alexandre Costa Rocha.
1. Energias renováveis. 2. Aprendizagem profunda. 3. Transferência de aprendizado. 4. Rede neural convolucional. 5. Previsão de irradiância solar. I. Título.
- CDD 620.1
-

VICTOR OLIVEIRA SANTOS

EFEITO DO VIÉS GEOGRÁFICO EM UMA REDE NEURAL PROFUNDA
ATRAVÉS DA TRANSFERÊNCIA DE APRENDIZADO ENTRE AS ESTAÇÕES DE
HANFORD (EUA) E PETROLINA (BRASIL)

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Mecânica, do Centro de Tecnologia da Universidade Federal do Ceará, como requisito parcial para a obtenção do Título de Mestre em Engenharia Mecânica. Área de Concentração: Processos, Equipamentos e Sistemas para Energias Renováveis.

Aprovada em 24/06/20222

BANCA EXAMINADORA

Prof. Dr. Paulo Alexandre Costa Rocha (Orientador)
Universidade Federal do Ceará (UFC)

Prof.^a Dra. Carla Freitas de Andrade
Universidade Federal do Ceará (UFC)

Prof. Dr. Elineudo Pinho de Moura
Universidade Federal do Ceará (UFC)

À minha família, amigos e colegas. A todos aqueles a quem este trabalho possa, de alguma forma, ajudar.

AGRADECIMENTOS

Aos meus pais, Maria do Carmo Oliveira Santos e Francisco de Assis dos Santos, por sempre acreditarem em mim, por fomentarem minha educação e me apoiarem sempre quando precisei.

À minha esposa Amanda. Obrigado por estar comigo durante esses anos, nos bons e maus momentos.

Ao meu orientador, Professor Dr. Paulo Alexandre Costa Rocha, pela paciência e excelência com a qual me orientou, pois sem elas este trabalho não seria possível.

Israel, Leandro, Pedro e Thales, obrigado pelos acalorados debates científicos diários.

À todas as professoras e professores que me guiaram desde o tempo de escola, até o ensino superior. O ensinamento proporcionado pelos senhores docentes foi imprescindível à minha formação acadêmica.

À FUNCAP pelo suporte financeiro.

Ao departamento de pós-graduação do curso de engenharia mecânica, por serem sempre solícitos e acessíveis.

A todos, muito obrigado.

*“We can only see a short distance ahead, but we
can see plenty there that needs to be done.”*

(Alan Turing)

RESUMO

Com o agravamento do aquecimento global, são necessários desenvolvimentos de estratégias que visam limitar a elevação da temperatura terrestre em no máximo 1,5 °C até o fim do século XXI. Para que isso seja possível, o desenvolvimento de tecnologias em energias renováveis e não poluentes é um pilar fundamental. Entretanto, a natureza estocástica e intermitente das fontes de energias renováveis torna a sua implementação junto a matriz energética já instalada complexa, impactando diretamente nos custos. Desta forma, é indispensável o conhecimento prévio do potencial de geração de energia elétrica em determinada área de interesse, para que seja então possível o emprego de estratégias para melhor gerir este tipo de energia. Assim, o presente trabalho busca prever a energia elétrica gerada analisando o efeito do viés geográfico através da transferência de aprendizado de uma rede neural convolucional para estimação de irradiâncias global horizontal e direta normal para as localidades de Hanford, na Califórnia, e Petrolina, em Pernambuco, utilizando imagens de satélite geoestacionário. Os resultados obtidos pela rede neural profunda foram comparados com modelos físicos de referência através de métricas adequadas para aferição de erro. A rede convolucional quando treinada e testada com dados de Hanford obteve resultados de RMSE para GHI e DNI de 88,17 W/m² e 178,40 W/m² respectivamente. Esses resultados estão em conformidade com os obtidos pelos modelos de referência Solcast e PSM, indicando a sua viabilidade da rede CNN-LSTM na estimação de irradiância solar, em especial para a global horizontal. Entretanto, quando a rede é treinada e testada com dados de localidades distintas, o viés geográfico presente nos dados mostra-se como um empecilho, não sendo possível obter resultados satisfatórios, quando comparados com os dos modelos de referência.

Palavras-chave: Energias renováveis. Aprendizagem profunda. Rede neural convolucional. Previsão de irradiância solar. Satélite GOES-16. Transferência de aprendizado.

ABSTRACT

Due to the worsening of global warming, the development of new strategies aimed to damper the rise of Earth's temperature to a maximum of 1,5°C by the end of 21st century. To comply with this goal, the development of technologies focused on renewable and less pollutant energies is crucial. However, the stochastic and sporadic behaviour makes this type of energy hard to merge together with the already implemented energy mix, making a direct impact on its implementation costs. Therefore, it is imperative to previously know the solar irradiance at a place of interest, so better management strategies for this type of energy could be employed. In this way, the present study seeks to forecast the solar energy analysing the geographic bias effect in the learning transfer by the assessment of the capacity of a convolutional neural network to estimate solar global horizontal and solar direct normal irradiances using geostationary satellite imagery for the location of Hanford, California. The hybrid CNN-LSTM model, when trained and tested with data from Hanford, achieved values for RMSE of 88,17 W/m² e 178,40 W/m² respectively. These results are in accordance with those obtained by the reference models Solcast and PSM. However, when the CNN-LSTM model is trained and test with data from different locations, the geographic bias showed to be a hindrance, preventing the convolutional network to achieve satisfactory results, when compared with those from the reference models.

Keywords: Renewable energies. Deep learning. Convolutional neural network. Solar irradiance estimation. GOES-16 satellite. Learning transfer

LISTA DE FIGURAS

Figura 1 – Participação de cada fonte de energia renovável na geração de energia elétrica no ano de 2019 em termos percentuais.	17
Figura 2 – Gráfico à esquerda mostra a variação da produção de energia renovável para cada fonte analisada. Gráfico à direita mostra a variação na produção para cada fonte de energia entre o período de 2018 e 2019 em unidade de TWh.	18
Figura 3 – Evolução da produção científica anual para previsão de energia solar.	21
Figura 4 – Termos em tendências nos últimos 10 anos.	22
Figura 5 – Produção científica por autor.	23
Figura 6 – Periódicos com mais publicações no período analisado.	23
Figura 7 – Produção científica por país para o período analisado.	24
Figura 8 – Ângulo de zênite solar, θ_z , para uma determinada região.	25
Figura 9 – Diferentes tipos de irradiância solar. A irradiância global tem como componentes DNI e DHI, sendo modelado em função delas juntamente com o ângulo de zênite.	25
Figura 10 – Comparação entre diferentes paradigmas aplicados no campo de inteligência artificial.	28
Figura 11 – Comparação entre (a) neurônio biológico e (b) neurônio artificial.	28
Figura 12 – Esquematização de uma rede MLP, onde cada círculo representa um neurônio artificial.	29
Figura 13 – Exemplo de hierarquia espacial identificada pela rede convolucional.	31
Figura 14 – Exemplo de camada de convolução.	32
Figura 15 – Exemplo de camada de <i>max pooling</i>	32
Figura 16 – Exemplo de rede recorrente convencional RNN.	33
Figura 17 – Exemplo de rede recorrente LSTM.	34
Figura 18 – Fluxograma da metodologia utilizada.	41
Figura 19 – A frota de satélites geoestacionários GOES para geomonitoramento e sua área de cobertura.	42
Figura 20 – Diferentes resoluções para cada canal do satélite geoestacionário GOES-16 sobre a localidade de Hanford, Califórnia.	43
Figura 21 – Estações componentes da rede brasileira SONDA. Destaca-se a estação de Petrolina, em Pernambuco.	44

Figura 22 – Estações componentes da rede estadunidense SOLRAD.	45
Figura 23 – Dados do satélite GOES-16 não normalizados à esquerda e normalizados à direita.	47
Figura 24 – Exemplo de imagem com raio igual à 6 km.	48
Figura 25 – Exemplo de rede neural híbrida CNN-LSTM.	49
Figura 26 – Gráfico de dispersão das irradiâncias (a) GHI e (b) DNI previstas pelo modelo CNN-LSTM usando dados de Hanford para treino e teste.	54
Figura 27 – Gráfico de dispersão das irradiâncias (a) GHI e (b) DNI previstas pelo modelo CNN-LSTM na forma PH.	55
Figura 28 – Gráfico de dispersão dos dados previstos pelo modelo Solcast e mensurados <i>in loco</i> pela estação de Hanford para (a) GHI e (b) DNI.	57
Figura 29 – Gráfico de dispersão dos dados previstos pelo modelo PSM e mensurados <i>in loco</i> pela estação de Hanford para (a) GHI e (b) DNI.	58
Figura 30 – Gráfico de dispersão das irradiâncias (a) GHI e (b) DNI previstas pelo modelo CNN-LSTM na forma HP ₁	59
Figura 31 – Gráfico de dispersão das irradiâncias (a) GHI e (b) DNI previstas pelo modelo CNN-LSTM na forma HP ₂	60
Figura 32 – Gráfico em forma de alvo para resultados de previsão para Hanford. Imagem (a) apresenta resultados para GHI e imagem (b) apresenta resultados para DNI.	60
Figura 33 – Gráfico em forma de alvo para resultados de previsão para Petrolina. Imagem (a) apresenta resultados para GHI e imagem (b) apresenta resultados para DNI.	61
Figura 34 – Distribuição dos valores de ângulo de zênite para Hanford (à esquerda) e para Petrolina (à direita) durante o ano de 2018.	63
Figura 35 – Distribuição dos valores índice de céu claro para Hanford (à esquerda) e para Petrolina (à direita) durante o ano de 2018.	63

LISTA DE TABELAS

Tabela 1 – Escopo, banco de dados, termos e restrições impostas referentes à análise bibliométrica realizada com o <i>bibliometrix</i> (DONTHU <i>et al.</i> , 2021).	21
Tabela 2 – Configurações para os 16 canais do satélite GOES-16 (HILLGER; SCHMIT, 2004; PADULA; CAO, 2015; KALLURI <i>et al.</i> , 2018).	43
Tabela 3 – Resumo das informações das estações meteorológicas utilizadas.	46
Tabela 4 – Comparação entre as quantidades de observações para ambas estações meteorológicas.	46
Tabela 5 – Resumo das configurações utilizadas na rede CNN-LSTM.	49
Tabela 6 – Quantidade de dados utilizados para cada previsão de irradiância para PH e HP ₂	53
Tabela 7 – Quantidades de observações utilizadas para treino e teste da rede convolucional.	53
Tabela 8 – Resultados de RMSE e MBE para os modelos analisados em Hanford.	62
Tabela 9 – Resultados de RMSE e MBE para os modelos analisados em Petrolina (ROCHA; SANTOS, 2022).	62

LISTA DE ABREVIATURAS E SIGLAS

CAMS	Serviço de Monitoramento Atmosférico Copernicus, tradução do inglês <i>Copernicus Atmosphere Monitoring Service</i>
CNN	Rede Neural Convolucional, tradução do inglês <i>Convolutional Neural Network</i>
DNI	Irradiância Normal Direta, tradução do inglês <i>Direct Normal Irradiance</i>
GHI	Irradiância Horizontal Global, tradução do inglês <i>Global Horizontal Irradiance</i>
GOES	Satélite Ambiental de Operação Geostacionária, tradução do inglês <i>Geostationary Operational Environmental Satellite</i>
IRENA	Agência Internacional de Energia Renováveis, tradução do inglês <i>International Renewable Energy Agency</i>
LSTM	Memória de Curto Prazo Longa, tradução do inglês <i>Long Short-Term Memory</i>
MAE	Erro Médio Absoluto, tradução do inglês <i>Mean Absolute Error</i>
MBE	Erro Médio por Viés, tradução do inglês <i>Mean Bias Error</i>
MLP	Perceptron de Multi Camadas, tradução do inglês <i>Multi Layer Perceptron</i>
NASA	Administração Nacional da Aeronáutica e Espaço, tradução do inglês <i>National Aeronautics and Space Administration</i>
NOAA	Administração Nacional Oceânica e de Atmosfera, tradução do inglês <i>National Oceanic and Atmospheric Administration</i>
NREL	Laboratório Nacional de Energia Renovável, tradução do inglês <i>National Renewable Energy Laboratory</i>
NRSDB	Banco de Dados Nacional de Radiação Solar, tradução do inglês <i>National Solar Radiation Database</i>
PSM	Modelo Solar Físico, tradução do inglês <i>Physical Solar Model</i>
RMSE	Raiz do Erro Quadrático Médio, tradução do inglês <i>Root Mean Square Error</i>
RNN	Redes Neurais Recorrentes, tradução do inglês <i>Recurrent Neural Networks</i>
SOLRAD	Radiação Solar, tradução do inglês <i>Solar Radiation</i>
SONDA	Organização Nacional de Dados Ambientais
SURFRAD	Rede de Coleção de Radiação de Superfície, tradução do inglês <i>Surface Radiation Budget Network</i>
UTC	Tempo Universal Coordenado, tradução do inglês <i>Coordinated Universal Time</i>

LISTA DE SÍMBOLOS

BSh	Designação de clima semi-árido quente segundo classificação Köppen-Geiger
BSk	Designação para clima semi-árido frio segundo classificação Köppen-Geiger
C_H	Valor medido por um canal do satélite GOES-16
$\overline{C_H}$	Valor médio das medições de um canal do satélite GOES-16
G	Irradiância incidente em um local de interesse
$G_{0,cs}$	Irradiância de céu claro para um local de interesse
HH	Indica que a rede CNN-LSTM foi treinada com dados de Hanford (H) e testada com dados de Hanford (H)
HP ₁	Indica que a rede CNN-LSTM foi treinada com dados de Hanford (H) e testada com dados de Petrolina (P) utilizando todas as observações de suas respectivas estações meteorológicas
HP ₂	Indica que a rede CNN-LSTM foi treinada com dados de Hanford (H) e testada com dados de Petrolina (P) utilizando a mesma quantidade de observações para treino e teste
$K_{t,cs}$	Índice de céu claro
n	Quantidade total de medições realizadas por um canal do satélite GOES-16
PP	Indica que a rede CNN-LSTM foi treinada com dados de Petrolina (P) e testada com dados de Petrolina (P)
S_{CH}	Desvio padrão dos valores medidos por um canal do satélite GOES-16
x_{norm}	Valor normalizado para as medições realizadas para um canal do satélite GOES-16
PH	Indica que a rede CNN-LSTM foi treinada com dados de Petrolina (P) e testada com dados de Hanford (H)
\hat{y}_i	Valor previsto de irradiância pelo modelo analisado
\hat{y}	Valor real de irradiância medido <i>in loco</i> por uma estação meteorológica

SUMÁRIO

1	INTRODUÇÃO	16
1.1	Aquecimento global e seus impactos	16
1.2	Energias renováveis como medidas para combate ao aquecimento global e às mudanças climáticas	16
1.3	Objetivos	20
<i>1.3.1</i>	<i>Objetivo geral</i>	<i>20</i>
<i>1.3.2</i>	<i>Objetivos específicos</i>	<i>20</i>
2	REVISÃO BIBLIOGRÁFICA	21
2.1	Estado-da-arte	21
2.2	Conceitos teóricos de irradiância solar	24
2.3	Índice de céu claro $k_{t,cs}$	24
2.4	Inteligência artificial	26
2.5	Aprendizagem de máquina	27
2.6	Aprendizagem profunda	29
<i>2.6.1</i>	<i>Camadas de convolução, pooling e densa</i>	<i>31</i>
<i>2.6.2</i>	<i>Redes neurais recorrentes</i>	<i>33</i>
2.7	Modelos de estimação de radiação e irradiância solar	35
<i>2.7.1</i>	<i>Modelos físicos para previsão de energia solar</i>	<i>35</i>
<i>2.7.2</i>	<i>Modelos de aprendizado de máquinas para previsão de energia solar</i>	<i>36</i>
<i>2.7.3</i>	<i>Modelos de aprendizagem profunda para previsão de energia solar</i>	<i>37</i>
3	METODOLOGIA	41
3.1	Imagens de Satélite GOES-16	41
3.2	Estação Meteorológica de Petrolina	42
3.3	Estação Meteorológica de Hanford	44
3.4	Rede neural convolucional	46
3.5	Normalização dos dados	47
3.6	Tamanho da imagem	48
3.7	Previsão dos valores de irradiância pelo modelo CNN-LSTM	49
3.8	Modelos de referência	50
3.9	Métricas de avaliação	51

3.10	Transferência de aprendizado	52
3.10.1	<i>Treino e teste com dados de Hanford</i>	52
3.10.2	<i>Treino com dados Petrolina e teste com dados de Hanford</i>	52
3.10.3	<i>Treino com dados de Hanford e teste com dados de Petrolina</i>	53
4	RESULTADOS	54
4.1	Resultado para previsão de irradiâncias: treino e teste com dados de Hanford	54
4.2	Resultados para previsão de irradiâncias: treino com dados de Petrolina e teste com dados de Hanford	55
4.3	Comparação com modelos de referência	56
4.4	Resultados para previsão de irradiâncias: treino com dados de Hanford e teste com dados de Petrolina	58
4.5	Viés geográfico da rede CNN-LSTM e transferência de aprendizado	62
5	CONCLUSÃO	65
	REFERÊNCIAS	67

1 INTRODUÇÃO

O desenvolvimento de tecnologias que possibilitem o uso de energias renováveis e não emissoras de gases poluentes é, atualmente, tópico importante de pesquisas. Devido ao agravamento do efeito estufa e suas consequências para a sociedade (como as mudanças climáticas), tem-se o entendimento que não é possível combater a elevação da temperatura terrestre sem uma mudança na forma de obtenção de energia elétrica, passando principalmente pela evolução de fontes de energias eólica e solar, sendo esta última fonte o tópico de estudo deste trabalho.

1.1 Aquecimento global e seus impactos

O efeito estufa é um fenômeno natural e indispensável para o surgimento da vida como temos conhecimento na Terra. Entretanto, ações antrópicas vêm potencializando os efeitos deste fenômeno de forma a por em risco o equilíbrio do ecossistema e da biodiversidade terrestre. O aquecimento global e os seus efeitos colaterais têm sido tópico de constante debate tanto na ciência quanto na sociedade. Os efeitos do aquecimento global na fauna e flora terrestre e marítima têm sido identificados há pelo menos duas décadas (WALTHER *et al.*, 2002). Ainda, a mudança climática ocasionada pelo aquecimento global pode ser notada de forma mais constante e intensa nos últimos anos através de eventos climáticos mais extremos, da redução na produção de alimentos e do impacto direto na qualidade de vida e saúde da população humana (VOHRA *et al.*, 2021; DUFFY *et al.*, 2019; ZHANG; ZHOU, 2019; WAGNER, 2020; MANISALIDIS *et al.*, 2020). Além disso, os gases poluentes e potencializadores do efeito estufa emitidos por ações antrópicas põem em risco a população global: segundo a Organização Mundial da Saúde, 90% das pessoas vivem em locais com baixa qualidade do ar, e em torno de 4,2 milhões de pessoas morrem todos os anos devido a poluição atmosférica (ORGANIZAÇÃO MUNDIAL DA SAÚDE, 2022), tendo seu risco potencializado por particulados finos vindo da queima de combustíveis fósseis (VOHRA *et al.*, 2021).

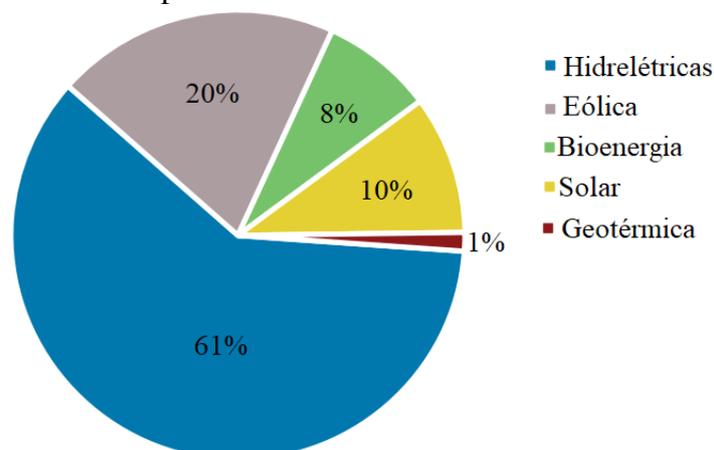
1.2 Energias renováveis como medidas para combate ao aquecimento global e às mudanças climáticas

A sociedade global vem buscando impor medidas mais restritivas em relação à emissão de poluentes a fim de reduzir os impactos do aquecimento terrestre. Uma das ações

de combate ao aumento da temperatura mais conhecidas é o Acordo de Paris de 2015. Este pacto, assinado pelas principais economias do mundo, tem como objetivo limitar o aumento da temperatura terrestre em no máximo de 1,5 °C através da redução da emissão dos gases do efeito estufa a partir do ano de 2020 até o fim do século XXI (NAÇÕES UNIDAS, 2015). Para que este objetivo seja alcançado, mudanças na matriz energética mundial devem ser feitas, passando, impreterivelmente, pelo desenvolvimento de energias renováveis e não poluentes a fim de reduzir a dependência de combustíveis fósseis.

Apesar dos riscos apontados pela emissão de gases poluentes, os combustíveis fósseis corresponderam a 95% do fornecimento de energia global em 2020 (SAIDI; OMRI, 2020). Ainda, a emissão de CO₂ devido à atividades antrópicas atingiu na década de 2010 a 2019 a sua maior emissão média anual (IPCC, 2021). Porém, é de se esperar que este cenário mude nos anos vindouros: segundo dados da Agência Internacional de Energia Renováveis, tradução direta do inglês de Agência Internacional de Energia Renováveis, tradução do inglês *International Renewable Energy Agency* (IRENA), a produção de energias renováveis cresceu mundialmente durante o período de 2018 e 2019, variando positivamente em 5,5% produzindo um total de 6.963 TWh (IRENA, 2021). A seguir, a figura 1 mostra a participação de cada fonte de energia renovável na produção de energia elétrica, e a figura 2 mostra a variação na produção de cada fonte de energia comparando os anos de 2018 e 2019.

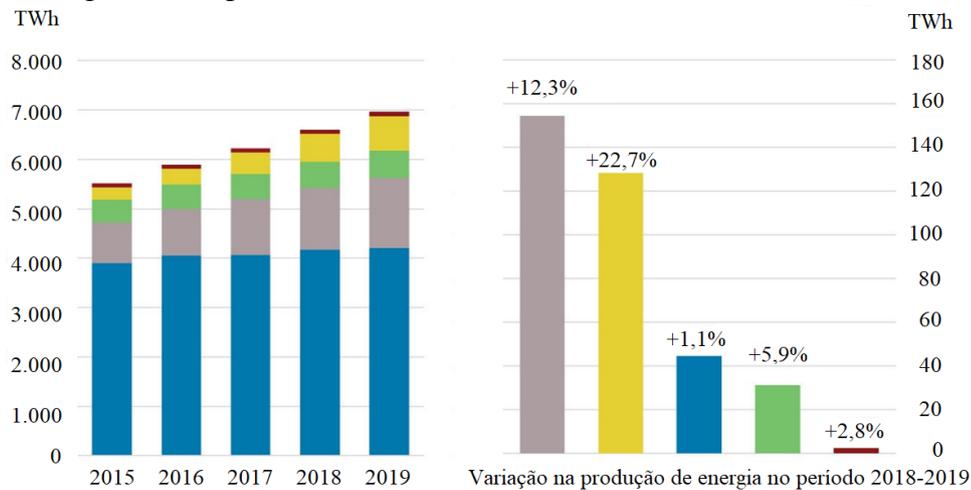
Figura 1 – Participação de cada fonte de energia renovável na geração de energia elétrica no ano de 2019 em termos percentuais.



Fonte: Adaptado de IRENA (2021)

A figura 1 indica que a principal fonte de energia renovável na Terra é a obtida

Figura 2 – Gráfico à esquerda mostra a variação da produção de energia renovável para cada fonte analisada. Gráfico à direita mostra a variação na produção para cada fonte de energia entre o período de 2018 e 2019 em unidade de TWh.



