Deep Learning Models to Estimate and Predict the Solar Irradiation in Brazil

Wesley A. Souza[1](✉), Augusto M. S. Alonso[2], Luiz G. R. Bernardino[3],
Marcelo F. Castoldi[1], Claudionor F. Nascimento[4],
and Fernando P. Marafão[3]

[1] DAELE, Federal University of Technology - Paraná (UTFPR), Cornélio Procópio,
PR, Brazil
wesleyangelino@utfpr.edu.bra
[2] EESC, University of São Paulo (USP), São Carlos, SP, Brazil
[3] ICTS, São Paulo State University (UNESP), Sorocaba, SP, Brazil
[4] DEE, Federal University of São Carlos (UFSCar), São Carlos, SP, Brazil

Abstract. Solar irradiation is the backbone of photovoltaic power technologies and its quantization allows to optimize energy generation. However, solar irradiation can be difficult to detect, mostly due to the design and disposition of sensors, as well as their high cost. To address this limitation, this paper proposes a deep neural network-based model to estimate global solar irradiation by only relying on weather data, focusing on applications targeting the Brazilian territory. The model uses a deep neural network trained with data from the Brazilian National Institute of Meteorology (INMET), which includes 606 nationwide weather stations and over 39 million hourly records of meteorological variables cataloged from years 2010 to 2022. Thus, in this paper i) a deep neural network is used to estimate irradiation, and ii) a long short-term memory is used to predict solar irradiation considering different time granularities: 5 min, 30 min, 6 h, and 1 day. The results show a small error between the measured irradiation data and the calculated results with regard to the following six meteorological variables: time, temperature, relative humidity, wind speed, precipitation, and atmospheric pressure. Moreover, experimental validations conducted using a weather station set up by the authors demonstrate that the proposed models can accurately predict solar irradiation. Thus, the developed model stands as a promising approach for applications within the Brazilian perspective, improving the efficiency and reliability of solar energy generation.

Keywords: Solar irradiation · Deep learning · Weather station · Weather quantities

List of Abbreviations

AI   Artificial Intelligence
ANFIS   Adaptive Neuro-Fuzzy Inference System

ANN	Artificial Neural Network
ARMA	Autoregressive Moving Average
A-P	Angstrom-Prescott
DL	Deep Learning
DNN	Deep Neural Network
GAN	Generative Adversarial Networks
GEP	Gene Expression Programming
GRNN	Generalized Regression Neural Network
GRU	Gated Recurrent Unit
INMET	Brazilian National Institute of Meteorology
LSTM	Long Short-term Memory
MAD	Median Absolute Deviation
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MLP	Multilayer Perceptron
MSE	Mean Squared Error
PV	Photovoltaic
RBF	Radial Basis Function
RBFNN	Radial Basis Neural Network
RF	Random Forest
RNN	Recurrent Neural Network
SVM	Support Vector Machine
WRNN	Wavelet Recurrent Neural Networks

1 Introduction

The availability of energy supply is crucial for the economic and social development of any country, and producing electricity by means of sustainable generation sources has been playing a pivotal role in meeting the expected world demand for energy. Photovoltaic- (PV) and wind-based energy sources are examples of alternative resources that have supported the reduction of fossil fuel usage and decreased the need for nuclear power installations. Consequently, due to the advancement of such technologies, the world is becoming less dependent on power generation means, which may harm the environment or human life while producing electricity [1]. Hence, such alternative sources are shedding light on the new energy paradigm [2].

Particularly in Latin America, PV power plants have presented an accentuated growth since the past decades; for instance, now accounting for 2.47% of Brazil's total power generation [3]. PV-based technologies rely on the physical nature of solar radiation, depending directly on the irradiation performance to attain efficient energy conversion and/or utilization. The more one knows about the irradiation patterns in a certain territory, the more adequate the PV-based energy can be processed. Thus, quantifying solar irradiation is of paramount importance in several scenarios, such as power generation and utility markets, heat load distribution in buildings [4], PV system analysis and installation [5], agricultural applications [6], as well as irrigation systems [7].

Solar irradiation, measured in W/m^2, varies throughout the day at any given geographic location, mainly due to the earth's movement and the chaotic effects of the atmosphere [8]. With regards to Brazil, it has the world's highest potential for solar energy generation, given that a large part of its territory is located in the equatorial and tropical zones [9].

Measuring such irradiation is challenging due to its cost, maintenance requirements, and technical demands for calibration of sensors [10,11]. Several approaches have been proposed in the literature to address this issue, considering empirical models, mathematical formulations, and satellite-based data. Recently, the development of artificial intelligence (AI) algorithms to predict and estimate solar irradiation profiles at specific geographical locations has also been commonly evidenced [4–7,12,13]. Gao, Miyata and Akashi highlighted in [14] that most of the solar-related research findings available worldwide applied long short-term memory (LSTM), autoregressive moving average (ARMA), and multilayer perceptron (MLP) as a basis for solar irradiation forecasting algorithms.

Many research works have focused on artificial neural networks (ANN) for either predicting or forecasting solar irradiation, such as the one conducted by Yadav and Chandel [15], Shaddel, Javan, and Baghernia [16], and Zhang et al. [8]. In particular, Zhang et al. [8] and Salazar et al. [9] reviewed and compared various models, such as MLP, radial basis function (RBF), and wavelet recurrent neural networks (WRNN), in terms of estimation type and time scale.

For what concerns the Brazilian perspective, ANN-based applications focused on solar irradiation quantization are still limited, being particularly targeted only for a small number of areas or regions. Some research efforts have investigated ANN techniques for solar irradiation prediction in local scenarios, such as in Fortaleza - Ceará [17], Seropédica - Rio de Janeiro [18], Petrolina - Pernambuco [9,19], and Botucatu - Sao Paulo [20]. However, Brazil is a geographically extensive country with a significant north-to-south extension, and it lacks research on a generalist model to predict solar irradiation throughout the entire region.

Motivated by such a scientific gap, this paper presents two main contributions to fulfill the need to estimate solar irradiation in any location within the Brazilian territory. First, a model based on a deep-learning neural network (DNN) is developed to estimate solar irradiation based on the following attributes: daytime, temperature, humidity, atmospheric pressure, wind speed, and hourly precipitation. As a second contribution, a dynamic model is proposed to forecast daily irradiation based on locations' latitude, longitude, and month of the year.

