



ISSN ONLINE: 2447-0228



PHOTOVOLTAIC GENERATION PREDICTION USING THE DEEP LEARNING LONG SHORT TERM MEMORY MODEL

Reinier Herrera Casanova*¹ and Lesyani León Viltres²

^{1,2} Central University "Marta Abreu" of Las Villas. Santa Clara, Cuba.

¹ <http://orcid.org/0000-0002-3303-2063> , ² <http://orcid.org/0000-0003-1822-2480> 

Email: *herrercasanovareinier@gmail.com, lesyani.leon@gmail.com

ARTICLE INFO

Article History

Received: December 15th, 2021

Accepted: February 21th, 2022

Published: February 28th, 2022

Keywords:

Photovoltaic system,
PV power prediction,
LSTM,
Deep learning.

ABSTRACT

Solar photovoltaic energy is a renewable, clean and safe energy source that is currently used worldwide. However, it presents a dynamic and intermittent behavior, caused by the variation of climatic conditions. Due to this, it has been necessary to develop different methods for the prediction of the energy generated in photovoltaic systems. The present work is focused on analyzing the prediction of the power generated in a photovoltaic plant connected to the grid, by means of the Long Short-Term Memory (LSTM) deep learning model. In order to carry out the study, a database obtained from the photovoltaic plant of the Central University "Marta Abreu" of Las Villas (UCLV) with a nominal installed power of 1.1 MW is used. Initially, the correlation between the different variables with respect to the photovoltaic power generated is analyzed, then the LSTM model is implemented to make the prediction. The results obtained show that the predictions made for different time horizons and for days with different behavior are adequate, which demonstrates the effectiveness of this prediction method.



Copyright ©2022 by authors and Galileo Institute of Technology and Education of the Amazon (ITEGAM). This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

I. INTRODUCTION

Currently, the planet shows an increase in energy needs driven by the technological and economic development of society. One of humanity's major problems is its dependence on fossil fuels, as they cause a strong environmental impact, in addition to various changes in the economic sphere. The challenge is to ensure that renewable energy sources gradually replace traditional fossil fuels. The main advantage of renewable energies is their lower environmental impact, as they reduce the number of pollutants in the atmosphere, as well as their less concentrated territorial distribution. They are continuous and inexhaustible energy sources, making them the alternative of the future [1].

One of the most widely used renewable sources today is solar energy. Photovoltaic (PV) systems connected to electrical grids are one of the distributed generation technologies with the greatest impact and growth in recent years. In fact, the world's annual solar PV capacity has increased exponentially over the last ten years. It is now the cheapest type of energy in a large number of countries due to substantially lower production costs of PV modules [1].

Atmospheric variables, such as solar irradiance, temperature, humidity and cloud properties, can directly and indirectly influence PV power generation. The dependence of the electrical energy generated in a photovoltaic farm on weather conditions, and the high variability of these conditions, make the problem of predicting the energy generated in a photovoltaic farm a complex task [2].

The future time period for PV generation prediction, or the duration between the actual time and the effective time of prediction is the forecast horizon [3]. Some researchers propose three categories for the forecast horizon: short term (up to 24 hours), medium term (1 day-1month), and long term (1month-1year). Others have added a fourth category based on the requirements of the decision-making process for smart grids or microgrids, aptly named very short-term or ultra-short-term (less than 1 hour) forecast horizon. However, so far there is no universally agreed classification criterion [4].

The large-scale penetration of PV in today's power systems requires forecasting models to operate the power grid economically and reliably [5]. Accurate solar forecasting eliminates the impact of uncertainty of solar PV power production, improves system

stability, increases the penetration level of the PV system, and reduces the maintenance cost of auxiliary devices. In addition, it is a powerful tool that helps power system operators and designers to model and manage solar PV plants efficiently [6].