Fonte: Adaptado de IRENA (2021)

por meios hídricos, correspondendo a 61% do total produzido em 2019. Com destaque, são apresentadas também as energias eólica e solar, contabilizando 30% de toda energia renovável produzida mundialmente para este mesmo ano. A figura 2 mostra a variação na produção de energia para cada fonte discriminada, destacando-se ambas energias solar e eólica como as fontes renováveis de maior crescimento em produção, entre 2018 e 2019, com 22,7% e 12,3% respectivamente. A relevância adquirida por essas fontes de energia dentre as demais disponíveis é ainda mais impactante quando tomado o período entre 2011 e 2020: a capacidade instalada para energia eólica passou de 220 GW para 732 GW, um aumento de 330%. No mesmo período, a capacidade instalada de energia solar foi de 74 GW para 716 GW, representando uma variação de 967% (IRENA, 2021).

Dentro do contexto brasileiro, as consequências do aquecimento global e das mudanças climáticas podem ser percebidos, nos últimos anos, em forma do agravamento das secas que atingem importantes bacias hidrográficas do Brasil (HUNT *et al.*, 2022; FREITAS *et al.*, 2022). As secas põem em risco a produção energética brasileira, uma vez que a matriz nacional é predominantemente baseada em fontes hídricas e correspondem a 109.429 GW (57%) da capacidade total instalada (MINISTÉRIO DE MINAS E ENERGIA, 2022). As secas obrigaram o país a retornar a utilização de combustíveis fósseis e biocombustíveis como fontes principais de energia, através das termoelétricas. Essa tecnologia dispendiosa e altamente poluente, foi responsável pela produção de 13,2% da energia brasileira em julho de 2021, sendo o maior

valor atingido até então (GETIRANA *et al.*, 2021). Assim, a diversificação da matriz energética nacional é necessária para que o país consiga passar pelos períodos de seca sem ficar dependente em demasia de combustíveis fósseis. Desta forma, seguindo a tendência mundial, fontes de energias eólica e solar vêm sendo constantemente ampliadas. No período entre fevereiro de 2021 e fevereiro de 2022, a capacidade de energia eólica instalada no Brasil passou de 17.202 GW para 21.490 GW (crescimento de 24,9%), enquanto a capacidade de energia solar evoluiu de 8.230 GW para 14.004 GW (aumento de 70,2%) (MINISTÉRIO DE MINAS E ENERGIA, 2022).

Apesar de trazer inegáveis vantagens ambientais, energias renováveis possuem uma característica negativa inerente a elas: sua natureza estocástica e intermitente. A energia solar, por exemplo, é fortemente limitada pelas nuvens, capazes de bloquear grande parte da radiação solar. Outros fatores locais como quantidade de aerossóis na atmosfera, o clima e a geografia também tem impacto direto na limitação da produção de energia vinda do sol (GUEYMARD, 2010; LARSON *et al.*, 2020; RAJAGUKGUK *et al.*, 2020). Faz-se necessária, então, a estimação precisa da irradiância solar em determinado local de interesse, uma vez que tal informação possibilita melhorar o gerenciamento e fornecimento de energia, permitindo um planejamento mais eficiente e, conseqüentemente, reduzindo os seus custos (KAUR *et al.*, 2016; LAW *et al.*, 2016).

Desta forma, a fim de aprofundar o conhecimento científico em relação a estimação de irradiância solar, o presente estudo se propõe a verificar a generalização dos resultados obtidos pelo modelo de previsão por redes convolucionais desenvolvido por Rocha e Santos (2022) através da influência do viés geográfico nas localidades de Hanford e Petrolina. Assim, imagens do Satélite Ambiental de Operação Geoestacionária, tradução do inglês *Geostationary Operational Environmental Satellite* (GOES) série R, também conhecido como GOES-16, foram usadas para estimar valores de Irradiância Horizontal Global, tradução do inglês *Global Horizontal Irradiance* (GHI), e Irradiância Normal Direta, tradução do inglês *Direct Normal Irradiance* (DNI), para as localidades de Hanford e Petrolina. O viés geográfico foi analisado através do treino e teste da rede profunda com dados de uma mesma localidade, e com a mistura dos dados entre Hanford e Petrolina (ou seja, a rede profunda é treinada com dados de um lugar e testada para realizar previsão para outro local). Os resultados obtidos pelo modelo convolucional

foram então comparados os obtidos por modelos de referência Solcast e Modelo Solar Físico, tradução do inglês *Physical Solar Model* (PSM). As métricas usadas para comparação dos resultados foram Erro Médio por Viés, tradução do inglês *Mean Bias Error* (MBE), e Raiz do Erro Quadrático Médio, tradução do inglês *Root Mean Square Error* (RMSE).

1.3 Objetivos

1.3.1 Objetivo geral

Esta dissertação tem como objetivo geral analisar o viés geográfico na previsões de irradiância solar do modelo desenvolvido por Rocha e Santos (2022), utilizando os dados da estação de Hanford, no estado da Califórnia, juntamente com os dados de Petrolina por meio de transferência de aprendizado.

1.3.2 Objetivos específicos

Os objetivos específicos são:

1. Aquisição de dados de imagem do satélite GOES-16 de todos os 16 canais
2. Aquisição de dados mensurados localmente para GHI e DNI para ambos locais de Petrolina e Hanford
3. Elaboração, ajustes e testes de uma rede neural profunda para modelagem da irradiância solar com uso de imagens de satélite
4. Treinamento e teste da rede profunda com dados de uma mesma localidade e com mistura dos dados de Petrolina e Hanford
5. Validação dos resultados obtidos a partir de comparação com modelos de referência baseados em abordagens físicas para estimação de energia solar

O restante do presente trabalho é dividido da seguinte forma: no capítulo 2, são apresentados o estado-da-arte atual para o campo de pesquisa referente a previsão de energia solar e a revisão bibliográfica. A metodologia utilizada no trabalho é apresentada no capítulo 3. Por fim, os resultados obtidos e as conclusões são apresentados nos capítulos 4 e 5, respectivamente.

2 REVISÃO BIBLIOGRÁFICA

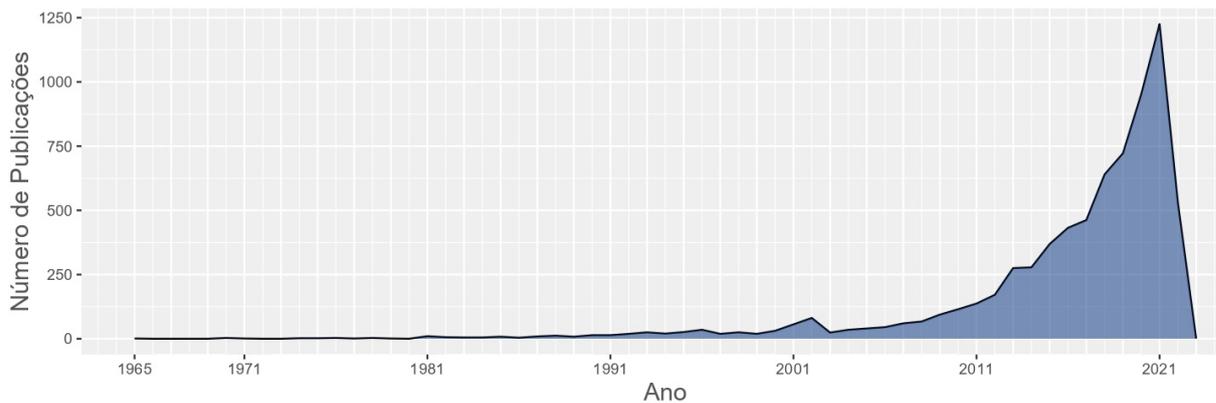
2.1 Estado-da-arte

Para contextualizar o estudo aqui desenvolvido e entender o estado-da-arte da área de previsão e modelagem de energia solar, uma análise bibliométrica foi realizada. O programa utilizado foi o *bibliometrix*¹, uma biblioteca em linguagem R desenvolvida por Aria e Cuccurullo (2017)². Para a data de referência 05 de maio de 2022, a pesquisa realizada na base de dados Scopus, juntamente com os parâmetros discriminados pela tabela 1 e a metodologia sugerida por Donthu *et al.* (2021), foram retornados 7.371 documentos publicados entre os anos de 1965 e 2023.

Tabela 1 – Escopo, banco de dados, termos e restrições impostas referentes à análise bibliométrica realizada com o *bibliometrix* (DONTU *et al.*, 2021).

Definição do escopo da pesquisa	Base de dados	Termos pesquisados	Restrições impostas à pesquisa	Realização da análise bibliométrica
Previsão de energia solar e modelagem	Scopus	<i>satellite, solar, irradiation, irradiance, remote sensing, radiation, model, modeling, nowcast e nowcasting</i>	Língua: inglesa Documentos: artigos científicos e de revisão Áreas pesquisadas: energia, matemática e ciências da computação	<i>Software bibliometrix</i>

Figura 3 – Evolução da produção científica anual para previsão de energia solar.



Fonte: o autor.

Adquirindo as informações dos documentos retornados pela pesquisa, a análise bibliométrica pôde ser realizada. Pela figura 3, pode-se observar que a produção científica atingiu seu pico no ano de 2021, contabilizando 1.225 trabalhos publicados. Ainda pela figura 3, nota-se que há pelo menos 19 anos, existe um crescente volume de publicações dentro da área

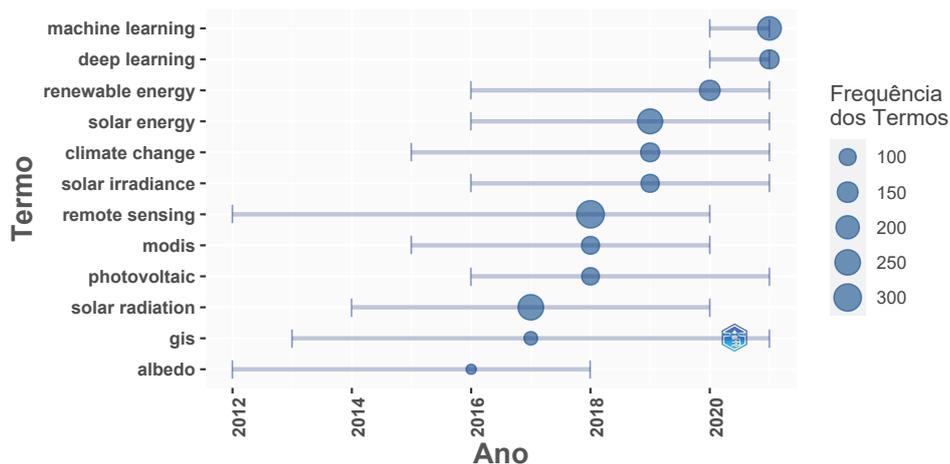
¹ O código fonte e o pacote para a biblioteca podem ser encontrados em <<https://www.bibliometrix.org/>>.

² O código utilizado na análise bibliométrica pode ser obtido neste [link](#).

analisada de modelagem e previsão de energia solar. Isso é coincidente com a problemática do aquecimento global e da mudanças climáticas causadas pela elevação da temperatura terrestre. Assim, é de se esperar que a comunidade científica tenha interesse em compreender melhor e desenvolver tecnologias relacionadas a energia solar, uma vez que ela é uma fonte de energia elétrica renovável importante para reduzir os efeitos do efeito estufa.

A figura 4 apresenta quais os termos de maior destaque entre o período de 2012 e 2022. Nela, é possível notar que os assuntos referentes a aprendizagem de máquinas (identificados pelas palavras em inglês *machine learning* e *deep learning*) usados em estudos sobre energia solar estão atualmente em evidência, representando assim o estado-da-arte deste campo de pesquisa.

Figura 4 – Termos em tendências nos últimos 10 anos.

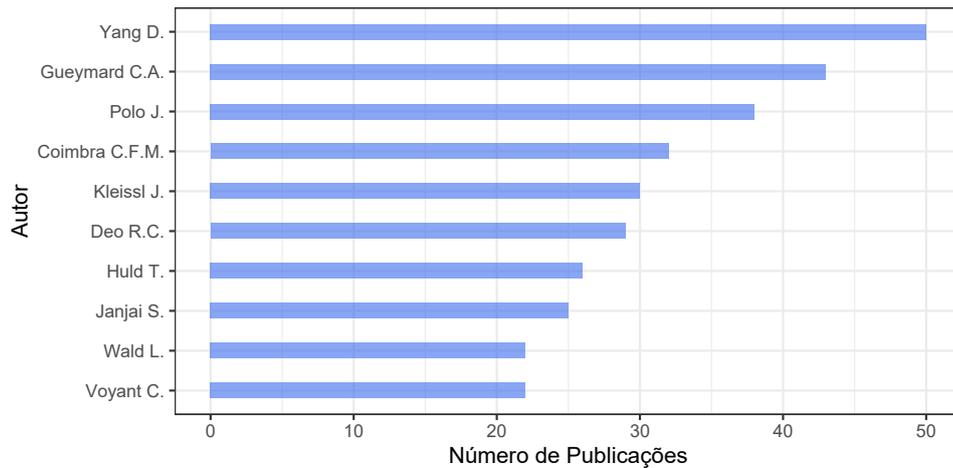


Fonte: O autor.

A figura 5 expõem os autores com mais publicações no período abordado da análise bibliométrica, enquanto a figura 6 mostra os periódicos com mais trabalhos na área analisada. Sendo esses autores e periódicos os que mais produzem publicações dentro da área de interesse, é de se esperar que trabalhos de maior relevância para previsão e modelagem de energia solar sejam produzidos e publicados por eles.

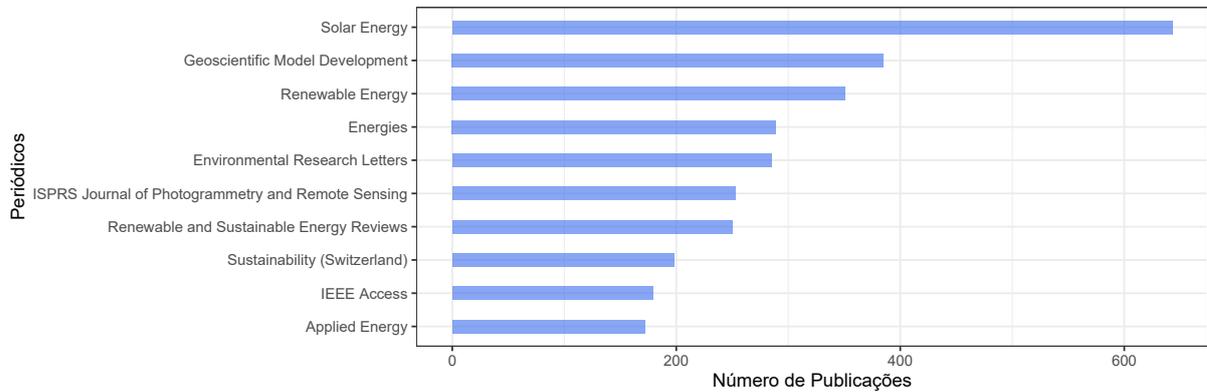
Por fim, a figura 7 mostra os 15 países que mais produziram trabalhos científicos na área analisada. Percebe-se que há uma concentração nas produções científicas dos Estados Unidos e China, totalizando aproximadamente 2.500 publicações. Ainda, é possível concluir que Europa e Ásia são outros polos na pesquisa sobre energia solar. O Brasil ocupa posição de

Figura 5 – Produção científica por autor.



Fonte: O autor.

Figura 6 – Periódicos com mais publicações no período analisado.

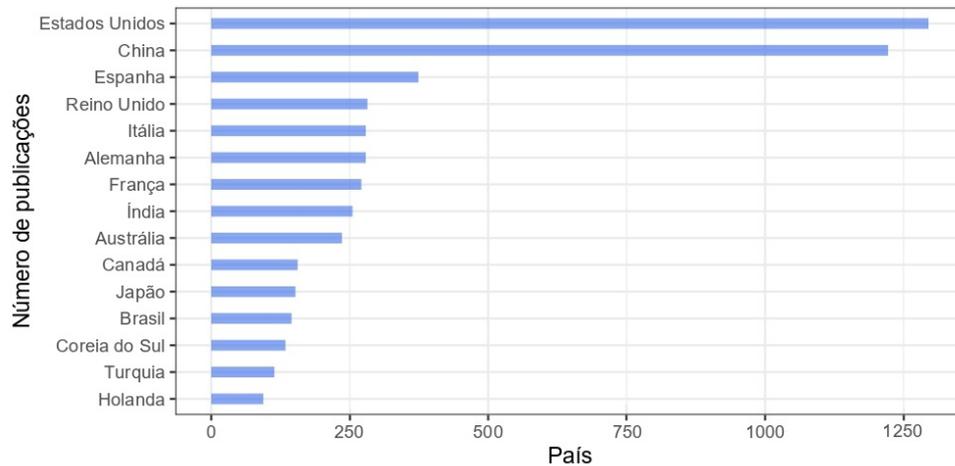


Fonte: O autor.

destaque dentro do contexto regional da América-Latina, sendo o país com mais publicações, contabilizando 145 trabalhos.

A análise bibliométrica realizada permite concluir que, até a data de referência, pesquisas relacionadas a energia solar vêm aumentando em quantidade de trabalhos publicados, totalizando aproximadamente 1.250 publicações em 2021. Isto demonstra um interesse crescente da comunidade científica no desenvolvimento de fontes de energias renováveis e limpas, sendo este um reflexo da necessidade de se buscar combater os efeitos do aquecimento global e das mudanças climáticas. Desta forma, espera-se que este tópico continue sendo de extrema relevância nos próximos anos. Complementa-se ainda que o desenvolvimento de tecnologias relacionadas a energia solar atualmente são baseadas em ferramentas que usam inteligência artificial, indicando papel de destaque desta abordagem na área de energias renováveis.

Figura 7 – Produção científica por país para o período analisado.



Fonte: O autor.

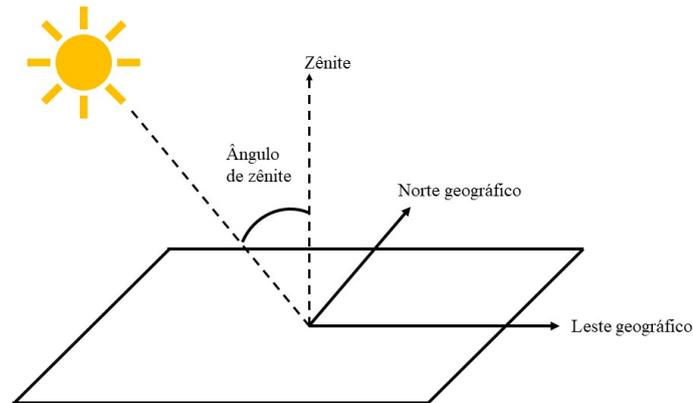
2.2 Conceitos teóricos de irradiância solar

A irradiância solar que atinge a Terra pode ser classificada em três tipos distintos (DUFFIE; BECKMAN, 2013). O primeiro é a irradiância normal direta (DNI). A DNI é a irradiância que atinge o solo terrestre com uma incidência normal (perpendicular) à superfície, sem dispersão da radiação solar. Quando há dispersão, chama-se de irradiância difusa a taxa de energia que atinge o solo terrestre em determinada área. Por fim, há ainda a irradiância horizontal global (GHI), que contabiliza a taxa de energia total que incide em determinada área vinda de todas as direções, incluindo DNI e a irradiância difusa. Ainda é capaz que parte da energia do sol seja completamente refletida, não chegando a atingir a superfície. Vale ressaltar que o ângulo que o Sol faz com a Terra em relação ao zênite e posição geográfica do local observado também são importantes na determinação das irradiâncias incidentes. Uma esquematização exemplificando o ângulo de zênite é apresentada na figura 8, enquanto a figura 9 mostra os tipos de irradiância solar incidentes na superfície da Terra.

2.3 Índice de céu claro $k_{t,cs}$

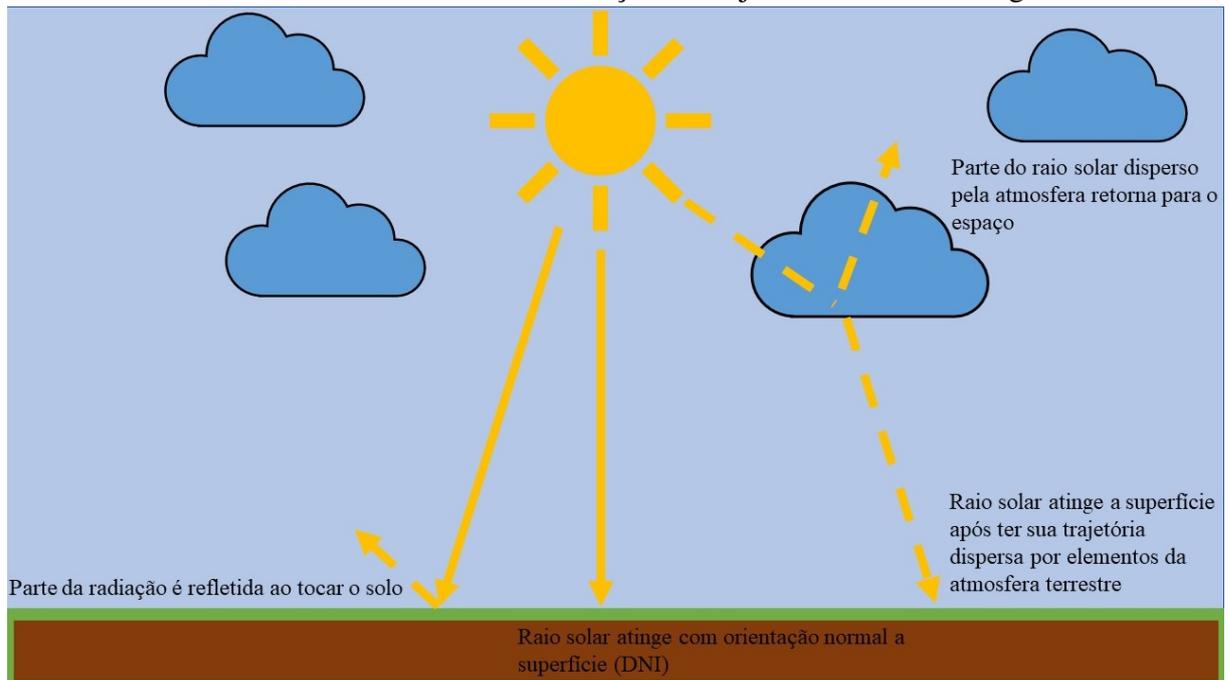
As irradiâncias descritas acima não possuem valores fixos ao longo do tempo em uma mesma região. De fato, os movimento celestiais como a rotação e translação terrestre são essenciais para determinar as irradiâncias que atingem a Terra. Ainda, a localização geográfica e incidência de nuvens são outros fatores determinantes na quantidade de energia solar que irá incidir no solo terrestre (LARSON *et al.*, 2020). Dessa forma, a modelagem para previsão dessas irradiâncias não é uma tarefa tão simples. Entretanto, o uso do índice de céu claro, $k_{t,cs}$,

Figura 8 – Ângulo de zênite solar, θ_z , para uma determinada região.



Fonte: Duffie e Beckman (2013).

Figura 9 – Diferentes tipos de irradiância solar. A irradiância global tem como componentes DNI e DHI, sendo modelado em função delas juntamente com o ângulo de zênite.



Fonte: Duffie e Beckman (2013).

busca remover efeitos climáticos devido às estações, além de tornar os dados de irradiância estacionários (MARINHO, 2020). O modelo de céu claro, na forma do índice $k_{t,cs}$, busca estimar a irradiância incidente sob a condição de céu claro (condição ideal para o céu) para uma determinada localidade em função do ângulo de zênite e outros fatores atmosféricos e geográficos (STEIN *et al.*, 2012). O equacionamento do índice $k_{t,cs}$ é apresentado na equação 2.1 (DUFFIE; BECKMAN, 2013).

$$k_{t,cs} = \frac{G}{G_{0,cs}} \quad (2.1)$$

Onde G representa a irradiância na superfície em estudo e $G_{0,cs}$ é a irradiância para a localidade sob a condição de céu claro. Modelos baseados em condição de céu claro possuem boa aceitação entre pesquisadores, sendo tópico de estudos ao longo dos anos (INEICHEN, 2006; BADESCU *et al.*, 2012; MARQUEZ; COIMBRA, 2013).