A DNN is used as a regressor for the former model, and its performance is compared with other estimation techniques. The latter (i.e., dynamic) model uses a Recurrent Neural Network (RNN) based on the LSTM principle, allowing solar irradiation prediction for different time scales ranging from 5 min to 24 h. It is worth highlighting that this study utilizes data from 606 meteorological stations managed by the Brazilian National Institute of Meteorology (INMET) to train and evaluate the proposed models, considering the datalogging period between 2010 and 2022.

This paper reads as follows. Section 2 presents the process of choosing and implementing DNN and LSTM for the estimation and forecasting process of solar irradiation. Section 3 presents information about the database provided by INMET, the data munging process, and the results attained from the ANNs' model. In addition, it is presented performance analyses for short, medium, and long-term estimation and forecasting of irradiation, also considering comparisons with previous works from the literature. Section 4 presents the final considerations about the proposed models, their limitations, and future work proposals.

2   Artificial Neural Network Techniques for Solar Irradiation

ANNs are bio-inspired computational models capable of representing complex knowledge, maintenance, and generalization processes using the relationship between input and output data [6,12]. The basic unit of an ANN is the neuron, and models known as synapses interconnect the multiple units of neurons. A tuning value is associated with such synapses comprising the ANN, known as the weight factor.

The first ANNs were idealized in 1943 [21]; however, practical models were implemented in applications only after 1986, with the construction of an MLP with backpropagation [22]. An MLP is an ANN architecture comprising an input layer, one or more hidden layers, and an output layer, as shown in Fig. 1a. The MLP uses the supervised learning concept called backpropagation in the training process and can solve the nonlinearity of the input data to perform pattern recognition or estimation [6].

Since the MLP milestone, deep learning (DL) has progressed because of the computational evolution in the last decades and the possibility of increasing the number of hidden layers and neurons [23], as shown in Fig. 1b. Therefore, new architectures of ANNs were developed for several purposes, such as regression, supervised classification, computer vision, speech recognition, natural language processing, and audio detection [24].

Many research efforts present techniques for solar irradiation forecast. For instance, Wang et al. [25] conducted a study on daily solar radiation prediction comparing three ANN architectures: the MLP, generalized regression neural network (GRNN), and radial basis function neural network (RBFNN). The models were developed using as attribute input the air temperature, relative humidity, air pressure, water vapor pressure, and sunlight duration measured from 12 weather stations in different climate zones. Based on the results, they found that the MLP and RBFNN models provide better accuracy than GRNN.
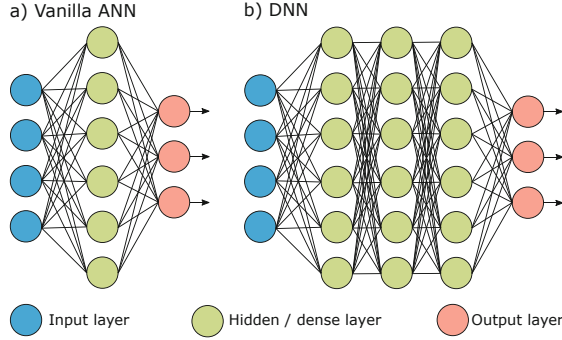
Fig. 1. ANN and DNN comparison.

Belaid and Mellit [26] developed a method that uses a support vector machine (SVM) to predict daily and monthly global average solar irradiation in an arid climate (Ghardaia, Algeria), taking as input the temperature, maximum sunshine duration, and the extraterrestrial solar radiation. For these quantities, the correlation coefficient ranged from 0.894 to 0.896, and the prediction error of approximately 7.5%.

Mehdizadeh et al. [27] conducted a study comparing gene expression programming (GEP), ANN, adaptive neuro-fuzzy inference system (ANFIS), and 48 empirical equations to estimate daily solar radiation in Kerman, Iran. The authors reported that the scenarios based on meteorological parameters and sunlight in ANFIS and ANN showed better accuracy than empirical models.

For the Brazilian scenario, ANN-based applications relating to the solar irradiation context are still limited. However, some studies have presented ANN approaches for solar irradiation prediction focusing on the regions of Fortaleza - Ceará [17] and Seropédica - Rio de Janeiro [18], achieving an accuracy of 89.7%. Salazar et al. [9] have developed a time series-based method to identify the solar irradiation in the equatorial near-zone and obtained a median absolute deviation (MAD) equivalent to 1.4% in the validation at a weather station installed in Petrolina - Pernambuco - Brazil. Carneiro et al. [19] used an ensemble learning method based on crest regression achieving mean absolute percentage error (MAPE) values of 14.191% also in Petrolina - Pernambuco - Brazil. Silva et al. [20] applied SVM, Angstrom-Prescott (A-P), and ANNs to estimate solar irradiation: the first achieved the best result while comparing to the A-P and ANN models, achieving a R^2 of 0.806.

Based on some studies found in the literature [17,25,27], ANNs provide significant capacity to predict solar irradiation. ANNs can estimate solar irradiation based on meteorological quantities and predict future irradiation based on historical events. Thus, this paper presents both models for obtaining solar irradiation, with the steps depicted in Fig. 2 and detailed in the Subsects. 2.1, 2.2 and discussed in Sect. 3.
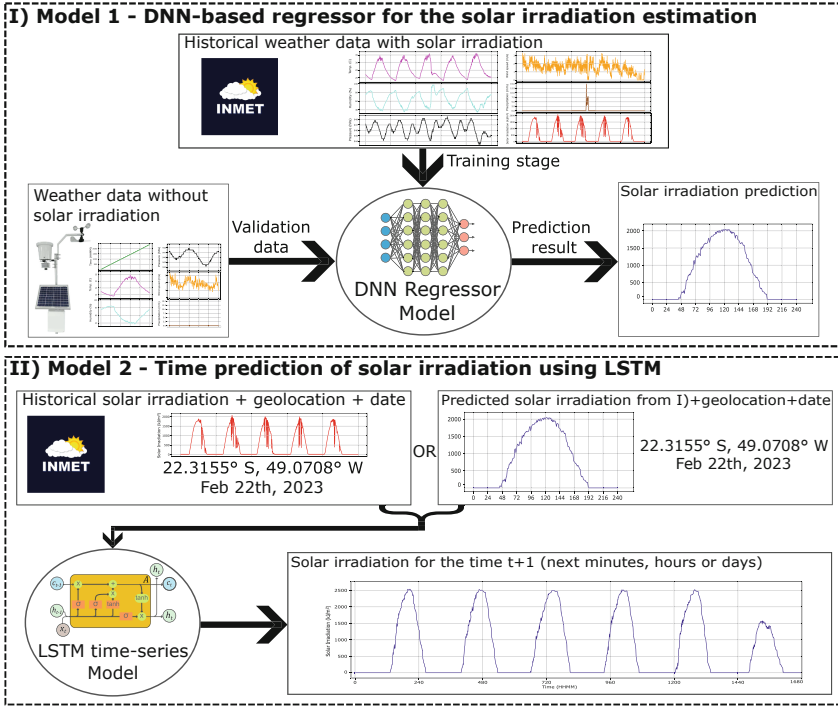
Fig. 2. The proposed ANN models.