Since the very emergence of photovoltaic plants, several techniques have been developed to achieve predictive models that contribute to the improvement of the management of these plants. Currently, intensive work is being done in the application of artificial intelligence tools for the development of these types of models, based on the proven capacity of these techniques in the handling of information contained in large volumes of data obtained from the systems under study.

The photovoltaic plant of the Central University "Marta Abreu" of Las Villas (UCLV), in the province of Villa Clara in Cuba, has a nominal power of 1.1 MW, was put into operation in 2019 and is located in the southwest area of the Faculty of Electrical Engineering of the Central University "Marta Abreu" of Las Villas, at 22.4° north latitude and 79.96° west longitude.

This photovoltaic plant has the necessary characteristics to develop prediction models that can contribute to the above mentioned. In particular, it has a large group of historical measurements of the following variables: solar irradiance, ambient temperature, temperature of the photovoltaic modules and power generated. In other words, we have the necessary elements to carry out the study in this plant. Therefore, the main objective of this work is to predict the power generated in the UCLV photovoltaic plant using the Long Short-Term Memory (LSTM) deep learning model.

II. THEORETICAL REFERENCE

II.1 PREDICTIVE MODELS BASED ON ARTIFICIAL INTELLIGENCE TECHNIQUES WITH MACHINE LEARNING

Different international publications have addressed the issue of predicting the different variables associated with photovoltaic systems using artificial intelligence techniques.

Reference [7] presents a hybrid model for the long-term prediction of the photovoltaic power generated in a photovoltaic installation based on Artificial Neural Network (ANN) and fuzzy logic. It uses temperature, dew point, wind speed and direction, and solar irradiance as input variables. The proposed model is compared with other prediction models and in all cases presented a superior performance with a Mean Absolute Percentage Error (MAPE) value of 29.60 %.

In reference [8] a Neural Network Ensemble (NNE) prediction model trained by Particle Swarm Optimization (PSO) is proposed to predict the day-ahead power in a smart grid. The model uses as inputs historical data of PV power, solar irradiance, wind speed, temperature and humidity. The performance of the model was measured against five other prediction methods and the NNE method was superior with a MAPE value of 9.75%.

Reference [9] presents a Support Vector Regression (SVR) model to predict the power output of a PV plant for a short-term time horizon. It uses as inputs PV power measurements and solar irradiance forecast from the Numerical Weather Prediction (NWP). The model is able to generate good predictions for clear and cloudy sky conditions with a Root Mean Square Error (RMSE) value of less than 15%. The results obtained are compared with a physical model.

II.2 PREDICTIVE MODELS BASED ON DEEP LEARNING ARTIFICIAL INTELLIGENCE TECHNIQUES

Most conventional approaches to solar power forecasting are not capable of digging deep into the time series and uncovering implicit and relevant information. With the huge data of the modern power system, the use of conventional approaches is not adequate to ensure accurate prediction. Deep learning approaches are becoming increasingly popular due to their good ability to describe dependencies in the time series. Recently, deep learning approaches have emerged as powerful tools that enable complicated pattern recognition, regression and prediction analysis [10], [11].

Reference [12] proposes an LSTM deep learning model for predicting PV power. The proposed model is compared with two other models and proved to be the best performing model with an RMSE value of less than 21%.

Reference [13] presents a LSTM model coupled with a deep neural network. The model is used to predict the load and PV power generated in a smart grid. The performance of the model is compared with other models and a satisfactory result is obtained.

In [14], a Deep Convolutional Neural Network (DCNN) model is used to predict the power of a PV system. The accuracy of the model is compared with a persistent model and a SVR model resulting superior with a Normalized Mean Absolute Percentage Error (nMAPE) value of 11.80%.

Reference [15] proposes a forecasting algorithm to predict PV power generation using a LSTM neural network and a synthetic weather forecast. The proposed model is compared with several machine learning models: a Recurrent Neural Network (RNN), a Generalized Regression Neural Network (GRNN) and an Extreme Learning Machine (ELM) and the results of the LSTM model were superior in all cases analyzed.

