

Forecasting hourly global solar radiation with an Artificial Neural Network

Previsão da radiação solar global horária a partir de uma rede neural artificial

Ronald Rocha de Jesus¹, Elias Fernandes de Sousa², Antônio J. Silva Neto³

RESUMO

Muitos estudos têm sido realizados para estimar a radiação solar global, uma vez que a determinação por equipamentos manuais ou automáticos envolve custos elevados. De forma geral, os modelos de previsão da radiação solar têm por finalidade a previsão da radiação solar diária e são desenvolvidos a partir de dados de temperatura. Estudos sobre a radiação solar horária são escassos. A determinação da radiação solar horária pode aumentar a precisão em algumas áreas de pesquisa, como a agricultura de precisão. Além disso, as variações horárias da radiação solar computam intrinsecamente atributos locais que interferem no processo da radiação solar como a topografia, detalhes do relevo e composição atmosférica. O objetivo desse estudo foi desenvolver uma Rede Neural Artificial para a previsão da radiação global horária a partir de dados de temperatura e da radiação solar incidente no topo da atmosfera ao meio dia. Para isso, foram utilizados dados de monitoramento da estação automática do Instituto Nacional de Meteorologia – INMET, localizadas em Campos dos Goytacazes/RJ, às coordenadas 41,35° Longitude O e 21,71° Latitude S. O modelo adotado foi uma Rede Neural Artificial de múltiplas camadas (*Multilayer Perceptron*). O desempenho do modelo foi avaliado empregando Erro Quadrático Médio, o Erro Médio Absoluto e R^2 .

Palavras-chave: Amplitude Térmica. Incidência Solar. Inteligência Artificial

ABSTRACT

Many studies have been carried out to estimate the global solar radiation, since the determination by manual or automatic equipment involves high costs. In general, solar radiation prediction models are intended to predict daily solar radiation and are developed based on temperature data. Studies on hourly solar radiation are scarce. The determination of hourly solar radiation can increase accuracy in some areas of research, such as precision agriculture. In addition, the hourly variations of the solar radiation intrinsically compute local attributes that interfere in the process of solar radiation, such as topography, relief and atmospheric composition. The goal of this study was to develop an Artificial Neural Network to predict hourly solar radiation from temperature data and incident solar radiation at the top of the atmosphere at noon. For this, monitoring data from the automatic station of the National Institute of Meteorology, located in Campos dos Goytacazes, Rio de Janeiro State, Brazil, at coordinates 41.35° W Longitude and 21.71° S Latitude, were used. The adopted model was an Artificial Neural Network with multiple layers (*Multilayer Perceptron*). The performance of the model was evaluated by Mean Squared Error, Mean Absolute Error and R^2 .

Keywords: Thermal Amplitude. Solar Incidence. Artificial Intelligence

¹ PhD student in Modeling and Technology for the Environment Applied to Water Resources - Instituto Federal de Educação, Ciência e Tecnologia Fluminense - IFF.

<https://orcid.org/0000-0003-2634-3100>

engenheiroronald@gmail.com

² PhD in Plant Production - Universidade Estadual do Norte Fluminense Darcy Ribeiro – UENF.

<https://orcid.org/0000-0002-9323-3203>

sousa.elias.fernandes@gmail.com

³ PhD in Mechanical and Aerospace Engineering – Universidade do Estado do Rio de Janeiro – UERJ.

1. INTRODUCTION

Solar energy is our planet's main source of energy. Photosynthesis is responsible for most of life. When light penetrates the atmosphere and the water column (from the sea, rivers, lakes, etc.), part of it is absorbed into the physiological and biochemical processes of plants and part is converted into heat. Plants convert light energy into chemical energy, which allows them to produce biomass (ASSIS & MENDEZ, 1989; HARARI, 2021).

Furthermore, current population growth rates have created pressure to increase food and energy production. The generation of electricity from solar radiation represents an important tool for energy transition, offering more ecologically appropriate and sustainable solutions (RIBEIRO, 2011).

Therefore, the study of the potential of global solar radiation (GSR) incident on the Earth's surface is extremely important, as it has impact on research on electricity production, agriculture, and climate change, for example (AULER & MINUZZI, 2022).

The GSR is an essential input variable in the construction of crop growth models, since photosynthesis is primarily driven by solar radiation. Variations in hourly global solar radiation (HGSR) impact plant physiological processes and, consequently, the climatological hydric balance. The lack of solar radiation data is a limiting factor in the development of adequate models, with solar radiation being an important independent variable for these models (TRNKA *et al.*, 2019). Thornton, Running & White (1997) warned of gaps in global solar radiation monitoring, considering that many stations primarily cover temperature and weather data. The relationship between stations monitoring solar radiation and changes in temperature monitoring and forecasting reached 1:500 on a global scale at the time of the study. The use of models focused on the optimal use of water and energy resources can lead to productivity gains and cost reductions for agriculture, especially in countries where agriculture is intensive and water resources are scarce. These often presuppose the use of fossil fuels for pumping, further incurring environmental additional costs (AL MAMUM *et al.*, 2025).

Regarding electricity generation, the model proposed in this article seeks to provide more discrete estimates of solar radiation. This is because, based on variations in incident solar radiation, it is possible, for example, to estimate the electricity generation potential for a given location and compare it with local electricity consumption data. Furthermore, it is possible to verify the solutions and potential of a given project based on temperature data.