2.4 Inteligência artificial

Desenvolvimento de novos componentes de computador, juntamente com novos algoritmos e a necessidade de trabalhar com elevadas quantidade de dados, fizeram com que o campo da aprendizagem de máquina se desenvolvesse na década de 2010 (CHOLLET; ALLAIRE, 2018). Atualmente, a inteligência artificial se faz presente no dia-a-dia de forma recorrente: de assistentes de bancos, passando por filmes sugeridos em plataformas de *streaming*, até carros autônomos. Para melhor entender a atual situação da aprendizagem profunda (abordagem utilizada neste trabalho para a previsão de irradiância solar), a história do desenvolvimento desta tecnologia é apresentada a seguir.

"Podem os computadores serem programados para pensar por conta própria?" foi a indagação que cientistas da computação no período pós-guerra mundial fizeram, buscando automatizar tarefas cognitivas até então inerentes a seres-humanos (CHOLLET; ALLAIRE, 2018). A princípio, os pesquisadores acreditavam que elevar o nível de inteligência das máquinas seria possível somente com o uso de programação simbólica: um paradigma de programação onde regras pré-estabelecidas de como manipular dados são passadas diretamente à máquina em forma de combinação e manipulação de símbolos por processos de inferência, que por sua vez retorna as respostas esperadas (SMOLENSKY, 1987; HOEHNDORF; QUERALT-ROSINACH, 2017; CHOLLET; ALLAIRE, 2018). Dessa forma, pesquisadores adeptos desta abordagem de programação acreditavam que seria possível obter um modelo similar ao de um neurônio humano, à medida que mais detalhadas fossem as regras implementadas em forma de manipulação desses símbolos (SMOLENSKY, 1987; HOEHNDORF; QUERALT-ROSINACH, 2017).

Esta abordagem, que se assemelha ao comportamento cognitivo humano, manteve-se

muito popular nas décadas subsequentes, principalmente até o fim da década de 80, (HOEHNDORF; QUERALT-ROSINACH, 2017; CHOLLET; ALLAIRE, 2018). Entretanto, este paradigma provou-se pouco útil quando implementada para tarefas mais complexas e abstratas (como análise de imagens e reconhecimento de voz), uma vez que programação simbólica produz sistemas rígidos, incapazes de determinar regras implícitas dentro das tarefas passadas a estes sistemas (SMOLENSKY, 1987; CHOLLET; ALLAIRE, 2018).

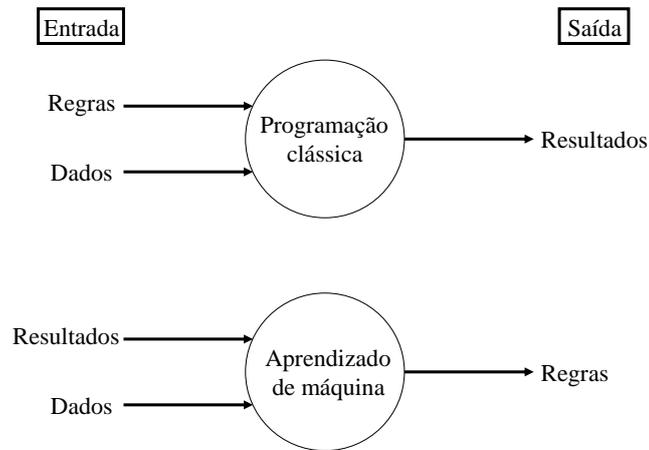
2.5 Aprendizagem de máquina

O aprendizado de máquina foi concebido para suprir as limitações da programação simbólica. Este novo paradigma surgiu a partir da indagação de se as máquinas são capazes de ir além do que lhes é ordenado. Allan Turing baseou-se neste questionamento em seu trabalho *Computing Machinery and Intelligence*, onde foi desenvolvido o chamado Teste de Turing, que tem como objetivo determinar se uma máquina possui ou não inteligência (TURING, 1950).

O aprendizado de máquina tem como princípio a criação de um algoritmo que seja capaz de identificar as estruturas estatísticas que determinam as regras de conjunto de dados, através dos resultados fornecidos a ela por meio de treinamento, ao invés de serem programadas explicitamente. Uma comparação entre programação simbólica clássica e a proposta pela aprendizado de máquina é mostrado pela figura 10, ilustrando as diferentes entradas e saídas para cada situação.

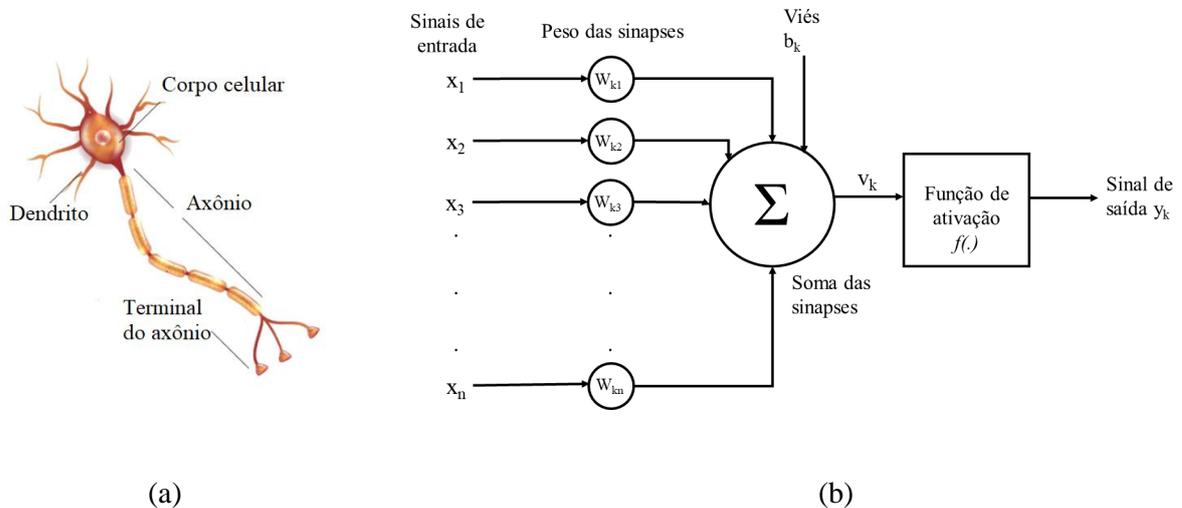
Diversos modelos se utilizam do aprendizado de máquinas. Um dos mais conhecidos, talvez seja o modelo de redes neurais artificiais. O primeiro neurônio artificial, elemento fundamental na composição das redes neurais artificiais, pode ser identificado desde a década de 40, quando McCulloch e Pitts (1943) se inspiraram no neurônio biológico humano para conceber um modelo artificial análogo. Uma comparação entre neurônios artificial e biológico é ilustrada pela figura 11. O trabalho de o McCulloch e Pitts foi aperfeiçoado por diversos outros pesquisadores nos anos subsequentes, destacando-se os trabalho de Hebb (1949) (postulação da primeira regra de aprendizado auto organizado) e Rosenblatt (1958) (propôs o neurônio artificial como o primeiro modelo de aprendizagem supervisionada) (HAYKIN, 2009). Uma rede neural, então, é composta por diversos neurônios artificiais distribuídos paralelamente e capazes de armazenar informação em seus pesos sinápticos (HAYKIN, 2009).

Figura 10 – Comparação entre diferentes paradigmas aplicados no campo de inteligência artificial.



Fonte: adaptado de Chollet e Allaire (2018)

Figura 11 – Comparação entre (a) neurônio biológico e (b) neurônio artificial.

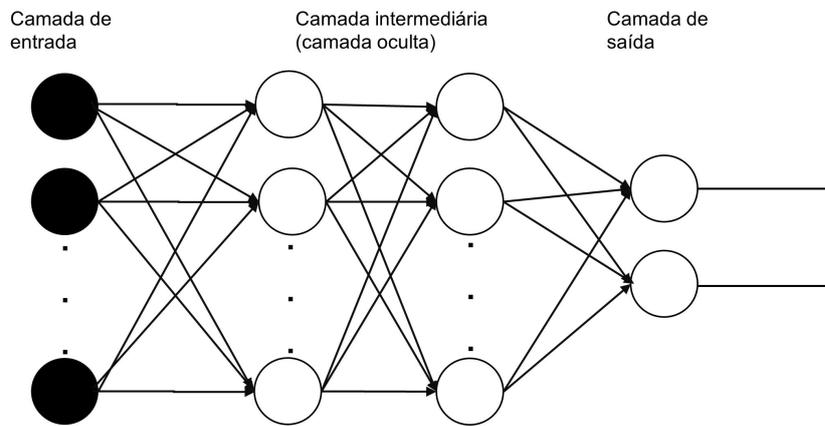


Fonte: o autor.

A configuração mais clássica e utilizada para modelos de redes neurais artificiais é o chamado Perceptron de Multi Camadas, tradução do inglês *Multi Layer Perceptron* (MLP) (MARINHO, 2020). As redes MLP são compostas por camadas ocultas, no sentido de que elas não recebem diretamente os dados de entrada e nem fornecem os dados de saída (HAYKIN,

2009), e sua estrutura é mostrada na figura 12.

Figura 12 – Esquemática de uma rede MLP, onde cada círculo representa um neurônio artificial.



Fonte: o autor

Por muito tempo, redes neurais sofreram com o problema de dissipação do gradiente, o que impossibilitava redes maiores serem elaboradas. Foi a partir do desenvolvimento do chamado *backpropagation*, um método iterativo de redução de erros por meio de gradiente descendente, que redes neurais maiores puderam ser treinadas (SCHMIDHUBER, 2015; CHOLLET; ALLAIRE, 2018). A partir desse marco, redes mais complexas surgiram, sendo o primeiro trabalho a se utilizar de *backpropagation* aplicados a redes neurais atribuído a Werbos (1982). Trabalhos relacionados foram elaborados nos anos subsequentes, podendo destacar o de Cun *et al.* (1989), que usou ideias até então recentes de redes neurais convolucionais para identificação de números manuscritos (CHOLLET; ALLAIRE, 2018; KUMARI; TOSHNIWAL, 2021).

A flexibilização trazida por este novo paradigma, graças a sua não-linearidade, capacidade adaptativa e computação robusta (HAYKIN, 2009), propulsionou a popularidade do aprendizado de máquina nos anos subsequentes, permitindo que uma nova área mais especializada pudesse ser desenvolvida: a aprendizagem profunda.

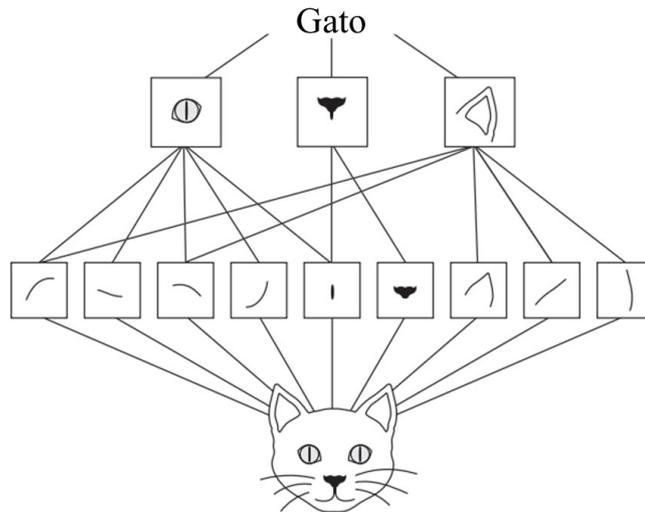
2.6 Aprendizagem profunda

Métodos convencionais de aprendizado de máquina requerem engenharia meticulosa e domínio sobre a área para construir um modelo capaz de processar dados de entrada brutos, retornando uma representação satisfatória da informação a ser utilizada para previsão de valores

ou classificação de padrões (LECUN *et al.*, 2015). A aprendizagem profunda vem para melhorar a abordagem até então vigente, reduzindo a quantidade de hiperparâmetros (ALBAWI *et al.*, 2017). Nesse novo método, múltiplas camadas permitem que o algoritmo aprenda, independentemente de programação prévia, particularidades inerentes do conjunto de dados de entrada através da análise da informação através de suas sucessivas camadas (LECUN *et al.*, 2015; CHOLLET; ALLAIRE, 2018; ROCHA; SANTOS, 2022). Essa ferramenta tem se mostrado eficaz em problemas envolvendo reconhecimento de imagens e de voz, além de estar presente em assistentes digitais e carros autômatos (LECUN *et al.*, 2015; CHOLLET; ALLAIRE, 2018).

Dentro do campo de aprendizagem profunda, vem se destacando a chamada Rede Neural Convolutiva, tradução do inglês *Convolutional Neural Network* (CNN). Nas CNNs, os dados de entrada possuem múltiplas dimensões (chamados de tensores), como no caso de vídeos ou imagens, sendo muito utilizado em problemas que envolvam visão computacional (LECUN *et al.*, 2015; CHOLLET; ALLAIRE, 2018). Diferentemente das abordagens anteriores, que se utilizam de camadas densamente conectadas, como no caso de redes neurais para a etapa de aprendizagem dos padrões dos dados de entrada, a CNN usa operações de convolução para esta finalidade: assumindo-se uma imagem como informação de entrada, camadas densas identificam padrões considerando todos os pixels fornecidos; enquanto na abordagem convolutiva, padrões locais característicos de cada imagem são reconhecidos pela máquina (CHOLLET; ALLAIRE, 2018; ALBAWI *et al.*, 2017). Isso se deve à configuração da CNN: camadas sucessivas de convolução e de *pooling*, que agem como filtros nos dados fornecidos e extraem informações relevantes e mais complexas, além de identificar padrões e formas comuns em cada mapa característico gerado a cada camada (FENG *et al.*, 2022). Para redes convolutivas que analisam imagens, isso permite que dizer que CNN identificam combinações locais de arestas, que por sua vez possuem padrões que se propagam por todas as camadas da rede, o que torna a orientação da imagem irrelevante devido a sua hierarquia espacial (LECUN *et al.*, 2015; ALBAWI *et al.*, 2017; CHOLLET; ALLAIRE, 2018). A figura 13 exemplifica a hierarquia entre as partes componentes de uma imagem, onde a figura do felino pode ser obtida através de formas mais simples, que quando combinadas, constituem formas mais complexas.

Figura 13 – Exemplo de hierarquia espacial identificada pela rede convolucional.



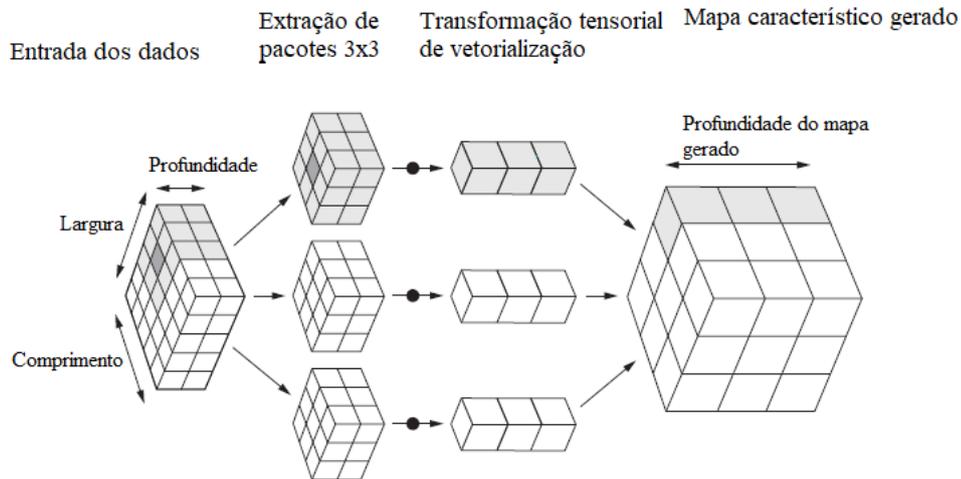
Fonte: adaptado de Chollet e Allaire (2018).

2.6.1 Camadas de convolução, pooling e densa

Os componentes que formam a CNN são a camada de convolução, a camada de *pooling* e a camada densa. Na camada de convolução, a operação matemática de convolução é aplicada nos dados de entrada. É nesta camada que a rede consegue identificar as relações hierárquicas locais e invariáveis de espaço em imagens, permitindo analisar as áreas de maior relevância em cada camada (LECUN *et al.*, 2015; ALBAWI *et al.*, 2017). Nesta camada, a rede neural opera nos dados de entrada (que são tensores com três ou mais dimensões), extraindo pacotes de informações de regiões específicas por todo o conjunto por meio de janelas. Essas janelas percorrem todo o tensor e têm como parâmetros o seu tamanho e o passo com o qual ela se move. As informações coletadas nesse processo sofrem os mesmos tipos de transformações tensoriais, significando que as conexões possuem o mesmo peso no mapa característico formado (LECUN *et al.*, 2015; ALBAWI *et al.*, 2017; CHOLLET; ALLAIRE, 2018). Os mapas característicos gerados são tensores que atuam como filtros codificando aspectos específicos dos dados analisados (CHOLLET; ALLAIRE, 2018). Ao final da convolução, a dimensão do tensor de entrada será maior do que a dimensão do mapa característico obtido. A figura 14 exemplifica o processo de convolução dentro da camada, sendo o tamanho da janela 3x3 com passo de 2. Nota-se que ao final do processo convolucional descrito pela figura 14 que a dimensão do mapa característico gerado é menor do que a dos dados iniciais.

Por sua vez, a camada de *pooling* tem como finalidade reduzir mais ainda a dimensão

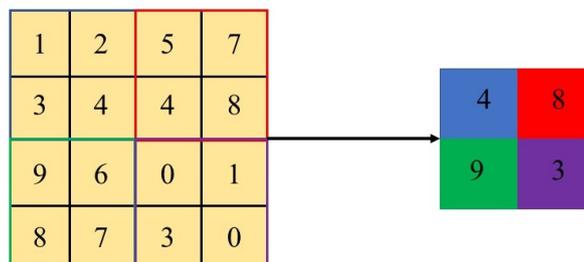
Figura 14 – Exemplo de camada de convolução.



Fonte: adaptado de Chollet e Allaire (2018).

dos mapas característicos, reduzindo a sua complexidade e permitindo filtragem das hierarquias espaciais (ALBAWI *et al.*, 2017; CHOLLET; ALLAIRE, 2018). O tipo de *pooling* mais comum é o *max pooling*, consistindo em retornar o maior valor de um mapa característico contido em uma janela, como elucidado na figura 15. Assim como na convolução, *max pooling* possui como parâmetros o tamanho da janela juntamente com o passo de seu movimento. Um exemplo de aplicação desta camada é mostrado pela figura 15, onde os maiores valores do mapa característico dentro da janela analisada são retornados, diminuindo o tamanho do conjunto de dados gerado. Após as camadas de *max pooling*, há a camada de achatamento, que tem como função transformar o tensor obtido em um vetor.

Figura 15 – Exemplo de camada de *max pooling*.



Fonte: Chollet e Allaire (2018).

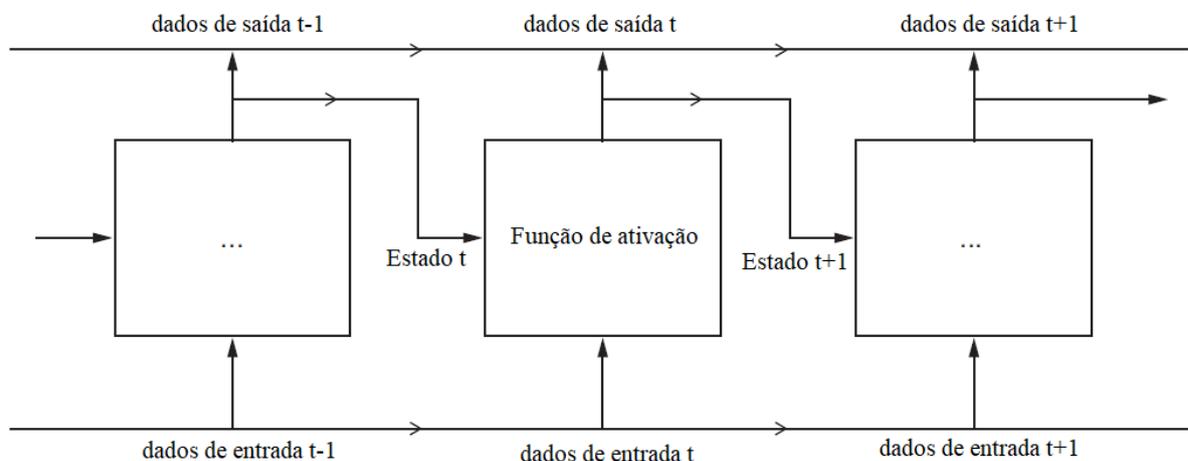
Por fim, há a camada densa composta por neurônios todos conectados entre si. A finalidade desta camada é tentar mapear as relações entre os dados de entrada a ela fornecida (CHOLLET; ALLAIRE, 2018). Como discutido anteriormente, a principal diferença entre uma

camada densa e uma de convolução, é que a primeira busca analisar todos os dados fornecidos, enquanto a segunda buscar pontos locais (CHOLLET; ALLAIRE, 2018). Assim, de forma resumida, uma rede convolucional é composta de alternância entre camadas de convolução e *pooling* seguida de uma camada de achatamento e uma camada densa que irá retornar o resultado obtido pelo modelo gerado pela rede.

2.6.2 Redes neurais recorrentes

Redes Neurais Recorrentes, tradução do inglês *Recurrent Neural Networks* (RNN), são redes neurais profundas que se baseiam no mesmo princípio da inteligência humana que analisa informações parte a parte, mantendo conhecimento prévio do que já foi apresentado e atualizando-o a cada nova informação fornecida (LECUN *et al.*, 2015; CHOLLET; ALLAIRE, 2018). Assim, redes recorrentes possuem um laço interno de repetição que serve para armazenar informação previamente obtida, sendo capazes de mapear sequência de dados de entrada e saída (ou seja, capazes de identificar relações no conjunto de dados) (BENGIO *et al.*, 1994). Esse tipo de rede neural é especialmente capaz de modelar séries temporais, que contêm informação intrínseca que não pode ser identificada por uma rede neural mais simples (KUMARI; TOSHNIWAL, 2021). Entretanto, armazenar e aprender informações em uma rede recorrente mostrou-se difícil para grandes intervalos de tempo, devido, novamente, ao problema de dissipação e explosão do gradiente (BENGIO *et al.*, 1994; LECUN *et al.*, 2015; CHOLLET; ALLAIRE, 2018; KUMARI; TOSHNIWAL, 2021). Um exemplo de rede recorrente é apresentado na figura 16.

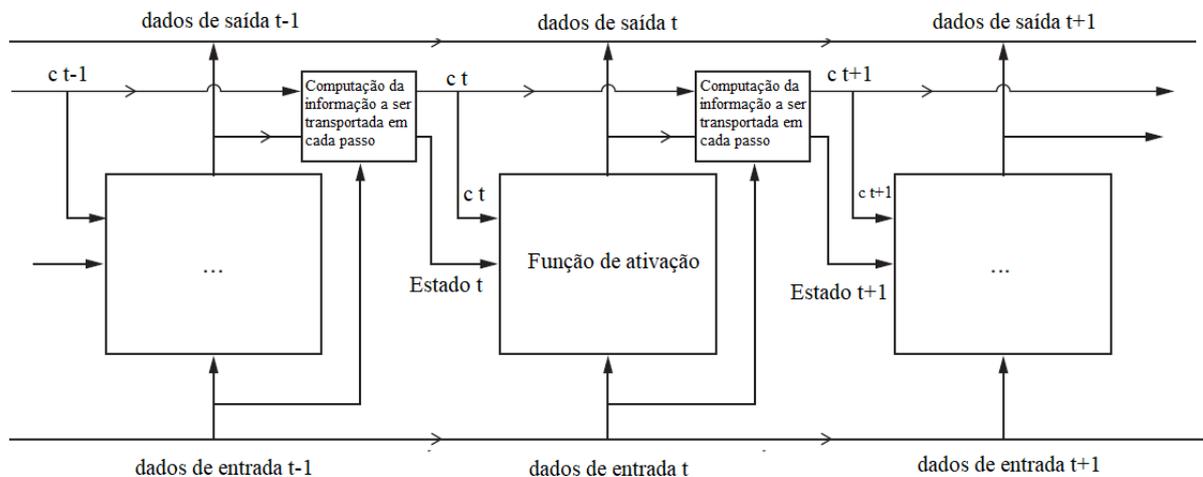
Figura 16 – Exemplo de rede recorrente convencional RNN.