III. MATERIALS AND METHODS

III.1 RECURRENT NEURAL NETWORKS

Recurrent neural networks (RNN) [16] are a type of neural networks in which the connections between units form a directed cycle. This creates an internal state of the network that allows it to exhibit dynamic temporal behavior. Unlike feed-forward neural networks, RNNs can use their internal memory to process arbitrary sequences of inputs.

The key idea of RNNs is to use sequential information. In a traditional neural network, all inputs and outputs are assumed to be independent of each other. However, for many applications, this assumption is not always true. RNNs are called recurrent because they repeat the same task for each element of a sequence, and the output depends on previous computations. In other words, RNNs have a memory that captures information about what has been computed so far [16].

Figure 1 illustrates a typical RNN structure, where x_t is the input at time step t , s_t is the hidden state at the at time step t , and o_t is the output at time step t . In detail, it is shown how an RNN is unfolded into a complete network. By unfolding the RNN, the network is presented in a full sequential format.

Note that the hidden state at time step t is also known as the network memory and is calculated from the hidden state at the previous time step and the input at the current time step, as suggested in the following equation:

$$s_t = f(U_{st} + W_{s_{t-1}}) \quad (1)$$

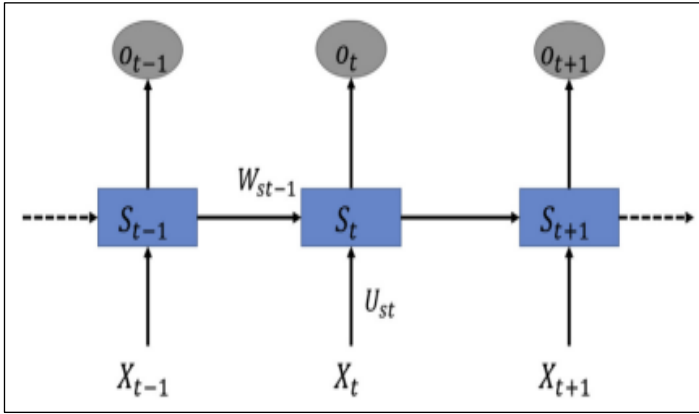


Figure 1: Basic structure of an RNN.
Source: [16].

RNNs use the same model to perform the sequence prediction at different time instants t . Due to this property, they can be used to process variable and large sequences. When working with RNNs, it is first necessary to select the type or architecture of RNN to be used, since the transition function and the handling of

the internal state of the network will depend on the RNN architecture [17].

After selecting the RNN architecture, it is recommended to preprocess the data beforehand, so that the model can use them efficiently. After preprocessing the data, the best model is selected based on a performance measure. While the RNN model has proven to be a powerful tool in terms of handling the data sequence assuming that the current time step depends on previous time steps; it also suffers from some limitations, as it has little ability to learn long-term dependencies due to rapid or explosive gradient decay as it propagates through layers or time instants. To address this problem, new RNN architectures have emerged among which are LSTM recurrent neural networks [17].

III.2 RECURRENT NEURAL NETWORKS OF THE LSTM TYPE

The Long Short Term Memory model (LSTM) [17], [18] arises to overcome the problem of gradient vanishing and thus to learn long term dependencies. In this model, the hidden layer nodes are replaced by special nodes called memory cells. Figure 2 shows the structure of a memory cell of the LSTM model.