Kim, Oh & Jeong (2016) report that a consistent HGSR database can significantly contribute to improving the energy performance analysis of buildings, for example.

Such research is usually expensive, since it presumes the use of equipment, data acquisition or recorders, and often a professional qualified to operate the equipment. A more economical alternative is to use models to estimate GSR, as they can be built from historical data and allow estimates to be made with very satisfactory accuracy (AULER & MINUZZI, 2022).

Usually, empirical models for estimating GSR are built from temperature data only. Hargreave & Samani (1985) proposed a model that estimates GSR based on daily temperature data, for example. In general, the models presented in the literature were developed with the objective of making predictions of the daily GSR. Therefore, studies focused on HGSR are scarce. Determining HGSR attributes a higher degree of precision to several studies focused on precision agriculture, in addition to considering intrinsically local characteristics, such as topography, relief and climate on a local scale, which affects atmospheric composition.

In this sense, this work aims to develop a model for predicting HGSR from air temperature data, in a manner analogous to that proposed by Hargreaves and Samani (1985). The model was built using an Artificial Neural Network (ANN).

2. RELATED WORKS

Although the studies to predict HGSR are few, many researchers have already used artificial intelligence to predict GSR. Some applied combined AI methods and empirical models, others presented studies comparing them (GAIRAA *et al.*, 2016).

Shboul *et al.* (2021), used an ANN to predict HGSR and wind speed. The model was composed of two Feed-forward Back-propagation Artificial Neural Network (FBANN). The first received monthly, daily and hourly data inputs and generated the values used in the second model, which forecasts HGSR and wind speed.

Siva Krishna Rao *et al.* (2018) developed and compared the results of six ANN architectures using Back-propagation learning algorithm for predicting monthly-mean daily GSR. Gairaa *et al.* (2016) proposed a combined model approach using a linear autoregressive moving average (ARMA) model and a nonlinear Artificial Neural Network (ANN) model to estimate daily GSR. One of the main motivations for creating the hybrid

model was that ARMA produces more accurate results on sunny days, while ANN responds better on cloudy days, according to the literature.

Antonopoulos *et al.* (2019) developed a comparative study between the application of the Hargraeves method, technology and multi-linear regression methods (MLR) and ANN to predict GSR. These researchers obtained similar results for the methods applied and, therefore, they concluded that the complexity in implementing the ANN and the insignificant gain in accuracy do not justify its implementation, in their specific case.

Fan *et al.* (2017) conducted a study to predict daily global solar radiation using 20 empirical models. Ten simple models were used based on temperature data, including the model proposed by Hargreave & Samani (1985). The remaining (complex) models use parameters other than temperature data, such as relative humidity, precipitation, and vapor pressure deficit, because according to Quej *et al.* (2016), in humid regions, temperature variation correlates with these meteorological variables. The average values obtained by Fan *et al.* (2017) for R^2 in the simple models was 0.65 and in the complex models, 0.75.

Feng *et al.* (2019) used four empirical models to predict daily global solar radiation, and the results for R^2 ranged from 0.827 to 0.829. The authors also used artificial neural networks in the study. It is important to note that the studies did not address hourly global radiation, which demonstrates the contribution of this work. As reported in the initial section, research focused on forecasting hourly solar radiation is scarce. Although the unit is distinct, which implies a change in methodology and increases the complexity of the method, it is possible to conclude that the model proposed in this work is promising.

Studies using artificial neural networks, such as that by Feng *et al.* (2019), have also been developed for forecasting daily global solar radiation, reinforcing the finding that there is a methodological gap that this study can contribute to filling.

3. DATA PROCESSING AND NORMALIZATION

Before applying the model, it is necessary to make some corrections and adjustments, such as removing gaps in monitoring, sorting and removing data that is not relevant to the modeling (null cells, for example) and other anomalies that can interfere with the results, such as outliers. Alanazi *et al.* (2016) warns that if the variables have different units and scales, it is recommended that a data normalization process be employed. After normalization, the ANN input values present minimum and maximum values, usually a dimensionless scale from 0 to 1, which facilitates the application of the algorithm and

improves the accuracy of the model (ALANAZI *et al.*, 2016; SINGH *et al.*, 2015). Shi *et al.* (2021) emphasize that normalization tends to maintain the proportionality between the values of the data set, thereby increasing the accuracy of the model.

For data normalization, the methodology adopted was developed in a manner analogous to the Hargreaves-Samani model, Eq. (1), which uses daily data on temperature and solar radiation incident at the top of the atmosphere to estimate global daily radiation. However, the objective was to use the hourly data so that it is possible to estimate the hourly global radiation.

$$R_s = K_{rs} (T_{\max} - T_{\min})^2 R_a \quad (1)$$

where R_s is the incident solar radiation (MJ / (m² day)); K_{rs} the empirical coefficient (dimensionless); $T_{\max}$ is the maximum air temperature (°C); $T_{\min}$ the minimum air temperature (°C); R_a is the solar radiation incident at the top of the (MJ / (m² day)).