2.1 Model 1 - DNN-Based Regressor for the Solar Irradiation Estimation

We assume this model as linear regression, in which a numerical value (target value or dependent variable) is obtained as a function of input values (attributes or independent variables), as presented in Eq. (1). Target values are continuous, meaning they can take any numerical value within the real number domain. In the literature, linear regression is used in various applications, such as stock market price forecasting, house price forecasting, sales forecasting, and others [28]. With regards to linear regression applications, using DNNs as regressors is helpful since they can learn the complex relationship between attributes and the target, mainly due to the presence of the activation function in each layer [6].

$$Y = \beta_0 + \sum_{n=1}^{N} \beta_n X_n + \epsilon \tag{1}$$

where:
Y is the numerical value of the dependent variable. It is this value that is wanted to be predicted;
β_0 is the intercept on the Y-axis when all input attributes are zero;

β_n are the fitness coefficients for the attribute n. In ANN's case, these values are calculated to indicate the effects that each attribute causes for the most accurate prediction of Y;

X_n is the n-th independent variable;

N is the number of independent variables in the regression model;

ϵ is the model error, which shows the difference between the real and the predicted value;

One must take into account the following considerations to build the linear regression model using DNN:

– Build a sequential ANN architecture;
– Define the quantity and neurons of the dense layers;
– Assign a performance metric (loss function) based on numerical error calculation, such as the mean absolute error (MAE), which is calculated according to Eq. (2);
– Defining the output layer with a single neuron, having as activation function the linear function [$f(x) = x$];

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i| \tag{2}$$

in which:

i is the sample;

n is the total number of samples;

y_i is the true or real value of the dependent variable;

$\hat{y}_i$ is the value of the dependent variable predicted by the regression model.

The second model is a forecasting process for future events, and it is based on another ANN architecture, as presented in the following subsection.

2.2   Model 2 - Time Prediction of Solar Irradiation Using LSTM

Vanilla ANN cannot perform time series prediction, depending on a previous data history to predict the next instant [29]. On the other hand, RNNs are well-known for achieving solid results in many applications with time series and sequential data [30]. The most well-known RNN structures, such as the LSTM and the gated recurrent unit (GRU), can capture the long-term temporal dependencies in variable-length samples [31]. Another distinguishing characteristic of RNNs is that they share parameters across each network layer. In addition, while feed-forward networks have different weights across each node, RNNs share the same weight parameter within each network layer. Such weights are still adjusted through the backpropagation and gradient descent approaches to facilitate reinforcement learning.

Since the previous outputs obtained during training leave an information base, the RNN model supports predicting future outputs as a function of the input attributes (X_t). Note that this occurs with the help of the previous outputs (h_t), as presented in Fig. 3.
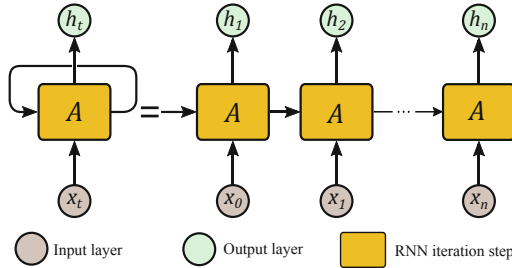
Fig. 3. Introducing the iterative process of an RNN.

LSTM is an RNN technique that can learn long-term dependencies, especially in sequential or seasonal prediction problems. LSTM has feedback connections that can process the entire data sequence and single data points. Each iteration of the LSTM network presents the data vector as input and two output data for each iteration:

– X_t is the input vector;
– C_t is the memory state cell, which maintains its state over time, considered as an output with memory;
– h_t is the time series output value.

Information can be added to or removed from the state C_t, regulated by input, forgetting, and output gates, presented after the layer applications shown in Fig. 4. These gates allow information to flow in and out of the cell, thus allowing memory propagation to the next iteration. The sigmoid layers (Fig. 4) present output numbers between zero and one, in which the former means that "nothing should be carried forward", and the latter means that "everything should be carried forward".

For constructing the solar irradiation prediction model, the number of steps represents the input layer (which corresponds to hourly data) and the attributes. We have the solar irradiation output in the output layer.
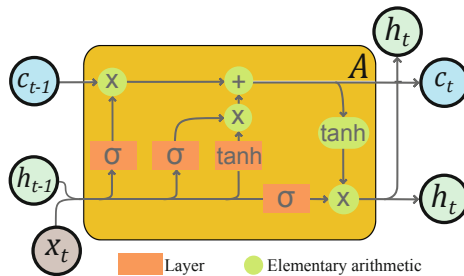
**Fig. 4.** Iterative process of an LSTM network.

Considering the structures of the two models, Sect. 3 presents the construction of the ANNs to estimate and predict solar irradiation and perform the models proposed in this work.