Na figura 16, a rede RNN usa os dados de saída de um determinado estado no tempo $t-1$. Esses dados são utilizados, então, no estado atual t como argumento da função de ativação da RNN.

Para tentar solucionar o problema de dissipação e explosão do gradiente, Hochreiter e Schmidhuber (1997) propuseram a chamada Memória de Curto Prazo Longa, tradução do inglês *Long Short-Term Memory* (LSTM). Em sua arquitetura composta pela chamada célula de memória, a LSTM é capaz de armazenar informação por mais tempo, contornando o problema de dissipação do gradiente, sendo capaz de determinar quais dados podem ser dispensados e quais devem ser armazenados e atualizados, tudo isto sendo feito por operações tensoriais (KUMARI; TOSHNIWAL, 2021). Esta arquitetura já se mostrou ser promissora tanto em reconhecimento de fala, quanto em aplicações de previsão de irradiância solar (LECUN *et al.*, 2015; SHI *et al.*, 2015; KUMARI; TOSHNIWAL, 2021; ROCHA; SANTOS, 2022). Um exemplo de rede profunda LSTM é apresentado na figura 17.

Figura 17 – Exemplo de rede recorrente LSTM.



Fonte: adaptado de Chollet e Allaire (2018).

Pela figura 17, observa-se uma arquitetura similar a apresentada pela figura 16 para uma rede RNN convencional. Na figura 17, o símbolo c representa a informação carregada obtida de uma célula de memória. Assim, no caso da LSTM, a informação carregada c é uma combinação entre dados de entrada, dados de saída e de dados de uma célula de memória anterior ao passo atual. Dessa forma, em um passo t , a informação carregada c pode ser usada como argumento da função de ativação da rede LSTM.

2.7 Modelos de estimaco de radiao e irradincia solar

Diversos modelos de previso de energia solar foram desenvolvidos no ltimo sculo. Tais modelos podem ser divididos em dois conjuntos principais: modelos fsicos (indo de sries temporais, como modelos numricos de previso de clima, regresso numrica, imagens do cu e de satlite); o segundo grupo principal compreende modelos usando aprendizagem de mquinas (VOYANT *et al.*, 2017; YANG *et al.*, 2018; RAJAGUKGUK *et al.*, 2020).

2.7.1 Modelos fsicos para previso de energia solar

O primeiro modelo fsico pode ser identificado desde o ano de 1924, quando o cientista sueco Angstrom props o uso de uma relao linear entre a radiao global horizontal, a radiao global para dia claro e o tempo total de incidncia de raios solares como estimadores para previso da quantidade de irradiaco solar em um dia (ANGSTROM, 1924). Passado-se quase um sculo desde Angstrom, diversos outros modelos preditivos baseados em abordagens fsicas foram desenvolvidos. O trabalho de Hassan *et al.* (2018), realizado no Egito, compreende uma reviso e um estudo de caso sobre variados modelos de previso independentes, ou seja, modelos que no necessitam de registros solares anteriores nem parmetros meteorolgicos. A concluso obtida pelos autores foi que, apesar de ser uma abordagem simples e menos precisa quando comparados com o estado-da-arte dentro deste campo cientfico, tal modelo oferece uma boa estimaco para localidades sem acesso a dados prvios de incidncia solar. No trabalho de Paulescu *et al.* (2014), os autores propuseram que a estimaco da irradiaco global horizontal na Romnia fosse realizada usando um indicador binrio indicando se o sol est ou no encoberto por nuvens. A abordagem proposta pelos autores indicou que, quando comparada com o modelo de referncia utilizado, possui melhor preciso. Em Zhu *et al.* (2019), os pesquisadores utilizaram um modelo hbrido para estimaco da irradiaco direta normal nos Estados Unidos. O modelo hbrido proposto usa ambos modelos baseados em condioes de cu limpo juntamente com correo de erro. Desta forma, uma melhoria de at 70% na mtrica raiz do erro mdio quadrtico normalizado.

Modelos de previso fsicos tambm podem usar informaoes obtidas por satlite para estimar incidncia solar. No trabalho de Nonnenmacher *et al.* (2014), um modelo baseado

em imagem de satélite para estimação de irradiância chamado SUNY (PEREZ *et al.*, 2002) foi usado para determinar DNI em uma localidade dos Estados Unidos. Tal modelo proposto obteve valor médio de 31,00% para a RMSE, 19,98% para Erro Médio Absoluto, tradução do inglês *Mean Absolute Error* (MAE), e 11,75% para MBE quando comparado com medições feitas *in loco*. Em Larson *et al.* (2020), os pesquisadores desenvolveram um novo modelo para estimação em tempo real das propriedades de nuvens a partir de dados do satélite GOES-16. Este modelo mostrou-se uma alternativa viável após ser validado através de estimação de radiação solar nos Estados Unidos. Em um estudo realizado na Índia, Kamath e Srinivasan (2020) usaram um modelo semi-empírico baseado no modelo de céu limpo para estimar irradiância global horizontal a partir de dados de satélite. O modelo proposto obteve melhoria de 40% comparado com modelos de referência.

Outra possível implementação de modelos físicos é baseada em imagens do céu para estimação de irradiância e radiação solar. No trabalho de Alonso-Montesinos e Batlles (2015), desenvolvido na Espanha, uma câmera para captura de imagens do céu foi utilizada. O modelo proposto mostrou-se como uma opção possível para estimação de radiação solar. Em Nouri *et al.* (2021), um modelo usando imagens do céu e modelos de persistência para previsão de irradiação foi proposto também para uma localidade espanhola. O modelo desenvolvido foi capaz de reduzir em até 11,70% o RMSE quando comparado com os modelos de referência analisados. Ainda no país ibérico, o trabalho desenvolvido por Kuhn *et al.* (2019) propôs um sistema de previsão baseado em imagens das sombras das nuvens projetadas no solo para estimação de DNI. Esta abordagem mostrou-se praticável, pois foi capaz de melhorar ambas métricas RMSE e MAE em 29,01% e 35,13%, respectivamente, em comparação com o referencial analisado.

2.7.2 Modelos de aprendizado de máquinas para previsão de energia solar

O segundo grupo principal consiste em modelos que utilizam aprendizado de máquinas. O trabalho de Rocha *et al.* (2019) utilizou um conjunto de dados com 14 anos de informações sobre incidência de radiação diária, semanal e mensal no nordeste brasileiro. O modelo proposto usou redes neurais artificiais para previsão de radiação e obteve bons resultados quando considerados os fenômenos naturais *El Niño* e *La Niña*, cujos impactos no clima local são relevantes. Os resultados de 6,42% para o erro quadrático médio normalizado e 20,83%

para o erro médio absoluto percentual foram obtidos quando considerado uma estimaco mensal de radiao. Na abordagem proposta por Marzo *et al.* (2017), a previso de radiao solar em reas desrticas do Chile, Israel, frica do Sul e Austrlia foi realizada usando dados de radiao extraterrestre e temperaturas como valores de entrada para um modelo de rede neural artificial usando MLP. O algoritmo proposto obteve valores mdios de 13,00% e 4,00% para raiz do erro quadrtico mdio normalizado e erro mdio por vis normalizado, respectivamente.

Em Cornejo-Bueno *et al.* (2019), a radiao solar sobre a Espanha foi estimada usando dados de satlite. Os autores analisaram diferentes modelos de aprendizado de mquina, como vetores de suporte de regresso, perceptron de multicamadas, mquinas de aprendizado extremo e processos gaussianos. Melhor resultado obtido no estudo foi usando o modelo de aprendizado extremo, que foi capaz de superar os modelos de referncia estudados. Tambm usando dados de satlites, Koo *et al.* (2020) estimaram GHI e mapearam a irradiao solar sobre a Coreia do Sul. Os pesquisadores usaram uma combinao entre redes neurais artificiais usando MLP, obtendo um resultado de 54,44 W/m² para RMSE (com valor normalizado de 12,93%). Em Chen *et al.* (2021), os autores usaram dados do satlite GOES-16 para estimaco de radiao solar sobre os Estados Unidos. O modelo de florestas aleatrias iterativa proposto pelos pesquisadores foi capaz de superar outros modelos de aprendizado de mquinas como a floresta aleatria convencional e a regresso linear mltipla. O trabalho provou que no so o modelo de floresta aleatria iterativo  uma abordagem precisa para estimaco de radiao, como tambm provou que o satlite GOES-16  uma ferramenta confivel para mapeamento de radiao solar. Modelos de aprendizagem de mquina podem, ainda, ser usados para estimaco do total de energia eltrica gerada em uma localidade, como demonstrado no estudo realizado por Jebli *et al.* (2021). Neste trabalho, o modelo de redes neurais artificiais obteve melhores resultados de previso de gerao de energia em Marrocos, obtendo valores de 3,49 J/m² e 1,94 J/m² para RMSE e MAE, respectivamente.

2.7.3 Modelos de aprendizagem profunda para previso de energia solar

Modelos de aprendizado profundo vm chamando a ateno dos pesquisadores de energias renovveis (RAJAGUKGUK *et al.*, 2020; ALKHAYAT; MEHMOOD, 2021). Dentre os modelos existentes, a CNN, que a princpio foi desenvolvido como um algoritmo capaz de

reconhecer vozes e imagens (LECUN *et al.*, 2015), tem sido utilizado com sucesso para previsão tanto de energia solar quanto eólica. No trabalho de Bastos *et al.* (2021), estudo usando dados obtidos no estado da Bahia, foi mostrado que a rede convolucional é capaz de prever adequadamente velocidade de vento a partir de dados espaço-temporais. No trabalho de Feng e Zhang (2020), os autores desenvolveram um modelo de rede convolucional chamado de SolarNet. Tal modelo foi capaz de prever GHI nos Estados Unidos usando imagens do céu, obtendo melhores resultados para RMSE normalizado e MBE normalizado em 8,85% e 0,35% respectivamente, quando comparado com os modelos referência. Uma abordagem similar foi desenvolvida no trabalho de Huertas-Tato *et al.* (2021), onde o modelo de rede neural convolucional teve como dados de entrada múltiplas imagens do céu para estimação de GHI na Espanha. O modelo proposto superou os modelos usados como referência obtendo valores de 17,75% para raiz do erro quadrático médio normalizado e 12,76% para erro médio absoluto normalizado. No trabalho de Lima *et al.* (2022), os autores compararam o desempenho de um modelo de aprendizagem profunda na forma de uma rede recorrente LSTM com uma abordagem de vetores de suporte de regressão. Ao fim do trabalho, eles concluíram que a rede profunda tanto obteve menores erros nas previsões, quanto gerou resultados mais precisos, com uma taxa de acerto de 92,1% .

Modelos de aprendizagem profunda do tipo convolucional podem ser modificados para melhorar suas previsões, através da implementação em conjunto de uma rede neural de memória recorrente. Este tipo de configuração híbrida (CNN-LSTM) consegue melhorar o desempenho da previsão de irradiância em até 37% quando comparado com outros modelos de aprendizagem profunda convencionais (KUMARI; TOSHNIWAL, 2021). Em Ghimire *et al.* (2019), os autores usam uma rede neural híbrida, formada por uma rede convolucional e uma rede recorrente para previsão de radiação solar na Austrália para um horizonte de 30 minutos. O modelo proposto alcançou resultados superiores aos dos modelos de referência, com valores de 1,51% para RMSE e 4,67% para erro médio absoluto percentual. O trabalho de Zang *et al.* (2020) também usa uma configuração híbrida do tipo CNN-LSTM na previsão de irradiância global horizontal nos Estados Unidos para horizonte de 60 minutos. Na abordagem proposta, a rede convolucional extrai dados espaciais a partir de parâmetros meteorológicos e a rede LSTM extrai dados temporais a partir de dados de irradiância global horizontal. Quando comparado com outros modelos de aprendizagem profunda usados como referência, o modelo proposto pelos autores obteve melhores resultados, atingindo 42,4 W/m² para valor médio de MAE,

melhorando as previsões obtidas em média de 3,04%.

Em Gao *et al.* (2020), previsão de irradiância solar para horizonte de 60 minutos nos Estados Unidos e Argélia foi proposto usando novamente a rede híbrida CNN-LSTM juntamente com pré-processamento dos dados utilizados: os dados foram, primeiramente, decompostos em frequências características mais simples, com subsequente extração de informação pelo modelo híbrido. De acordo com os autores, o modelo obteve resultado de 38,49 W/m² para RMSE. Em Huang *et al.* (2021), os autores também fazem uso de pré-processamento dos sinais, utilizando a técnica de decomposição por transformações de *wavelet* (em inglês *wavelet packet decomposition*) antes de alimentarem a rede neural híbrida por eles proposta. Os resultados de 32,10 W/m² para RMSE e 15,48% para RMSE normalizado obtidos pelo modelo dos autores foram melhores do que os obtidos pelos modelos de referência. Em Rocha e Santos (2022), os autores propuseram um modelo baseado em CNN-LSTM, além do modelo de árvores aleatórias XGBoost, cujos dados de entrada são imagens de satélite GOES-16. As imagens foram usadas para previsão de irradiâncias horizontal global e normal direta na cidade de Petrolina, em Pernambuco. Os resultados obtidos mostraram que os modelos propostos pelos autores foram capazes de melhorar as previsões em 2,89% e 1,70% para CNN-LSTM e XGBoost, respectivamente, quando comparados com o modelos físicos de referência.

Pela revisão bibliográfica acerca das diferentes abordagens possíveis para previsão de energia solar, conclui-se que há uma variedade de modelos disponíveis. A escolha do modelo ideal irá depender para cada situação: desde o tipo de dados a serem utilizados como informação de entrada, até o local no qual se deseja estimar os valores solares. Entretanto, é possível notar, juntamente com a análise bibliométrica apresentada, que modelos de aprendizado de máquinas e de aprendizado profundo possuem, atualmente, posição de destaque dentro do campo científico em estudo. Desta forma, o presente trabalho se propõe a verificar a capacidade de generalização da previsão de um modelo de aprendizagem profunda que se utiliza de imagens do satélite geostacionário GOES-16 para estimação de irradiâncias GHI e DNI na localidade de Hanford, na Califórnia, referente ao ano de 2018, através dos efeitos do viés geográfico. As motivações que levaram à escolha por este modelo se baseia em duas vantagens deste tipo de abordagem em comparação com modelos físicos (LECUN *et al.*, 2015; ROCHA; SANTOS, 2022): (1) modelos de aprendizagem profunda são capazes de identificar relações complexas não-lineares entre os

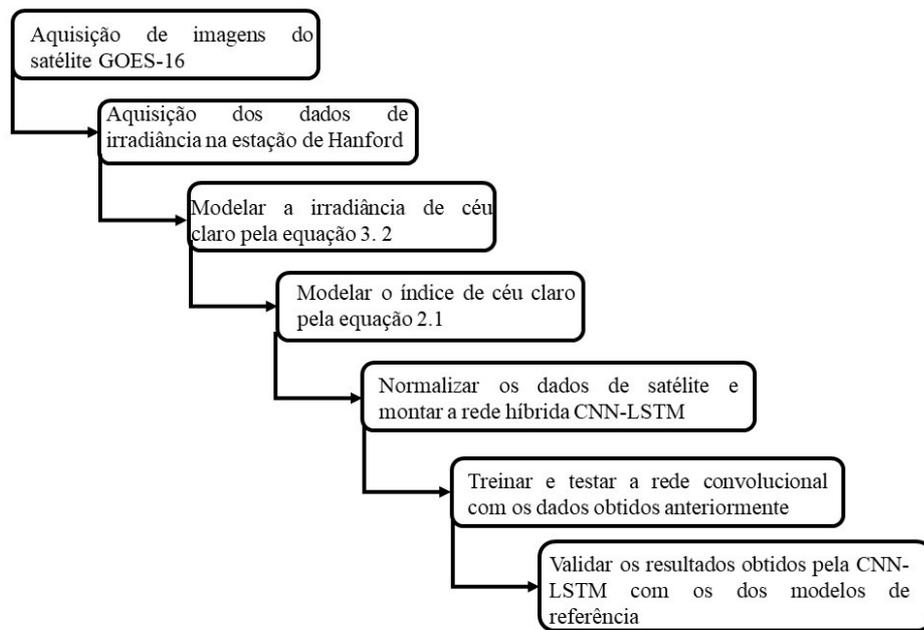
dados usados, sem programação manual prévia nem entendimento físico dos dados; (2) como o modelo CNN-LSTM usa imagens de satélite como dados de entrada, então é virtualmente possível estimar irradiância solar para qualquer ponto coberto pelo satélite.

Espera-se que o trabalho aqui desenvolvido consiga fomentar e ampliar o conhecimento no campo de previsão de energia solar com o uso de imagens de satélite em um modelo de aprendizagem profundo. Isso permitirá que, futuramente, a estimação dessa fonte de energia será possível para qualquer local sob monitoramento remoto de satélite, facilitando o planejamento energético para uma região de interesse.

3 METODOLOGIA

A metodologia utilizada neste trabalho é apresentada e explicada a seguir e pode ser vista de forma resumida pelo fluxograma da figura 18

Figura 18 – Fluxograma da metodologia utilizada.



Fonte: o autor.

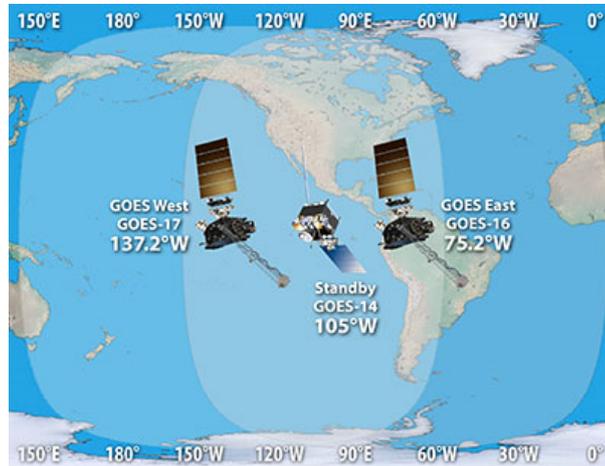
3.1 Imagens de Satélite GOES-16

Para realização das previsões de irradiância para Hanford e Petrolina, primeiramente foi realizada a aquisição das imagens do satélite GOES-16. O satélite GOES-16, juntamente com os satélites GOES-14 e GOES-17, faz parte da frota de satélites estadunidenses responsáveis pelo geomonitoramento terrestre sendo o mais avançado dentre eles. O satélite GOES-16 foi posto em órbita no dia 19 de novembro de 2016 e está sob gerenciamento das agências de Administração Nacional Oceânica e de Atmosfera, tradução do inglês *National Oceanic and Atmospheric Administration* (NOAA) e de Administração Nacional da Aeronáutica e Espaço, tradução do inglês *National Aeronautics and Space Administration* (NASA), ambos órgãos governamentais dos Estados Unidos da América¹. Ainda, o satélite está localizado na longitude 75,2° O, a uma altitude de aproximadamente 36.000 km, sendo capaz de monitorar todo o continente americano,

¹ Mais informações sobre o projeto GOES-R podem ser encontradas em <<https://www.nasa.gov/content/goes-r/index.html>> e <<https://www.goes-r.gov>>.

o Oceano Atlântico e parte da costa oeste africana, como mostra a figura 19.

Figura 19 – A frota de satélites geostacionários GOES para geomonиторamento e sua área de cobertura.



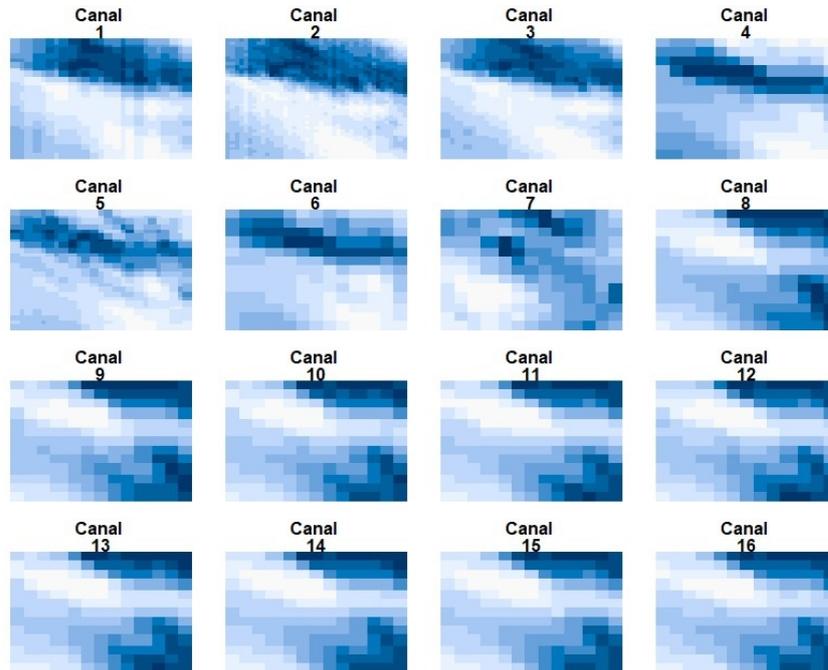
Fonte: NOAA (2022a)

O satélite GOES-16 é o atual estado-da-arte dos satélites de geomonиторamento estadunidenses, fornecendo através de seus 16 canais de imagens informações do clima terrestre, do oceano, do meio-ambiente, dentre outros (SCHMIT *et al.*, 2017). Para manter a mesma resolução temporal utilizada em Rocha e Santos (2022), as imagens de *full disk* foram coletadas em intervalos de 15 minutos. As configurações de cada um dos 16 canais são mostradas pela tabela 2 (HILLGER; SCHMIT, 2004; PADULA; CAO, 2015; KALLURI *et al.*, 2018): desde a resolução espacial de cada canal, o comprimento de onda que cada um detecta e seus respectivos valores mínimos e máximos de medição. Ainda, a figura 20 ilustra a resolução dos 16 canais do satélite GOES-16, onde a maior resolução é a do canal 2, com 0,5 km. Canais 1, 3 e 5 possuem resolução de 1 km. Os demais possuem resolução de 2 km.

3.2 Estação Meteorológica de Petrolina

A cidade de Petrolina, localizada no estado brasileiro de Pernambuco, encontra-se no clima semi-árido, ou BSh segundo a classificação Köppen-Geiger (PEEL *et al.*, 2007). Tal bioma é conhecido por sua elevadas temperaturas, chuvas escassas e irregulares, além de elevada incidência de raios solares (em média 2.800 horas/ano) (SILVA *et al.*, 2010). Além disso, a cidade está localizada nas coordenadas 9° 23' 34" S, 40° 30' 28" O, a 387 metros acima do nível do mar, perto do Equador terrestre e perto do nadir do GOES-16 (ou seja, próximo ao centro das

Figura 20 – Diferentes resoluções para cada canal do satélite geostacionário GOES-16 sobre a localidade de Hanford, Califórnia.



Fonte: o autor

Tabela 2 – Configurações para os 16 canais do satélite GOES-16 (HILLGER; SCHMIT, 2004; PADULA; CAO, 2015; KALLURI *et al.*, 2018).

