Predictive modeling based on Deep Machine Learning models coupled to Discrete Wavelet Transform applied to short-term solar radiation forecasting

Lucas O. M. da Silva[1], Geovane D. da Silva[1], Elber C. A. C. Lopes[1], Gabrielle M. da Silva[1], Frede O. Carvalho[1], Hendersson S. Wanderley[2]

[1]*Technology Center, Federal University of Alagoas*
Av. Lourival Melo Mota, S/N, 57072-900, Maceió/Alagoas, Brazil
lucas.mendes149@gmail.com, geovanedomi@gmail.com, elbercaio@gmail.com, gabriellemelo439@gmail.com, fredecarvalho@yahoo.com.br
[2]*Department of Environmental Sciences, Rural Federal University of Rio de Janeiro*
BR-465, Km 7, 23.897-000, Seropédica/Rio de Janeiro, Brazil
henderson@ufrrj.br

Abstract. Forecasting solar radiation is important for analyzing the feasibility of power generation and optimizing the operation. In this context, approaches using simple Machine Learning models, such as Artificial Neural Network (ANN) and Support Vector Machine (SVM) were made, but could not satisfy the performance requirements in complex scenarios. Therefore, it was necessary to use hybrid models with the Discrete Wavelet Transform (DWT). This paper intended to implement solar radiation forecasting using ANN, SVM, Long-Short-Term-Memory Network (LSTM) and their hybrids coupled with DWT as a preprocessing technique using Python. The results obtained were compared to the Autoregressive Integrated Moving Average model (ARIMA). The data of solar radiation was collected by the automatic weather Salinas station of Nova Friburgo (RJ), the Holdout validation was used to evaluate the prognosis performances using Root-Mean-Square error (RMSE) and Coefficient of Determination (R^2). The results revealed that the simple models had worse results than ARIMA, and that hybrid models had superior performance, especially the DWT-LSTM which had an R^2 of 0.994 and an RMSE of 0.516. In addition, Python showed to be powerful as an Open Source tool for implementation of robust models that are useful for applications in science and engineering.

Keywords: Solar radiation forecasting, Machine Learning, Discrete Wavelet Transform.

1 Introduction

In the Brazilian energy industry, it is common that the levels of hydroelectric energy drop during drought periods, making energy rationing necessary and increasing costs. In order to attenuate this problem, the development of other energy sources has been the target of studies, such as solar energy. According to Akarslan and Hocaoglu [1], the solar energy production is dependent on daily radiation levels and it is influenced by different meteorological factors, which make the solar radiation forecasting a relevant tool to analyze not only the viability of energy generation, but also to optimize the operation.

The study of solar radiation data is performed by an univariate time series analysis and machine learning techniques has been widely used for this kind of analysis, such as Artificial Neural Networks (ANN) and Support Vector Machines (SVM), but the time series for solar radiation has such complexity in its non-stationary behavior where there are multiple variations in frequencies during its entire time domain that it needs to be processed and simplified, therefore, the models are often combined with preprocessing techniques, such as Discrete Wavelet Transform (TWD). (Shamshirband et al. [2]). Also, recurrent neural networks are widely used for this analysis, but these are subject to some problems, especially in cases where it is necessary to retain information in the long term (Husein and Chung [3]). In order to avoid these problems, Long-Short-Term-Memory (LSTM) networks were proposed, introducing the concept of a memory cell, capable of remembering long-term information

(Sherstinsky [4]). Wang et al. [5] proposed a DL (Deep Learning) model based on wavelet decomposition (WD) and Long Short-Term Memory (LSTM) for day-ahead solar irradiance forecasting in Python Language. The raw solar irradiance sequence was decomposed into several subsequences via discrete wavelet transformation. Since the extracted features of each subsequence are also time series data, they are individually transported to LSTM to construct the subsequence forecasting model.

Therefore, this paper intended to develop and evaluate a Python implementation for short-term solar radiation forecasting using ANN, SVM and LSTM, as well as their hybrids coupled to the DWT as a preprocessing technique using data collected from 2013 to 2014 by the automatic weather Salinas station of Nova Friburgo (RJ), provided by the Department of Environmental Sciences of Rural Federal University of Rio de Janeiro.