3 Methodology and Results

3.1 INMET Meteorological Station Data

The study comprised within this paper uses the meteorological database from the Brazilian National Institute of Meteorology (INMET), which is available at [32]. In this database, namely BDMEP, each data sample corresponds to the collection of meteorological variables collected at every hour or every six hours, being separated into individual files for each of the 606 meteorological stations distributed throughout Brazil, as shown in Fig. 5. Each file is composed of a header containing information about each meteorological station, as well as a structured set of samples of the collected meteorological variables. The dataset used in this study considers the interval between 01/01/2010 and 31/12/2022, as presented in Table 1.
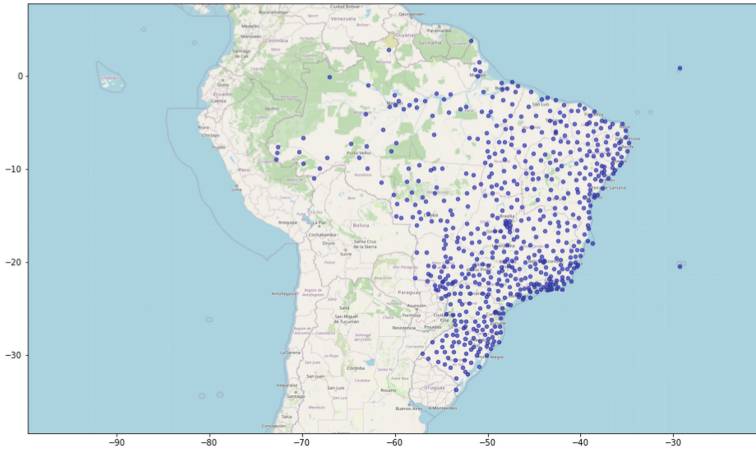
Fig. 5. Localization of the meteorological stations used in this work.

Table 1. Composition of each file in the BDMEP dataset.

Name	Unit/Format
Header	
Region of the country	
State	
Code (WMO)	
Latitude	°
Longitude	°
Altitude	m
Foundation date	dd/mm/yy
Data	
Date	YYYY-MM-DD
Time	HHMM UTC
Hourly precipitation	mm
Atmospheric pressure at station level	mB
Maximum atmospheric pressure in the previous hour	mB
Minimum atmospheric pressure in the previous hour	mB
Global radiation	kJ/m^2
Air temperature	°C
Dew point temperature	°C
Maximum temperature in the previous hour	°C
Minimum temperature in the previous hour	°C
Maximum dew point temperature in the previous hour	°C
Minimum dew point temperature in the previous hour	°C
Maximum relative humidity in the previous hour	%
Minimum relative humidity in the previous hour	%
Relative humidity	%
Wind direction	°
Maximum wind gust	m/s
Wind speed	m/s

Data preprocessing and cleaning are considered relevant step, as it enhances the quality of the information, assists in decision-making, and improves the machine learning model [33]. As a first step in data preprocessing, only momentary quantities that do not depend on the previous time were considered, resulting in 11 quantities. Furthermore, the data from all 606 meteorological stations were merged, resulting in 39,656,352 samples.

Subsequently, data cleaning was performed [34], with the removal of samples with reading errors, missing data, duplicate data, and removal of outliers, considering the empirical rule of 3σ [35], resulting in 36,433,601 samples, which represents approximately 91.87% of the initial dataset.

After the data preprocessing and cleaning step, the data is used in the training stage for modeling and constructing the solar irradiation estimation tool, as presented in Subsect. 3.3.

Considering the data presented in Sect. 3.1, the method and the results of the studies are presented in the following subsections. We consider two main scenarios for the results: *i)* the estimation of solar irradiation based on indirect meteorological quantities; and *ii)* the prediction of solar irradiation based on geolocation and date.

3.2 Estimation of Solar Radiation Based on Other Meteorological Variables Applying Model 1

In this first scenario, the focus is given to the solar radiation estimation applying Model 1, presented in Subsect. 2.1, and the target data was initially normalized using the *Z-score* technique [36]. After that, we conducted a *k*-fold cross-validation ($k = 5$) analysis to evaluate the efficacy of deep learning-based models on the BDMEP dataset. Cross-validation is a widely recognized technique for assessing machine learning model performance [37]. In 5-fold cross-validation, the dataset is divided into five equal subsets, where four subsets are employed for model training, and the remaining subset is employed for model validation. This procedure is repeated five times, using a different subset for validation. The model's generalization performance is accurately assessed by averaging the performance metrics over the five folds. Cross-validation aids in ensuring that the models do not overfit and can effectively generalize to new data.

The DNN architecture is presented in Fig. 6. As input data for the DNN, six variables were considered: hour, precipitation, atmospheric pressure, temperature, humidity, and wind speed. The output layer corresponds to the value of solar radiation. In all intermediate layers, the ReLu activation function was used [38], and the linear function $f(x) = x$ was used in the output layer. The optimizer of the model is the "Adamax".
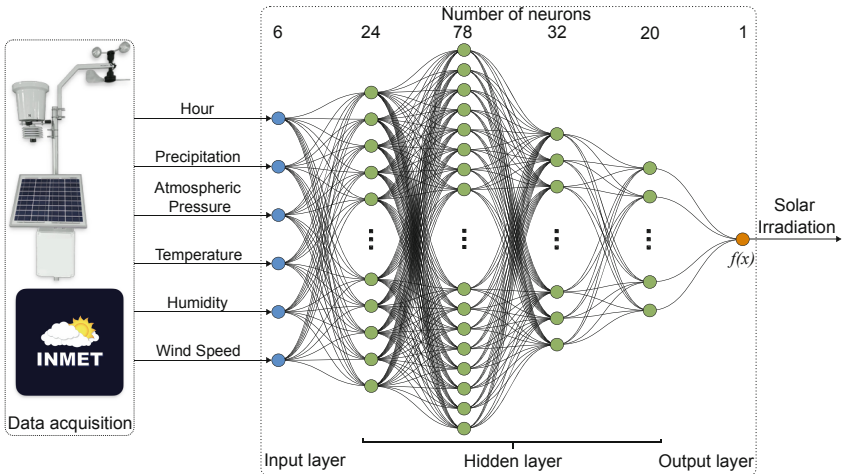

Fig. 6. DNN architecture with the best parameters.

In order to present the best results, a grid search was performed, which is a technique that searches for the best parameters for the machine learning model. For the DNN, a grid search was performed for the number of neurons (n) in the i-th hidden layer and the number of layers (i), where $n = 20, 24, 28, ..., 66, 72, 78$ and $i = 2, 3, 4, 5$. At this stage, k-fold cross-validation was also considered, with $k = 10$ [37].