To develop the HGSR model some mathematical procedures must be performed. The daily temperature amplitude (A_{term}) and the hourly temperature difference (DT_h), Eq. (2), were calculated. Some tests were carried out with the aim of observing the relationship between these new variables and the hourly global solar radiation.

$$DT_h = T_{a_h} - T_{b_h} \quad (2)$$

where T_{a_h} is the hourly air temperature (°C); and T_{b_h} is the base temperature (°C), according to Eq. (3),

where,

$$\begin{aligned} \text{if, } 0 < H < H_{sr} &\rightarrow T_{b_h} = T_{a_h} \\ \text{if, } H_{ss} < H < 24 &\rightarrow T_{b_h} = T_{a_h} \\ \text{if, } H_{sr} < H < H_{ss} &\rightarrow T_{b_h} = T_{a_h} + \frac{(T_{a_{ss}} - T_{a_{sr}}) \times (H - H_{sr})}{(H_{ss} - H_{sr})} \end{aligned} \quad (3)$$

where, T_{b_h} is the base temperature (°C).

$T_{a_{ss}}$ is the air temperature at sunset (°C).

$T_{a_{sr}}$ is the air temperature at sunrise (°C).

The H is the measurement time.

According to the proposed model, T_{bh} will only to be different from the hourly air temperature when H is between sunrise time (H_{sr}) and the time of sunset (H_{ss}).

The A_{term} is defined as the difference between the maximum temperatures (T_{max}) and minimal (T_{min}) daily, as presented next.

$$A_{term} = T_{max} - T_{min} \quad (4)$$

As recommended by Alanazi *et al.* (2016), to define the normalized input variables, temperature data related to temperature amplitude (A_{term}), and the hourly temperature difference were used. For the radiation, the solar radiation incident at the top of the atmosphere at noon (Ra) was used. Furthermore, tests were carried out during the process and, empirically, the equations for the final procedure for normalizing the variables were defined. The objective of applying these equations is to facilitate the use of data by the ANN and improve the performance of the predictions. Therefore, for the normalization of radiation data, the following equations are used:

$$Ra_{norm} = \sqrt{Ra^{0,1}} \quad (5)$$

$$RT = \sqrt{\left(\frac{DT_h}{A_{term}}\right)^{0,4}} \quad (6)$$

$$HGSR_{norm} = \sqrt{HGSR^{0,4}} \quad (7)$$

where $HGSR_{norm}$ is the normalized hourly global radiation; $HGSR$ is the hourly global radiation; RT is the standardization of the relationship between thermal amplitude and air temperature; DT_h is the difference between air temperature (T_a) and base temperature (T_b); A_{term} is the temperature range, the difference between the daily maximum (T_{max}) and minimum (T_{min}) temperatures.

The independent variables adopted were Ra_{norm} and RT , and the dependent variable is the $HGSR_{norm}$ and, from these parameters, it is possible to construct an ANN and evaluate its ability to estimate the global hourly radiation.

4. MODEL EVALUATION METRICS

To evaluate the model's performance, three criteria were adopted: the Mean Squared Error (MSE) and the Mean Absolute Error (MAE) and the R^2 (R-squared), which are expressed by, respectively,

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (\text{best value} = 0; \text{worst value} = +\infty) \quad (8)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (\text{best value} = 0; \text{worst value} = +\infty) \quad (9)$$

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \quad (\text{worst value} = 0; \text{best value} = +1) \quad (10)$$

where $\hat{y}_i$ is the predicted output value; y_i is the target value; n is the number of observations and $\bar{y}$ is the mean of target values.

5. UTILIZED DATA

The data used were obtained from the database of the automatic station of the National Institute of Meteorology - INMET of Campos dos Goytacazes, Rio de Janeiro State, Brazil, located at coordinates 41.35°W Longitude and 21.71°S Latitude. The monitoring covers the period from 2011 to 2021 for the parameters Global Radiation (KJ/m²), hourly dry bulb air temperature (°C), dew point temperature (°C), maximum temperature in the previous hour (°C), minimum temperature in the previous hour (°C), maximum dew temperature in the previous hour (°C), minimum dew temperature in the previous hour (°C) and the time of measurement.

6. DEVELOPMENT OF THE ARTIFICIAL NEURAL NETWORK

The Artificial Neural Network (ANN) is a simplification of brain structure, where neurons are represented by nodes and synapses by weighted links. ANNs have, at least, two layers: one input layer and one output layer. In addition to these, the network can have intermediate layers, called hidden layers (HAYKIN, 2001). Also, according to Haykin (2001), the first layer receives the input data from the model and the number of neurons is linked to the number of independent variables or parameters adopted. The inner layers, or hidden layers, are where the data is processed. To define the number of neurons in the hidden layer, there are different approaches. Silva (2005) proposed a technique for defining the number of neurons in the hidden layer based on Kolmogorov's Theorem, which correlates the number of neurons in the hidden layer (N_{hidden}) with the number of inputs (N_{inputs}), as presented next.

$$N_{\text{hidden}} = 2 N_{\text{inputs}} \quad (11)$$

Haykin (2001) states that the output layer can has as many outputs as there are dependent variables in the model. On the other hand, the number of hidden layers can be defined randomly, depending on the topology adopted in the ANN and tests, basically consisting of trial and error, to adjust the model.