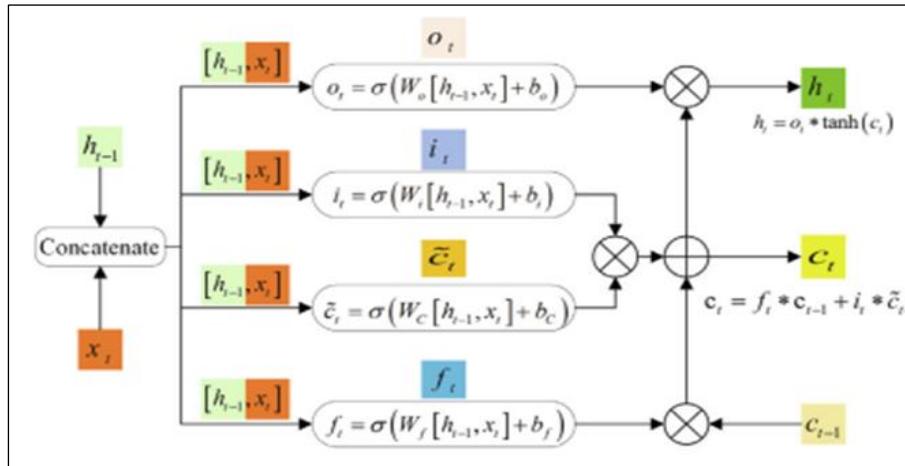


Figure 2: Structure of a memory cell of the LSTM model.
Source: [18].

Each memory cell has a recurring connection with a fixed weight, ensuring that the gradient as it propagates through time does not rapidly diminish or explode. A memory cell is composed of single nodes in a specific connection pattern. Between the nodes that make up a memory cell are a series of gates that are responsible for managing the flow of information in the unit. The components of a memory cell are detailed below [18]:

1. Input node: the input node performs the linear combination of the input vector x_t and the output of the hidden layer at the previous instant h_{t-1} . This node delivers new information to the memory cell.
2. Internal state (c_t): is the main node of the LSTM model, it uses a linear activation function and fixed weights. Since a fixed weight is used in the internal state recurrence, the error propagates through time without the problems of gradient disappearance or explosion.
3. Input gate (i_t): is the first gate used by the LSTM network, it controls the flow of information entering the memory cell, if the value of the gate is zero, no new information enters the memory cell, on the other hand if it is one, all new information enters the memory cell.

4. Forgetting gate (f_t): this gate allows to restrict the information kept in the internal state of the memory cell.
5. Output gate (o_t): the output gate controls the flow of information out of the memory cell. The value finally delivered by a memory cell is given by the internal state of the unit and the output gate.
6. Candidate memory cell (\tilde{c}_t): the LSTM model needs to calculate the candidate memory cell \tilde{c}_t , its calculation is similar to that of the three gates (input, forgetting and output), but uses a \tanh function as the activation function with a range of values between $[-1,1]$.
7. Hidden state (h_t): the \tanh function ensures that the value of the hidden state element is between $[-1,1]$. Note that when the output gate is approximately one, the information in the memory cells is passed to the hidden state to be used by the output layer; and when the output gate is approximately zero, the information in the memory cells is retained by itself.

Given an input sequence $\{x_0, x_1, \dots, x_t\}$ and using the memory cell components, the basic behavior of the LSTM model is described by the following equations:

$$i_t = \sigma(x_t \cdot w_{xi} + h \cdot w_{hi} + b_i) \quad (2)$$

$$f_t = \sigma(x_t \cdot w_{xf} + h_{t-1} \cdot w_{hf} + b_f) \quad (3)$$

$$o_t = \sigma(x_t \cdot w_{xo} + h_{t-1} \cdot w_{ho} + b_o) \quad (4)$$

$$\tilde{c}_t = \tanh(x_t \cdot w_{xc} + h \cdot w_{hc} + b_c) \quad (5)$$

$$c_t = \tilde{c}_t \cdot i_t + f \cdot c_{t-1} \quad (6)$$

$$h_t = \tanh(c_t) \cdot o_t \quad (7)$$

Where $w_{xc}, w_{xi}, w_{xf}, w_{xo}$ and $w_{hc}, w_{hi}, w_{hf}, w_{ho}$ are the weights connecting the layers, t is the time instant of the sequence, $t-1$ is the previous time instant, h_{t-1} is the output value of the hidden layer of the network at the previous time instant; b_c, b_i, b_f, b_o are the bias parameters, \tanh and σ are the hyperbolic and sigmoid tangent activation functions respectively.