With the DNN configuration presented in Fig. 6, the training process was performed with a limit of 100 epochs. Additionally, the stabilization of MAE was considered as the stopping criterion. Figure 7 presents the learning curve of the DNN implemented in this study.
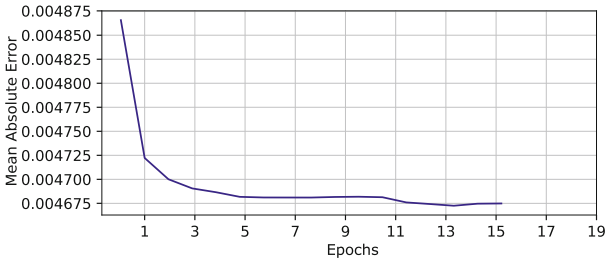

Fig. 7. DNN learning curve.

For the test data, the data for the year 2022 (i.e., until 10/30/2022) were considered, which can be accessed at [32]. Therefore, through such data, the mean absolute error equivalent to 9.34 kJ/m^2 was obtained in the study. To verify the system's dynamics in estimating solar irradiation, the results are presented in Fig. 8 for two real scenarios: 1) Xanxerê-SC-Brazil station, which is a member of BDMEP; and 2) IBAURU9 station, located in Bauru-SP-Brazil, which was developed by the authors [39] and used as a scenario of data not seen previously by the DNN model.

The model could predict irradiation behavior, as demonstrated by comparing the actual and estimated values presented in Fig. 8. On days with maximum solar radiation, the model could follow the approximate trend (i.e., note the estimated and actual curves in the lower graphs of Fig. 8). Furthermore, the model could still follow the irradiation reduction in the location on cloudy or rainy days, even though it presented a more significant error in the estimation.

3.3   Solar Irradiation Forecasting Based on Geolocation and Date Applying the Model 2

At this stage, the data presented in Sect. 3.1 was considered for Model 2, presented in Subsect. 2.2, being then normalized using the Z-score technique, which has the advantage of using a common normalization for variables with different standard deviations [40]. The training data corresponded to data between 2010
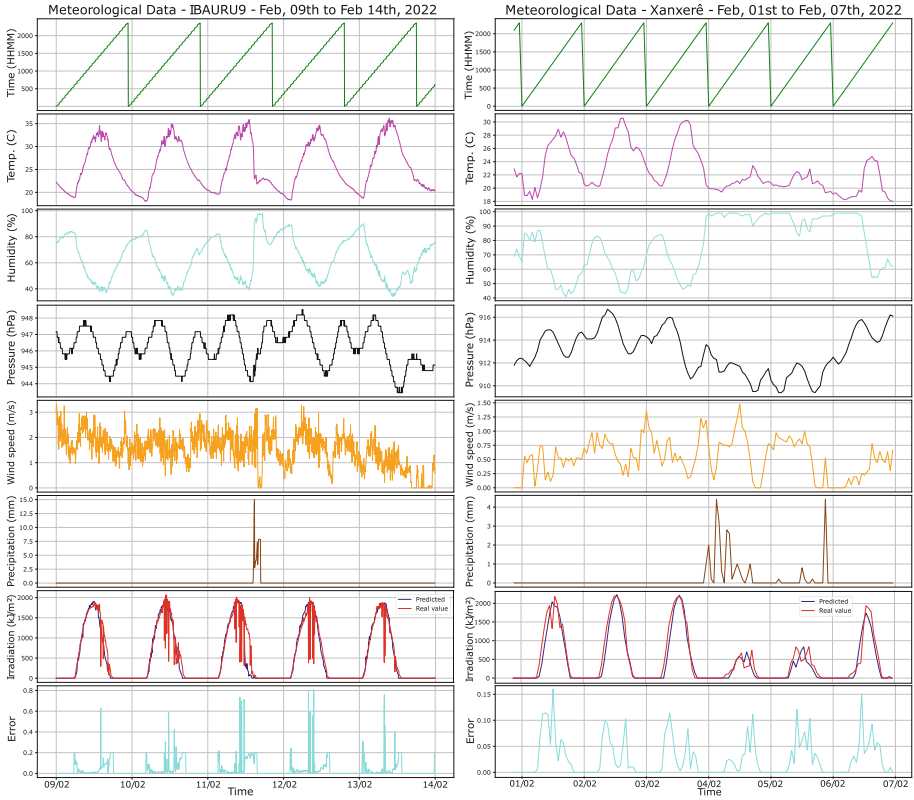
Fig. 8. Result of solar irradiation estimation for Bauru-SP-Brazil and Xanxerê-SC-Brazil stations.
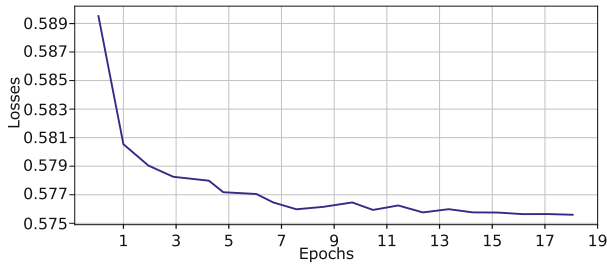
**Fig. 9.** LSTM learning curve.

and 2021, while the data from 2022 was used for testing. The target inputs are the solar irradiation information, the geolocation, and the date. The output of the model is the predicted irradiation at the next time step t. Moreover, the learning curve is shown in Fig. 9, presenting a final loss of 0.576.

Different time granularity was considered in this result. For the prediction for the next 5 min, the model presents an MAE equivalent to 18.91 kJ/m^2; for the next 30 min, 33.15 kJ/m^2; for the next hour 43.64 kJ/m^2; and for the prediction for 24 h, the MAE corresponds to 91.52 kJ/m^2. Figure 10 shows the response of the model for the next hour of irradiation prediction, and the system can track the real solar irradiation. A valley is visible around 17:00 due to the appearance of a cloud, which caused a significant decrease in solar irradiation. The model did not follow the real value, but it did follow the trend when the sun returned.
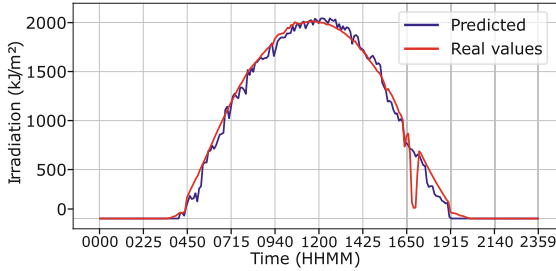
Fig. 10. One-day prediction results from the LSTM.

