IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 26 November 2023, accepted 4 January 2024, date of publication 25 January 2024, date of current version 7 February 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3358339

RESEARCH ARTICLE

Neural Networks Forecast Models Comparison for the Solar Energy Generation in Amazon Basin

ANDRÉ LUÍS FERREIRA MARQUES[1], MÁRCIO JOSÉ TEIXEIRA[2], FELIPE VALENCIA DE ALMEIDA[1], AND PEDRO LUIZ PIZZIGATTI CORRÊA[1]

[1]Polytechnic School, University of São Paulo, São Paulo 05508-010, Brazil
[2]Institute of Physics, University of São Paulo, Sao Pāulo 05508-090, Brazil

Corresponding author: André Luís Ferreira Marques (4004andre3003@gmail.com)

The authors acknowledge the support of the Research Centre for Greenhouse Gas Innovation (RCGI), hosted by University of Sao Paulo (USP) and sponsored by FAPESP (grants #2014/50279-4 and #2020/15230-5, #2022/07974-0) and Shell Brasil, and the strategic importance of the support given by Brazil's National Oil, Natural Gas and Biofuels Agency (ANP) through the R&D levy regulation. The work of Felipe Valencia de Almeida is sponsored by the National Council for Scientific and Technological Development (CNPq), grant #140253/2021-1.

ABSTRACT Deep learning has grown among the prediction tools used within renewable energy options. Solar energy belongs to the options with the lowest atmosphere impact after considering their limitations. In the last five years, Brazil has seen the expansion of wind and solar options almost all over the country, and to preserve the Amazon rainforest, the use of solar energy has helped large and small cities towards a greener future. The novelty of this research covers the use of Deep Learning with data from twelve cities in the state of Amazonas to forecast solar irradiation (W.h/m^2) within 30 days. The data input came from ground stations, as much as possible, and NASA satellite models, with a daily time aggregation. The types of neural networks considered are Long Short-Term Memory (LSTM), a Multi-Layer Perceptron (MLP), and an LSTM Gated Recurrent Unit (GRU). Among the metrics used to check the algorithm's performance, the Mean Absolute Percentage Error (MAPE) indicates that the values of this research are coherent with other scenarios to forecast solar energy; the boundary conditions were not the same, however. The lowest MAPE was observed in the city of Labrea with the LSTM GRU.

INDEX TERMS Deep learning, long short-term memory, multi-layer perceptron, data science, Amazon basin, solar energy.

I. INTRODUCTION

Monitoring greenhouse gases (GHG) has proved its value for climate change evaluation, encouraging the development of various new technologies within this goal. Fossil fuels are the leading cause of environmental impacts, such as the overall temperature rise and more frequent droughts. Therefore, renewable energy options have been considered in the energy matrix upgrade because they have a smaller carbon impact or Carbon Dioxide (CO_2) emission [1]. Solar energy has been one of the main options due to its flexibility and easier build-up and maintenance.

Some scenarios have been investigated on the Amazon Basin Forest evolution, with some considering the current deforestation rate as progressive and with a non-recovery profile [2]. On the other hand, many other researchers point out that the rainforest can recover, as already observed in the region. No matter the scenario, the use of fossil fuels in the region shall be dimmed with solar energy, which may be tailor-made for each specific area, following a worldwide trend. This work focuses on solar power prediction in the Amazon Basin using Artificial Intelligence throughout neural networks.

From a technical perspective, the region represents a harsh environment with high temperatures and air humidity, dense vegetation, diverse fauna, and flora [3]. Equally important, the people density is minimal, and thus, the ground stations are not abundant. In addition, the existing ones have shown technical problems in measuring, storing, and transmitting data for remote processing. In most cases, these

The associate editor coordinating the review of this manuscript and approving it for publication was Bing Li.

activities have been done manually. Therefore, in our research, these boundary conditions have motivated us to use satellite sources, which employ math methods and analysis to present their results [4].

A. RELATED WORKS

Data Science (DS) has used math algorithms and tools for several applications and fields, such as Deep Learning (DL) and even not-so-complex methods. Technical studies have shown existing correlations between solar radiation and meteorological variables and their links to the digital control of smart solar panels and energy storage devices. For instance, the local mean temperature influences the solar panel's performance due to photovoltaic material reasons. DS has been employed to predict solar energy for different locations and times. The prediction time horizons comprise intra-hour, day-ahead (short-term), month (medium-term) and year (long-term) for different applications, such as energy load or energy decision-making.

References [5] and [6] provide a broad view of the neural networks used to predict solar and wind energy. Most of the neural networks are Long Short-Term Memory (LSTM) in different groups, such as bidirectional [7], with wavelet activation kernels [8], in autoencoder structure [9], with Deep Recurrent Neural Network (DRNN) [10], with Gated Recurrent Unit (GRU) [11], with Convolution Neural Network (CNN) [12] and with Multi-Layer Perceptron (MLP) [13]. There are worth noting groups are known as 'hybrid models,' such as Deep Belief Network + Feedforward neural network (DBN + FNN) [14], LSTM + Deep Generative Model (DGM) [15], CNN + MLP + LSTM [16], Discrete wavelet Transform (DWT) + LSTM [17], Stationary Wavelet Transform (SWT) +LSTM+DNN [18], Wavelet Package Decomposition (WPD) +LSTM [19], DBN+ Autoregressive Integrated Moving Average (ARIMA) [20], and Genetic Algorithm Network (GAN) + CNN [21]. In all these sets, optimization techniques are used, like Dropout [22], Adam, Rectified Linear Unit (ReLU), greedy algorithms, cross-validation, adaptive and colony algorithms, Grid Search, Maximal Overlap Discrete Wavelet Packet Transform (MODWPT) [23], Nadam, backward elimination algorithm, among others.

Regarding forecast time interval or 'horizon,' most references work up to one day, with some cases of 15, 30, 60, or 90 minutes. Few cases deal with a couple of weeks, months, or a year. The preprocessing tools considered in most cases were data normalization, removing outliers, changing data resolution, data augmentation, correlation analysis, WPD [24], Virtual Machine Systems (VMS) [25], Complete Ensemble Empirical Mode Decomposition Adaptive Noise (CEEMDAN) [26], Variational Mode Decomposition (VMD) [27], K-means clustering, Principal Component Analysis (PCA), Graph Construction and Empirical Mode Decomposition (EMD) [28]. There were a few cases with no preprocessing. In terms of input data, Table 1 shows the main parameters taken, associated with the statistics metrics used by the references.

TABLE 1. -Input parameters and statistics metrics used by references.

Reference	Input parameter	Statistic metrics
[29]	Diffuse solar radiation	RRMSE, MAE
[30]	Humidity, wind speed, temperature & pressure	RMSE, MAE, MSE
[31]	Surface meteorological measurements	MAbE, RMSE
[32]	Minimum and maximum values of temperature, humidity, wind speed and solar irradiation	MSE, MAE, RMSE and MAPE
[33]	Wind speed, sun height, ambient temperature	MAE, RMSE, NMBE
[34]	Wind speed, dew-point, relative humidity, wind direction, outdoor air-dry bulb temperature	RMSE, NMBE, R^2, CV(RMSE)
[35]	Hourly solar radiation	R^2 and RMSE
[36]	Clear-sky solar irradiation model, cloud index, reflectivity values	MBE, MAE, RMSE and R^2
[37]	Historical meteorological data	nRMSE

where:

- RRMSE: Relative Root Mean Square Error
- MAE: Mean Absolute Error
- RMSE: Root Mean Square Error
- MSE: Mean Square Error
- MABE: Mean Absolute Error
- MAPE: Mean Absolute Percentual Error
- NMBE: Normalized Mean Bias Error
- R^2: Correlation Coefficient
- CV(RMSE): Coefficient Variation RMSE
- MBE: Mean Bias Error
- nRMSE: Normalized Root Mean Square Error

The countries involved in the works above were the USA, Egypt, China, Algeria, India, France, Japan, Spain, Nigeria, Bangladesh, South Africa, Australia, Iran, Malaysia, and Singapore. As seen, a diverse set of boundary conditions in terms of latitude, longitude, year temperature variation, vegetation coverage surroundings, and human development indicators, among others.