In the model proposed in this work, initially, both approaches were used. First, the methodology proposed by Silva (2005) was adopted to define the number of neurons in the hidden layers. Subsequently, configuration tests were carried out in a manner analogous to the approach suggested by Haykin (2001) for the definition of hidden layers. Finally, to define the optimal architecture, tests were performed with different parameter configurations and their respective performances were evaluated. The parameters used in the tests to define the final ANN structure were number of neurons in the hidden layer, different training dataset ratios, learning rates and training algorithms. To perform the comparative tests, fixed parameters were defined, while other parameters were changed in each version. The default configuration was 6 neurons in the hidden layers (N_{hidden}), 20% of samples were segregated for testing, a learning rate of 0.001, and the Adam algorithm was chosen as the optimization algorithm. Adam algorithm is a Stochastic Gradient Descent method proposed by Kingma & Ba (2014).

7. RESULTS

7.1 Defining the number of neurons in the hidden layers

To define the number of neurons in the hidden layers, a comparative analysis of different compositions was performed. The number of neurons used in the performance analysis varied from 6 to 14. As shown in Table 1, the performance of the ANN composed of 12 neurons in each hidden layer, evaluated through MSE, MAE and R^2 , presented the best results, except for MSE Testing and, therefore, it was used in the final ANN.

Dataset	Statistical parameter	Number of Neurons				
		6	8	10	12	14
Training	MSE	0.0102	0.0099	0.0099	0.0095	0.0096
Testing	MSE	0.0091	0.0098	0.0093	0.0092	0.0094
Training	MAE	0.0779	0.0765	0.0766	0.0749	0.0755
Testing	MAE	0.0716	0.0733	0.0718	0.0711	0.0718

Training	R^2	0.8016	0.8016	0.8016	0.7997	0.7997
Testing	R^2	0.7997	0.7997	0.7997	0.8016	0.8016

Table1. Performance with different numbers of neurons in the hidden layer.

7.2 Defining training dataset size

The performance evaluation of different dataset sizes was investigated and presented in Table 2. Test compositions of 20%, 30%, 40% and 50% were tested. The results for MSE Training datasets of 10%, 20% and 30% were similar (0.0095, 0.0094 and 0.0095, respectively). Training and testing datasets of 20% also presented slightly better results than the others. However, the Testing dataset of 30% presented the best result for R^2 (0.8016) and better performance in the other comparisons. Therefore, the percentage of the testing dataset adopted in the final ANN was composed of 30% of the data for training and 70% for testing.

Table 2. Performance with different training data set ratios.

Dataset	Statistical parameter	Percentage of testing datasets				
		10	20	30	40	50
Training	MSE	0.0095	0.0094	0.0095	0.0099	0.0096
Testing	MSE	0.0106	0.0106	0.0092	0.0092	0.0098
Training	MAE	0.0751	0.0744	0.0749	0.0761	0.0749
Testing	MAE	0.0752	0.0757	0.0711	0.0721	0.0739
Training	R^2	0.7996	0.8038	0.7997	0.7974	0.8032
Testing	R^2	0.8020	0.7949	0.8016	0.8004	0.7946

7.3 Definition of initial learning rate

The results obtained by the model did not show variations for R^2 (0.8016 and 0.7997 for training and testing, respectively). However, learning rates of 0.001 and 0.003 produced very close results and were slightly better than the other rates tested, as shown in Table 3. Therefore, the learning rate chosen to be used in the final ANN was 0.001.

Table 3. Performance with different initial learning rates.

Dataset	Statistical parameter	Initial learning rate				
		0.001	0.003	0.005	0.010	0.015
Training	MSE	0.0095	0.0096	0.0097	0.0096	0.0096
Testing	MSE	0.0092	0.0088	0.0088	0.0089	0.0092
Training	MAE	0.0749	0.0753	0.0756	0.0752	0.0753
Testing	MAE	0.0711	0.0710	0.0710	0.0707	0.0714
Training	R^2	0.8016	0.8016	0.8016	0.8016	0.8016
Testing	R^2	0.7997	0.7997	0.7997	0.7997	0.7997

7.4 Optimization algorithms

Four optimization algorithms were tested, including Adam, Stochastic Gradient Descent (SGD), Adagrad and Root Mean Square Propagation (RMSprop). According to Dogo *et al.* (2018), these algorithms are among the most widely used in deep learning nowadays. That's why they were chosen for performance testing. Among them, the one that performed best was Adam, which despite presenting R^2 equal to the other algorithms, returned better results in the error evaluation metrics MSE and MAE (Table 4).

Therefore, as an optimizer, Adam was applied. This is a gradient based stochastic optimization method, widely used in ANNs, in several areas of research, and is considered an update of other algorithms that use Gradient Descent (KINGMA & BA, 2014). Furthermore, if necessary, Adam updates the parameters of the learning model and its application aims to minimize the cost function in the machine learning process.

Table 4. Performance with different training algorithms.

Dataset	Statistical parameter	Optimization algorithm			
		Adam	SGD	Adagrad	RMSprop
Training	MSE	0.0095	0.0506	0.0511	0.0101
Testing	MSE	0.0092	0.0488	0.0488	0.0094
Training	MAE	0.0749	0.1796	0.1806	0.0773
Testing	MAE	0.0711	0.1719	0.1722	0.0720
Training	R^2	0.8016	0.8016	0.8016	0.8016
Testing	R^2	0.7997	0.7997	0.7997	0.7997

7.5 Optimal neural network

Finally, the architecture that showed the best results combined was a multilayer ANN or MLP (Multilayer Perceptron), densely connected, where every neuron is connected to the previous layer, composed of the inputs of the model, five hidden layers, composed of 12 neurons each and an output layer (Figure 1), learning rate 0.001, with a 70/30 ratio for training and testing data, and Adam as optimization algorithm.