Canal	Comprimento de onda (μm)	Resolução (km)	Mínimo valor para o canal $\text{mW}/(\text{m}^2 \cdot \text{sr} \cdot \text{cm}^{-1})$	Máximo valor para o canal $\text{mW}/(\text{m}^2 \cdot \text{sr} \cdot \text{cm}^{-1})$
1	0,47	1	-0,57338338	17,77488538
2	0,64	0,5	-0,83799012	25,97769412
3	0,86	1	-0,90205983	27,96385483
4	1,38	2	-0,85959404	26,64741504
5	1,61	1	-0,79255993	24,56935793
6	2,25	2	-0,48595030	15,06445930
7	3,90	2	-0,020	25,0
8	6,18	2	-0,25	28,1
9	6,95	2	-0,45	45,3
10	7,34	2	-0,50	80,2
11	8,50	2	-0,65	135,4
12	9,61	2	-0,80	109,2
13	10,35	2	-0,85	184,4
14	11,20	2	-0,85	200,0
15	12,30	2	-0,85	213,2
16	13,30	2	-2,65	171,7

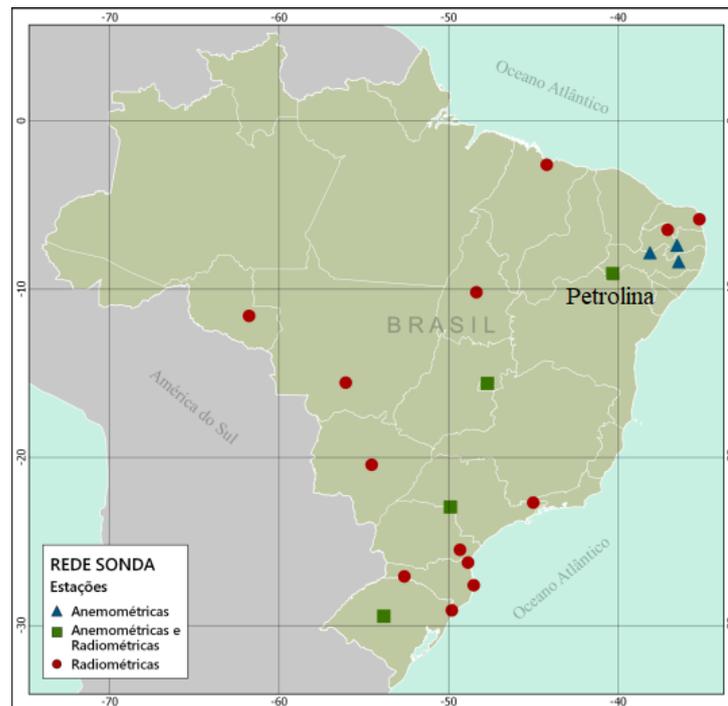
imagens capturadas pelo satélite).

Os dados de irradiância para a localidade de Petrolina são fornecidos pelo Organização Nacional de Dados Ambientais (SONDA)². A rede SONDA tem 20 estações meteorológicas espalhadas pelo território brasileiro, abrangendo todas as suas cinco regiões geográficas e empre-

² Mais informações acerca do projeto SONDA podem ser encontradas em <<http://sonda.ccst.inpe.br/index.html>>.

gando um rigoroso controle de qualidade sobre os dados obtidos garantindo sua confiabilidade (MARTINS *et al.*, 2005; SALAZAR *et al.*, 2020). Dessa forma, o projeto é capaz de aprimorar e aumentar as informações acerca de energias eólica e solar brasileiras. Os dados coletados possuem resolução temporal de 1 minuto, compreendendo o ano de 2018 exceto pelos meses de agosto, setembro, outubro e novembro. Assim, para que a mesma resolução temporal seja mantida entre as imagens de satélite e as medições fornecidas para Petrolina, GHI e DNI foram calculadas para intervalos de 15 minutos usando média móvel. Os dados desta estação SONDA já foram validados com sucesso em estudos anteriores, sendo usadas tanto para previsão de energia eólica (ALENCAR *et al.*, 2017), quanto para derivação de série temporal de para irradiâncias (SALAZAR *et al.*, 2020) e na previsão de irradiância solar (ROCHA; SANTOS, 2022). A figura 21 mostra a localização de todas as estações que formam a rede SONDA no mapa político brasileiro, com destaque para Petrolina.

Figura 21 – Estações componentes da rede brasileira SONDA. Destaca-se a estação de Petrolina, em Pernambuco.



Fonte: adaptado de INPE (2022)

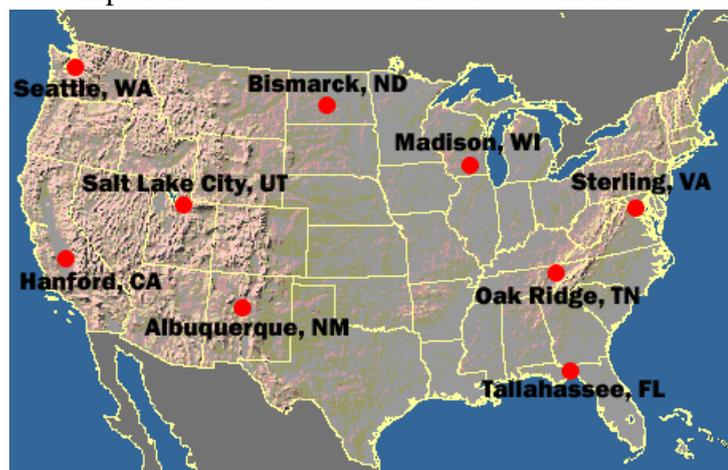
3.3 Estação Meteorológica de Hanford

A segunda estação meteorológica é localizada em Hanford, no estado norte-americano da Califórnia. As suas coordenadas geográficas são $33^{\circ} 23' 34''$ N, $40^{\circ} 30' 28''$ O, e fica a 73

metros acima do nível do mar. A classificação Köppen-Geiger para esta cidade é BSk (PEEL *et al.*, 2007), indicando um clima semi-árido frio, possuindo invernos frios com neve e vento, enquanto o verão é quente e seco (CANAN *et al.*, 2020).

A estação de Hanford faz parte da rede de monitoramento estadunidense Radiação Solar, tradução do inglês *Solar Radiation* (SOLRAD), que juntamente com a Rede de Coleção de Radiação de Superfície, tradução do inglês *Surface Radiation Budget Network* (SURFRAD), ambos gerenciados pelo NOAA, monitoram a incidência de raios solares na parte continental dos Estados Unidos da América³. Os dados fornecidos são de acesso público, tendo sido validados em estudos prévios (YANG, 2018; MATSUNOBU *et al.*, 2021). Neste trabalho, os dados utilizado compreendem o ano de 2018 e possuem resolução temporal de 1 minuto. Novamente, visando a mesma resolução temporal entre estações e imagens de satélite, GHI e DNI foram calculadas para intervalos de 15 minutos usando média móvel. A seleção da estação de Hanford foi baseada em três condições: (1) ser de uma região sob monitoramento do satélite GOES-16; (2) possuir dados e irradiância GHI e DNI para o ano de 2018; e (3) não estar tão adentro do continente. A figura 22 mostra as estações que compõem o programa SOLRAD. Um resumo das informações de cada estação é apresentado na tabela 3.

Figura 22 – Estações componentes da rede estadunidense SOLRAD.



Fonte: NOAA (2022b)

³ Mais informações sobre os projetos SOLRAD e SURFRAD podem ser encontradas em <<https://gml.noaa.gov/grad/surfrad/overview.html>> e <<https://gml.noaa.gov/grad/solrad/solradsites.html>>.

Tabela 3 – Resumo das informações das estações meteorológicas utilizadas.

Estação	Órgão responsável	Coordenadas geográficas	Altitude (m)	Classificação Köppen-Geiger	Resolução temporal (minutos)
Petrolina	SONDA	09° 04' 08" S, 40° 19' 11" O	387	BSh	1
Hanford	SOLRAD	36° 18' 49" N, 119° 37' 54" O	73	BSk	1

3.4 Rede neural convolucional

A configuração da rede neural convolucional utilizada foi obtida por meio de teste topológico conforme descrito no trabalho de Rocha e Santos (2022). Primeiramente, foi determinada a melhor configuração para as camadas de convolução e *max pooling*. O teste topológico começou com apenas uma repetição entre essas camadas até uma máximo de 3 repetições (devido a redução de dimensionalidade dos tensores utilizados, este valor não pôde ser maior), onde os dados de saída de uma camada precedente são usados como entrada na camada subsequente. Os autores chegaram a conclusão que a melhor configuração é dada por 3 pares de camadas de convolução e *max pooling*, na forma $3 \times (\text{convolução} + \text{max pooling}) \rightarrow$ camada de achatamento \rightarrow camada densa.

Em seguida, foi determinada a melhor configuração para a camada densa. Várias configurações para esta camada foram testadas contendo desde 0 nós até 4.096 nós, incluindo uma configuração de duas camadas densas com 512 nós cada. O melhor resultado para os testes foi obtido para uma configuração contendo 2.048 nós. Após a camada densa, foi escolhida a melhor configuração para a camada recorrente LSTM. Novamente, devido a redução dimensional, um máximo de duas camadas LSTM seguidas foi testado. Os melhores resultados foram obtidos para um configuração de rede contando com 2 camadas LSTM cada uma com 32 nós. Para a etapa de treinamento e teste do modelo, os dados foram divididos na forma de 70% das observações usadas para treino, e os 30% restante para teste, conforme mostrado na tabela 7.

Tabela 4 – Comparação entre as quantidades de observações para ambas estações meteorológicas.

Estação	Quantidade total de medições	Quantidade usada para treino	Quantidade usada para teste
Hanford	10.229	7.161	3.068
Petrolina	5.862	4.104	1.758

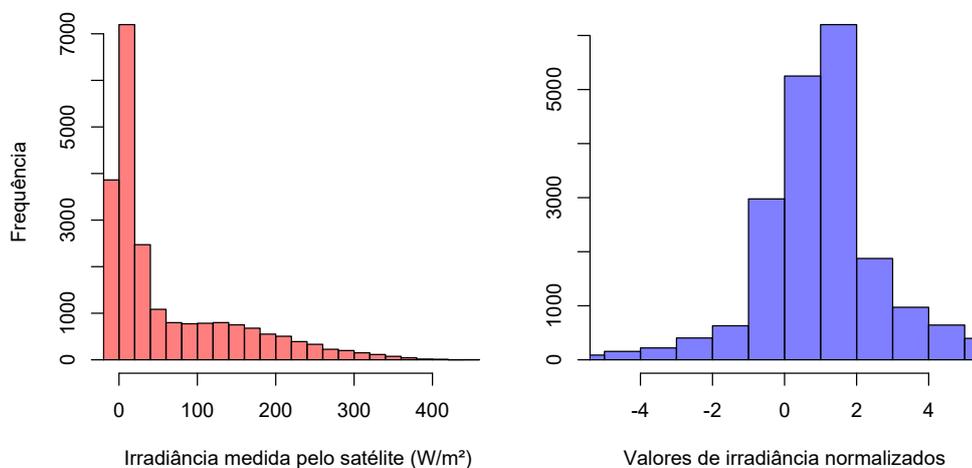
3.5 Normalização dos dados

O pré-processamento dos dados do satélite também foi estudado, em um total de 8 normalizações propostas. Os melhores resultados foram obtidos para dados de entrada que foram normalizados conforme a equação 3.1 (ROCHA; SANTOS, 2022).

$$x_{\text{norm}} = \frac{C_H - \overline{C_H}}{S_{C_H} \cdot \cos(\theta_z)} \quad (3.1)$$

onde C_H é o valor medido pelo canal do satélite, $\overline{C_H}$ é o valor médio da medição do canal, S_{C_H} é o desvio padrão e θ_z é o ângulo de zênite. A normalização dos dados em pesquisas sobre previsão de energia solar já se mostrou como uma ferramenta importante na melhoria dos resultados obtidos por modelos de aprendizado profundo (HUANG *et al.*, 2021; RODRÍGUEZ *et al.*, 2022). A normalização permite que os dados sejam analisados dentro de uma mesma ordem de grandeza, além de promover similaridade entre os dados em uma distribuição estatística, que neste caso é uma distribuição gaussiana (ROCHA; SANTOS, 2022). A comparação entre os dados antes e depois da normalização é apresentada pela figura 23, onde na imagem vermelha à esquerda estão os dados sem normalização, enquanto na imagem azul à direita os dados estão normalizados pela equação 3.1. Percebe-se que após a normalização, a distribuição assume uma forma próxima a de uma distribuição gaussiana.

Figura 23 – Dados do satélite GOES-16 não normalizados à esquerda e normalizados à direita.

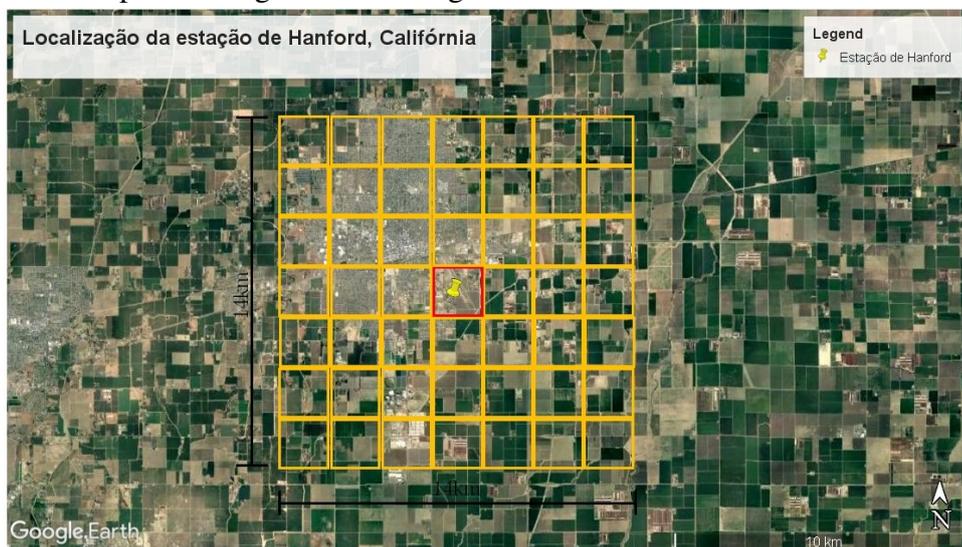


Fonte: o autor.

3.6 Tamanho da imagem

Ainda durante os testes de configuração do modelo CNN-LSTM, os autores realizaram testes para determinar o melhor tamanho de imagem de satélite, indo desde uma imagem representando a exata localização da estação, até uma imagem com *raio* de 14 km. Neste caso, a palavra raio designa uma imagem quadrada ao invés de uma circular e contabiliza a distância do local da estação até o final da imagem utilizada, considerando a menor resolução de 2 km para os canais do satélite. Assim, o tensor resultante possui dimensão na forma $(x * 2 + 2) \text{ km} \times (x * 2 + 2) \text{ km} \times 16$, sendo x o raio da imagem para cada um dos 16 canais (ROCHA; SANTOS, 2022). O melhor resultado foi obtido usando uma imagem com raio igual à 12 km. Um exemplo de imagem de raio igual à 6 km é apresentada na figura 24. Para este exemplo considerando a localização da estação de Hanford, tem-se que cada quadrado possui lado igual à 2 km (de acordo com a menor resolução de canal do GOES-16). Desta forma, a partir do quadrado central, a imagem contabiliza um raio de 6 km para todos os lados do quadrado maior, resultando numa imagem de tamanho $14 \text{ km} \times 14 \text{ km}$.

Figura 24 – Exemplo de imagem com raio igual à 6 km.

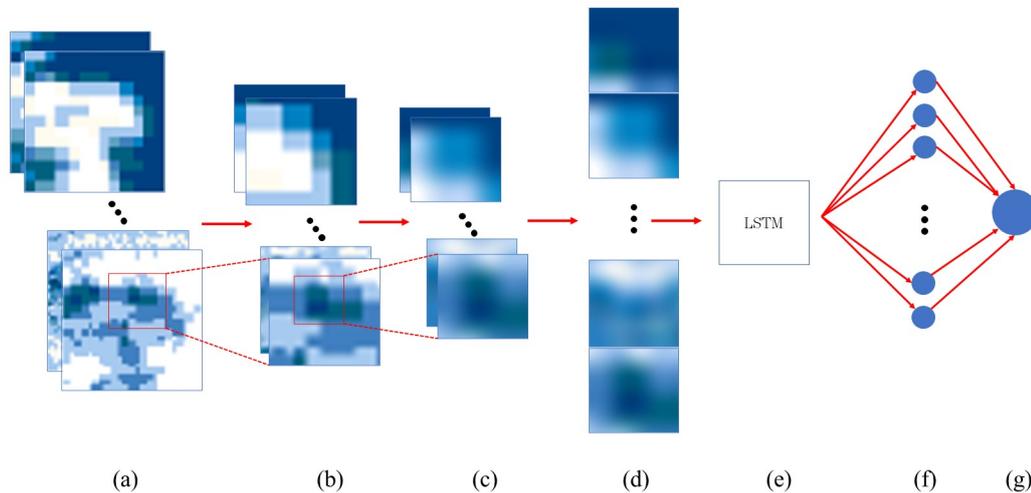


Fonte: Google Earth (2022)

A tabela 5 resume as configurações utilizadas para a CNN-LSTM, o tamanho da imagem e normalização utilizada nos dados obtidos pelo satélite. A figura 25 exemplifica uma rede híbrida CNN-LSTM similar a utilizada por este trabalho. Na imagem 25, (a) representa os dados de entrada dos 16 canais do satélite; (b) representa os dados após a camada de

convolução; (c) representa os dados após a camada de *max pooling*; (d) representa os dados após o achatamento; (e) representa a camada LSTM; (f) representa a camada densa; e (g) representa o resultado para o valor estimado obtido pela rede.

Figura 25 – Exemplo de rede neural híbrida CNN-LSTM.



Fonte: Rocha e Santos (2022)

Tabela 5 – Resumo das configurações utilizadas na rede CNN-LSTM.

Configurações utilizadas para a CNN-LSTM	
Camada de convolução	3 camadas
Camada de <i>max pooling</i>	3 camadas
Camada LSTM	2 camadas com 32 nós cada
Camada densa	1 camada com 2.048 nós
Dimensão do tensor de entrada	26 x 26 x 16
Normalização dos dados	Equação 3.1

3.7 Previsão dos valores de irradiância pelo modelo CNN-LSTM

Na rede desenvolvida no presente trabalho, a previsão das irradiâncias GHI e DNI é realizada por meio do índice de céu claro $k_{t,cs}$, conforme apresentado anteriormente na seção 2. Para tal, há inicialmente o cálculo da irradiância para céu claro, $G_{0,cs}$, para GHI e DNI baseado no modelo polinomial de terceiro grau apresentado a seguir pela equação 3.2 (MARQUEZ; COIMBRA, 2013).

$$G_{0,cs} = C_3(\cos\theta_z)^3 + C_2(\cos\theta_z)^2 + C_1(\cos\theta_z) + C_0 \quad (3.2)$$

Onde C_3 , C_2 , C_1 e C_0 são constantes e θ_z é o ângulo de zênite. Os valores das constantes são encontrados por uma função não linear de mínimos quadrados nativa da linguagem R, onde o modelo foi implementado. Ainda, como as resoluções temporais das imagens de satélite e das medições pelas estações meteorológicas são distintas, uma média móvel para GHI e DNI em intervalos de 15 minutos foi utilizada para permitir concordância entre os dados analisados (ROCHA; SANTOS, 2022).

Determinados os valores de $G_{0,cs}$ para os diferentes ângulos de zênite da localidade de Hanford, a equação 2.1 é usada para determinar o índice de céu claro para as imagens de satélite. Por sua vez, o modelo CNN-LSTM estima valores de $k_{t,cs}$ a partir dessas mesmas imagens. De posse desses índices juntamente com os valores de $G_{0,cs}$, é possível realizar finalmente a previsão da irradiância G para o local de interesse utilizando novamente a equação 2.1.

3.8 Modelos de referência

Inicialmente, esperava-se o uso dos modelos de previsão de irradiância de referência Solcast, Serviço de Monitoramento Atmosférico Copernicus, tradução do inglês *Copernicus Atmosphere Monitoring Service* (CAM5), e PSM fornecido pelo Banco de Dados Nacional de Radiação Solar, tradução do inglês *National Solar Radiation Database* (NRSDB). Porém, devido a localização de Hanford, o modelo de referência CAM5 não pôde ser utilizado no presente estudo, já que é um serviço europeu e que não cobre a totalidade da região continental dos Estados Unidos. Assim, os resultados para a estimacão de irradiância para Hanford obtida pela rede CNN-LSTM foi comparada com os resultados de Solcast e PSM.

Solcast⁴ é um serviço privado para previsão de irradiância solar. Possui abrangência global e se utiliza de um modelo físico semi-empírico combinando dados de satélite e equações de transferência de radiação (BRIGHT, 2019). Seus dados já foram utilizados com êxito e validados em estudos como Bright (2019), Yang e Bright (2020) e Mabasa *et al.* (2021). Este

⁴ Mais informações sobre o serviço prestado por Solcast podem ser encontradas em <<https://solcast.com/>>.

serviço oferece resoluções temporais de 5 até 60 minutos, sendo as medidas para uma resolução de 15 minutos utilizadas neste trabalho.

O PSM é um modelo físico de previsão de irradiância fornecido pelo NRSDB⁵. Possui acesso livre e é resultado de um esforço colaborativo entre órgãos estadunidenses como NOAA, NASA e o Laboratório Nacional de Energia Renovável, tradução do inglês *National Renewable Energy Laboratory* (NREL), abrangendo não somente os Estados Unidos e América do Norte, mas também países do Caribe e da América do Sul. Seu modelo preditivo para a localidade de Hanford foi o PSM versão 3 (SENGUPTA *et al.*, 2018). A sua resolução espacial é de 4 km x 4 km e resolução temporal é de 30 minutos (ROCHA; SANTOS, 2022). Este banco de dados já foi validado por diferentes estudos como Habte *et al.* (2017), Sengupta *et al.* (2018) e Yang (2018). Como PSM possui uma resolução temporal de meia hora, uma média móvel de 30 minutos foi implementada para analisar os resultados de irradiância obtidos para este modelo.

3.9 Métricas de avaliação

Os resultados obtidos pelos modelos analisados precisam ser avaliados de forma quantitativa. A quantificação dos resultados através de métricas que se utilizam dos erros entre os valores previstos e os reais mensurados *in loco* ajuda a determinar o desempenho de um modelo. Dessa forma, é necessário escolher quais métricas utilizar para este fim. Os trabalhos de Yang *et al.* (2018) e Kumari e Toshniwal (2021) listam que RMSE e MBE são métricas populares entre os pesquisadores da área de energia solar. O RMSE pode ser compreendido como uma distância euclidiana entre dois pontos: o valor estimado pelo modelo analisado e o valor real fornecido pela estação meteorológica. Dessa forma, quanto menor for a distância entre o valor previsto e o valor real, melhor será o desempenho do modelo. Por sua vez, MBE indica tendência de superestimação (para valores positivos) ou subestimação (para valores negativos) das irradiâncias estimadas pelos modelos de previsão (YANG *et al.*, 2018; KUMARI; TOSHNIWAL, 2021; ROCHA; SANTOS, 2022). A formulação para RMSE e MBE são apresentados, respectivamente, pelas equações 3.3 e 3.4 (YANG *et al.*, 2018).

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (3.3)$$

⁵ Mais informações sobre o serviço oferecido por NRSDB podem ser encontradas em <<https://nsrdb.nrel.gov/>>

$$\text{MBE} = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i) \quad (3.4)$$

Para ambas equações 3.3 e 3.4, \hat{y}_i , y_i e n representam, respectivamente, o valor de irradiância previsto pelo modelo, o valor real de irradiância medido pela estação meteorológica e a quantidade total de medições. Ambas equações possuem unidades em $[\text{W}/\text{m}^2]$.

3.10 Transferência de aprendizado

A transferência de aprendizado é a capacidade do modelo de utilizar conhecimento previamente adquirido e conseguir aplicá-lo em uma nova tarefa. Para tal, o modelo é testado para diferentes configurações de dados de treino e teste. Originalmente, o conjunto de treino e teste usado em Rocha e Santos (2022) foi composto somente por dados de Petrolina. Agora, é proposto o estudo da transferência de aprendizagem na forma de treinamento treino e teste usando dados de Hanford e uma combinação entre Hanford e Petrolina. Assim, pretende-se analisar se o modelo de rede convolucional consegue usar o que foi aprendido durante a fase de treinamento na fase de teste e se há efeito de algum viés geográfico que por ventura venha a influenciar nos resultados obtidos e no desempenho da generalização do modelo convolucional. Antes de apresentar as configurações de treino e teste, fica convencionado que de agora em diante, a primeira letra é referente ao conjunto de treino e a segunda ao conjunto de teste. Dessa forma, HH indica treino e teste com dados de Hanford.

3.10.1 Treino e teste com dados de Hanford

Para a configuração HH, os dados de Hanford foram divididos obedecendo a divisão de 70% dos dados para treino e 30% para teste, a mesma proporção usada somente para treino e teste com dados de Petrolina (ROCHA; SANTOS, 2022).