The goal of LSTM neural networks is to learn when to let new information into the internal state and when to let information out of the memory cell. Back-Propagation Through Time (BPTT) is used for learning, and the weights of the gates and the input and output nodes are adjusted [18].

In practice, the LSTM model has a better ability to learn long-term dependencies compared to simple RNNs.

III.3 FORECAST MODEL PERFORMANCE

Performance estimation is critical for assessing the accuracy of a model's predictions. Common tools include: Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE). MAE estimates the average significance of the errors in a forecast data set, averages the differences between actual observations and predicted outcomes across the entire test sample, and assigns all individual discrepancies equal weight. Similarly, RMSE estimates the mean value of the error by the square root of the average of the squared differences between the predicted values and the actual observations. It is therefore more robust in dealing with large deviations that are especially undesirable, giving the researcher the ability to identify and eliminate outliers. However, both average metrics (MAE and RMSE) can vary from zero to infinity. In contrast, MAPE is a standard prediction technique that measures prediction accuracy and justifies the diversity of predictions for real data sets [1]. The equations for these metrics are as follows:

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_j - t_j| \quad (8)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_j - t_j)^2} \quad (9)$$

$$MAPE = \frac{1}{N} \sum_{i=1}^N \frac{|y_j - t_j|}{t_j} \times 100\% \quad (10)$$

Where y_j and t_j are the corresponding predicted and measured values of PV power and N is the number of test samples.

III.4 PROPOSED METHODOLOGY

Figure 3 shows a summary of the methodology used to predict the power generated in a photovoltaic installation using the LSTM model.

This process takes into account several fundamental steps that include obtaining historical data from the PV plant and preprocessing them to eliminate outliers in each of the time series.

Then, a data filtering process is applied in order to reduce the computational load of the prediction models and the training time; in this process, all the measurements corresponding to the nighttime hours when the plant does not generate active power are removed from the database. Therefore, only 14 measurements are analyzed each day, corresponding to the plant's working hours between 6:00 am and 20:00 pm. A statistical analysis of the data is also performed, including the correlation analysis between the meteorological variables and the power generated, in order to define the inputs of the prediction model.

Subsequently, data normalization is performed to avoid distortion or deviation of the results and to achieve a more accurate prediction, then the data is divided into training data and test data. Then the characteristics of the model are defined and its training is performed, after which the proposed model is validated and tested. Several statistical indicators (MAE, RMSE, MAPE) are used to quantify the accuracy of the developed model. Finally, the designed LSTM model can be used for energy production prediction.