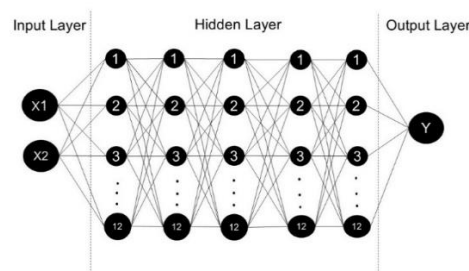


Figure 1. Structure of the developed ANN.

In the hidden layers, a hyperbolic tangent sigmoid activation function was used, with a linear activation function applied at the output layer. The sigmoid is a non-linear activation function, which allows the model to achieve greater generalization and adaptability. Its output is comprised of the interval from 0 to 1 (SHARMA, 2023). In Szandala (2021), it's highlighted that the linear activation function is the most basic, being employed in the output layer, since it does not alter the output of a neuron, it only transmits the result.

8. DISCUSSION

After normalization of RT and Ra, and the dependent variable (target variable) HGSR, besides splitting the data in two sets, one for training and the other for testing and validation as defined by the methodology for building an optimal model. The total sample size was 46,216 divided into a ratio of 70% for training and 30% for model testing as defined in the previous section. The training batches (batch size) were of the order of 32.

In addition, the Dropout data regularization technique was implemented, the purpose of which is to avoid overfitting, which consists of removing the weights, in some neurons, randomly, in a percentage that can be defined from 0 to 1 (BADOLA *et al.*, 2020). Some values were tested and the one that returned the best result was 0.2, that is, the weights were removed in 20% of the neurons in the hidden layer in which the dropout was applied.

As shown in Figure 2, the model was used with 50 epochs, as it presented stability in a relatively short period, denoting the good fit between the predictions from training to the test for the MSE and MAE performance evaluation metrics. The cost function presented very satisfactory results, being 0.0092 for the MSE, while the MAE was 0.0711 for testing error.

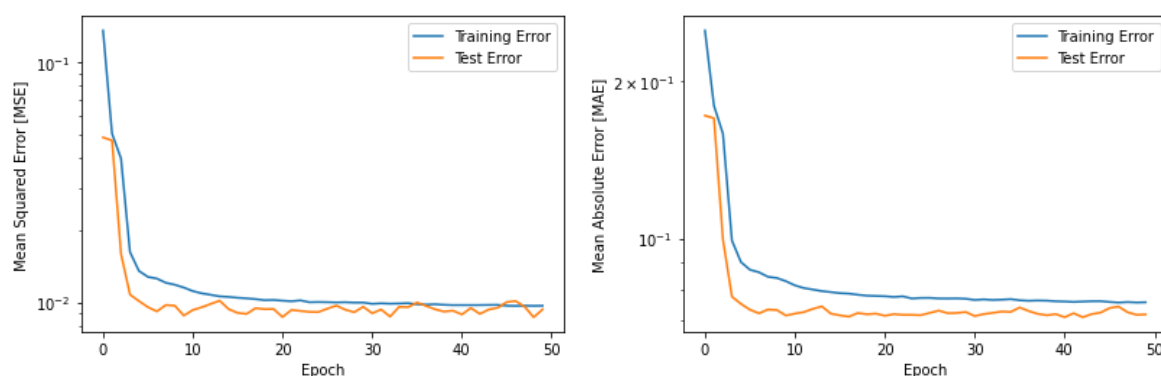


Figure 2. Graph of the MSE and MAE against training and testing epochs.

The training presented good alignment with the test data, indicating that the partitioning of the lots was adequate and that the algorithm was successful in the predictions in both procedures.

The mean for the MSE testing was 0.0101, while the standard deviation was 0.053 and 0.0188 for the MAE Testing. Table 5 shows the statistical data related to the error of the forecasts both for training and testing.

Table 5. MSE and MAE obtained with the ANN.

Statistics	MSE Training	MSE Testing	MAE Training	MAE Testing
Mean	0.0119	0.0101	0.0824	0.0745
Standard deviation	0.0078	0.0053	0.0188	0.0138
Minimal	0.0097	0.0086	0.0758	0.0703
25%	0.0098	0.0090	0.0764	0.0714
50%	0.0101	0.0093	0.0775	0.0719
75%	0.0107	0.0095	0.0801	0.0731
Maximum	0.0603	0.0468	0.1922	0.1695

In addition to the metrics applied to evaluate the ANN performance, a graphical comparison was made between the first 150 HGSR measurements obtained at the stationary stations and the predictions made by the model. The samples used were from the daytime period, which is why they presented values above zero. In Figure 3, it is possible to infer that the predictions follow the trend of the historical data, corroborating the results obtained by the model performance evaluation metrics. In addition, the statistics of the results obtained presented in Table 6 confirm that the model prediction returned data very close to the historical measurements.