B. RESEARCH CONTRIBUTION AND ORGANIZATION

The novelty of this research is the application of NN on renewable energy (solar) forecasting in the Brazilian part of the Amazon basin (12 cities) within a decade (2013-2023), aligned with the current trends and needs to change the energy matrix in the world and to preserve the environment. As far as the authors know, the current research is the sole done

within the context of solar energy, as a renewable option, in that region covering a broad area. Equally important, this study's contribution is to use the NN techniques, following a technical procedure to increase the prediction performance, by monitoring the math error, and keeping the attention to overfitting or the introduction of bias. Another contribution of the research is the data analysis regarding the correlation between meteorological variables and solar irradiance, within an extended time, related to the solar energy worldwide spread, which includes the occurrence of the 'El-Niño' phenomenon in that area (2015-2016), affecting the rain and temperature patterns. It is also worth noting that, during this covered period of the research, the occurrence of fires in that region has increased, which may impact the solar energy.

This work is organized as follows. Section II presents and comments on the materials and methods, focusing on the data gathering and its exploratory analysis, among others. Section III presents a brief description of the theory and calculations of the neural networks taken here. The results are presented in Section IV. Sections V and VI show the comments about these results and the overall conclusions of the research, respectively.

II. MATERIALS AND METHODS

Among the main targets of DS, identifying insights hits the first place by learning based on data and carrying out tasks of prediction, classification, or clustering. DS uses a broad set of algorithms, such as Decision Trees (DT) and Neural Networks (NN). The term 'supervised learning' deals with the ML category, where the algorithm develops its learning in two separate phases: training and testing and handling specific data groups. The 'unsupervised learning' means a self-learning approach, where the algorithm learns without the split mentioned between training and test. It is worth noting that DS also employs a third data group, or validation group, in many cases. In this research, we use the supervised data method.

A. DATA ANALYSIS AND PROCESSING

This research focused on data from the Amazonas state, the largest in Brazil, having the rainforest as its primary landscape vegetation, and the city of Manaus is the capital. In this work, 12 cities of that state were selected to have the solar incidence studied, with different stages of deforestation nearby, different human occupations, and economic levels. Table 2 presents their initial data, such as latitude, longitude, altitude (m), and population [39]. This research follows the standard ISO 6709 for the latitude and longitude representation because it allows better math data processing. In this research, most cities are classified as 'small,' except Manaus, which is 'large.' All the cities are located near a main river and have rainforests in their surroundings. This pattern looks different in the case of Manaus, which has industrial facilities around it. Figure 1 shows the map localization, respectively.

TABLE 2. Geographic coordinates of the 12 selected cities [39].

City	Latitude	Longitude	Altitude	Population
Barcelos	-0.97	-60.92	40.00	27,638
Benjamin Constant	-4.38	-70.03	65.00	44,873
Coari	-4.08	-63.13	46.00	86,713
Codajas	-3.84	-62.06	32.00	29,691
Eirunepe	-6.67	-69.87	104.00	36,121
Iaurete	0.61	-69.18	120.00	3,000
Labrea	-7.25	-64.83	61.00	47,685
Manaus	-3.10	-60.02	61.25	2,255,903
Manicore	-5.82	-61.30	50.00	57,405
Parintins	-2.63	-56.73	29.00	116,439
S.G.Cachoeira	-0.12	-67.06	90.00	47,031
Tefe	-3.83	-64.70	47.00	59,250

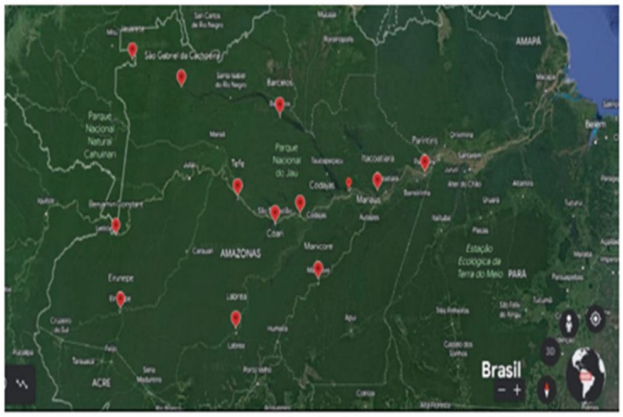

FIGURE 1. Localization of the twelve cities in the Amazonas state.

For solar incidence, data may come from ground stations or satellite sources, and the best ones come from the first, where the radiation measurements are taken with specific and fixed instruments. The local input variables are solar radiation incidence, relative humidity, wind speed and direction, mean temperature, and rain precipitation. Initially, we looked for data from the Brazilian National Institute of Meteorology (INMET) website [40]. After a detailed survey, the weather stations presented different data availability statuses, not allowing straight and continuous use by the DL tools. Therefore, we shifted to the NASA data products, using models CERES and MERRA2 [41], along with data from ground stations for side comparisons. The selected elapsed time covered the interval from January 2013 to March 2023, linked to another research in the region [42]. The dataset was obtained as comma-separated values (CSV), with 3741 observations in a daily time aggregation. For each city's description and units of the variables are shown in Table 3.

After an initial EDA, each city dataset was checked, treated, and made ready for the DL application. The use of boxplot graphs helped to understand the data distribution, showing the mean, the quartiles, and the outliers. The twelve cities' names are on the horizontal axis, and the vertical axis

TABLE 3. Technical variables for solar energy forecast.

Variable	Unit	Description
ALLSKY_SFC_SW_DWN	kW-hr/m2 (day)	The total solar irradiance incident (direct plus diffuse) on a horizontal plane at the surface of the earth under all sky conditions.
WS10M	m/s	The average wind speed 10 meters above the surface of the earth.
WD10M	degrees	The average wind direction 10 meters above the surface of the earth.
PS	kPa	The average surface pressure at the surface of the earth.
RH2M	dimensionless	The ratio of the actual partial pressure of water vapor to the partial pressure at saturation, expressed in percent (%), at 2 meters above the surface of the earth.
T2M	Celsius	The average air (dry bulb) temperature 2 meters above the surface of the earth.

shows the technical variable in the analysis. The following figures show the overall data distribution, starting with the target variable 'ALLSKY_SFC_SW_DWN' (Figure 2) in kW.h/m^2. The variable's mean value is close to 4.7 kW.h/m^2, and there are outliers only in the lower bound of the set. There is no significant clustering pattern identified specifically for this variable.

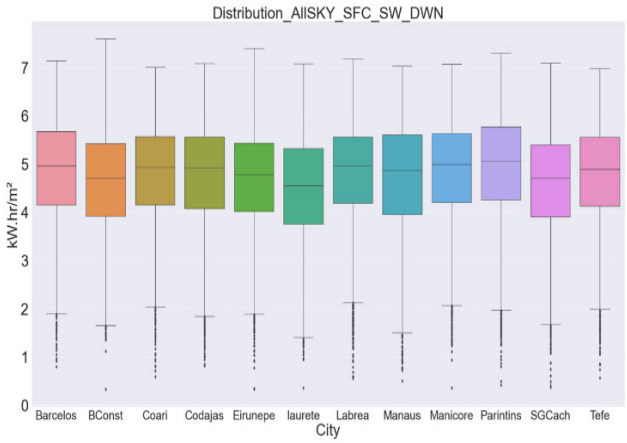

FIGURE 2. 'ALLSKY_SFC_SW_DWN' data distribution among the 12 cities.