3.10.2 Treino com dados Petrolina e teste com dados de Hanford

Como apresentado anteriormente, a diferença entre a quantidade de dados obtidos para as localidades é 4.367 observações. A princípio, para a configuração PH, planejou-se realizar o treino e teste com todos os dados de ambas as estações. Entretanto, devido à limitações de *hardware*, a placa de vídeo utilizada não possuía memória suficiente para analisar os 10.229

dados de Hanford na parte de teste. Dessa forma, foi necessário usar da mesma quantidade de observações para ambas etapas de treino e teste. Para que isso fosse possível, ambos bancos de dados foram analisados e os pontos observados para um mesmo horário de acordo com o Tempo Universal Coordenado, tradução do inglês *Coordinated Universal Time* (UTC), presente tanto em Hanford quanto em Petrolina foi selecionado. Ao fim dessa filtragem, para previsão de DNI foram utilizados 3.035 observações e para GHI 3.111 observações usando a configuração PH, como apresentado pela tabela 6.

Tabela 6 – Quantidade de dados utilizados para cada previsão de irradiância para PH e HP₂.

Estação	Total de observações	Observações usadas para previsão de DNI	Observações usadas para previsão de GHI
Hanford	10.229	3.035	3.111
Petrolina	5.862	3.035	3.111

3.10.3 Treino com dados de Hanford e teste com dados de Petrolina

Para esta configuração, duas abordagens foram analisadas: treino e teste utilizando todos os dados das estações (10.229 para Hanford e 5.862 para Petrolina); e treino e teste usando a mesma quantidade de observações, como descrito pela tabela 6. Dessa forma, fica convencionando que HP₁ é referente a treino com Hanford (H) e teste com Petrolina (P) usando todos os dados das estações. Enquanto HP₂ designa que foram usadas as mesmas quantidades de dados para treino e teste da rede neural. A tabela 7 apresenta de forma resumida as diferentes configurações utilizadas.

Tabela 7 – Quantidades de observações utilizadas para treino e teste da rede convolucional.

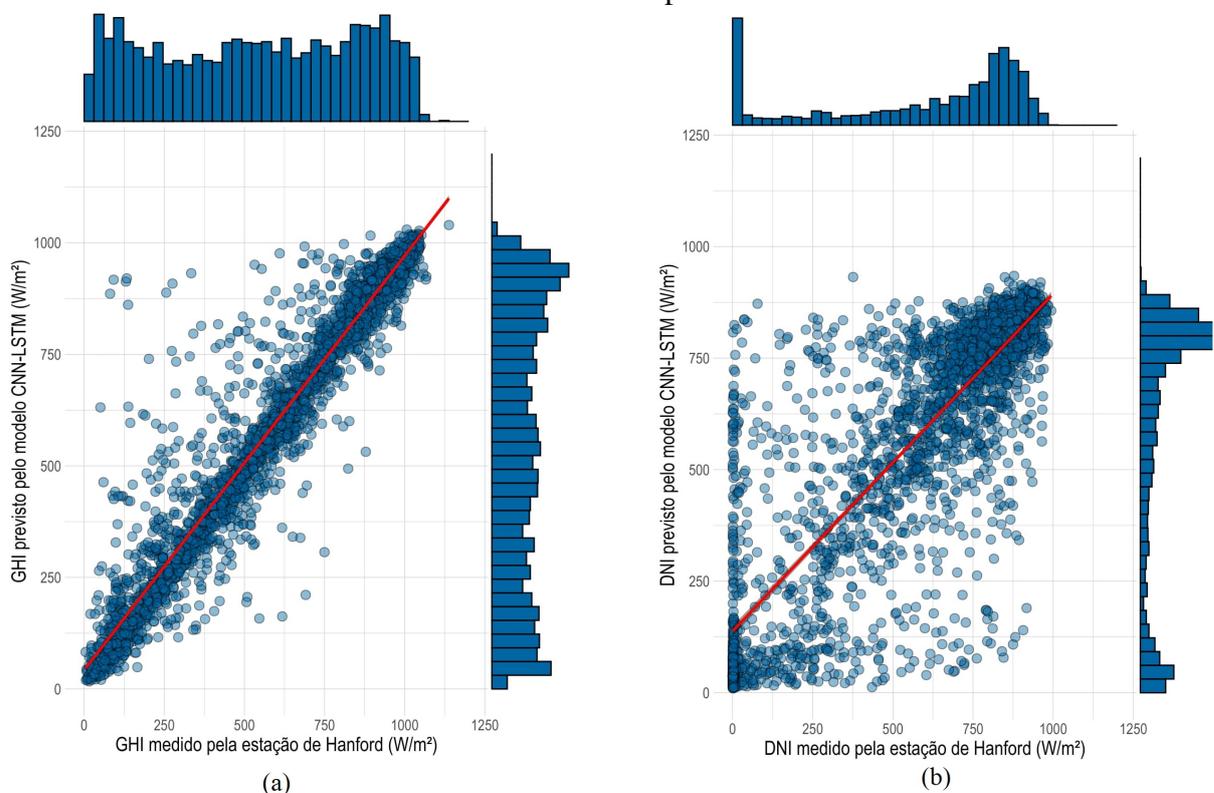
Configuração (treino/teste)	Quantidade para treino	Quantidade para teste
HH	7.161	3.068
PH	3.035 (DNI)	3.035 (DNI)
	3.111 (GHI)	3.111 (GHI)
HP ₁	10.229	5.862
	3.035 (DNI)	3.035 (DNI)
HP ₂	3.111 (GHI)	3.111 (GHI)
	4.104	1.758
PP		

4 RESULTADOS

4.1 Resultado para previsão de irradiâncias: treino e teste com dados de Hanford

Para configuração HH, a rede CNN-LSTM obteve valores de RMSE iguais à 88,17 W/m^2 e 178,40 W/m^2 para GHI e DNI, respectivamente. Para MBE, o modelo obteve para GHI 3,90 W/m^2 e -3,10 W/m^2 para DNI. Os gráficos de dispersão comparando as previsões e os valores reais de irradiância são apresentados pela figura 26.

Figura 26 – Gráfico de dispersão das irradiâncias (a) GHI e (b) DNI previstas pelo modelo CNN-LSTM usando dados de Hanford para treino e teste.



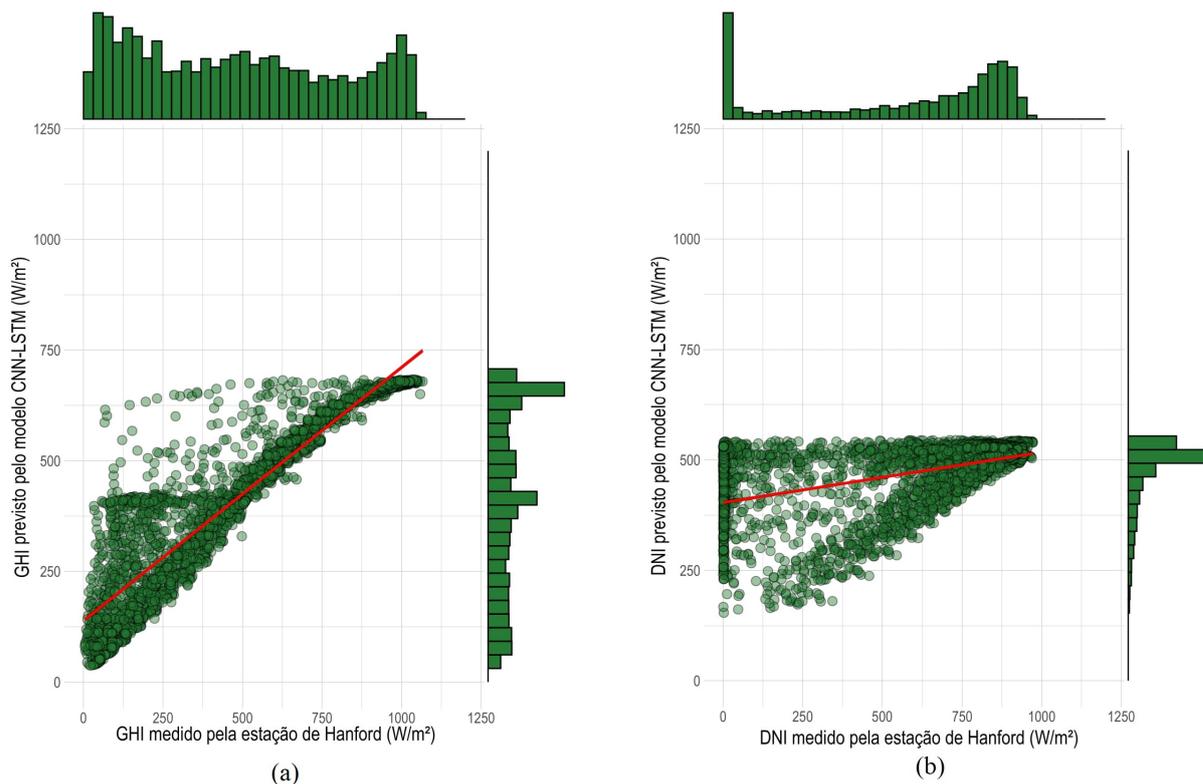
Fonte: o autor.

Para previsão de irradiância horizontal global, os dados ficaram mais agrupados ao redor da linha de regressão. Isso indica que o modelo usado para previsão gerada pela CNN-LSTM consegue prever bem este tipo de irradiância. Para o caso da DNI, os dados estão dispostos de forma mais dispersa ao longo da área do gráfico, indicando que, embora ainda haja correlação positiva entre dados previstos e reais, esta irradiância não foi tão bem modelada e prevista quanto a anterior. Isso é corroborado pela métrica RMSE apresentada anteriormente, onde o menor valor foi atingido pelo GHI. Essa diferença pode ser explicada pela própria natureza da irradiância

normal direta: ela é muito sensível à variações climáticas, ocorrência de aerossóis, chuvas e nuvens dependendo fortemente da direção do raio solar incidente na superfície (GUEYMARD, 2010; PADOVAN *et al.*, 2014; ROCHA; SANTOS, 2022). Assim, irradiância direta normal possui um comportamento mais inesperado, tornando-a de difícil modelagem e previsão. Por sua vez, GHI leva em conta toda irradiância que chega à superfície terrestre por todas direções, facilitando sua previsão (ROCHA; SANTOS, 2022).

4.2 Resultados para previsão de irradiâncias: treino com dados de Petrolina e teste com dados de Hanford

Figura 27 – Gráfico de dispersão das irradiâncias (a) GHI e (b) DNI previstas pelo modelo CNN-LSTM na forma PH.



Fonte: o autor.

A figura 27 mostra os gráficos de dispersão das irradiâncias previstas usando uma configuração PH. Para GHI, houve maior dispersão no entorno da linha de regressão quando comparado para treino e teste com apenas a localidade de Hanford, como mostrado pela figura 26. Ainda por inspeção visual, é notável que a linha de regressão possui menor inclinação

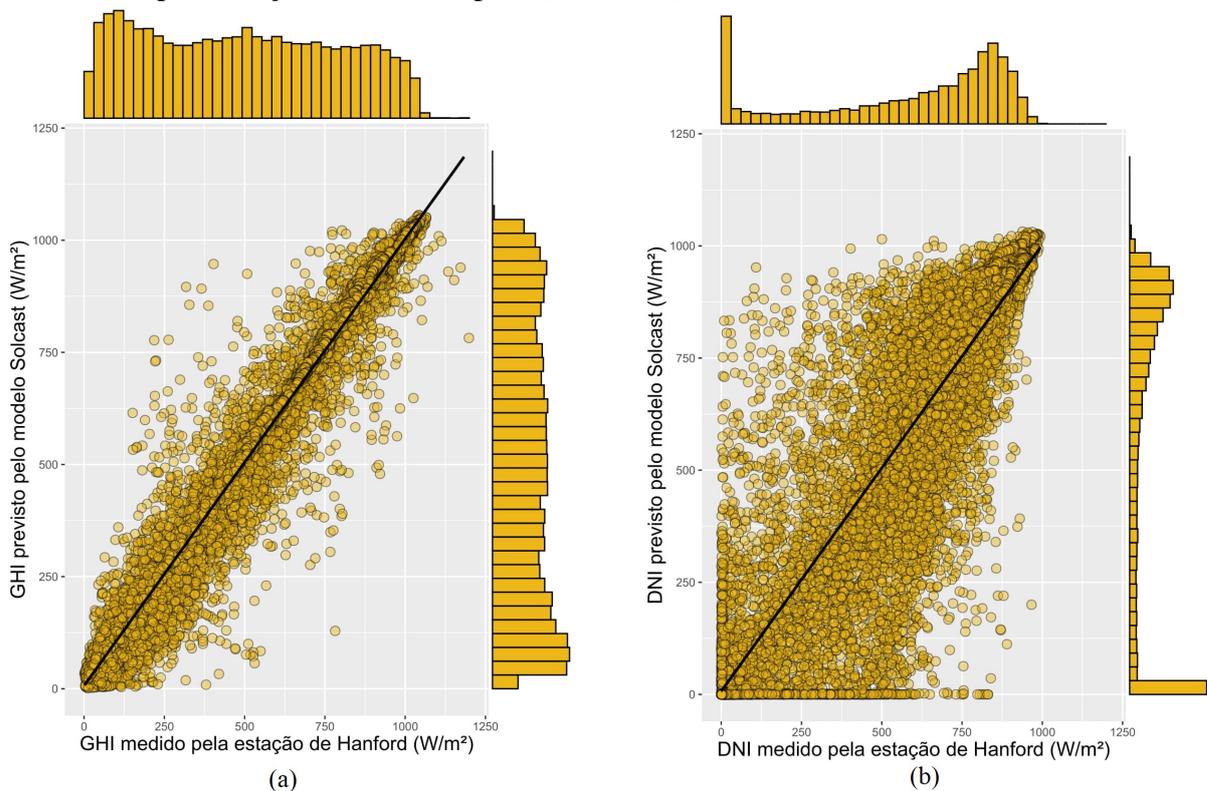
indicando que os dados previstos e os mensurados pela estação possuem uma correlação menor. O mesmo é visto para o caso do DNI, dessa vez de forma mais acentuada. Os valores das métricas corroboram com as conclusões apresentadas: para GHI, RMSE teve valor de 141,54 W/m^2 e MBE valor de $-75,49 W/m^2$; enquanto DNI teve RMSE de 318,92 W/m^2 e MBE de $-92,99 W/m^2$. Isso significa uma piora de 60,53% e 78,77% para RMSE de GHI e DNI, respectivamente. O desempenho da métrica MBE também foi inferior nesse caso, indicando que o modelo treinado com dados de Petrolina passa a subestimar significativamente irradiâncias previstas para Hanford, quando comparado com a configuração HH.

4.3 Comparação com modelos de referência

Em relação aos modelos de referência, Solcast obteve RMSE de 57,88 W/m^2 para irradiância global horizontal e 152,30 W/m^2 para direta normal. O modelo PSM foi o que obteve o pior desempenho na previsão de GHI quando comparado com os demais, atingindo um valor de 117,56 W/m^2 ; já o RMSE para o DNI foi de 169,03 W/m^2 , possuindo o segundo menor valor dentre os modelos analisados. Os gráficos de dispersão para os modelos de Solcast (GHI e DNI) e PSM (GHI e DNI) são apresentados pelas figuras 28 e 29. Ressalta-se que para GHI estimado pelo PSM, que possui a menor resolução temporal e conseqüentemente capta menos detalhes da variação de irradiância, o resultado obtido por este serviço de estimacão foi pior quando comparado o RMSE com os demais modelos analisados. Novamente, o mesmo comportamento para dados previstos de DNI observados na previsão da rede híbrida ocorre nos resultados obtidos pelos Solcast e PSM, mostrando que esta irradiância é de modelagem mais complexa para os modelos de referências.

Considerando a métrica MBE, Solcast obteve os valores de $-4,94 W/m^2$ e $-4,82$ para GHI e DNI, respectivamente. Por fim, PSM contabilizou os valores de 2,88 W/m^2 para GHI e $-11,89 W/m^2$ para DNI. Como discutido previamente, a métrica MBE descreve a tendência de um modelo a superestimar ou subestimar as suas previsões. Dessa forma, ambos modelos CNN-LSTM e PSM superestimaram suas previsões para GHI, enquanto Solcast tendeu a subestima-la. No caso do DNI, todos os modelos foram unânimes em subestimar esse tipo de irradiância, com destaque para o PSM que obteve o menor valor de MBE, indicando que é o modelo que mais subestima o DNI.

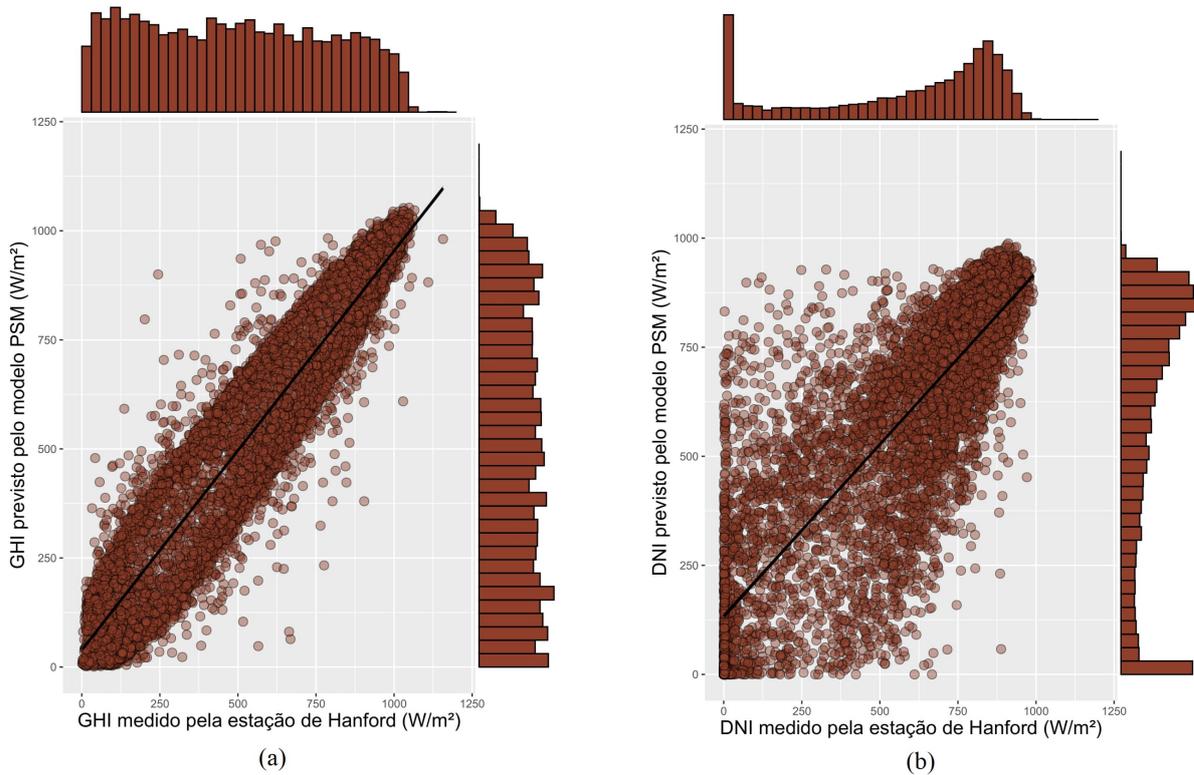
Figura 28 – Gráfico de dispersão dos dados previstos pelo modelo Solcast e mensurados *in loco* pela estação de Hanford para (a) GHI e (b) DNI.



Fonte: o autor.

Observando os resultados obtidos pelo modelo CNN-LSTM para a localidade de Petrolina (ROCHA; SANTOS, 2022), o RMSE obtido foi de 161,89 W/m² para GHI e 238,22 W/m² para DNI. Quando comparado com o resultado para previsão para Hanford, é notável que para ambas irradiâncias a métrica RMSE foi menor na localidade estadunidense. Isso pode ser explicado pelo fato de que os dados disponibilizados pela estação meteorológica de Hanford possui 74,45% a mais de observações, abrangendo todo o ano de 2018. Por sua vez, a estação SONDA em Petrolina não disponibiliza dados para os meses de agosto a novembro, o que corresponde em torno de 33% de todo ano de 2018. Assim, sendo os dados de irradiância uma série temporal onde um valor subsequente depende de valores anteriores a ele, a rede neural que utiliza os dados de Hanford é capaz de analisar todo o comportamento de irradiação ao longo do ano, identificando de forma superior a relação entre os dados fornecidos à ela e, conseqüentemente, melhorando seu desempenho.

Figura 29 – Gráfico de dispersão dos dados previstos pelo modelo PSM e mensurados *in loco* pela estação de Hanford para (a) GHI e (b) DNI.



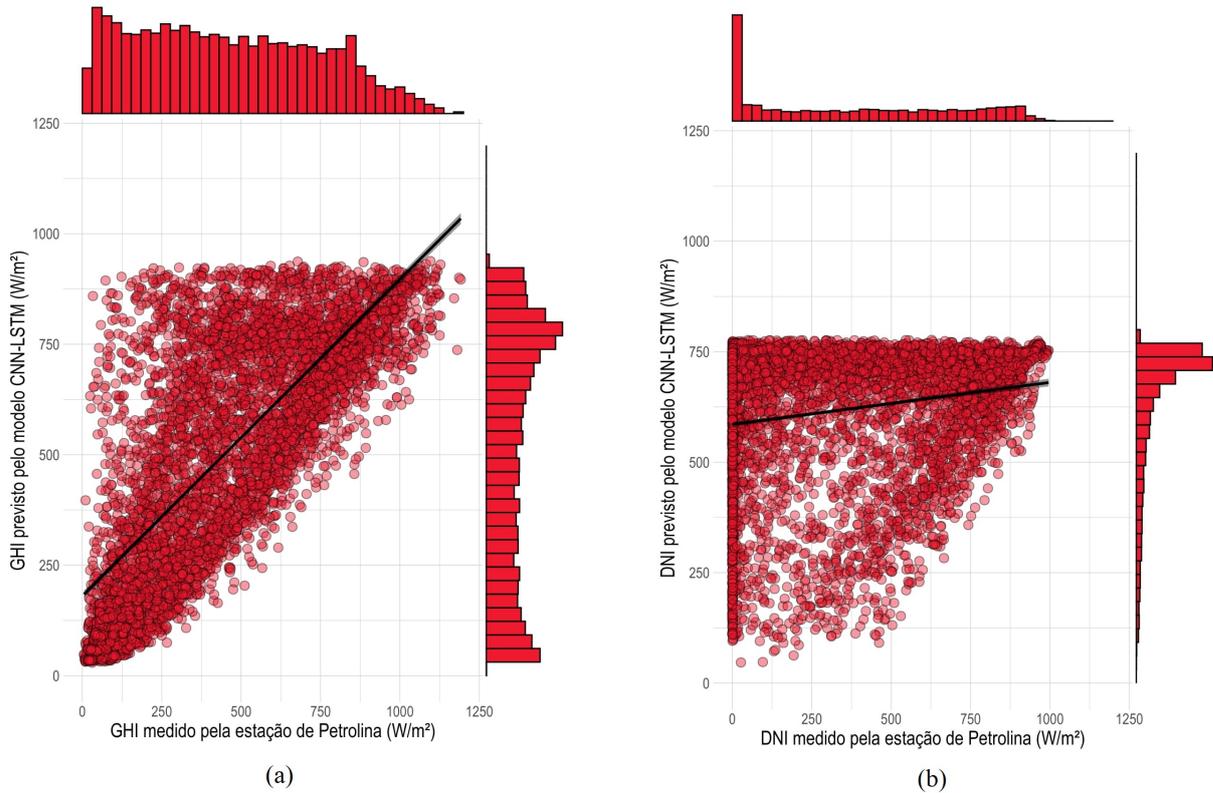
Fonte: o autor.