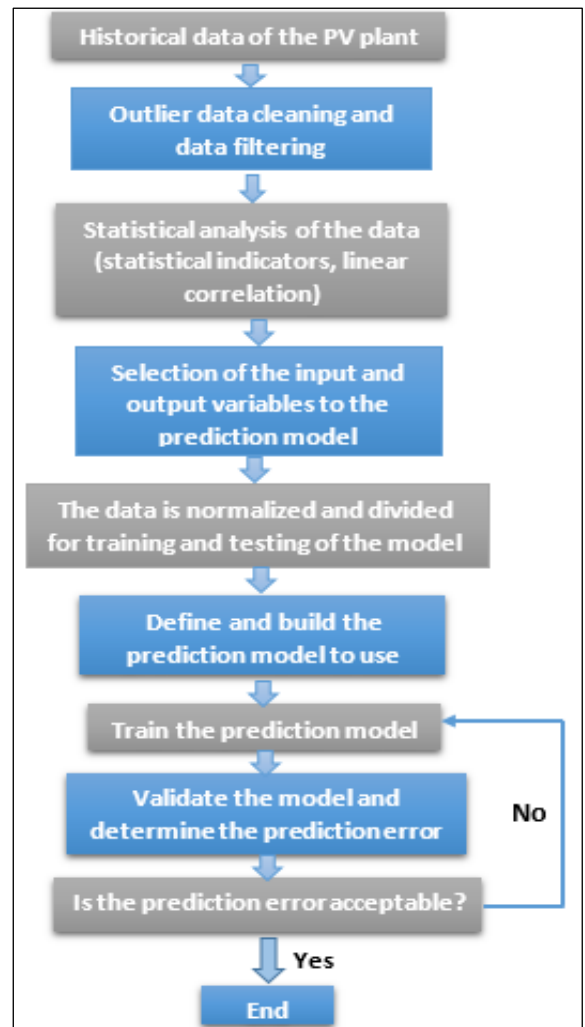


Figure 3: Methodology for PV generation prediction. Source: Authors, (2021).

IV. RESULTS AND DISCUSSIONS

IV.1 STATISTICAL ANALYSIS OF THE DATA

A database was compiled containing measurements of active power generated (expressed in MW), solar radiation (expressed in W/m²), PV module temperature (expressed in °C), ambient temperature (expressed in °C), wind speed (expressed in

m/s) and wind direction (expressed in 0°-360°). Those measurements were taken with a temporal resolution of one hour

from July 2018 to April 2021. Figure 4 shows a fragment of the database.

Date	Wind direction	Wind speed	Solar irradiance	T modulo	T ambient	PV power
'17-jul-2018 06:00:00'	109	0	0	22.5	23.9	0
'17-jul-2018 07:00:00'	114.2	0	83.5	25.3	29	0.09
'17-jul-2018 08:00:00'	103.8	1.8	340.5	32	32.1	0.31
'17-jul-2018 09:00:00'	88.8	2.6	566.4	35.1	32.4	0.54
'17-jul-2018 10:00:00'	90.6	2.6	762.5	38	32.6	0.75
'17-jul-2018 11:00:00'	84.8	1.8	970.1	42.4	34.1	0.86
'17-jul-2018 12:00:00'	96	1.1	1118.9	41.8	34.5	0.96
'17-jul-2018 13:00:00'	92.5	1.1	1116.2	45.2	34.9	0.94
'17-jul-2018 14:00:00'	91.8	1.1	969.5	46	36.8	0.88
'17-jul-2018 15:00:00'	93.4	1.8	301.4	39.5	34.9	0.29
'17-jul-2018 16:00:00'	50.4	1.8	594.1	43.1	35.4	0.54
'17-jul-2018 17:00:00'	82	1.8	373	39	35.8	0.33
'17-jul-2018 18:00:00'	82.3	1.8	22.7	31.2	31.2	0.01
'17-jul-2018 19:00:00'	57.4	0	0	28.5	29.6	0
'18-jul-2018 06:00:00'	118.9	0	0	22.3	24.1	0
'18-jul-2018 07:00:00'	138.2	1.1	86.3	25.6	29.8	0.09
'18-jul-2018 08:00:00'	131.6	1.8	369.6	31.6	31.1	0.33
'18-jul-2018 09:00:00'	114.6	1.1	566.4	36.8	33	0.53
'18-jul-2018 10:00:00'	111.8	1.8	748.4	41.1	33.1	0.72
'18-jul-2018 11:00:00'	15	1.1	947.1	47.9	33.8	0.82
'18-jul-2018 12:00:00'	58.5	1.8	1040.5	54.2	34.8	0.86
'18-jul-2018 13:00:00'	178.6	1.8	1026.9	52.1	35.8	0.9

Figure 4: Fragment of the UCLV photovoltaic plant database.

Source: Authors, (2021).

Tables 1, 2, 3, 4, 5 and 6 summarize the main statistical indicators of the variables active power generated, solar irradiance, PV module temperature, ambient temperature, wind speed and wind direction for each year that was present in the database.

When analyzing the results shown in the tables above, it can be said that in general the quality of the data was acceptable, although there were values that at first glance could be considered erroneous.