Considering that the data were normalized for modeling, as described in Section 4, and that the values for the output layer, after standardization, varied between 0.1584 and 1.3403, the average difference between the absolute values of the differences between the real values and those obtained by the model, MAE, was 0.0780. Therefore, confirming the accuracy of the model.

However, MAE is less sensitive to outliers. On the other hand, MSE is a metric used to calculate the mean of the squares of the differences between the values predicted in the modeling and the values originally observed and it has the ability to amplify errors (CHAO *et al.*, 2022). Therefore, the two metrics complement each other in the analysis. The predictions made by the ANN were adequate to the level of accuracy required, since the lower the MAE and MSE, the more accurate the model's predictions are CHAO *et al.* (2022).

In any case, the suggested model must be improved and new studies must be carried out with the aim of minimizing errors, increasing reliability and expanding the applicability of

results in other regions, in accordance with the proposal to fill the knowledge gap regarding data availability from HSRG.

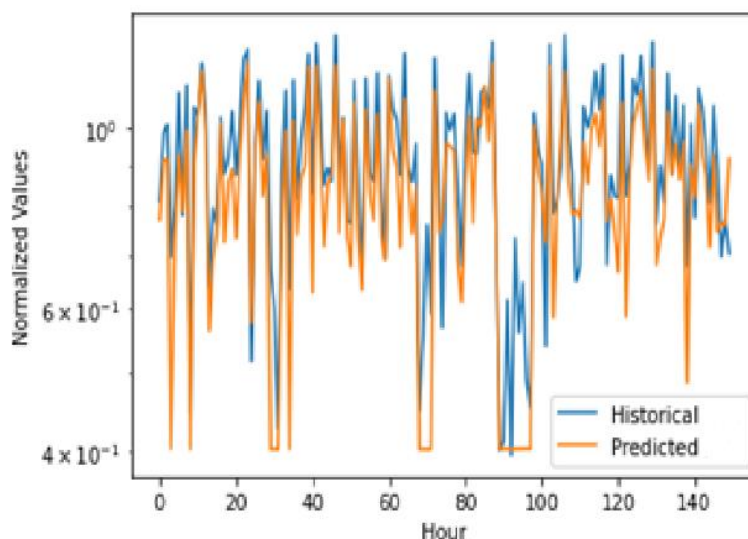


Figure 3. Historical and predicted HSGR by the model for 150 initial samples.

As shown in Table 6, with the exception of the value generated for the minimum data predicted by the ANN, the statistical analysis of the results reveals that the model performs satisfactorily, which was also corroborated by the performance evaluation metrics MSE, MAE and R^2 .

The results obtained for the ANN output layer (target variable) were analyzed by means of linear regression, with the aim of verifying the fit of the regression line obtained by the model and the real output data. Linear regression was therefore applied to the total samples for Rad (Figure 4) and the results confirmed the good relationship between the regression line established by the model and the data, with R^2 being 0.8016 for training and 0.7997 for testing.

Table 6. Statistics of model results and real data.

Statistics	Historical HSGR	Predicted HSGR
Mean	1.0099	1.0178
Standard deviation	0.2238	0.2006
Minimal	0.1584	0.3998
25%	0.8821	0.9182
50%	1.0642	1.0885
75%	1.1838	1.1761
Maximum	1.3403	1.2353