4.4 Resultados para previsão de irradiâncias: treino com dados de Hanford e teste com dados de Petrolina

O espalhamento mostrado pela figura 30 mostra que para para HP_1 há ainda mais espalhamento dos pontos ao redor da linha de regressão quando comparado com os demais resultados obtidos. Novamente, destaca-se de forma negativa a previsão proposta pelo modelo para DNI, que dessa vez mostra menos correlação entre os dados mensurados em Petrolina e os previstos pela CNN-LSTM. No treino e teste do tipo HP_1 , GHI obteve RMSE de 200,18 W/m² e MBE de 49,72 W/m². Já para DNI, RMSE foi de 410,07 W/m² e MBE de 245,46 W/m². Quando comparado com os resultados obtidos pelo modelo proposto em Rocha e Santos (2022), a piora do RMSE foi de 23,65% para GHI e 72,14% para DNI. Novamente, o uso de dados de uma estação para previsão de irradiância em outra não obteve um desempenho satisfatório: o modelo tem sua capacidade de previsão fortemente depreciada, não sendo capaz de gerar resultados bons.

Respostas similares foram obtidas para HP_2 , como mostra a figura 31. Dessa vez,

Figura 30 – Gráfico de dispersão das irradiâncias (a) GHI e (b) DNI previstas pelo modelo CNN-LSTM na forma HP_1 .

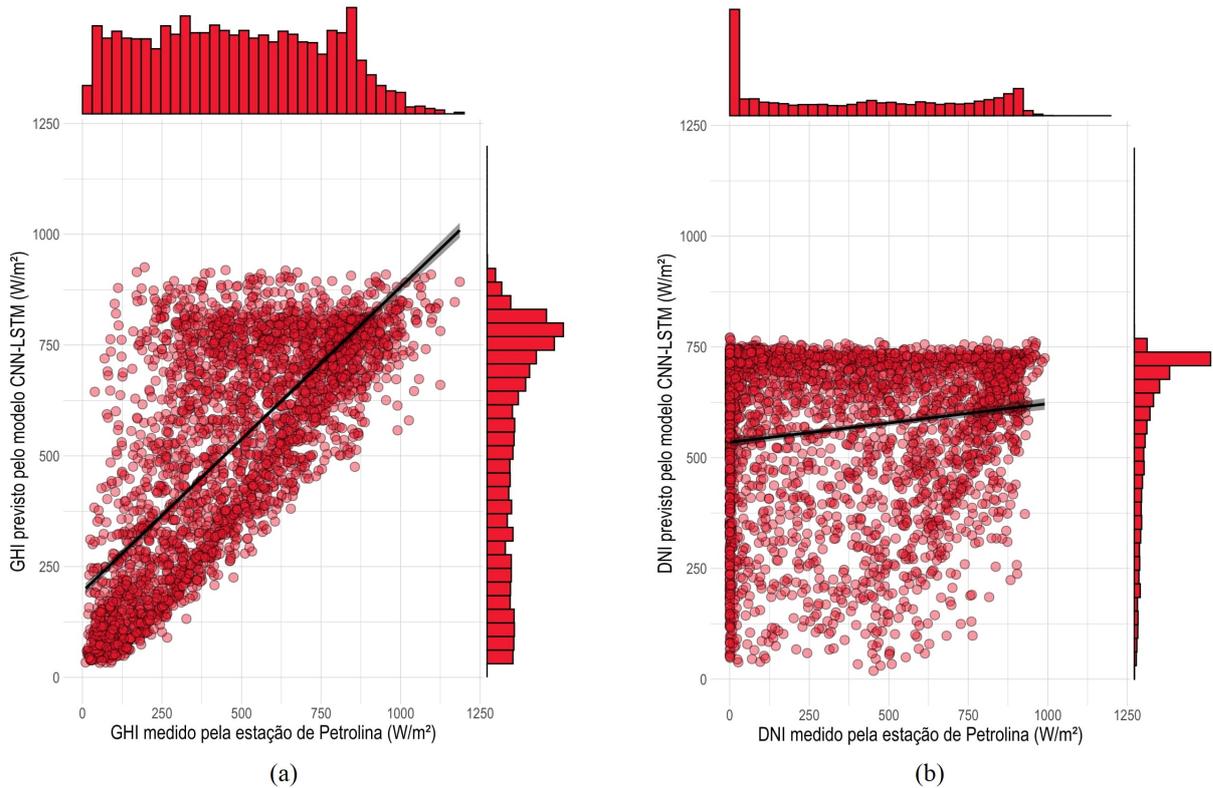


Fonte: o autor.

houve leve melhora nos resultados quando comparados com HP_1 : RMSE e MBE para GHI foram $196,60 \text{ W/m}^2$ e $43,48 \text{ W/m}^2$, respectivamente. Enquanto para DNI, o modelo obteve RMSE de $381,87 \text{ W/m}^2$ e MBE de $163,73 \text{ W/m}^2$. A melhora para treino e teste usando a mesma quantidade de observações, entretanto, não é suficiente para garantir bom desempenho do modelo, uma vez que os resultados ainda são inferiores aos obtidos pelo modelo proposto por Rocha e Santos (2022): HP_2 obteve piora de 21,44% e 60,30% para RMSE de GHI e DNI, respectivamente.

Para melhor verificar o desempenho dos modelos utilizados, as figuras 32 e 33 são apresentadas. Nelas, a relação entre as métricas RMSE e MBE para cada um dos modelos é disposta em um plano cartesiano, onde quanto mais próximo da origem estiver o resultado de métrica, melhor desempenho o modelo obteve. Para figura 32, são apresentados os resultados obtidos por este trabalho para previsão de irradiância pela rede CNN-LSTM para localidade de Hanford pelas configurações HH e PH, comparando com os modelos de referência. Já na figura 33, os resultados foram obtidos pela rede convolucional para as configurações PP, HP_1 e HP_2 .

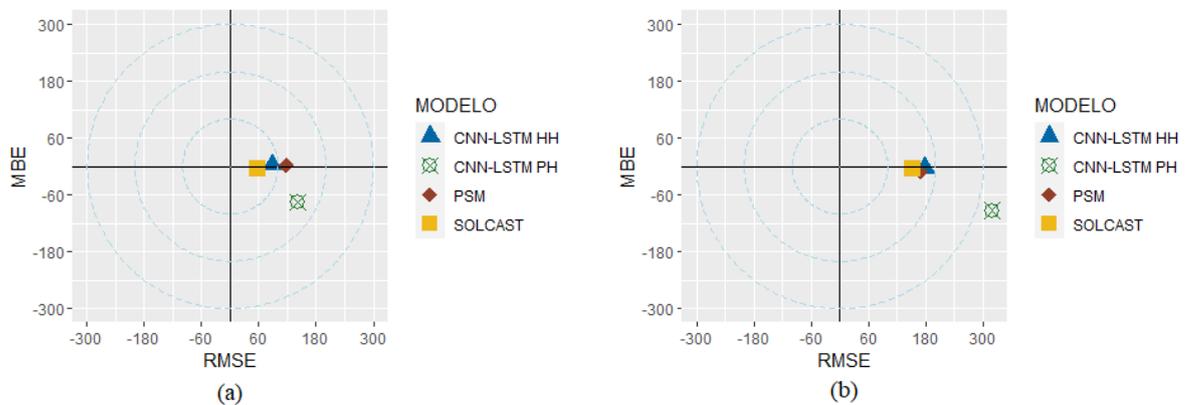
Figura 31 – Gráfico de dispersão das irradiâncias (a) GHI e (b) DNI previstas pelo modelo CNN-LSTM na forma HP₂.



Fonte: o autor.

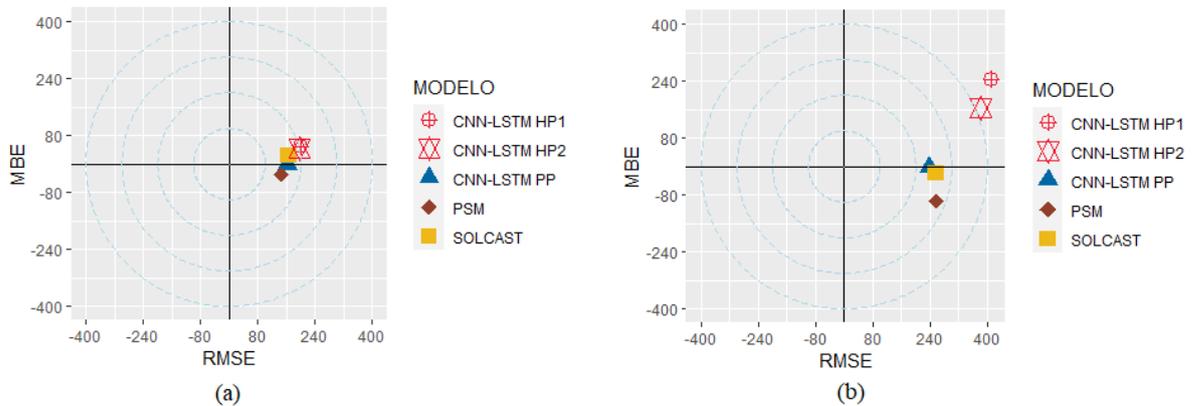
Ressalta-se ainda que na figura 33, os resultados de PP e dos modelos Solcast e PSM, foram obtidos no trabalho de Rocha e Santos (2022), conforme apresentado pela tabela 9.

Figura 32 – Gráfico em forma de alvo para resultados de previsão para Hanford. Imagem (a) apresenta resultados para GHI e imagem (b) apresenta resultados para DNI.



Fonte: o autor.

Figura 33 – Gráfico em forma de alvo para resultados de previsão para Petrolina. Imagem (a) apresenta resultados para GHI e imagem (b) apresenta resultados para DNI.



Fonte: o autor.

No caso de previsão para Hanford, pela figura 32 é possível visualizar mais claramente que os resultados obtidos pelo modelo Solcast foram os melhores para ambas irradiâncias analisadas. Para o caso GHI, os modelos se encontram mais distantes entre si, demonstrando que uma diferença mais acentuada nos valores obtidos. No caso do DNI, já é possível notar que os modelos se encontram mais agrupados, indicando convergência entre os modelos para esta irradiância. Ainda para Hanford, quando analisado o resultado obtido por PH, fica mais evidente que o desempenho dessa configuração foi insatisfatório, uma vez que estes ficou mais distante da origem. Considerando as previsões para Petrolina, pela figura 33 novamente é possível verificar que treinamento com dados de Hanford e teste com dados de Petrolina não obtiveram bom desempenho. As configurações HP₁ e HP₂ foram as que ficaram mais longe da origem dos gráficos, em especial no caso de previsão de DNI.

Sabendo então que o desempenho do modelo CNN-LSTM com treino e teste usando locais distintos não alcançou desempenho satisfatório, a tabela 8 agrupa os melhores resultados obtidos para previsão de irradiância para a localidade de Hanford (HH), enquanto a tabela 9 agrupa os resultados obtidos para Petrolina (PP), conforme em Rocha e Santos (2022).

Pela tabel 8, é possível destacar que a rede convolucional conseguiu melhor desempenho quando comparado com o PSM, reduzindo em 25,00% o RMSE para GHI. Entretanto, Solcast conseguiu um resultado superior, obtendo o melhor desempenho nesta área. Considerando DNI, a rede convolucional obteve o maior erro, não sendo capaz de superar os serviços de referência. Entretanto, pode-se afirmar que para DNI o modelo obtido pela CNN-LSTM foi mais

Tabela 8 – Resultados de RMSE e MBE para os modelos analisados em Hanford.

Hanford							
Modelo	Resolução temporal (minutos)	GHI			DNI		
		RMSE (W/m ²)	MBE (W/m ²)	Variação RMSE	RMSE (W/m ²)	MBE (W/m ²)	Variação RMSE
CNN-LSTM	15	88,17	3,90	-	178,40	-3,10	-
Solcast	15	57,88	-4,94	52,33%	152,30	-4,82	17,14%
PSM	30	117,56	2,88	-25,00%	169,03	-11,89	5,54%

preciso quando analisado seu valor de MBE, sendo ainda o que menos subestima previsão pra irradiância normal direta.

Tabela 9 – Resultados de RMSE e MBE para os modelos analisados em Petrolina (ROCHA; SANTOS, 2022).

Petrolina							
Modelo	Resolução temporal (minutos)	GHI			DNI		
		RMSE (W/m ²)	MBE (W/m ²)	Variação RMSE	RMSE (W/m ²)	MBE (W/m ²)	Variação RMSE
CNN-LSTM	15	161,89	-0,33	-	238,22	-3,66	-
Solcast	15	165,91	22,89	-2,42%	256,85	18,77	-7,25%
PSM	30	147,23	-29,86	9,96%	257,75	-98,37	-7,58%

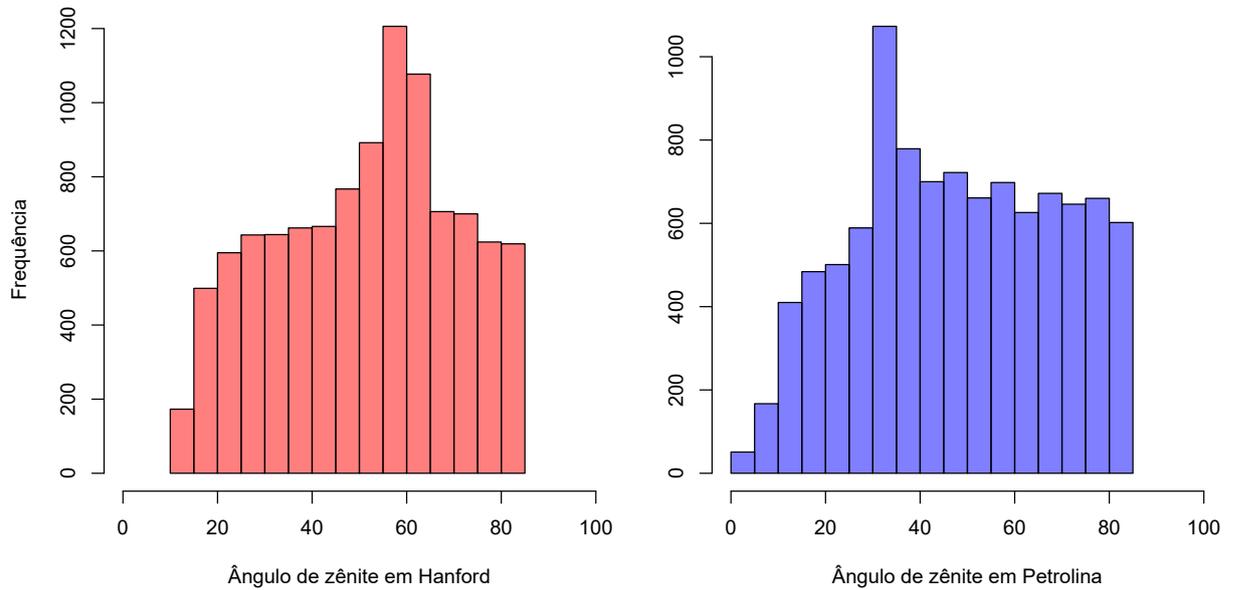
Comparando as tabelas 8 e 9, é possível notar que todos os modelos obtiveram resultados melhores em termos de RMSE para Hanford do que para Petrolina. Novamente, isso se dá ao fato de haver mais dados medidos *in loco* na estação de Hanford, permitindo melhor desempenho de previsão tanto do modelo CNN-LSTM quanto dos usados como referencial. Para o caso de Petrolina, é possível afirmar que a rede convolucional híbrida se saiu melhor para uma localidade com dados mais dispersos entre si, no sentido de que menos medições de irradiância estavam disponíveis, com destaque para previsão de DNI.

4.5 Viés geográfico da rede CNN-LSTM e transferência de aprendizado

Como apresentado anteriormente, o desempenho da rede neural na previsão de irradiâncias caiu substancialmente quando utilizados dados para treino e teste de localidade distintas. Isso pode ser explicado pelo chamado viés geográfico dos dados, que não foi transferido de forma satisfatória durante a etapa de treinamento. Como mostrado pela tabela 3, Petrolina e Hanford estão localizadas em hemisférios diferentes. Isso impacta diretamente na incidência solar em cada localidade, uma vez que para cada uma delas o ângulo de zênite também será distinto. A influência da localização de ambas estações na previsão de irradiância pode ser

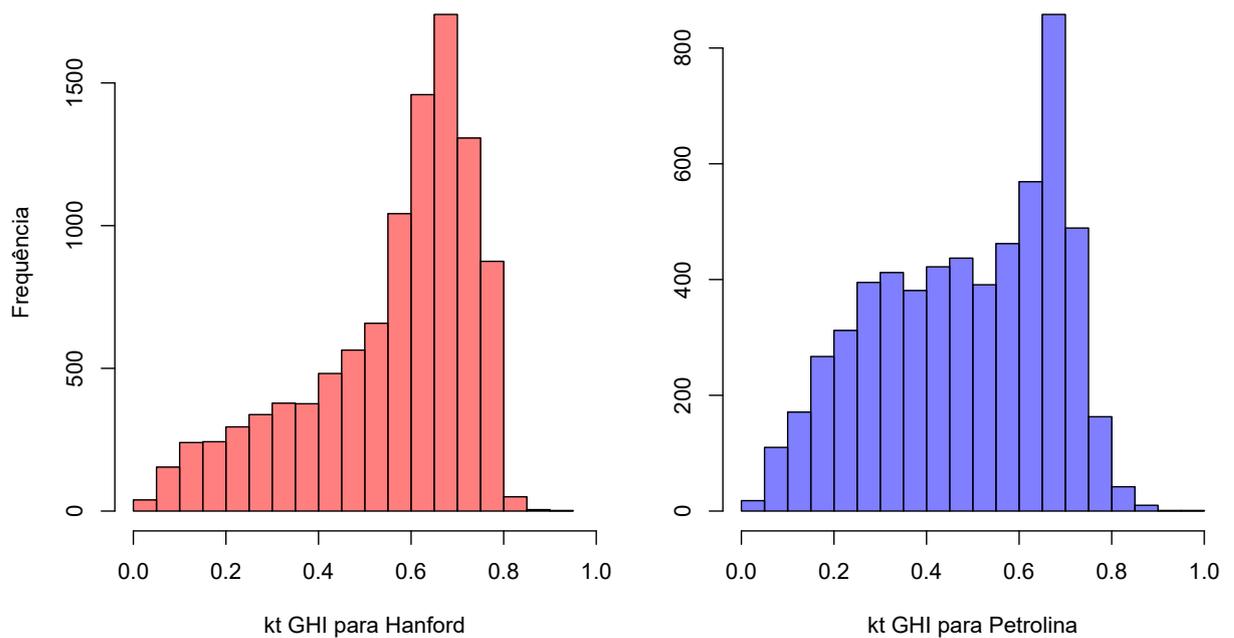
observada melhor nas figuras 34 e 35.

Figura 34 – Distribuição dos valores de ângulo de zênite para Hanford (à esquerda) e para Petrolina (à direita) durante o ano de 2018.



Fonte: o autor.

Figura 35 – Distribuição dos valores índice de céu claro para Hanford (à esquerda) e para Petrolina (à direita) durante o ano de 2018.



Fonte: o autor.

A figura 34 mostra que em Hanford, a incidência solar, em sua maior parte durante o ano de 2018, se deu com uma inclinação em torno de 60° , enquanto em Petrolina esse valor ficou mais próximo de 30° . Uma vez que pela equação 3.2 utilizada para modelagem do irradiância de céu claro é dada em função do ângulo de zênite, é de se esperar que os valores obtidos para o índice de céu claro também sejam distintos para ambas localidades. Isto é ilustrado pela figura 35. Por ela, é possível ver a diferença entre os valores desse índice: enquanto ambas localidades possuem maior frequência de concentrada entre 0,6 e 0,8, a distribuição em Petrolina possui comportamento mais uniforme durante o ano de 2018. Isso pode ser explicado pela própria localização de Petrolina, que está em uma zona tropical próxima ao Equador terrestre e menos propensa à variações climáticas e, conseqüentemente, a menores variações de ângulo de zênite. Enquanto isso, Hanford está em uma zona temperada, sendo assim mais sujeito às variações das estações do ano.

Desta forma, pode-se afirmar que a transferência de aprendizado foi afetada, no sentido de que o modelo não foi capaz de identificar satisfatoriamente as características geográficas de cada local. Isto refletiu em um desempenho inferior da rede convolucional quando comparada com modelo obtido por treino e teste usando dados de uma mesma região, indicando que o viés geográfico é mais difícil de ser modelado para regiões distintas entre si, prejudicando a transferência de aprendizado e a generalização do modelo.

5 CONCLUSÃO

Neste trabalho, verificou-se o efeito do viés geográfico através da transferência de aprendizado no desempenho da rede convolucional desenvolvida por Rocha e Santos (2022) para a localidade de Hanford, na Califórnia. A rede possui uma configuração híbrida, combinando rede neural convolucional (CNN) e rede neural recorrente (LSTM) na previsão de de GHI e DNI usando imagens do satélite GOES-16 para a localidade de Hanford, Califórnia. Os resultados obtidos pela rede foram comparados com previsões realizadas por modelos físicos de referência Solcast e PSM utilizando as métricas RMSE e MBE. O viés geográfico foi analisado por meio de treino e teste para diferentes configurações, na forma HH, PH, HP₁ e HP₂.

O melhor resultado do modelo CNN-LSTM foi obtido para treino e teste com dados da estação de Hanford (HH). Neste caso, para GHI o RMSE foi de 88,17 W/m² e para DNI o valor obtido foi 178,40 W/m². Quando comparados com modelos de referência, Solcast teve melhor desempenho para ambas irradiâncias analisadas, conseguindo atingir valores de RMSE iguais à 57,88 W/m² e 152,30 W/m² para GHI e DNI, respectivamente. O desempenho do modelo PSM pode ser destacado de forma negativa para previsão de GHI em Hanford, sendo esta a irradiância com maior valor de RMSE com 117,56 W/m². Para DNI, este serviço obteve o segundo melhor desempenho, sendo superado apenas pelo Solcast. Entretanto, o modelo CNN-LSTM foi o que menos subestimou DNI de acordo com o MBE, que foi igual à -3,90 W/m², sendo um modelo mais preciso do que os demais.

Apesar do bom desempenho da rede para o caso HH, resultados satisfatórios não foram obtidos para as demais configurações de treino e teste (PH, HP₁ e HP₂). De fato, houve piora no desempenho do modelo convolucional quando treino e teste foram realizados com dados para estações dissimilares. Isso pode ser explicado pelo viés geográfico inerente de cada estação: Petrolina e Hanford estão em regiões muito distintas entre si, o que influencia diretamente no modelo de previsão obtido pela CNN-LSTM. Isso reflete diretamente no modelo utilizado para previsão de irradiância de céu claro e, conseqüentemente, nos valores de índice de céu claro e de irradiância incidente em cada local. Dessa forma, pode-se concluir que o viés geográfico possui papel de destaque na previsão de irradiância solar, não sendo possível transferir de forma satisfatória o aprendizado adquirido durante a fase de treinamento entre locais muito distintos entre si.

Comparando os resultados obtidos para Hanford (HH) e os obtidos em Petrolina (PP), é notável como para a primeira localidade os modelos obtiveram melhores desempenhos. Isso pode ser justificado pelo fato de Petrolina não possuir dados para todos os meses do ano de 2018. Ainda para Petrolina, considerando DNI, o modelo CNN-LSTM conseguiu melhorar em média 7,41% os resultados quando comparados com modelos físicos de referência. Isso pode indicar que para situações com menos dados disponíveis (como no caso de Petrolina), o modelo tem melhor desempenho para estimação de DNI.

A partir dos resultados apresentados, conclui-se que o modelo estudado conseguiu desempenho satisfatório para ambas localidades de Petrolina e Hanford (quando treinado e testado com dados de uma mesma localidade), mostrando ser uma alternativa viável aos modelos físicos Solcast e PSM. Entretanto, a rede neural CNN-LSTM analisada neste trabalho não é capaz de prever satisfatoriamente irradiância quando treinada com dados de um local dissimilar do qual se deseja realizar a previsão, mostrando que há realmente influência de um viés geográfico durante a fase de treinamento.