Figure 11 shows the 24-hour forecast made by the model for a sequence of seven days. There is also a slight delay between the predicted and the actual signal, but the forecast had adequate generation tracking for days of full solar irradiation. On the sixth day, there was a cloudy and rainy day, and the model had a higher error but still captured the decrease in irradiation for the day. On cloudy days, solar irradiation is diffuse, elusive and typically between 10 and 25 percent of its normal value on sunny days [41]. The proposed model tries to adjust values according to historical data without high accuracy, but presents a decrease in the solar irradiation estimation.
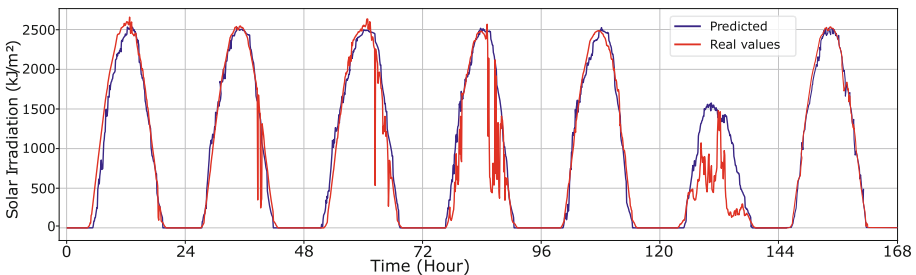
Fig. 11. 7-day prediction results from the LSTM.

3.4 Method Comparison and Discussion

This section considers techniques commonly found in the literature aiming at comparing the two proposed scenarios. The data used for testing and validating the methodologies was obtained from INMET, being identical to that presented in Sect. 3.1. A Macbook Pro, model A1990, with an Intel i9 9980H processor, 16GB of RAM and a Radeon Pro 560×4GB video card, was used to compute the comparison analyses of the results.

Model 1 - Comparison. According to Gao, Miyata, and Akashi [14], deep learning-based models have demonstrated outstanding ability for predicting solar irradiation, with LSTM demonstrating superior assertiveness performance compared to other techniques. However, it is necessary to verify the performance of the two scenarios proposed in this paper using methodologies from the literature. In addition to the DNN proposed in this project, traditional machine-learning regression techniques used in the literature for estimating solar irradiation from meteorological quantities were contrasted. Support Vector Machine (SVM) [42,43], Random Forest (RF) [42,44], and MLP [6,42,45,46] are the techniques used in the comparison with the results displayed in Table 2. The results indicate that the DNN approach had a lower MAE for the prediction scenarios. The MLP, which is another neural network architecture, presented the second-best performance. Moreover, the RF approach had the worst results, confirming the findings of Gao, Miyata, and Akashi [14]. On the other hand, while SVM has the best efficacy in training and testing regarding computational time, its error rate is nearly twice that from the DNN.

Table 2. Model 1 - Regression model comparison to estimate the solar irradiation.

Method	MAE	train(s)	test(s)
DNN - Model 1	**14.86**	74,736	102.3
MLP	18.42	20,375	75.4
RF	104.89	17,759	**21.2**
SVM	28.25	**13,302**	245.7

Note in Table 2 that Model 1 provided results comparable to the current state of the art, presenting the lowest MAE of all evaluated methods. Thus, based on such results, the DNN is recommended. One of the DNN's disadvantages, however, is that it has one of the highest computational costs. As demonstrated in Sect. 3.3, another disadvantage is the behavior on cloudy and rainy days. A potential solution to such an issue is to either use generative adversarial networks (GAN) or consider balanced data for what concerns sunny, cloudy, and rainy days, aiming at providing the model with generalized adjustments.

Model 2-Comparison. The efficacy of four statistical and machine learning tools for global solar irradiation forecasting was analyzed and compared during this study. Gao, Miyata, and Akashi [14] demonstrate that ARMA [47,48], MLP [49], and LSTM [14,50,51] are the most popular and appropriate models for the prediction of future solar irradiation. The Kalman filter [52] was also used to perform such a comparison. Table 3 displays the results taking into account various future time granularities, such as 5 min, 30 min, 6 h, and 1 day. Note that the Kalman filter produced the best results for the 5 min and 1 day baselines, while the LSTM produced the best results for 30 min and 6 h instances. With regards to neural network-based architectures, the LSTM and MLP demanded the most time to train, while the Kalman filter presented the faster training process. During the test, ARMA had the fastest time. Hence, the Kalman filter is recommended for the smallest and largest granularities. On the other hand, based on the data analyzed in this paper, the LSTM is recommended for the intermediate granularities of 30 min and 6 h. The presented comparative analyses corroborate the findings of Yu, Cao, and Zhu [50], who concluded that the LSTM is not recommended for a 24-hour granularity model.

Table 3. Model 2 - Time-series prediction comparison of solar irradiation.

	MSE					
Method	5 min	30 min	6 h	1 day	train(s)	test(s)
LSTM - Model 2	18.91	**33.15**	**43.64**	191.52	6,971.3	23.55
ARMA	22.37	36.54	61.11	102.54	533.5	**12.50**
Kalman's filter	**18.11**	35.04	48.31	**122.50**	**101.4**	88.52
MLP	58.44	98.104	114.88	157.30	7,392.8	45.77

At last, note that Model 2 produced results that are similar or better than achieved by the state-of-the-art, with regards to the mean squared error (MSE): a ranking near the best one was obtained for 5 min; it was the lowest one for the 30-minutes and 6-hour predictions; although it was the worst performance for the 1-day forecast. As an inherent disadvantage, Model 2 presents one of the highest computational costs for training. Nonetheless, Model 2's performance could be improved by taking into account additional input attributes, such as the weather data displayed in Model 1. Moreover, minimizing the number of input samples may enhance training performance.