2 Materials and methods

2.1 Artificial Neural Networks (ANN) and Long-Short-Term-Memory Networks (LSTM)

Artificial Neural Networks (ANN) are learning machines based on the human brain, they are formed by a complex network of computational cells, called artificial neurons, interconnected in the shape of layers. Each artificial neuron has an activation function that calculates the intensity of the weighted inputs and generates the output to be passed to the next neurons. There are several activation functions, including Rectified Linear Unit (ReLu), hyperbolic tangent (tanh) and logistic, which were used in this paper. By the learning process, the ANNs can create a relation between the input and output data, establishing weights and biases, this is usually done by gradient descendent methods. This allows the ANN to assimilate the behavior and patterns of the data and also forecasting and classifying data (Husein and Chung [3]; Haykin [6]; Matsumoto [7]).

There are many architectures for ANNs depending on the type of application. For time series analysis, it is common to use recurrent neural networks. In this type of ANN, the outputs are feedback on the network in the form of a hidden state, this feature is fundamental to give to the network the capacity of memorization of previous data and establishing relations between the current and past data (Wang et al. [5]).

The recurrent neural networks are subject to some problems, such as a possibility of no memory persistence, mainly in cases where it is necessary to retain information in long-short. Furthermore, the vanishing gradient problem is also a common problem and it happens in the learning stage, caused by activation function saturation, making the gradient extremely small and becoming the network training slow and tricky (Wang et al. [5]; Husein and Chung [3])

In order to avoid those problems, Hochreiter and Schmidhuber [8] proposed a new type of recurrent neural networks, the Long-Short-Term-Memory (LSTM), introducing the concept of the memory cell, capable of remembering long-term information. All data manipulation is doing by a set of gates: the forgotten, input and output gates, the first two actuates together to decide whether and how the data must be remembered or forgotten in the memory cell and the last combines the information of input with the memory cell to provide the output and pass this on the hidden state. The cell can learn at arbitraries time intervals, being able to act in time series with long intervals and in with delayed data (Husein and Chung [3]; Liang et al. [9])

2.2 Support Vector Machines (SVM)

SVM are supervised machine learning algorithms developed by Vapnik, based on the theory of statistical learning and in mathematical optimization techniques that can be used either for classification or regression problems. The concept elucidated by Liu and Lu [10] is that the data can be mapped nonlinearly in a dimensional space bigger than the original dimensions in with it is possible to solve a linear regression problem and use it either to infer what is needed or to separate the data in classes.

Machine Learning techniques are often used in complex non-linear processes, with is the case of short-term solar radiation forecasting, a regression problem, therefore, there is a need to increase the dimension of the data to make it possible to apply linear process to the data, this is accomplished with a nonlinear mapping function called kernel function, which is normally chosen between the linear, sigmoid and Radial Basis Function (RBF). With the data in a dimensional space bigger than the original it is possible to use a mathematical object labeled hyperplane, a subspace with one less dimension than the dataset, in the regression function. Saraiva [11] explains that the hyperplane is based in weight vectors, where the nonzero ones correspond to the support vectors that estimate the hyperplane, the weight vector are calculated by the minimization of a function, called risk function, where this minimization is done using the Lagrange multipliers. The machine learning aspect of the technique relies on finding the optimal hyperplane to be used in the regression function with Huang and Wang [12] explains that this relies on choosing the optimal input feature subset for SVM, and how to set the best kernel parameters.

2.3 Discrete Wavelet Transform (DWT)

As explained by Adamowski and Chan [13], the wavelet transform is a mathematical technique used to process signals or time series, decomposing them through finite functions generated by dilations and translations of a function called mother wavelet, and by doing so, it is possible to acquire the characteristics of the non-stationary time series.

The Discrete Wavelet Transform acts by separating the frequency signals components by magnitude, into stable parts (low-frequency signals) and fluctuant parts (high-frequency signals), decomposing the signal respectively in approximation and detail coefficients as proposed by Wang [5]. With each new decomposition, only the approximation is used, acting like a filter to the frequencies, shaping the data to have a better behavior to be used in the forecasting model, as such, it however it is important to note that the number of decompositions used is not random, different authors established different methodologies to attain the level of decomposition represented by L, the one adopted in this work was from Nourani [14] as seen in eq. (1) where N is the number of data points within the time series. The number of decompositions used in this paper was 2.

$$L = int[log \;\; N] \tag{1}$$

2.4 Baseline and Hybrid models

The baseline model used was the autoregressive integrated moving average (ARIMA), one of the models utilized extensively in time series analysis. The model is composed of three parts: autoregressive (AR), integrated (I), and moving average (MA). Matsumoto [7] explains that the (AR) part establishes the dependency relation between the variable of interest in the current state and its observations in previous instants of time. The (I) part makes the series stationary by mediating the difference in observations at different time points. Finally, the (MA) part takes into account the dependence between the data and the regression errors. However, this approach brings some problems, as Namini et. al [15] explains, one of the problems is the difficulty of dealing with non-linear relationships between variables.