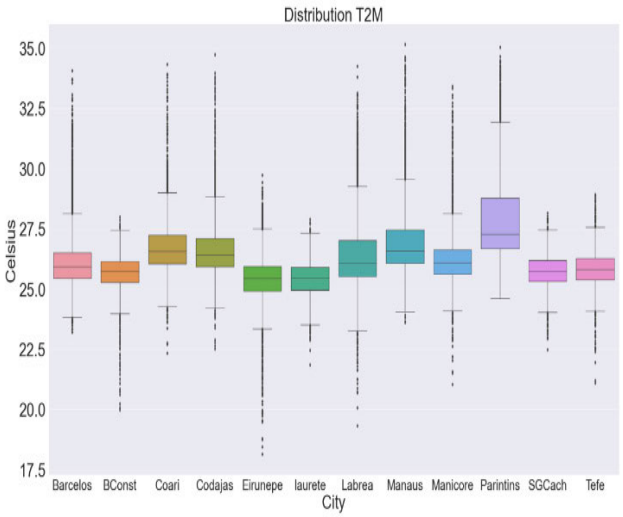

FIGURE 3. 'T2M' data distribution among the 12 cities.

The other variables shown here are the input variables for the NN, dealing with meteorological data. The mean temperature (°C) at 2m above the surface, or 'T2M', regarding the twelve cities is shown in Figure 3, where the mean temperature is around 26 °C, with outliers on both sides of the span. The city of Parintins presents the highest temperature, around 32 °C.

Figure 4 presents the data distribution of the wind speed (m/s), or 'WS10M', which is measured at 10 meters above the ground. The mean value among the cities is close to 0.8 m/s, which is considered small, except for the city of Parintins, which has a mean value close to 1.8 m/s. All the cases have outliers on both sides of the span.

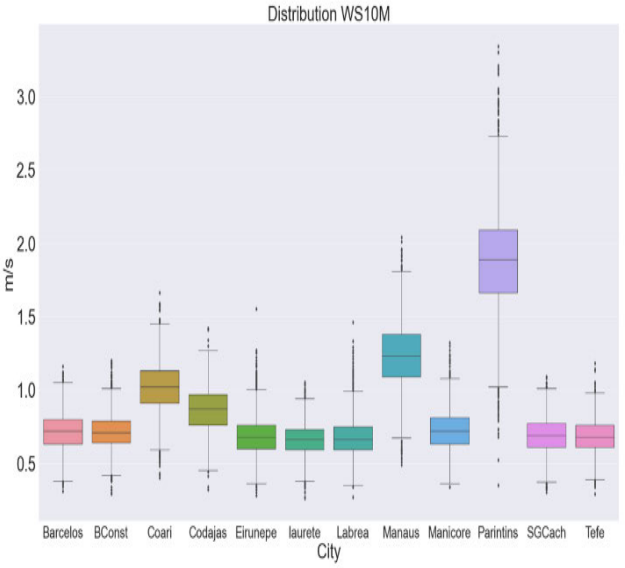

FIGURE 4. 'WS10 M' data distribution among the 12 cities.

The distribution of the relative humidity (%) at 2 meters above the ground (RH2M) is presented in Figure 5. At first

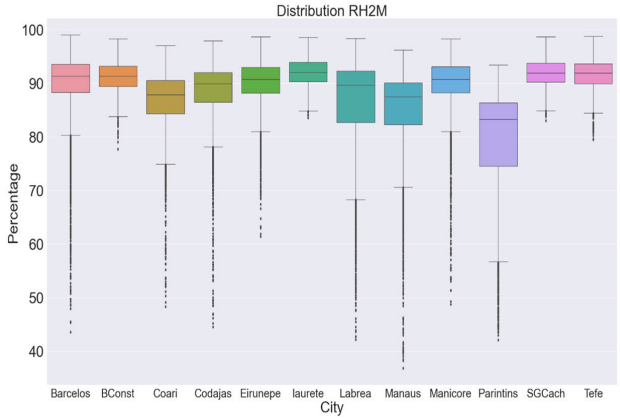

FIGURE 5. 'RH2M' data distribution among the 12 cities.

glance, Parintins has a different profile in comparison to the other cities. Its mean relative humidity is close to 83%, and the other remains around 92%. There are outliers only in the lower part of the span. The maximum value of the relative humidity for almost all the cities reaches the 100% mark.

Figure 6 presents the time-series of the target variable for the city of Barcelos. The pattern presented here is also seen in other cities, and the typical weather conditions throughout the year can explain this periodic profile. In that region, there are typically three seasons. The wet season lasts from November to March, the dry season covers the time interval between May and September, and the months of April and October comprise the 'intermediate' season [42]. The last twelve days do not have precise data, being read as '−999.0'. These data were substituted by the mean value of the whole time series to consider the most recent period. Consequently, this period has constant data, which may bind the forecasting. To avoid this odd boundary condition, we selected a thirty-day forecast period. This allowed us to consider the time oscillation observed over the past ten years, including the most recent time. This includes the occurrence

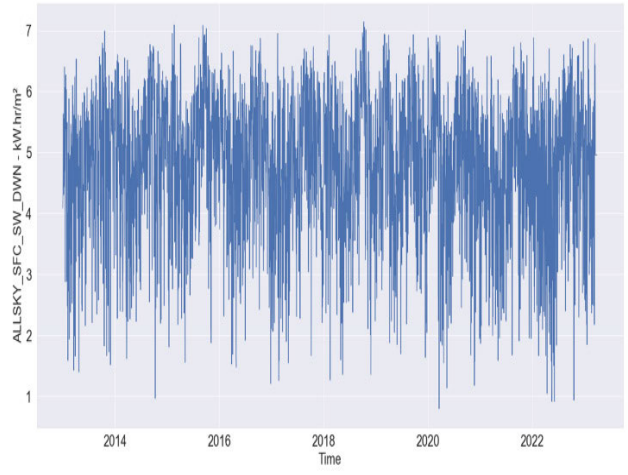

FIGURE 6. The variation of ALLSKY_SFC_SW_DWN in time for the Barcelos city.

of the 'El-Niño' phenomenon and the increase of the forest fires in the region, caused by natural and human actions.

Figure 7 shows the distribution of the values of the target variable for the Barcelos city, which is quite the same among the other cases. The distribution check is key to understand whether some kind of bias may be present or if there is a specific phenomenon indication. Also, the data distribution analysis helps the math models tuning and performance cross check because it provides an overall view on how the data behave in time or location. For instance, with Gaussian distribution profiles, no specific bias is foreseen. The data distribution seen in figure 9 indicates a Gaussian profile mostly, with a trend to hold data to the upper band. In a first glance, no major action shall be taken, but this may be reviewed later.

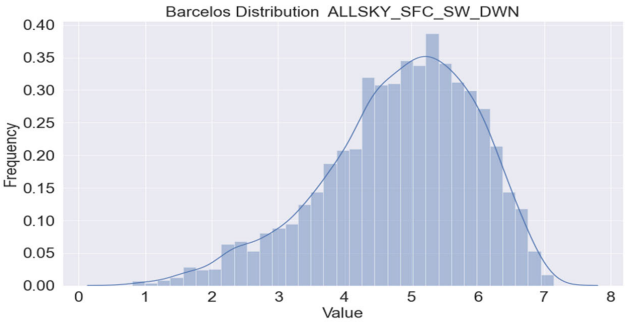

FIGURE 7. The distribution of ALLSKY_SFC_SW_DWN for the Barcelos city.

Figure 8 presents the correlation of the target variable (vertical axis) and T2M (horizontal axis) for the city of Barcelos, stratified by the three seasons discussed earlier, and the units are kW.h/m^2 and °C, respectively. Also, the profile seen here is representative of the other cases. Most of the data relays closer to 26 °C and 5 kW.hr/m^2, with a dense distribution towards the lower left corner of the figure. Towards the upper right corner, the distribution seems sparse, with data from the wet and intermediate seasons mostly. This figure shows a key parameter (T2M) for the operation of the solar panels and how it may be linked to the target variable along the year. The

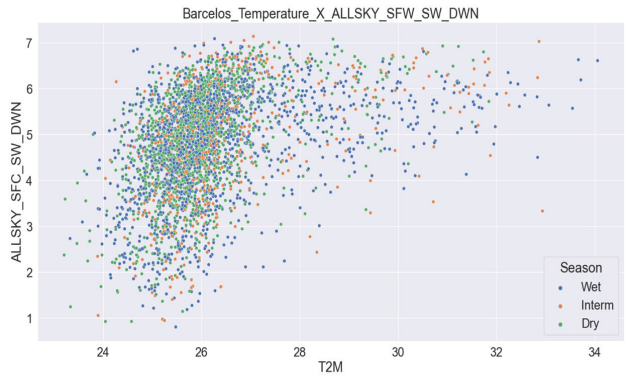

FIGURE 8. T2M and ALLSKY_SFC.

distribution of the higher temperatures values can be credited to the buildup of forest fires, from natural and human causes, and due to the 'El-Niño' phenomenon.