Among these were, for example, minimum values of ambient temperature and PV module temperature equal to zero, maximum values of ambient temperature higher than 40°C, solar irradiance values higher than 1300 W/m² and power values higher than 1.1 MW, evidently all the above mentioned measurements were outliers for each of the time series.

To solve these problems and improve data quality, an outlier cleaning process was applied that took into account the extreme limits of each time series. As for missing measurements, if these corresponded to small data segments (≤ 1 h) they were replaced by the average value of four observations, two points before and two points after to maintain the originality and length of the input sequences.

Finally, when there were large sequences (from several hours to weeks) of missing or defective data, imputation was performed, although this problem was practically absent in the available database.

Table 1: Statistical analysis of the active power generated variable.

Year	2018	2019	2020	2021	Total
Average	0.31	0.38	0.34	0.39	0.35
Standard deviation	0.33	0.33	0.32	0.32	0.33
Minimum	0	0	0	0	0
Maximum	1.03	1.05	1.05	1.01	1.03

Source: Authors, (2021).

Table 2: Statistical analysis of the solar irradiance variable.

Year	2018	2019	2020	2021	Total
Average	340.02	410.61	297.91	438.78	360.81
Standard deviation	370.34	369.99	357.28	380.41	370.80
Minimum	0	0	0	0	0
Maximum	1259.50	1354.20	1401.50	1382.60	1401.50

Source: Authors, (2021).

Table 3: Statistical analysis of the temperature of the PV modules variable.

Year	2018	2019	2020	2021	Total
Average	26.15	30.32	22.72	28.72	26.69
Standard deviation	11.99	8.19	14.39	8.26	12.29
Minimum	0	0	0	0	0
Maximum	59.30	48.90	49.10	50.40	59.30

Source: Authors, (2021).

Table 4: Statistical analysis of the ambient temperature variable.

Year	2018	2019	2020	2021	Total
Average	23.95	28.30	21.44	26.82	24.91
Standard deviation	11.73	5.97	13.03	6.20	10.53
Minimum	0	0	0	0	0
Maximum	37.50	39.50	38.60	40.60	40.60

Source: Authors, (2021).

Table 5: Statistical analysis of the wind speed variable.

Year	2018	2019	2020	2021	Total
Average	0.87	1.10	0.86	1.35	1
Standard deviation	1.10	1.13	1.11	1.23	1.14
Minimum	0	0	0	0	0
Maximum	8.70	10.20	7.20	6.40	10.20

Source: Authors, (2021).

Table 6: Statistical analysis of the wind direction variable.

Year	2018	2019	2020	2021	Total
Average	95.17	127.44	94.13	129.76	110.01
Standard deviation	88.72	97.53	97.43	98.89	97.55
Minimum	0	0	0	0	0
Maximum	359.20	359.40	359.40	359.20	359.40

Source: Authors, (2021).

The correlation analysis of the data was a very important aspect, since it allowed defining which variables were to be taken into account in a prediction model.

In this case, a database of the UCLV photovoltaic plant was available, which contained time series of several variables for a period of time close to 3 years. The variable to be predicted was the PV power generated at the facility. Therefore, in the correlation analysis, each of the variables was analyzed with respect to the PV power generated.

Table 7 shows the values obtained for the correlation of the different meteorological variables with respect to the photovoltaic power generated.

As can be seen, the highest correlation was presented by the solar irradiance variable, followed by module temperature, ambient temperature and wind speed. These correlation values were considered strong, therefore, these variables were selected as inputs to the prediction model.

The wind direction variable presented a very weak correlation and was not considered as an input to the prediction model.

Table 7: Correlation between PV power generated and meteorological variables.