4   Conclusion

Quantifying solar radiation is essential for a range of applications, from solar energy generation to building thermal management, agriculture, and irrigation. However, measuring this magnitude poses technical and operational challenges, making its real-time use difficult.

This work presented the use of a database of 606 Brazilian weather stations collected by INMET between 2010 and 2021 to develop two deep learning-based models. The first model is a DNN-based method that estimates solar radiation across Brazil based on hourly temperature, humidity, atmospheric pressure, wind speed, and precipitation data. The second model is an LSTM-based method that predicts future solar radiation for intervals of 5 min, 30 min, 1 h, and the entire day.

Our results demonstrate that both models accurately estimate and predict solar radiation. The first model has an MAE of 9.34 kJ/m^2, while the second model has an MAE of 1.89 kJ/m^2, 3.31 kJ/m^2, 4.36 kJ/m^2, and 31.52 kJ/m^2 for predicting the next 5 min, 30 min, 1 h, and 24 h, respectively.

Our findings demonstrate that DNN modeling can adequately identify solar radiation using indirect meteorological variables, while the LSTM model can adapt well to a prediction system, producing close-to-real results with geographic coordinates, the previous radiation level, and the month of the year as inputs. These results indicate the potential of deep learning-based methods for estimating and predicting solar radiation in Brazil, considering the successful performance of the model with over 39 million hourly data points from 606 weather stations nationwide.

Future work involves utilizing adversarial generative networks to improve the prediction performance on rainy or cloudy days and exploring RNA applicability in intelligent meter immersion to aid solar generation management and prediction.

References

1. Adedoyin, F. F., Bekun, F. V. & Alola, A. A. Growth impact of transition from non-renewable to renewable energy in the EU: The role of research and development expenditure. Renewable Energy 159, 1139–1145. ISSN: 0960–1481 (2020)
2. Brodny, J., Tutak, M.: Analyzing similarities between the european union countries in terms of the structure and volume of energy production from renewable energy sources. Energies **13** (2020) ISSN: 1996– 1073
3. EPE (Brasil). Balanço Energético Nacional 2022: Ano base 2021 In Portuguese. Tech. rep. (2022)
4. Wang, Z., Hong, T., Piette, M.A.: Building thermal load prediction through shallow machine learning and deep learning. Appli. Energy **263**, 11468 (2020). ISSN: 0306–2619
5. Notton, G., Voyant, C., Fouilloy, A., Duchaud, J.L., Nivet, M.L.: Some applications of ANN to solar radiation estimation and forecasting for energy applications. Appli. Sci. **9**, 2076–3417 (2019)
6. Antonopoulos, V.Z., Papamichail, D.M., Aschonitis, V.G., Antonopoulos, A.V. Solar radiation estimation methods using ANN and empirical models. Comput. Electr. Agricult. **160**, 160–167 (2019). ISSN: 0168–1699

7. Ali, B.: Comparative assessment of the feasibility for solar irrigation pumps in Sudan. Renew. Sustainable Energy Rev. **81**, 413–420 (2018). ISSN: 1364–0321

8. Zhang, J., Zhao, L., Deng, S., Xu, W., Zhang, Y.: A critical review of the models used to estimate solar radiation. Renew. Sustainable Energy Rev. **70**, 314–329 (2017). ISSN: 1364–0321

9. Salazar, G., Gueymard, C., Galdino, J.B., de Castro Vilela, O., Fraidenraich, N.: Solar irradiance time series derived from high-quality measurements, satellite-based models, and reanalyses at a near-equatorial site in Brazil. Renew. Sustainable Energy Rev. **117**, 109478 (2020). ISSN: 1364–0321

10. Rocha, A.B.d., Fernandes, E.d., Santos, C.A.C.d., Diniz, J.M.T., Junior, W.F.A.: Development of a real-time surface solar radiation measurement system based on the internet of things (IoT). Sensors **21**, 3836 (2021). ISSN: 1424–8220

11. Bayray, M., et al.: Measured solar irradiance data for resource assessment at four sites in Geba catchment, Tigray, North Ethiopia. Data Brief **40**, 107836 (2022). ISSN: 2352–3409

12. Khosravi, A., Nunes, R., Assad, M., Machado, L.: Comparison of artificial intelligence methods in estimation of daily global solar radiation. J. Cleaner Product. **194**, 342–358 (2018). ISSN: 0959–6526

13. Zang, H. et al. Estimation and validation of daily global solar radiation by day of the year-based models for different climates in China. Renewable Energy **135**, 984–1003 (2019). ISSN: 0960–1481

14. Gao, Y., Miyata, S., Akashi, Y.: Multi-step solar irradiation prediction based on weather forecast and generative deep learning model. Renewable Energy **188**, 637–650 (2022). ISSN: 0960–1481

15. Yadav, A.K., Chandel, S.: Solar radiation prediction using Artificial Neural Network techniques: a review. Renew. Sustain. Energy Rev. **33**, 772–781 (2014). ISSN: 1364–0321

16. Shaddel, M., Javan, D.S., Baghernia, P.: Estimation of hourly global solar irradiation on tilted absorbers from horizontal one using Artificial Neural Network for case study of Mashhad. Renew. Sustain. Energy Rev. **53**, 59–67 (2016). ISSN: 1364–0321

17. Barros Silva, A.W. et al. Methodology based on artificial neural networks for hourly forecasting of PV plants generation. IEEE Latin America Transactions **20**, 659–668 (2022)

18. Oliveira, J.M., Santos, D.D.S., Da Silva, R.M.: Predição da radiação solar global usando modelos de redes neurais artificiais. Revista Mundi Engenharia. Tecnologia e Gestão **4** (2019). ISSN: 2525–4782

19. Carneiro, T.C., Rocha, P.A., Carvalho, P.C., Fernández-Ramírez, L.M.: Ridge regression ensemble of machine learning models applied to solar and wind forecasting in Brazil and Spain. Appli. Energy **314**, 118936 (2022)

20. da Silva, M.B.P., Francisco Escobedo, J., Juliana Rossi, T., dos Santos, C.M., da Silva, S.H.M.G.: Performance of the Angstrom-Prescott Model (A-P) and SVM and ANN techniques to estimate daily global solar irradiation in Botucatu/SP/Brazil. J. Atmospheric Solar-Terrestrial Phys. **160**, 11–23. (2017). ISSN: 1364–6826