Considering the correlation between the RH2M and the target variable ALLSKY_SFC_SW_DWN, Figure 9 shows the overall profile for the town of Barcelos, which seems closer to all other cities. The relative humidity stays above 90% mostly, and the target variable is around 5 kW.hr/m^2. Unlike the last figure, the trend here goes towards the upper right corner, with a clear front aligned downwards. The RH2M sparse distribution relies on the left part of the figure, spanning from 50 to 80%, around the target variable figure of 6 kW.h/m^2. This variable impacts the solar panels' electrical hardware and maintenance. Regarding some technical points of attention, this high humidity standard may harm the electrical hardware used in that region, requiring constant maintenance and cross check.

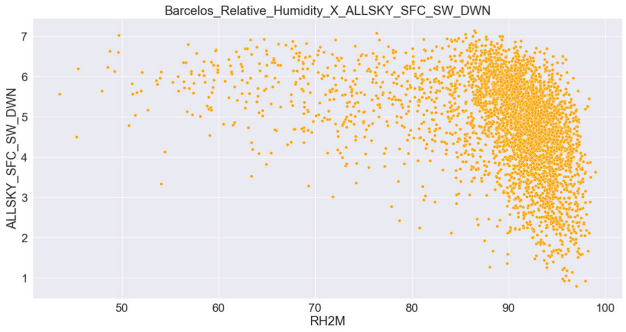

FIGURE 9. RH2m and ALLSKY_SFC_SW_DWN.

The graphs above helped to evaluate some feature engineering regarding the data distribution, focusing on the high-order variations. One point of attention is the outlier handling, which may be taken out if the relative quantity is below 3% [43], which is our case. Nonetheless, this research decided to retain all the data, to consider the influence of the fires increase and the 'El-Niño' phenomenon.

B. FEATURE IMPORTANCE

After checking the data, the most recent part showed values as '−999.00', indicating some problems in their measurement or calculation. These data were substituted by the mean value of the data series to preserve the data quantity as much as possible. This action may impact the model forecasting metrics once the data is constant after showing high variance in the training phase of the models.

The correlation matrix helps to understand the impact of each input variable on the target variable. The higher the correlation coefficient, the higher the variable importance. Figure 10 presents the correlation matrix, where the input variable ALLSKY_KT showed the highest impact as 0.926.

Another action to view the relative importance of the input variables is the feature importance from the use of an algorithm, such as Decision Trees. In this research, the Extra Tree Regressor was used for this purpose, and

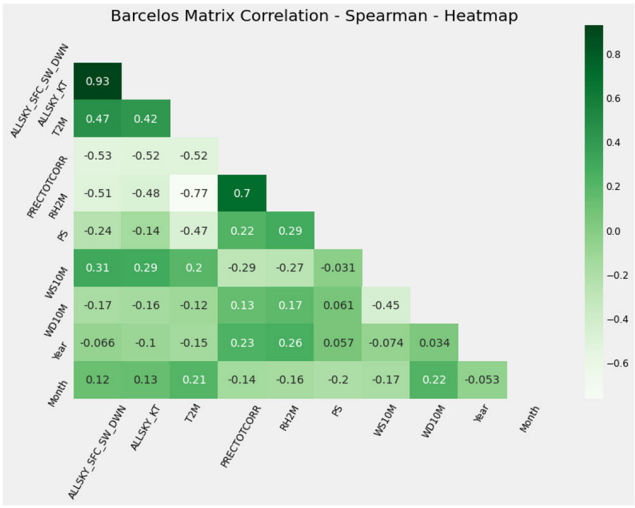

FIGURE 10. The correlation matrix with seven variables.

Figure 11 shows that the ALLSKY_KT variable has an importance above 80%. All other variables have importance lower than 10%.

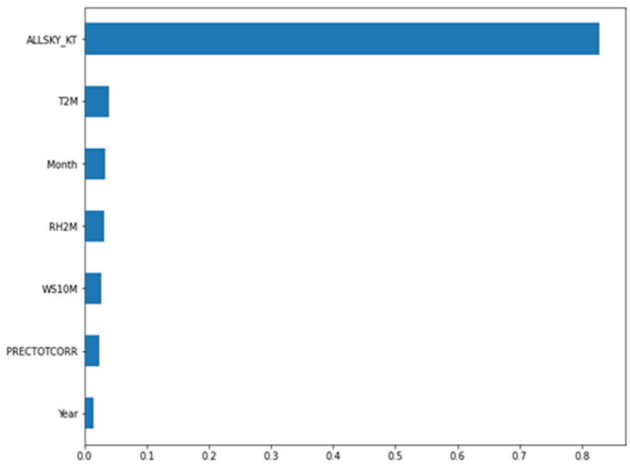

FIGURE 11. The feature importance of the input variables.

After this initial evaluation, the research adopted a decreasing number of features approach, starting with all the features and computing the associated model performance metrics. After this first case, the least important feature was taken out and the model metrics computed again, which was repeated progressively. The last case took the ALLSKY_KT as the only feature. The input variables set with the best model metrics was: ALLSKY_KT, T2M, RH2M, Month and WS10M, in descending order of the feature importance.

III. NEURAL NETWORKS BACKGROUND AND CALCULATIONS

The motivation to use the neural networks to forecast solar irradiance relies on the data variability discussed above and potential embedded phenomena present, which may not be

detected or explained easily. For instance, in some cities, the occurrence of local fires has been higher than in other areas. The 'El-Niño' phenomenon has been more intense along the years. Checking related works, several types of artificial intelligence techniques have been used, but the use of neural networks has increased, in comparison to other math tools, with special care to be taken while tuning the models, when the data distribution shall be revisited.

A. MULTI-LAYER PERCEPTRON NEURAL NETWORK (MLP)

Among the neural networks, the MLPs are one of the simplest in terms of architecture, which has tree layer types, with few layers of neurons generally, depending on the level of abstraction. The input layer receives the input data, followed by some hidden layers, and the results are available at the visible layer or output layer. The data has a table configuration in most of the uses, and MLPs can be assigned for forecasting or classification, which demands an input data treatment accordingly. As the architecture looks simply, it provides flexibility to many applications. For instance, an image and its pixels can be transformed into one long row of data and set into the input layer. In a prediction case with time series data, the lag observations can be configured to a long row of data for the input layer. Thus, the MLPs are employed as an alternative check reference to other types of neural networks by comparing the order of magnitude of their outcomes. Reference [44] presents the MLP architecture.

B. LONG SHORT-TERM MEMORY (LSTM)

In 1997 the LSTM was introduced by Hochreiter and Schmidhuber [47], being a special case of the RNNs and capable of learning long-term sequences. They are used for prediction problems with long input arrays in time series cases. They have a different architecture, normally more complex than the MLPs, aiming to manage data from past times in a longer time range. The core reason for its development was to handle long data sequences for a long-time interval or long-term dependence. LSTMs have feedback connections and operate with gates, which allow (or not) some data to take another way (in or out) while the neural network is still working. Memory management is feasible by a data bus, which is set alongside the configuration of a classic neural network. This 'parallel' data bus operates with data without neural processing, which is called 'cell state' in some cases. The fundamental action of LSTM can be seen as holding the required information and throwing away the information which is not useful for further prediction. Reference [46] shows the LSTM overall architecture. As it can be regarded, the pointwise multiplication operation and a sigmoid neural net layer, which assist the mechanism, manage the data throughout the gates, following a zero (nothing goes through) or one (everything goes through) logic. Taking a basic configuration, the LSTMs can have forward pass, backward pass, or bidirectional pass, depending on how the data will be allowed to go. Therefore, the LSTMs are more complex than MLPs.