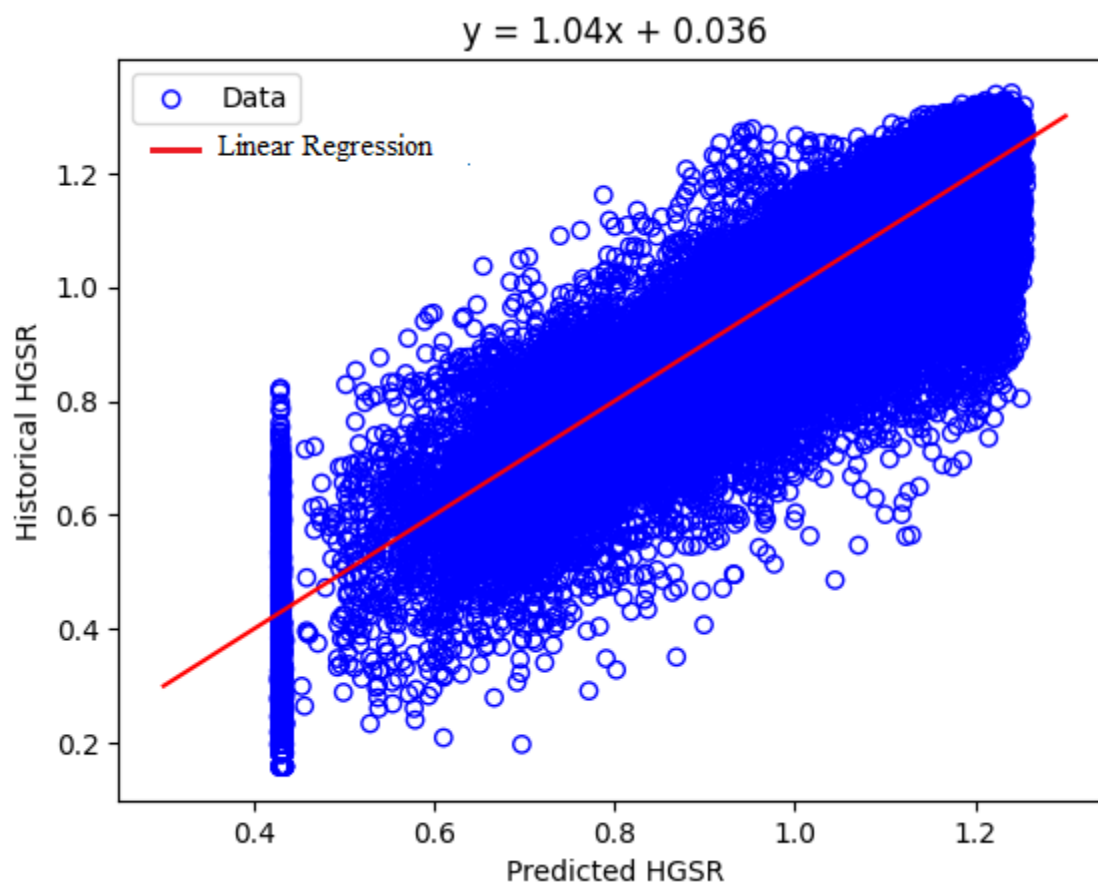


Figure 4. Linear regression for observed HGSR and model predicted HGSR.

9. CONCLUSIONS

Based on the results obtained, it is possible to infer that a densely connected MLP-type ANN is able to predict hourly solar radiation. The contribution may be important, since it would be possible to fill gaps in incomplete historical series or to implement the model in places where there is no history of measurement of global hourly radiation.

Despite the satisfactory results obtained in this study, it is important to emphasize that the implemented algorithm is relatively simple and can be adjusted and improved by adding techniques that allow an increase in performance and precision. Whether by reducing oscillations in forecast errors or by adjusting synaptic weights, back-propagation techniques, among others. In addition, other ANN architectures associated with different performance evaluation methods can be employed. Therefore, further studies should be carried out to improve the model.

Furthermore, it is important to consider that the model was developed using data from a single monitoring station and, consequently, implicitly considers local geographic and climatological aspects. Therefore, for the model to be applied more broadly, adjustments

and expansion of the database from different monitoring stations with diverse geographic and climatological features should be considered.

The next step in this work, therefore, is to use data from different locations and make the necessary adjustments to build a generalist version when it can be made available and shared in a public repository.

ACKNOWLEDGMENTS

The authors acknowledge the financial support provided by FAPERJ, Carlos Chagas Filho Foundation for Research Support of the Rio de Janeiro State, CNPq, National Council for Scientific and Technological Development, and CAPES, Coordination for the Improvement of Higher Education Personnel (Funding Code 001). Acknowledgments are also due to the Evosolar company for the interest in the work that is being conducted.

REFERENCES

- ALANAZI, M.; ALANAZI, A.; KHODAEI, A. **Long-term solar generation forecasting**. In: IEEE/PES Transmission and Distribution Conference and Exposition (T&D). 2016. Available at: <<https://ieeexplore.ieee.org/document/7519883>>. Accessed on: September 24, 2025.
- AI MAMUN, M. R., AHMED, A. K.; POMA, S. M. U.; HAQUE, M. M.; ASHIK-E-RABBANI, M. IoT-enabled solar-powered smart irrigation for precision agriculture. **Smart Agricultural Technology**, v. 10, p. 1-11, 2025.
- ANTONOPOULOS, V. Z.; PAPAMICHAIL, D. M.; ASCHONITIS, V. G.; ANTONOPOULOS, A. V. Solar radiation estimation methods using ANN and empirical models. **Computers and Electronics in Agriculture**, v. 160, p. 160–167, 2019.
- ASSIS, F.N., MENDEZ, M.E. Relationship between photosynthetically active radiation and global radiation. **Pesquisa Agropecuária Brasileira**, Brasília, v. 24, n.7, p. 797-800, 1989.
- AULER, F. R., MINUZZI, R. B. Estimation of daily global solar radiation based on air temperature for different cloudiness conditions in Florianópolis, Santa Catarina. **Journal of Environmental Analysis and Progress**, v. 7, n. 1, p. 026–032, 2022.
- BADOLA, A., NAIR, V. P., LAL, R. P. **An Analysis of Regularization Methods in Deep Neural Networks**. In: IEEE 17th India Council International Conference (INDICON). 2020. Available at: <<https://ieeexplore.ieee.org/document/9342192>>. Accessed on: September 24, 2025.
- CHAO, H.; ZHANG, J.; YAN, P. **Regression metric loss: Learning a semantic representation space for medical images**. In: 25th International Conference on Medical Image Computing and Computer Assisted Intervention, MICCAI. 2022. Available at: <<https://arxiv.org/abs/2207.05231>>. Accessed on: September 24, 2025.

DOGO E. M.; AFOLABI O. J.; NWULU N. I.; TWALA B.; AIGBAVBOA C.O. **A Comparative Analysis of Gradient Descent-Based Optimization Algorithms on Convolutional Neural Networks**. In: International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS). Available at: <<https://ieeexplore.ieee.org/document/8769211>>. Accessed on: September 24, 2025.

FAN, J.; CHEN, B; WU, L.; ZHANG, F.; LU, X.; XIANG, Y. Evaluation and development of temperature-based empirical models for estimating daily global solar radiation in humid regions. **Energy**, v. 144, p. 903-914, 2017.