As explained before, in the context of time series forecasting, machine learning models are widely utilized. However, these models may not fully satisfy situations where there are complex fluctuations and it is common to use data preprocessing techniques, such as DWT. This combination of methods of pre-processing data and obtaining better structures for neural networks have been widely used for the prognosis of time series, giving rise to those that are called hybrid systems. In this context, Python language has been standing out in the machine learning field, with powerful libraries that can satisfy the requirements for the implementation of such systems.

2.5 Evaluating models and Hyperparameters Selection

The standardization of data was applied in order to avoid the generation of weights with very different magnitudes in the SVM, ANN and LSTM models, increasing the efficiency of the training process. In order to evaluate the prognosis performances, Hold-out validation was used, with 70% of data for training and 30% for testing, using two different statistical metrics: Root-Mean-Square error (RMSE) and Coefficient of Determination (R^2).

In the time series analysis, it is necessary to define the numbers of lags, or time delays, to be considered in the models. This was done using the autocorrelation function (afc) (Flores et al. [16]) and the best lag value for the data studied was 3, once there is no significant autocorrelation for lags greater than 3.

Several model architectures were tested in order to obtain the best hyperparameters for each one and Table 1 summarizes this hyperparameters selection. The other parameters were used in their default values of the libraries Scikit-learn, Keras, Statsmodels, and PyWavelets. Several types of mother wavelet were tested, and it was observed that this had no significant impact in the performance of hybrid models, so it was chosen the Daubechies 1 (db1) which was marginally better than the other options.

Table 1. Hyperparameter selection for all models

Model	Hyperparameter	Best value	Search range
ARIMA	Difference Order	0	[0, 1, 2]
	Moving average order	1	[0, 1, 2, 3]
SVM	Kernel function	RBF	[RBF, linear, sigmoid]
	Regularization parameter (C)	1.0	[0.1, 1.0, 100, 1000]
ANN	Activation function	ReLU	[ReLU, tanh, logistic]
	Hidden neurons	10	[5, 10, 15, 20, 30]
LSTM	Optimizer	adam	[adam, sgd, rmsprop]
	Epochs	200	[150, 200, 300, 400]
	Hidden units	4	[1, 2, 4, 6, 8, 10]

3 Results and discussions

Table 2 summarizes the forecasting results of all models used and the metrics were computed on the data original scale. The hybrid models had a batter performance then the models without DWT and the baseline model, results expected based on the literature (Husein and Chung [3]; Wang et al. [5]). Also, the results showed the limitations of non-hybrid models in partial scenarios with complex fluctuations such as the solar radiation forecasting. The fig. 1 shows the comparison between results obtained without and with the DWT for the case of LSTM.

Table 2. Forecasting performance evaluation of the proposed models

Model	**R^2**	**RMSE**
ARIMA	0.401	5.154
SVM	0.375	5.275
ANN	0.374	5.287
LSTM	0.391	5.217
DWT-SVM	0.924	1.846
DWT-ANN	0.974	1.080
DWT-LSTM	**0.994**	**0.516**

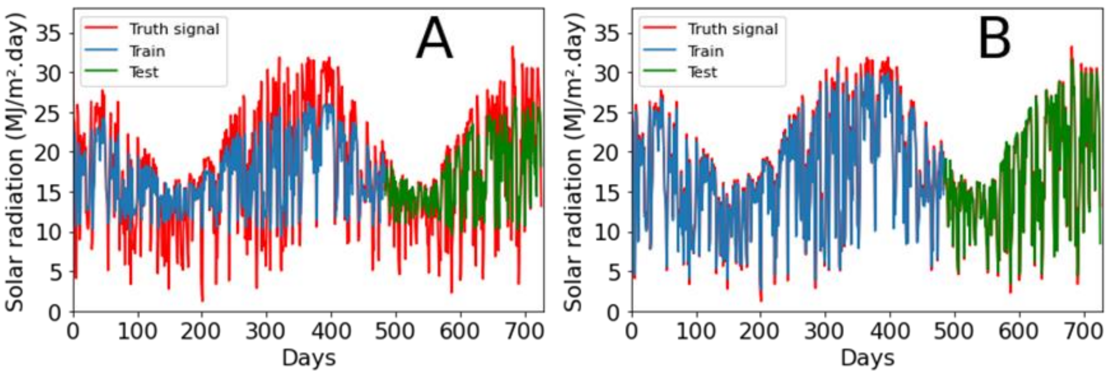

Figure 1. LSTM (A) and DWT-LSTM (B) forecasting performances

4 Conclusions

It was possible to observe the impact of DWT on the performance of all Machine and Deep Learning models in the prediction of short-term solar radiation, which demonstrates the potential of these hybrid models for problems like this one. Furthermore, the DWT-LSTM model showed the best performance as illustrated by R^2 of