C. LSTM_GRU

Dey and Salem [48], and Wang et al. [49] present the main concepts of GRUs. In brief, GRUs are like LSTMs, but with few key differences, showing performance improvements consequently. First, GRUs do not have an output gate, and thus, they have just two gates (input and target gate), one less than LSTMs, and they do not have any type of internal memory either. GRUs are then simpler and faster than LSTMs (around 30%), but they exhibit more difficulties in being adaptable. Figure 16 presents the architecture of the LSTM and GRU.

In this research, the three types of neural networks were employed for a prediction period of 30 days for each of the 12 stations or cities, according to a task flow, starting with the data input (3741 observations). The networks were used under a supervised approach, and the test set took the last 30 days, making the forecast target period. The train set took all the previous data (3711 observations).

The first step was the data check, cleaning and transform, and the scaling of the input variables between 0 and 1. The target variable was maintained with its original values. Next, the neural network architecture was defined in terms of the number of layers, type of activation function and optimizer, among others. For each network parameter, a list of possible values, or range, was considered in the 'GridSearch' analysis to minimize a loss function. Once the parameters were determined, the neural network was set to run and forecast.

The overfitting was checked under the criterion: no concern was taken if the train set error was 35% larger than the test set error. If not, the network architecture was reviewed in a feedback move. Another check was the order of magnitude of the output compared to similar references. The output was considered acceptable if the error metrics of the neural networks were within a 15% variation around the references. Otherwise, the same feedback move was taken to review the neural network's architecture.

D. METRICS

For practical reasons, this research took the mean absolute percentage error (MAPE), dimensionless, as the primary metric. The mean absolute error (MAE), in (kW.h/m^2), and the root mean squared error (RMSE), in (kW.h/m^2), was also calculated to allow the comparison with other similar references. The metrics inputs came from the test phase data group in the supervised approach: the target variable input taken from the NASA source [41], the calculated value by each neural network method, and the number of observations, which was equal to 30. The equations for each metric follow:

$$\mathrm{MAE} = (1/n) * \sum_{i=1}^{n} \left| y\,(i) - y^*\,(i) \right| \tag{1}$$

$$\mathrm{MAPE} = (1/n) * \sum_{i=1}^{n} \left| y\,(i) - y^*\,(i) \right| / y\,(i) \tag{2}$$

$$\mathrm{RMSE} = \sqrt{(1/n) * \sum_{i=1}^{n} \left[y\,(i) - y^*\,(i) \right]^2} \tag{3}$$

where:

1) $y(i)$ is the target variable input from the NASA source.
2) $y(i)^*$ is the forecasted value generated by the neural network.
3) n is the number of observations, or 30.

IV. RESULTS

Initially, the NN approaches used AWS SageMaker notebooks, Python libraries, with 4 GB and 2 vCPUs (ml.t3.medium), demanding around 40 hours of computing work to optimize resources and costs.

For the NN, a Grid Search was used to seek the best hyperparameters, considering the following items:

- Maximum number of iterations: 100
- Number of epochs: 50, 100
- Batch size: 32
- Parameter spaces:
  hidden layer sizes:
  (10,20,10),(10,30,10),(20,),(35,),(100,2)
- activation functions: 'tanh', 'relu', 'logistic', 'identity'
- solvers: 'sgd','adam','lbfgs'
- alpha factor: [0.0001, 0.0005, 0.001, 0.05]
- learning_rate: 'constant', 'invscaling', 'adaptive'.

The results of the forecast methods are expressed by the metrics taken from the three neural networks for the cities. The first set of results deals with the MLP, which demanded a computing time for each city, around 1100 seconds, as shown in Table 4. The second results are presented in Table 5, dealing with the second NN configuration (LSTM2), with a typical computing time of 250 seconds for each case. Finally, Table 6 shows the results for the third type of NN (LSTM_GRU), where the average computing time was 30 seconds.

TABLE 4. MLP metrics.

City	MAE	MAPE	RMSE
Barcelos	1.17	30.2	2.0
Benj.Constant	1.09	31.7	1.89
Coari	1.21	37.6	2.25
Codajas	1.21	52.2	1.65
Eirunepe	1.02	30.7	1.79
Iaurete	1.34	40.16	2.75
Labrea	0.61	19.5	1.05
Manaus	1.24	57.5	2.87
Manicore	1.56	39.4	3.21
Parintins	1.73	44.1	4.0
S.G.Cachoeira	1.28	67.21	1.7
Tefe	0.97	29.1	1.29

For the city of Codajas, Figure 12 shows the comparison between the actual (horizontal axis) and the predicted values (vertical axis), both measured in kW.hr/m^2 and using a LSTM_GRU architecture for the NN.

If the model had 100% accuracy, the orange plots should be aligned with the upper right diagonal of the figure (in red). As can be seen, there was a kind of 45° alignment, but with a certain distance from the diagonal, represented by the differences in the scales of the axes, meaning the prediction error is different from zero. However, 100% aligned in the diagonal means the occurrence of overfitting probably. Thus, the distribution of the points around the diagonal looks more interesting, meaning the model can handle new data, or variance, from unexpected phenomenon, such as the local fires. In this case, the model performance shall not be different.

TABLE 5. LSTM2 metrics.

City	MAE	MAPE	RMSE
Barcelos	1.14	24.2	1.79
Benj.Constant	1.22	29.5	2.04
Coari	1.21	32.3	2.22
Codajas	1.21	52.2	1.65
Eirunepe	1.73	37.2	3.61
Iaurete	1.30	31.6	2.43
Labrea	0.81	26.62	1.05
Manaus	1.21	41.86	2.06
Manicore	1.41	30.9	2.52
Parintins	1.14	24.2	1.79
S.G.Cachoeira	1.22	29.5	2.04
Tefe	1.21	32.3	2.22

TABLE 6. LSTM_GRU metrics.

City	MAE	MAPE	RMSE
Barcelos	1.27	20.5	0.75
Benj.Constant	1.64	26.6	1.08
Coari	1.31	25.2	0.89
Codajas	1.36	35.1	1,17
Eirunepe	3.09	35.1	1.58
Iaurete	2.11	28.3	1.15
Labrea	0.89	19.2	1.13
Manaus	1.96	47.5	1.12
Manicore	1.90	26.9	1.24
Parintins	1.14	19.9	0.81
S.G.Cachoeira	1.01	54	1.91
Tefe	1.11	27.3	1.96

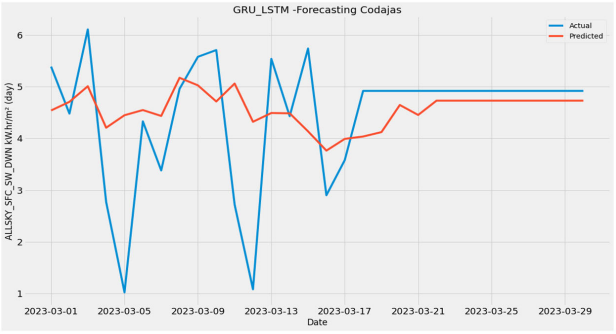

FIGURE 12. Combined plot of actual and predicted values for Codajas.

Considering all the cities and NN models, the RMSE stayed within the interval of 0.75 and 4.0. For the MAE metrics, the values detected were within the interval of 0.61 and 3.09. Lastly, the MAPE showed values between 19.2 and 67.21. These variations can be credited to the different meteorological input variables, such as cloud coverture, vegetation coverage and rain profile, among the different cities. The influence of the rainforest on local relative humidity, which is linked to low-altitude cloudiness, is well-documented. For instance, Manaus has an industrial surrounding, changing thus the relative humidity spatial distribution, which impacts the low altitude cloud coverture. The city of Parintins has a different input variables profile, as seen in the EDA, in terms of the wind speed, for instance. Although it could be straightforward to assume all the cities have a quite homogeneous set of boundary conditions, a more detailed analysis shows there are key differences among them, which may impact the forecasting. This subject shall be part of further research, considering the data with a more detailed view.