21. McCulloch, W.S., Pitts, W.: A logical calculus of the ideas immanent in nervous activity. Bull. Math. Biophys. **5**, 115–133 (1943). ISSN: 1522–9602

22. Rumelhart, D.E., Hinton, G.E., Williams, R.J.: Learning representations by back-propagating errors. Nature **323**, 533–536 (1986). ISSN: 1476–4687

23. Ng, A., Ngiam, J., Foo, C.Y., Mai, Y.: Deep learning. In: CS229 Lecture Notes, pp. 1–30 (2014)

24. Kim, K.G.: Book review: deep learning. hir **22**, 351–354 (2016)
25. Wang, L., et al.: Solar radiation prediction using different techniques: model evaluation and comparison. Renew. Sustain. Energy Rev. **61**, 384–397 (2016). ISSN: 1364–0321
26. Belaid, S., Mellit, A.: Prediction of daily and mean monthly global solar radiation using support vector machine in an arid climate. Energy Conversion Manag. **118**, 105–118 (2016). ISSN: 0196–8904
27. Mehdizadeh, S., Behmanesh, J., Khalili, K.: Comparison of artificial intelligence methods and empirical equations to estimate daily solar radiation. J. Atmospheric Solar-Terrestrial Phys. **146**, 215–227 (2016). ISSN: 1364–6826
28. Dixon, M.F., Halperin, I., Bilokon, P.: Introduction. In: Machine Learning in Finance, pp. 3–46. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-41068-1_1
29. Joshi, V., et al.: Accurate deep neural network inference using computational phase-change memory. Nat. Commun. **11**, 2473 (2020). ISSN: 2041–1723
30. Ismail Fawaz, H., Forestier, G., Weber, J., Idoumghar, L., Muller, P.-A.: Deep learning for time series classification: a review. Data Mining Knowl. Dis. **33**, 917–963 (2019) ISSN: 1573–756X
31. Murad, A., Pyun, J.-Y.: Deep recurrent neural networks for human activity recognition. Sensors **17**, 2556 (2017). ISSN: 1424–8220
32. INMET. Banco de Dados Meteorológicos do INMET In Portuguese (2022). https://bdmep.inmet.gov.br/
33. Ilyas, I.F., Chu, X.: Data Cleaning. Association for Computing Machinery, New York (2019). ISBN: 9781450371520
34. Van den Broeck, J., Argeseanu Cunningham, S., Eeckels, R., Herbst, K.: Data cleaning: detecting, diagnosing, and editing data abnormalities. PLOS Med. **2** (2005)
35. Zhao, Y., Lehman, B., Ball, R., Mosesian, J., de Palma, J.-F.: Outlier detection rules for fault detection in solar photovoltaic arrays. In: Proceedings of the IEEE Applied Power Electronics Conference and Exposition, pp. 2913–2920 (2013)
36. Patro, S., Sahu, K.K.: Normalization: A preprocessing stage. arXiv preprint arXiv:1503.06462 (2015)
37. Souza, W.A., et al.: Selection of features from power theories to compose NILM datasets. Adv. Eng. Inform. **52**, 101556 (2022). ISSN: 1474–0346
38. Eckle, K., Schmidt-Hieber, J.: A comparison of deep networks with ReLU activation function and linear spline-type methods. Neural Netw. **110**, 232–242 (2019). ISSN: 0893–6080
39. Tavares, K., Alonso, A.M.S., Souza, W.A.: Estimating solar irradiance system using meteorological quantities and deep neural networks (In Portuguese). Peer Rev. **5**, 225–238 (2023)
40. Zhang, Z., Cheng, Y., Liu, N.C.: Comparison of the effect of mean-based method and z-score for field normalization of citations at the level ofWeb of Science subject categories. Scientometrics **101**, 1679–1693 (2014). ISSN: 1588–2861
41. Armstrong, S., Hurley, W.: A new methodology to optimise solar energy extraction under cloudy conditions. Renew. Energy **35**, 780–787 (2010). ISSN: 0960–1481
42. Voyant, C., et al.: Machine learning methods for solar radiation forecasting: a review. Renew. Energy **105**, 569–582 (2017). ISSN: 0960–1481
43. Zendehboudi, A., Baseer, M., Saidur, R.: Application of support vector machine models for forecasting solar and wind energy resources: a review. J. Cleaner Prod. **199**, 272–285 (2018). ISSN: 0959–6526

44. Fouilloy, A., et al.: Solar irradiation prediction with machine learning: forecasting models selection method depending on weather variability. Energy **165**, 620–629 (2018). ISSN: 0360–5442
45. Ahmad, A., Anderson, T., Lie, T.: Hourly global solar irradiation forecasting for New Zealand. Solar Energy **122**, 1398–1408 (2015). ISSN: 0038–092X
46. Yousif, J.H., Kazem, H.A., Boland, J.: Predictive models for photovoltaic electricity production in hotweather conditions. Energies **10** (2017). ISSN: 1996–1073
47. Reikard, G.: Predicting solar radiation at high resolutions: a comparison of time series forecasts. Solar Energy **83**, 342–349 (2009). ISSN: 0038–092X
48. Yang, D., Jirutitijaroen, P., Walsh, W.M.: Hourly solar irradiance time series forecasting using cloud cover index. Solar Energy Solar Res. **86**, 3531–3543 (2012). ISSN: 0038–092X
49. Alfadda, A., Rahman, S., Pipattanasomporn, M.: Solar irradiance forecast using aerosols measurements: a data driven approach. Solar Energy **170**, 924–939 (2018). ISSN: 0038–092X
50. Yu, Y., Cao, J., Zhu, J.: An LSTM Short-term solar irradiance forecasting under complicated weather conditions. IEEE Access **7**, 145651–145666 (2019)
51. Srivastava, S., Lessmann, S.: A comparative study of LSTM neural networks in forecasting day-ahead global horizontal irradiance with satellite data. Solar Energy **162**, 232–247 (2018). ISSN: 0038–092X
52. Gupta, A., Gupta, K., Saroha, S.: A review and evaluation of solar forecasting technologies. In: Materials Today: Proceedings of International Conference on Materials and System Engineering, vol. 7, pp. 2420–2425. 2214–7853 (2021)