FENG, Y.; GONG, D.; ZHANG, Q.; JIANG, S.; ZHAO, L.; CUI, N. Evaluation of temperature-based machine learning and empirical models for predicting daily global solar radiation. **Energy Conversion and Management**, v. 198, n. 11780, 2019.

GAIRAA, K.; KHELLAF, A.; MESSLEM, Y.; CHELLALI, F. Estimation of the daily global solar radiation based on Box–Jenkins and ANN models: A combined approach. **Renewable and Sustainable Energy Reviews**, v. 57, p. 238–249, 2016.

HAYKIN, S. **Neural Networks - Principles and practices**. 2. ed. Porto Alegre: Bookman, 908 p., 2001.

HARARI, J. **Noções de Oceanografia**. Universidade de São Paulo – USP. Instituto Oceanográfico. 2021. Available at: <www.livrosabertos.abcd.usp.br/portadelivrosUSP/catalog/book/1086>. Accessed on: September 24, 2025.

HARGREAVES, G. H.; SAMANI, Z. A. Reference crop evapotranspiration from temperature. **Journal of Applied Engineering in Agriculture**, v. 1, n. 2, p. 96-99, 1985.

KIM, K.H.; OH, J.K.W.; JEONG, W. Study on Solar Radiation Models in South Korea for Improving Office Building Energy Performance Analysis. **Sustainability**, v. 8, n. 589, 2016.

KINGMA, D.P.; BA, J. **Adam: A Method for Stochastic Optimization**. In: 3rd International Conference for Learning Representations. 2015. Available at: <<https://arxiv.org/abs/1412.6980>>. Accessed on: September 24, 2025.

QUEJ V.H.; ALMOROX J.; IBRAKHIMOV M.; SAITO L. Empirical models for estimating daily global solar radiation in Yucatán Peninsula, Mexico. **Energy Conversion and Management**, v. 110, p. 448-56, 2016.

RIBEIRO, R. M. **Agroenergia na mitigação das mudanças globais, na segurança energética e na promoção social**. Dissertation (Master's Degree) – Universidade Federal de Viçosa, 2011. Available at: <<https://locus.ufv.br/server/api/core/bitstreams/6dce1b9f-fe62-4e23-bf27-56fcd281bf4f/content>>. Accessed on: September 24, 2025.

SHARMA, S. **Activation Functions in Neural Networks**. 2017. Available at: <https://towardsdatascience.com/activation-functions-neural-networks-1cbd9f8d91d6>. Accessed on: August 26, 2023.

SHBOUL, B.; AL-ARFI, I.; MICHAÏLOS, S.; INGHAM, D.; MA, L.; HUGHES, K. J.; POURKASHANIAN, M. A new ANN model for hourly solar radiation and wind speed prediction: A case study over the north & south of the Arabian Peninsula. **Sustainable Energy Technologies and Assessments**, v. 46, n. 101248, 2021.

SHI, J.; LEE, W.; LIU, Y.; YANG, Y.; WANG, P. **Forecasting power output of photovoltaic system based on weather classification and support vector machine**. In: IEEE Industry Applications Society Annual Meeting. 2011. Available at: <<https://ieeexplore.ieee.org/document/6168891>>. Accessed on: September 24, 2025.

SILVA, R. M. **Redes Neurais Artificiais aplicadas à Detecção de Intrusão em Redes TCP/IP**. 2005. Dissertation (Master's Degree) – Pontifícia Universidade Católica do Rio de Janeiro, 2005. Available at: <<https://www.maxwell.vrac.puc-rio.br/colecao.php?strSecao=resultado&nrSeq=7335&idi=1>>. Accessed on: September 24, 2025.

SINGH, V. P.; RAVINDRA, B.; VIJAY, V.; BHATT, M. S. **Forecasting of 5MW solar photovoltaic power plant generation using generalized neural network**. In: 39th National Systems Conference (NSC). 2015. Available at: <<https://ieeexplore.ieee.org/document/7489107>>. Accessed on: September 24, 2025.

SIVA KRISHNA RAO, K. D. V.; PREMALATHA, M.; NAVEEN, C. Analysis of different combinations of meteorological parameters in predicting the horizontal global solar radiation with ANN approach: A case study. **Renewable and Sustainable Energy Reviews**, v. 91, p. 248–258, 2018.

SZANDAŁA, T.; **Review and Comparison of Commonly Used Activation Functions for Deep Neural Networks**. In: Bio-inspired Neurocomputing. Studies in Computational Intelligence. 2021. Available at: <<https://arxiv.org/abs/2010.09458>>. Accessed on: September 24, 2025.

THORNTON, P.E.; RUNNING, S.W.; WHITE, M.A. Generating surfaces of daily meteorological variables over large regions of complex terrain. **Journal of Hydrology**, v. 190, p. 214-251, 2007.

TRNKA, M.; EITZINGER J.; KAPLER, P.; DUBROVSKÝ, M.; SEMERÁDOVÁ, D.; ŽALUD, Z.; FORMAYER, H. Effect of Estimated Daily Global Solar Radiation Data on the Results of Crop Growth Models. **Sensors**, v. 7, n. 10, p. 2330-2362, 2007.