The lowest RMSE obtained was for the city of Barcelos (0.75), when using the LSTM_GRU model. The lowest MAE is related to the city of Labrea (0.61) and the MLP model. In terms of MAPE, the lowest figure corresponds to the city of Labrea (19.2) again, using the LSTM_GRU approach.

V. DISCUSSION

The analysis of the results above considers the comparison against the same metrics of similar research, as far as possible, using more than one type of NN. It is worth noting that this research has yet to find fully equal cases in terms of boundary and initial conditions. For instance, most of the references aim for time intervals of hours or a day which is a short period. This research worked with forecast periods of 30 days, because it aims economic and feasibility evaluations. Another difference is the area conditions; this research dealt with rainforest environments, and most of the references deal with plain surface areas closer to deserts or urban areas.

Among the results, the lowest metrics found were indicated in Table 7 for the forecast time interval of 30 days.

TABLE 7. Lowest metrics obtained – city of Labrea.

NN	MAE	MAPE	RMSE
MLP	0.61	19.5	1.05
LSTM2	0.81	22.77	1.05
LSTM_GRU	0.89	19.2	0.75

The comparison to other references is condensed in Table 8, followed by comments.

Foo and Goh [50] use an evolutionary lean neural network to forecast solar irradiance in tropical weather conditions. In [51] authors use several machine learning (ML) algorithms, including MLP, LSTM, CNN, and hybrid models, in South Africa for a time interval of up to one hour. Finally, Ref. [52] concentrate on South Korea, using LSTM and CNN for time intervals of up to 6 hours.

TABLE 8. Metrics comparison between references and this research.

Reference	MAE	MAPE	RMSE
50	54.12	xx	104.08
51	51.56	xx	75.52
52	0.051	13.42	0.098
This work	0.61	19.2	0.75

When comparing the numbers presented in Table 8, it is evident that the current research results can be considered acceptable. This is because a longer time interval (one month) tends to result in larger error metrics, whereas shorter time intervals are associated with smaller error metrics, particularly when using the same data group. However, this does not apply to our case. In this research, the data condition for the last twelve days necessitated an unusual assumption to maximize data utilization. Although there is a small similarity with Ref. [37] for the tropical conditions the results differ reasonably. Nonetheless, even for references with the same type of forecasting time interval, the results differ. The peculiar neural network configurations may aid in explaining the diversity of these metrics.

Another evaluation done dealt with the statistical test, following the Analysis of Variance (ANOVA) [53], considering the Absolute Errors of each NN model, for the test phase. Thus, the 30 values of the Absolute Errors were taken, for each city and a statistical test was done to check the null hypothesis: the average Absolute Error was the same for the cities. Initially, other basic hypotheses were checked as the normal distribution of the Absolute Errors, in each city, and the associate variances shall be equal or close, shown by the ratio between the maximum standard deviation and the minimum standard deviation being smaller than 2, which was confirmed for every NN model.

For all NN models we had enough evidence that not all average Absolute Errors were the same for the cities, at significance level of 5%, with a p-value smaller than 1% for the MLP model and 4.6% for the LSTM cases.

The input data and Python notebooks used in this research are open-source and available on Zenodo platform [54], [55].

VI. CONCLUSION

After checking the models and metrics addressed above, the following conclusions can be reached:

This research has used neural networks to forecast solar energy for a thirty-day time horizon in rainforest surroundings. It took twelve cities of the Amazonas state considering the use of renewable energy to reduce the emissions of GHG, mainly in the cases of increased deforestation. The cities have quite similar boundary conditions, although key differences may explain the differences among the model differences, such as the relative humidity.

The lowest MAPE (19.2) was achieved using an LSTM_GRU model. The lowest MAE (0.61) came using an MLP model, and the lowest RMSE (0.75) was obtained with an LSTM_GRU model.

In terms of metrics, the LSTM_GRU model showed the best performance. The second was the MLP, and the last was the LSTM2 model.

For future work, the focus should be on examining meteorological input variables and their variations among different cities to understand their potential impact on forecast models. Furthermore, it is worth considering shorter time intervals, such as 15 days, in future research.

ACKNOWLEDGMENT

Authors acknowledge the support of the RCGI – Research Centre for Greenhouse Gas Innovation (23.1.8493.1.9), hosted by the University of São Paulo (USP) and sponsored by FAPESP – São Paulo Research Foundation (2020/15230-5) and sponsors, as well as the strategic importance of the support given by ANP (Brazil's National Oil, Natural Gas and Biofuels Agency) through the R&DI levy regulation. Márcio Teixeira is sponsored by FAPESP grant #2021/12954-5. Equally importantly, Felipe Almeida is sponsored by the National Council for Scientific and Technological Development (CNPq), grant #140253/2021-1. The authors also thank the support provided by the Ministry of Defense–Amazon Protection System (SIPAM).

REFERENCES

[1] World Economic Forum. (2022). *Four Innovations Preparing Cities for Climate Change*. [Online]. Available: https://www.weforum.org/agenda/2022/10/innovations-protect-cities-climate-change/

[2] M. Guimberteau, P. Ciais, A. Ducharne, J. P. Boisier, A. P. D. Aguiar, H. Biemans, H. De Deurwaerder, D. Galbraith, B. Kruijt, F. Langerwisch, G. Poveda, A. Rammig, D. A. Rodriguez, G. Tejada, K. Thonicke, C. Von Randow, R. C. S. Von Randow, K. Zhang, and H. Verbeeck, "Impacts of future deforestation and climate change on the hydrology of the Amazon basin: A multi-model analysis with a new set of land-cover change scenarios," *Hydrol. Earth Syst. Sci.*, vol. 21, no. 3, pp. 1455–1475, Mar. 2017, doi: 10.5194/hess-21-1455-2017.

[3] C. P. Castro, "Hydropower and the geopolitics of renewable energies in the Amazon Basin," *Ambiente Sociedade*, vol. 24, Nov. 2021, Art. no. e01291, doi: 10.1590/1809-4422asoc20200129r1vu2021l2ao.

[4] A. A. Prasad and M. Kay, "Prediction of solar power using near-real time satellite data," *Energies*, vol. 14, no. 18, p. 5865, Sep. 2021, doi: 10.3390/en14185865.

[5] G. Alkhayat and R. Mehmood, "A review and taxonomy of wind and solar energy forecasting methods based on deep learning," *Energy AI*, vol. 4, Jun. 2021, Art. no. 100060, doi: 10.1016/j.egyai.2021.100060.

[6] A. Michiorri, A. M. Sempreviva, S. Philipp, P. Perez-Lopez, A. Ferriere, and D. Moser, "Topic taxonomy and metadata to support renewable energy digitalisation," *Energies*, vol. 15, no. 24, p. 9531, Dec. 2022, doi: 10.3390/en15249531.

[7] F. R. Alharbi and D. Csala, "Wind speed and solar irradiance prediction using a bidirectional long short-term memory model based on neural networks," *Energies*, vol. 14, no. 20, p. 6501, Oct. 2021, doi: 10.3390/en14206501.

[8] S. Jafarzadeh Ghoushchi, S. Manjili, A. Mardani, and M. K. Saraji, "An extended new approach for forecasting short-term wind power using modified fuzzy wavelet neural network: A case study in wind power plant," *Energy*, vol. 223, May 2021, Art. no. 120052, doi: 10.1016/j.energy.2021.120052.

[9] A. Gensler, J. Henze, B. Sick, and N. Raabe, "Deep learning for solar power forecasting—An approach using AutoEncoder and LSTM neural networks," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, Budapest, Hungary, Oct. 2016, pp. 2858–2865, doi: 10.1109/SMC.2016.7844673.