REFERÊNCIAS

- ALBAWI, S.; MOHAMMED, T. A.; AL-ZAWI, S. Understanding of a convolutional neural network. In: IEEE. **2017 international conference on engineering and technology (ICET)**. [S.l.], 2017. p. 1–6.
- ALENCAR, D. Barbosa de; AFFONSO, C. de M.; OLIVEIRA, R. C. Limão de; RODRIGUEZ, J. L. M.; LEITE, J. C.; FILHO, J. C. R. Different models for forecasting wind power generation: Case study. **Energies**, Multidisciplinary Digital Publishing Institute, v. 10, n. 12, p. 1976, 2017.
- ALKHAYAT, G.; MEHMOOD, R. A review and taxonomy of wind and solar energy forecasting methods based on deep learning. **Energy and AI**, Elsevier, v. 4, p. 100060, 2021.
- ALONSO-MONTESINOS, J.; BATLLES, F. The use of a sky camera for solar radiation estimation based on digital image processing. **Energy**, Elsevier, v. 90, p. 377–386, 2015.
- ANGSTROM, A. Solar and terrestrial radiation. report to the international commission for solar research on actinometric investigations of solar and atmospheric radiation. **Quarterly Journal of the Royal Meteorological Society**, Wiley Online Library, v. 50, n. 210, p. 121–126, 1924.
- ARIA, M.; CUCCURULLO, C. bibliometrix: An r-tool for comprehensive science mapping analysis. **Journal of informetrics**, Elsevier, v. 11, n. 4, p. 959–975, 2017.
- BADESCU, V.; GUEYMARD, C. A.; CHEVAL, S.; OPREA, C.; BACIU, M.; DUMITRESCU, A.; IACOBESCU, F.; MILOS, I.; RADA, C. Computing global and diffuse solar hourly irradiation on clear sky. review and testing of 54 models. **Renewable and Sustainable Energy Reviews**, Elsevier, v. 16, n. 3, p. 1636–1656, 2012.
- BASTOS, B. Q.; OLIVEIRA, F. L. C.; MILIDIÚ, R. L. U-convolutional model for spatio-temporal wind speed forecasting. **International Journal of Forecasting**, Elsevier, v. 37, n. 2, p. 949–970, 2021.
- BENGIO, Y.; SIMARD, P.; FRASCONI, P. Learning long-term dependencies with gradient descent is difficult. **IEEE transactions on neural networks**, IEEE, v. 5, n. 2, p. 157–166, 1994.
- BRIGHT, J. M. Solcast: Validation of a satellite-derived solar irradiance dataset. **Solar Energy**, Elsevier, v. 189, p. 435–449, 2019.
- CANAN, F.; GOLASI, I.; FALASCA, S.; SALATA, F. Outdoor thermal perception and comfort conditions in the köppen-geiger climate category bsk. one-year field survey and measurement campaign in konya, turkey. **Science of the Total Environment**, Elsevier, v. 738, p. 140295, 2020.
- CHEN, J.; ZHU, W.; YU, Q. Estimating half-hourly solar radiation over the continental united states using goes-16 data with iterative random forest. **Renewable Energy**, Elsevier, v. 178, p. 916–929, 2021.
- CHOLLET, F.; ALLAIRE, J. J. **Deep Learning with R**. Shelter Island: Manning, 2018. 336 p.
- CORNEJO-BUENO, L.; CASANOVA-MATEO, C.; SANZ-JUSTO, J.; SALCEDO-SANZ, S. Machine learning regressors for solar radiation estimation from satellite data. **Solar Energy**, Elsevier, v. 183, p. 768–775, 2019.

CUN, Y. L.; JACKEL, L. D.; BOSER, B.; DENKER, J. S.; GRAF, H. P.; GUYON, I.; HENDERSON, D.; HOWARD, R. E.; HUBBARD, W. Handwritten digit recognition: Applications of neural network chips and automatic learning. **IEEE Communications Magazine**, IEEE, v. 27, n. 11, p. 41–46, 1989.

DONTHU, N.; KUMAR, S.; MUKHERJEE, D.; PANDEY, N.; LIM, W. M. How to conduct a bibliometric analysis: An overview and guidelines. **Journal of Business Research**, Elsevier, v. 133, p. 285–296, 2021.

DUFFIE, J. A.; BECKMAN, W. A. **Solar Engineering of Thermal Processes**. Noboken: Wiley, 2013. 910 p.

DUFFY, P. B.; FIELD, C. B.; DIFFENBAUGH, N. S.; DONEY, S. C.; DUTTON, Z.; GOODMAN, S.; HEINZERLING, L.; HSIANG, S.; LOBELL, D. B.; MICKLEY, L. J. *et al.* Strengthened scientific support for the endangerment finding for atmospheric greenhouse gases. **Science**, American Association for the Advancement of Science, v. 363, n. 6427, p. eaat5982, 2019.

FENG, C.; ZHANG, J. Solarnet: A sky image-based deep convolutional neural network for intra-hour solar forecasting. **Solar Energy**, Elsevier, v. 204, p. 71–78, 2020.

FENG, C.; ZHANG, J.; ZHANG, W.; HODGE, B.-M. Convolutional neural networks for intra-hour solar forecasting based on sky image sequences. **Applied Energy**, Elsevier, v. 310, p. 118438, 2022.

FREITAS, A. A.; DRUMOND, A.; CARVALHO, V. S.; REBOITA, M. S.; SILVA, B. C.; UVO, C. B. Drought assessment in São Francisco river basin, Brazil: characterization through SPI and associated anomalous climate patterns. **Atmosphere**, Multidisciplinary Digital Publishing Institute, v. 13, n. 1, p. 41, 2022.

GAO, B.; HUANG, X.; SHI, J.; TAI, Y.; ZHANG, J. Hourly forecasting of solar irradiance based on ceemdan and multi-strategy cnn-lstm neural networks. **Renewable Energy**, Elsevier, v. 162, p. 1665–1683, 2020.

GETIRANA, A.; LIBONATI, R.; CATALDI, M. **Brazil is in water crisis—it needs a drought plan**. [S.l.]: Nature Publishing Group, 2021.

GHIMIRE, S.; DEO, R. C.; RAJ, N.; MI, J. Deep solar radiation forecasting with convolutional neural network and long short-term memory network algorithms. **Applied Energy**, Elsevier, v. 253, p. 113541, 2019.

GOOGLE EARTH. **Localização geográfica da estação meteorológica de Hanford, na Califórnia**. 2022. Disponível em: <<https://earth.google.com/web/>>. Acesso em: 20 Maio 2022.

GUEYMARD, C. A. Progress in direct irradiance modeling and validation. In: **Proc. ASES Annual Conf. Phoenix, AZ, USA**. [S.l.: s.n.], 2010.

HABTE, A.; SENGUPTA, M.; LOPEZ, A. **Evaluation of the national solar radiation database (NSRDB): 1998-2015**. [S.l.], 2017.

HASSAN, M. A.; KHALIL, A.; KASEB, S.; KASSEM, M. Independent models for estimation of daily global solar radiation: A review and a case study. **Renewable and Sustainable Energy Reviews**, Elsevier, v. 82, p. 1565–1575, 2018.

- HAYKIN, S. S. **Neural networks and learning machines**. [S.l.]: Pearson, 2009. v. 3.
- HEBB, D. O. **The organization of behavior: A neuropsychological theory**. Nova Iorque: Psychology Press, 1949. 378 p.
- HILLGER, D. W.; SCHMIT, T. Quantization noise for goes-r abi bands. In: **Proceedings of the 13th Conference on Satellite Meteorology and Oceanography**. [S.l.]: American Meteorological Society, 2004.
- HOCHREITER, S.; SCHMIDHUBER, J. Long short-term memory. **Neural computation**, MIT Press, v. 9, n. 8, p. 1735–1780, 1997.
- HOEHNDORF, R.; QUERALT-ROSINACH, N. Data science and symbolic ai: Synergies, challenges and opportunities. **Data Science**, IOS Press, v. 1, n. 1-2, p. 27–38, 2017.
- HUANG, X.; LI, Q.; TAI, Y.; CHEN, Z.; ZHANG, J.; SHI, J.; GAO, B.; LIU, W. Hybrid deep neural model for hourly solar irradiance forecasting. **Renewable Energy**, Elsevier, v. 171, p. 1041–1060, 2021.
- HUERTAS-TATO, J.; GALVÁN, I. M.; ALER, R.; RODRÍGUEZ-BENÍTEZ, F. J.; POZO-VÁZQUEZ, D. Using a multi-view convolutional neural network to monitor solar irradiance. **Neural Computing and Applications**, Springer, p. 1–13, 2021.
- HUNT, J. D.; NASCIMENTO, A.; CATEN, C. S. ten; TOMÉ, F. M. C.; SCHNEIDER, P. S.; THOMAZONI, A. L. R.; CASTRO, N. J. de; BRANDÃO, R.; FREITAS, M. A. V. de; MARTINI, J. S. C.; RAMOS, D. S.; SENNE, R. Energy crisis in Brazil: Impact of hydropower reservoir level on the river flow. **Energy**, v. 239, p. 121927, 2022.
- INEICHEN, P. Comparison of eight clear sky broadband models against 16 independent data banks. **Solar Energy**, Elsevier, v. 80, n. 4, p. 468–478, 2006.
- INPE. **Localização das estações SONDA**. 2022. Disponível em: <<http://sonda.ccst.inpe.br/>>. Acesso em: 20 Maio 2022.
- IPCC. **Climate Change 2021: The Physical Science Basis**. 2021.
- IRENA. **Renewable Capacity Statistics 2021 International Renewable Energy Agency (IRENA), Abu Dhabi**. 2021.
- JEBLI, I.; BELOUADHA, F.-Z.; KABBAJ, M. I.; TILIOUA, A. Prediction of solar energy guided by pearson correlation using machine learning. **Energy**, Elsevier, v. 224, p. 120109, 2021.
- KALLURI, S.; ALCALA, C.; CARR, J.; GRIFFITH, P.; LEBAIR, W.; LINDSEY, D.; RACE, R.; WU, X.; ZIERK, S. From photons to pixels: processing data from the advanced baseline imager. **Remote Sensing**, Multidisciplinary Digital Publishing Institute, v. 10, n. 2, p. 177, 2018.
- KAMATH, H. G.; SRINIVASAN, J. Validation of global irradiance derived from insat-3d over india. **Solar Energy**, Elsevier, v. 202, p. 45–54, 2020.
- KAUR, A.; NONNENMACHER, L.; PEDRO, H. T.; COIMBRA, C. F. Benefits of solar forecasting for energy imbalance markets. **Renewable energy**, Elsevier, v. 86, p. 819–830, 2016.

- KOO, Y.; OH, M.; KIM, S.-M.; PARK, H.-D. Estimation and mapping of solar irradiance for korea by using coms mi satellite images and an artificial neural network model. **Energies**, Multidisciplinary Digital Publishing Institute, v. 13, n. 2, p. 301, 2020.
- KUHN, P. M.; GARSCHKE, D.; WILBERT, S.; NOURI, B.; HANRIEDER, N.; PRAHL, C.; ZARZARLEJO, L.; FERNÁNDEZ, J.; KAZANTZIDIS, A.; SCHMIDT, T. *et al.* Shadow-camera based solar nowcasting system for shortest-term forecasts. **Meteorologische Zeitschrift**, Borntraeger Science Publishers, 2019.
- KUMARI, P.; TOSHNIWAL, D. Deep learning models for solar irradiance forecasting: A comprehensive review. **Journal of Cleaner Production**, Elsevier, v. 318, p. 128566, 2021.
- LARSON, D. P.; LI, M.; COIMBRA, C. F. Scope: Spectral cloud optical property estimation using real-time goes-r longwave imagery. **Journal of Renewable and Sustainable Energy**, AIP Publishing LLC, v. 12, n. 2, p. 026501, 2020.
- LAW, E. W.; KAY, M.; TAYLOR, R. A. Calculating the financial value of a concentrated solar thermal plant operated using direct normal irradiance forecasts. **Solar Energy**, Elsevier, v. 125, p. 267–281, 2016.
- LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. **nature**, Nature Publishing Group, v. 521, n. 7553, p. 436–444, 2015.
- LIMA, M. A. F.; RAMÍREZ, L. M. F.; CARVALHO, P.; BATISTA, J. G.; FREITAS, D. M. A comparison between deep learning and support vector regression techniques applied to solar forecast in spain. **Journal of Solar Energy Engineering**, American Society of Mechanical Engineers Digital Collection, v. 144, n. 1, 2022.
- MABASA, B.; LYSKO, M. D.; MOLOI, S. J. Validating hourly satellite based and reanalysis based global horizontal irradiance datasets over south africa. **Geomatics**, Multidisciplinary Digital Publishing Institute, v. 1, n. 4, p. 429–449, 2021.
- MANISALIDIS, I.; STAVROPOULOU, E.; STAVROPOULOS, A.; BEZIRTZOGLOU, E. Environmental and health impacts of air pollution: a review. **Frontiers in public health**, Frontiers, p. 14, 2020.
- MARINHO, F. P. **Avaliação de métodos de estimativa da capacidade de carga de fundações diretas em solos não saturados**. 2020. 71 f. Dissertação (Mestrado em Engenharia Mecânica) — Centro de Tecnologia, Programa de Pós-Graduação em Engenharia Mecânica, Universidade Federal do Ceará, Fortaleza, 2020.
- MARQUEZ, R.; COIMBRA, C. F. Proposed metric for evaluation of solar forecasting models. **Journal of solar energy engineering**, American Society of Mechanical Engineers Digital Collection, v. 135, n. 1, 2013.
- MARTINS, F. R.; PEREIRA, E. B.; YAMASHITA, C.; PEREIRA, S. V.; NETO, S. M. Base de dados climático-ambientais aplicados ao setor energético-projeto sonda. **Proc. of XII Simpósio Brasileiro de Sensoriamento Remoto, INPE, Sao José dos Campos, Brazil**, 2005.
- MARZO, A.; TRIGO-GONZALEZ, M.; ALONSO-MONTESINOS, J.; MARTÍNEZ-DURBÁN, M.; LÓPEZ, G.; FERRADA, P.; FUENTEALBA, E.; CORTÉS, M.; BATLLES, F. Daily global solar radiation estimation in desert areas using daily extreme temperatures and extraterrestrial radiation. **Renewable Energy**, Elsevier, v. 113, p. 303–311, 2017.

- MATSUNOBU, L. M.; PEDRO, H. T.; COIMBRA, C. F. Cloud detection using convolutional neural networks on remote sensing images. **Solar Energy**, Elsevier, v. 230, p. 1020–1032, 2021.
- MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. **The bulletin of mathematical biophysics**, Springer, v. 5, n. 4, p. 115–133, 1943.
- MINISTÉRIO DE MINAS E ENERGIA. **Boletim Mensal de Monitoramento do Sistema Elétrico Brasileiro, Fevereiro de 2022**. 2022. Disponível em: <<https://www.gov.br/mme/pt-br/assuntos/secretarias/energia-eletrica/publicacoes/boletim-de-monitoramento-do-sistema-eletrico/2022>>. Acesso em: 04 Maio 2022.
- NAÇÕES UNIDAS. Acordo de Paris. **United Nations Treaty Collect**, p. 1–27, 2015.
- NOAA. **Mission overview**. 2022. Disponível em: <<https://www.goes-r.gov/mission/mission.html>>. Acesso em: 15 Maio 2022.
- NOAA. **SOLRAD Network**. 2022. Disponível em: <<https://gml.noaa.gov/grad/solrad/solradsites.html>>. Acesso em: 20 Maio 2022.
- NONNENMACHER, L.; KAUR, A.; COIMBRA, C. F. Verification of the suny direct normal irradiance model with ground measurements. **Solar Energy**, Elsevier, v. 99, p. 246–258, 2014.
- NOURI, B.; BLUM, N.; WILBERT, S.; ZARZALEJO, L. F. A hybrid solar irradiance nowcasting approach: combining all sky imager systems and persistence irradiance models for increased accuracy. **Solar RRL**, Wiley Online Library, 2021.
- ORGANIZAÇÃO MUNDIAL DA SAÚDE. **Air Pollution**. 2022. Disponível em: <https://www.who.int/health-topics/air-pollution#tab=tab_1>. Acesso em: 03 Maio 2022.
- PADOVAN, A.; COL, D. D.; SABATELLI, V.; MARANO, D. Dni estimation procedures for the assessment of solar radiation availability in concentrating systems. **Energy Procedia**, Elsevier, v. 57, p. 1140–1149, 2014.
- PADULA, F.; CAO, C. **CWG Analysis: ABI Max/Min Radiance Characterization and Validation**. [S.l.], 2015.
- PAULESCU, M.; MARES, O.; PAULESCU, E.; STEFU, N.; PACURAR, A.; CALINOIU, D.; GRAVILA, P.; POP, N.; BOATA, R. Nowcasting solar irradiance using the sunshine number. **Energy conversion and management**, Elsevier, v. 79, p. 690–697, 2014.
- PEEL, M. C.; FINLAYSON, B. L.; MCMAHON, T. A. Updated world map of the köppen-geiger climate classification. **Hydrology and Earth System Sciences**, v. 11, n. 5, p. 1633–1644, 2007.
- PEREZ, R.; INEICHEN, P.; MOORE, K.; KMIECIK, M.; CHAIN, C.; GEORGE, R.; VIGNOLA, F. A new operational model for satellite-derived irradiances: description and validation. **Solar Energy**, Elsevier, v. 73, n. 5, p. 307–317, 2002.
- RAJAGUKGUK, R. A.; RAMADHAN, R. A.; LEE, H.-J. A review on deep learning models for forecasting time series data of solar irradiance and photovoltaic power. **Energies**, Multidisciplinary Digital Publishing Institute, v. 13, n. 24, p. 6623, 2020.
- ROCHA, P. A.; SANTOS, V. O. Global horizontal and direct normal solar irradiance modeling by the machine learning methods xgboost and deep neural networks with cnn-lstm layers: a case study using the goes-16 satellite imagery. **International Journal of Energy and Environmental Engineering**, Springer, p. 1–16, 2022.

ROCHA, P. C.; FERNANDES, J.; MODOLO, A.; LIMA, R. P.; SILVA, M. V. da; BEZERRA, C. D. Estimation of daily, weekly and monthly global solar radiation using anns and a long data set: a case study of fortaleza, in brazilian northeast region. **International Journal of Energy and Environmental Engineering**, Springer, v. 10, n. 3, p. 319–334, 2019.

RODRÍGUEZ, F.; AZCÁRATE, I.; VADILLO, J.; GALARZA, A. Forecasting intra-hour solar photovoltaic energy by assembling wavelet based time–frequency analysis with deep learning neural networks. **International Journal of Electrical Power & Energy Systems**, Elsevier, v. 137, p. 107777, 2022.

ROSENBLATT, F. The perceptron: a probabilistic model for information storage and organization in the brain. **Psychological review**, American Psychological Association, v. 65, n. 6, p. 386, 1958.

SAIDI, K.; OMRI, A. The impact of renewable energy on carbon emissions and economic growth in 15 major renewable energy-consuming countries. **Environmental research**, Elsevier, v. 186, p. 109567, 2020.

SALAZAR, G.; GUEYMARD, C.; GALDINO, J. B.; VILELA, O. de C.; FRAIDENRAICH, N. Solar irradiance time series derived from high-quality measurements, satellite-based models, and reanalyses at a near-equatorial site in Brazil. **Renewable and Sustainable Energy Reviews**, Elsevier, v. 117, p. 109478, 2020.

SCHMIDHUBER, J. Deep learning in neural networks: An overview. **Neural networks**, Elsevier, v. 61, p. 85–117, 2015.

SCHMIT, T. J.; GRIFFITH, P.; GUNSHOR, M. M.; DANIELS, J. M.; GOODMAN, S. J.; LEBAIR, W. J. A closer look at the abi on the goes-r series. **Bulletin of the American Meteorological Society**, American Meteorological Society, v. 98, n. 4, p. 681–698, 2017.

SENGUPTA, M.; XIE, Y.; LOPEZ, A.; HABTE, A.; MACLAURIN, G.; SHELBY, J. The national solar radiation data base (nsrdb). **Renewable and Sustainable Energy Reviews**, Elsevier, v. 89, p. 51–60, 2018.

SHI, X.; CHEN, Z.; WANG, H.; YEUNG, D.-Y.; WONG, W.-k.; WOO, W.-c. Convolutional lstm network: A machine learning approach for precipitation nowcasting. In: CORTES, C.; LAWRENCE, N.; LEE, D.; SUGIYAMA, M.; GARNETT, R. (Ed.). **Advances in Neural Information Processing Systems**. [S.l.]: Curran Associates, Inc., 2015. v. 28.

SILVA, P. C. G. d.; MOURA, M. S. B. d.; KIILL, L. H. P.; BRITO, L. T. d. L.; PEREIRA, L. A.; SA, I. B.; CORREIA, R. C.; TEIXEIRA, A. H. d. C.; CUNHA, T. J. F.; FILHO, C. G. Caracterização do semiárido brasileiro: fatores naturais e humanos. In: EMBRAPA (Ed.). **Semiárido brasileiro: pesquisa, desenvolvimento e inovação**. Petrolina: Embrapa, 2010. cap. 1, p. 18–48.

SMOLENSKY, P. Connectionist ai, symbolic ai, and the brain. **Artificial Intelligence Review**, Springer, v. 1, n. 2, p. 95–109, 1987.

STEIN, J. S.; HANSEN, C. W.; RENO, M. J. **Global horizontal irradiance clear sky models: implementation and analysis**. [S.l.], 2012.

TURING, A. M. I.—COMPUTING MACHINERY AND INTELLIGENCE. **Mind**, LIX, n. 236, p. 433–460, 1950.

- VOHRA, K.; VODONOS, A.; SCHWARTZ, J.; MARAIS, E. A.; SULPRIZIO, M. P.; MICKLEY, L. J. Global mortality from outdoor fine particle pollution generated by fossil fuel combustion: Results from geos-chem. **Environmental Research**, Elsevier, v. 195, p. 110754, 2021.
- VOYANT, C.; NOTTON, G.; KALOGIROU, S.; NIVET, M.-L.; PAOLI, C.; MOTTE, F.; FOUILLOY, A. Machine learning methods for solar radiation forecasting: A review. **Renewable Energy**, Elsevier, v. 105, p. 569–582, 2017.
- WAGNER, D. L. Insect declines in the anthropocene. **Annual review of entomology**, Annual Reviews, v. 65, p. 457–480, 2020.
- WALTHER, G.-R.; POST, E.; CONVEY, P.; MENZEL, A.; PARMESAN, C.; BEEBEE, T. J.; FROMENTIN, J.-M.; HOEGH-GULDBERG, O.; BAIRLEIN, F. Ecological responses to recent climate change. **Nature**, Nature Publishing Group, v. 416, n. 6879, p. 389–395, 2002.
- WERBOS, P. J. Applications of advances in nonlinear sensitivity analysis. In: **System modeling and optimization**. [S.l.]: Springer, 1982. p. 762–770.
- YANG, D. A correct validation of the national solar radiation data base (nsrdb). **Renewable and Sustainable Energy Reviews**, Elsevier, v. 97, p. 152–155, 2018.
- YANG, D.; BRIGHT, J. M. Worldwide validation of 8 satellite-derived and reanalysis solar radiation products: A preliminary evaluation and overall metrics for hourly data over 27 years. **Solar Energy**, Elsevier, v. 210, p. 3–19, 2020.
- YANG, D.; KLEISSL, J.; GUEYMARD, C. A.; PEDRO, H. T.; COIMBRA, C. F. History and trends in solar irradiance and pv power forecasting: A preliminary assessment and review using text mining. **Solar Energy**, Elsevier, v. 168, p. 60–101, 2018.
- ZANG, H.; LIU, L.; SUN, L.; CHENG, L.; WEI, Z.; SUN, G. Short-term global horizontal irradiance forecasting based on a hybrid cnn-lstm model with spatiotemporal correlations. **Renewable Energy**, Elsevier, v. 160, p. 26–41, 2020.
- ZHANG, W.; ZHOU, T. Significant increases in extreme precipitation and the associations with global warming over the global land monsoon regions. **Journal of Climate**, American Meteorological Society, v. 32, n. 24, p. 8465–8488, 2019.
- ZHU, T.; XIE, L.; WEI, H.; WANG, H.; ZHAO, X.; ZHANG, K. Clear-sky direct normal irradiance estimation based on adjustable inputs and error correction. **Journal of Renewable and Sustainable Energy**, AIP Publishing LLC, v. 11, n. 5, p. 056101, 2019.