0.994 and RMSE of 0.516 MJ/(m².day). In a future work, we intend to validate this model with a larger database, making the prediction even more reliable.

To conclude, Python showed as a powerful programming and Open Source tool for implementation of these robust models that could be useful for among other applications in science and engineering in general. This programming tool also has several cloud platforms, which could assist both in the development of the model and to put it into production, enabling applications with a lower cost than other commercial softwares.

Authorship statement. The authors hereby confirm that they are the sole liable persons responsible for the authorship of this work, and that all material that has been herein included as part of the present paper is either the property (and authorship) of the authors, or has the permission of the owners to be included here.

References

[1] E. Akarslan and F. Hocaoglu. O. *A novel adaptive approach for hourly solar radiation forecasting*. Renewable Energy, v. 87, p.628-633, 2016.
[2] S. Shamshirband et al. *Estimating the diffuse solar radiation using a coupled support vector machine-wavelet transform model*. Renewable and Sustainable Energy Reviews, v. 56, p.428-435, 2016.
[3] M. Husein and I. Chung. *Day-Ahead Solar Irradiance Forecasting for Microgrids Using a Long Short-Term Memory Recurrent Neural Network: A Deep Learning Approach.* Energies, v. 12, p. 1856, 2019
[4] A. Sherstinsky. *Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network*. Physica D: Nonlinear Phenomena, v. 404, p. 132306, 2020.
[5] F. Wang et al. Wavelet decomposition and Convolutional LSTM Networks Based Improved Deep Learning Model for Solar Irradiance Forecasting. *Appl. Sci.* 2018, *8*, 1286; doi:10.3390/app8081286.
[6] S. Haykin. *Neural Networks and Learning Machines*. 3ª ed. Pearson Prentice Hall, 2009.
[7] D. K. F. Matsumoto. *Estudo em séries temporais financeiras utilizando redes neurais recorrentes.* Universidade Federal de Alagoas, Maceió, 2019. http://200.17.114.109/handle/riufal/6813
[8] S. Hochreiter and J. Schmidhuber. *Long Short-Term Memory. Neural Computation*, 9(8), 1735–1780, 1997. https://doi.org/10.1162/neco.1997.9.8.1735
[9] X. Liang et al. *LSTM with wavelet transform based data preprocessing for stock price prediction*. Mathematical Problems in Engineering, v. 2019, 2019. https://doi.org/10.1155/2019/1340174
[10] M. Liu and J. Lu. *Support vector machine: an alternative to artificial neuron network for water quality forecasting in an agricultural nonpoint source polluted river*. Environ Sci Pollut Res 21:11036–11053, DOI 10.1007/s11356-014-3046-x, 2014.
[11] S. V. Saraiva. *Estudo de um controlador preditivo baseado em sistemas inteligentes*. Final Paper. Universidade Federal de Alagoas, Maceió, 2017.
[12] C.L. Huang and C.J. Wang. *A GA-based feature selection and parameters optimization for support vector machines*. Expert Syst Appl 31: 231–240, 2006.
[13] J. Adamowski, and H. F. Chan. *A wavelet neural network conjunction model for groundwater level forecasting*. Journal of Hydrology 407, p. 28 – 40, 2011
[14] V. Nourani et al. O. *Applications of hybrid wavelet–Artificial Intelligence models in hydrology: A review*. Journal of Hydrology, p. 358-377, 2014.
[15] S. Siami-Namini et. al. *A Comparison of ARIMA and LSTM in Forecasting Time Series. Proceedings* - 17th IEEE International Conference on Machine Learning and Applications, ICMLA 2018, 1394–1401, 2019. https://doi.org/10.1109/ICMLA.2018.00227
[16] J. H. F. Flores et al. *Autocorrelation and partial autocorrelation functions to improve neural networks models on univariate time series forecasting*. World Congress on Computational Intelligence, Brisbane, Australia, 2012.