[10] A. Alzahrani, P. Shamsi, M. Ferdowsi, and C. Dagli, "Solar irradiance forecasting using deep recurrent neural networks," in *Proc. IEEE 6th Int. Conf. Renew. Energy Res. Appl. (ICRERA)*, San Diego, CA, USA, Nov. 2017, pp. 988–994, doi: 10.1109/ICRERA.2017.8191206.

[11] Z.-K. Feng, Q.-Q. Huang, W.-J. Niu, T. Yang, J.-Y. Wang, and S.-P. Wen, "Multi-step-ahead solar output time series prediction with gate recurrent unit neural network using data decomposition and cooperation search algorithm," *Energy*, vol. 261, Dec. 2022, Art. no. 125217, doi: 10.1016/j.energy.2022.125217.

[12] D. Cannizzaro, A. Aliberti, L. Bottaccioli, E. Macii, A. Acquaviva, and E. Patti, "Solar radiation forecasting based on convolutional neural network and ensemble learning," *Expert Syst. Appl.*, vol. 181, Nov. 2021, Art. no. 115167, doi: 10.1016/j.eswa.2021.115167.

[13] R. Nematirad and A. Pahwa, "Solar radiation forecasting using artificial neural networks considering feature selection," in *Proc. IEEE Kansas Power Energy Conf. (KPEC)*, Manhattan, KS, USA, Apr. 2022, pp. 1–4, doi: 10.1109/KPEC54747.2022.9814765.

[14] X. He, Y. Wang, Y. Zhang, X. Ma, W. Wu, and L. Zhang, "A novel structure adaptive new information priority discrete grey prediction model and its application in renewable energy generation forecasting," *Appl. Energy*, vol. 325, Nov. 2022, Art. no. 119854, doi: 10.1016/j.apenergy.2022.119854.

[15] M. Gao, J. Li, F. Hong, and D. Long, "Day-ahead power forecasting in a large-scale photovoltaic plant based on weather classification using LSTM," *Energy*, vol. 187, Nov. 2019, Art. no. 115838, doi: 10.1016/j.energy.2019.07.168.

[16] E. M. Al-Ali, Y. Hajji, Y. Said, M. Hleili, A. M. Alanzi, A. H. Laatar, and M. Atri, "Solar energy production forecasting based on a hybrid CNN-LSTM-transformer model," *Mathematics*, vol. 11, no. 3, p. 676, Jan. 2023, doi: 10.3390/math11030676.

[17] F. Wang, Y. Yu, Z. Zhang, J. Li, Z. Zhen, and K. Li, "Wavelet decomposition and convolutional LSTM networks based improved deep learning model for solar irradiance forecasting," *Appl. Sci.*, vol. 8, no. 8, p. 1286, Aug. 2018, doi: 10.3390/app8081286.

[18] X. Guo, Y. Mo, and K. Yan, "Short-term photovoltaic power forecasting based on historical information and deep learning methods," *Sensors*, vol. 22, no. 24, p. 9630, Dec. 2022, doi: 10.3390/s22249630.

[19] N. L. M. Jailani, J. K. Dhanasegaran, G. Alkawsi, A. A. Alkahtani, C. C. Phing, Y. Baashar, L. F. Capretz, A. Q. Al-Shetwi, and S. K. Tiong, "Investigating the power of LSTM-based models in solar energy forecasting," *Processes*, vol. 11, no. 5, p. 1382, May 2023, doi: 10.3390/pr11051382.

[20] L. Fara, A. Diaconu, D. Craciunescu, and S. Fara, "Forecasting of energy production for photovoltaic systems based on ARIMA and ANN advanced models," *Int. J. Photoenergy*, vol. 2021, pp. 1–19, Aug. 2021, doi: 10.1155/2021/6777488.

[21] Q. Li, D. Zhang, and K. Yan, "A solar irradiance forecasting framework based on the CEE-WGAN-LSTM model," *Sensors*, vol. 23, no. 5, p. 2799, Mar. 2023, doi: 10.3390/s23052799.

[22] G. M. Yagli, D. Yang, and D. Srinivasan, "Ensemble solar forecasting and post-processing using dropout neural network and information from neighboring satellite pixels," *Renew. Sustain. Energy Rev.*, vol. 155, Mar. 2022, Art. no. 111909, doi: 10.1016/j.rser.2021.111909.

[23] M. El-Hendawi and Z. Wang, "An ensemble method of full wavelet packet transform and neural network for short term electrical load forecasting," *Electr. Power Syst. Res.*, vol. 182, May 2020, Art. no. 106265, doi: 10.1016/j.epsr.2020.106265.

[24] P. Li, K. Zhou, X. Lu, and S. Yang, "A hybrid deep learning model for short-term PV power forecasting," *Appl. Energy*, vol. 259, Feb. 2020, Art. no. 114216, doi: 10.1016/j.apenergy.2019.114216.

[25] Z. Pang, F. Niu, and Z. O'Neill, "Solar radiation prediction using recurrent neural network and artificial neural network: A case study with comparisons," *Renew. Energy*, vol. 156, pp. 279–289, Aug. 2020, doi: 10.1016/j.renene.2020.04.042.

[26] K. Li, W. Huang, G. Hu, and J. Li, "Ultra-short term power load forecasting based on CEEMDAN-SE and LSTM neural network," *Energy Buildings*, vol. 279, Jan. 2023, Art. no. 112666, doi: 10.1016/j.enbuild.2022.112666.

[27] S. Netsanet, D. Zheng, W. Zhang, and G. Teshager, "Short-term PV power forecasting using variational mode decomposition integrated with ant colony optimization and neural network," *Energy Rep.*, vol. 8, pp. 2022–2035, Nov. 2022, doi: 10.1016/j.egyr.2022.01.120.

[28] F.-F. Li, S.-Y. Wang, and J.-H. Wei, "Long term rolling prediction model for solar radiation combining empirical mode decomposition (EMD) and artificial neural network (ANN) techniques," *J. Renew. Sustain. Energy*, vol. 10, no. 1, Jan. 2018, Art. no. 013704, doi: 10.1063/1.4999240.

[29] L. Feng, A. Lin, L. Wang, W. Qin, and W. Gong, "Evaluation of sunshine-based models for predicting diffuse solar radiation in China," *Renew. Sustain. Energy Rev.*, vol. 94, pp. 168–182, Oct. 2018, doi: 10.1016/j.rser.2018.06.009.

[30] D. Sadeghi, A. Golshanfard, S. Eslami, K. Rahbar, and R. Kari, "Improving PV power plant forecast accuracy: A hybrid deep learning approach compared across short, medium, and long-term horizons," *Renew. Energy Focus*, vol. 45, pp. 242–258, Jun. 2023, doi: 10.1016/j.ref.2023.04.010.
[31] L. Yang, Q. Cao, Y. Yu, and Y. Liu, "Comparison of daily diffuse radiation models in regions of China without solar radiation measurement," *Energy*, vol. 191, Jan. 2020, Art. no. 116571, doi: 10.1016/j.energy.2019.116571.
[32] F. Nawab, A. S. A. Hamid, A. Ibrahim, K. Sopian, A. Fazlizan, and M. F. Fauzan, "Solar irradiation prediction using empirical and artificial intelligence methods: A comparative review," *Heliyon*, vol. 9, no. 6, Jun. 2023, Art. no. e17038, doi: 10.1016/j.heliyon.2023.e17038.
[33] Y. El Mghouchi, E. Chham, E. M. Zemmouri, and A. El Bouardi, "Assessment of different combinations of meteorological parameters for predicting daily global solar radiation using artificial neural networks," *Building Environ.*, vol. 149, pp. 607–622, Feb. 2019, doi: 10.1016/j.buildenv.2018.12.055.
[34] E. L. Segarra, H. Du, G. R. Ruiz, and C. F. Bandera, "Methodology for the quantification of the impact of weather forecasts in predictive simulation models," *Energies*, vol. 12, no. 7, p. 1309, Apr. 2019, doi: 10.3390/en12071309.
[35] J. Chen, W. Zhu, and Q. Yu, "Estimating half-hourly solar radiation over the continental United States using GOES-16 data with iterative random forest," *Renew. Energy*, vol. 178, pp. 916–929, Nov. 2021, doi: 10.1016/j.renene.2021.06.129.
[36] Q. Dai and X. Fang, "A simple model to predict solar radiation under clear sky conditions," *Adv. Space Res.*, vol. 53, no. 8, pp. 1239–1245, Apr. 2014, doi: 10.1016/j.asr.2014.01.025.
[37] C. N. Obiora, A. Ali, and A. N. Hassan, "Predicting hourly solar irradiance using machine learning methods," in *Proc. 11th Int. Renew. Energy Congr. (IREC)*, Hammamet, Tunisia, Oct. 2020, pp. 1–6, doi: 10.1109/IREC48820.2020.9310444.
[38] S. Sivarajah. (2020). *The 'Generic' Data Science Life-Cycle*. Accessed: Jan. 15, 2023. [Online]. Available: https://towardsdatascience.com/stoend-to-end-data-science-life-cycle-6387523b5afc
[39] *IBGE and Cidades e Estados*. Accessed: Jan. 15, 2023. [Online]. Available: https://www.ibge.gov.br/cidades-e-estados/am.html
[40] *INMET Dados*. Accessed: Jan. 15, 2023. [Online]. Available: https://tempo.inmet.gov.br/TabelaEstacoes/A001
[41] *NASA and The POWER Project*. Accessed: Jan. 15, 2023. [Online]. Available: https://power.larc.nasa.gov/
[42] P. Artaxo et al., "The Green Ocean Amazon Experiment (GoAmazon2014/5) observes pollution affecting gases, aerosols, clouds, and rainfall over the rain forest," *Bull. Amer. Meteorol. Soc.*, vol. 98, no. 5, pp. 981–997, May 2017, doi: 10.1175/bams-d-15-00221.1.
[43] G. Fisch. *Clima Da Amazônia*. Accessed: Jan. 15, 2023. [Online]. Available: http://climanalise.cptec.inpe.br/~rclimanl/boletim/cliesp10a/fish.html
[44] J. Osborne and A. Overbay, "The power of outliers (and why researchers should ALWAYS check for them)," *Pract. Assess. Res. Eval.*, vol. 9, no. 1, p. 6, 2004.
[45] H. A. Afan, A. I. A. Osman, Y. Essam, A. N. Ahmed, Y. F. Huang, O. Kisi, M. Sherif, A. Sefelnasr, K.-W. Chau, and A. El-Shafie, "Modeling the fluctuations of groundwater level by employing ensemble deep learning techniques," *Eng. Appl. Comput. Fluid Mech.*, vol. 15, no. 1, pp. 1420–1439, Jan. 2021, doi: 10.1080/19942060.2021.1974093.
[46] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.
[47] G. Van Houdt, C. Mosquera, and G. Nápoles, "A review on the long short-term memory model," *Artif. Intell. Rev.*, vol. 53, no. 8, pp. 5929–5955, Dec. 2020, doi: 10.1007/s10462-020-09838-1.
[48] R. Dey and F. M. Salem, "Gate-variants of gated recurrent unit (GRU) neural networks," in *Proc. IEEE 60th Int. Midwest Symp. Circuits Syst. (MWSCAS)*, Boston, MA, USA, Aug. 2017, pp. 1597–1600, doi: 10.1109/MWSCAS.2017.8053243.
[49] L. Wang, X.-D. Zhang, J.-W. Tang, Z.-W. Ma, M. Usman, Q.-H. Liu, C.-Y. Wu, F. Li, Z.-B. Zhu, and B. Gu, "Machine learning analysis of SERS fingerprinting for the rapid determination of mycobacterium tuberculosis infection and drug resistance," *Comput. Struct. Biotechnol. J.*, vol. 20, pp. 5364–5377, Sep. 2022, doi: 10.1016/j.csbj.2022.09.031.
[50] Y. W. Foo and C. Goh, "Solar irradiance forecasting in tropical weather using an evolutionary lean neural network," in *Proc. IEEE Congr. Evol. Comput. (CEC)*, Kraków, Poland, Jun./Jul. 2021, pp. 490–497, doi: 10.1109/CEC45853.2021.9504875.
[51] C. N. Obiora, A. N. Hasan, and A. Ali, "Predicting solar irradiance at several time horizons using machine learning algorithms," *Sustainability*, vol. 15, no. 11, p. 8927, Jun. 2023, doi: 10.3390/su15118927.
[52] W. Lee, K. Kim, J. Park, J. Kim, and Y. Kim, "Forecasting solar power using long-short term memory and convolutional neural networks," *IEEE Access*, vol. 6, pp. 73068–73080, 2018, doi: 10.1109/ACCESS.2018.2883330.
[53] S. F. Sawyer, "Analysis of variance: The fundamental concepts," *J. Manual Manipulative Therapy*, vol. 17, no. 2, pp. 27E–38E, Apr. 2009, doi: 10.1179/jmt.2009.17.2.27e.
[54] A. Marques, 2023, "12_datasets_solar_energy_Amazonas_state_neural _networks, *Zenodo*, doi: 10.5281/zenodo.8090459.
[55] A. Marques, 2023, "Three_jupyter_notebooks_neural_networks_solar _energy_forecasting_12_cities_Amazonas_state_Brazil," *Zenodo*, doi: 10.5281/zenodo.8090563.

ANDRÉ LUÍS FERREIRA MARQUES received the bachelor's degree in naval engineering (propulsion and hydrodynamics) and the master's degree in nuclear engineering from the University of São Paulo, in 1990 and 1995, respectively, and the master's degree in mechanical engineering and the Engineering degree in nuclear engineering from the Massachusetts Institute of Technology, in 1998. He is currently pursuing the Ph.D. degree in computer engineering with the University of São Paulo, with a focus on the application of data science in the energy matrix transition with renewable sources. After serving in the Brazilian Navy for 40 years, with the development and research of nuclear systems, dealing with experimental work, hardware construction, and operation. He is a Data Scientist.

MÁRCIO JOSÉ TEIXEIRA received the B.Sc. degree in physics and the M.Sc. degree in particle physics from the Institute of Physics, University of São Paulo (USP), in 1994 and 1997, respectively, and the Ph.D. degree in communications and information systems from the State University of Campinas (UNICAMP), in 2021.

With over 15 years of experience, he has led projects and innovation initiatives in multinational companies, such as Ericsson and America Movil, primarily focusing on mobile internet and the Internet of Things (IoT). After a sabbatical period (2018–2021), he received the Ph.D. degree. Since December 2021, he has been a Postdoctoral Fellow with the Research Center for Greenhouse Gas Innovation (RCGI), USP. His research interests include mobile communications and the application of machine learning in environmental sciences.

Dr. Teixeira is a member of the Brazilian Physical Society (SBF) and the American Geophysical Union (AGU).

FELIPE VALENCIA DE ALMEIDA received the bachelor's and master's degrees in computer engineering from the University of São Paulo, in 2018 and 2020, respectively, where he is currently pursuing the Ph.D. degree, with a focus on research around high-performance programming and emphasis on the application of accelerators implemented in FPGAs and heterogeneous architectures.

PEDRO LUIZ PIZZIGATTI CORRÊA received the bachelor's and master's degrees in computer science from the University of São Paulo, in 1987 and 1992, respectively, and the Ph.D. degree in electrical engineering from the Polytechnic School, University of São Paulo, in 2002. He held a Postdoctoral Fellowship in data science with the University of Tennessee, in 2015. Since 2017, he has been an Associate Professor with the Computer Engineering and Digital Systems Department (PCS), Polytechnic School, University of São Paulo, Brazil. His current research interests include distributed databases, data science, computer system modeling, distributed system architecture, computing in biodiversity, agricultural automation, and electronic government.