Input variables	Correlation coefficient with respect to photovoltaic power generated				
	Year 2018	Year 2019	Year 2020	Year 2021	Total
Solar irradiance	0.9868	0.9916	0.9902	0.9875	0.9890
PV module temperature	0.7898	0.7781	0.7752	0.7819	0.7812
Ambient temperature	0.7242	0.6840	0.6683	0.6890	0.6964
Wind speed	0.4367	0.4148	0.4204	0.4317	0.4259
Wind direction	0.0946	0.1041	0.1094	0.1032	0.1028

Source: Authors, (2021).

IV.2 PREDICTION RESULTS USING THE DEEP LEARNING MODEL LSTM

To perform the prediction of the power generated at the UCLV photovoltaic plant, a deep learning model LSTM was implemented, currently recommended in a considerable number of publications, due to its advantages and its proven effectiveness in learning long-term dependencies compared to simple RNNs.

The model that was implemented contained two hidden layers and each layer possessed 200 memory cells. For its training the initial learning rate was set to 0.05, the Adaptive Estimation of Momentum (ADAM) algorithm was used as the optimizer, the number of training epochs was 250. In the training process of the LSTM model, different variants of data splitting were tested. In this process, one of the options that provided the best results was the division of the data according to the seasons of the year. In other words, the measurements of one year of work of the photovoltaic plant were taken and divided into periods of three months so that each period corresponded to a season of the year under study. For the training of the LSTM model, 90% of the station data was used and the remaining 10% was used to test its performance. The model was implemented in MATLAB 2019a and run on a computer with Intel(R) Core(TM) i3 CPU at 2.4 GHz and 8 GB of memory.

To analyze the performance of the proposed model, predictions of the power generated in the PV system were made for three days of different behavior (sunny day, cloudy day and partially sunny day). In this case, a short-term prediction was made (for 14 hours of the following day). The results of these predictions are shown in Figures 5, 6 and 7 respectively.

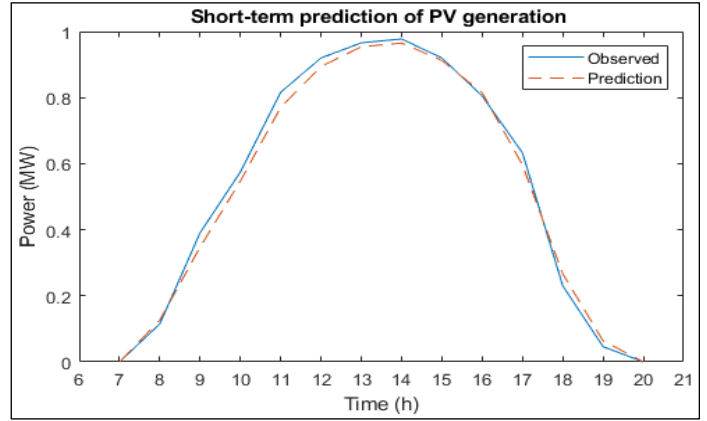


Figure 5: Short-term forecast of PV generation for a sunny day in July 2020.

Source: Authors, (2021).

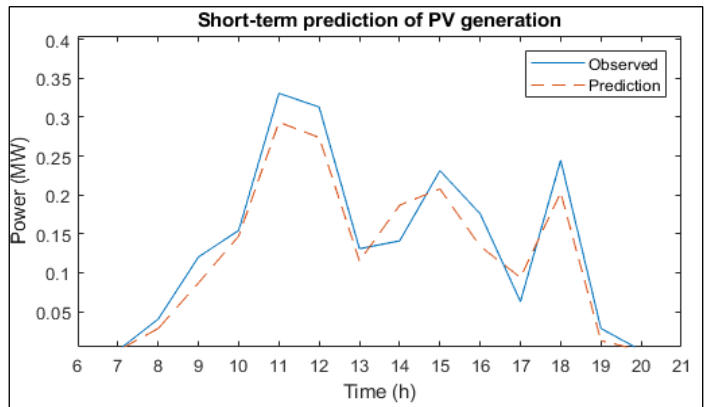


Figure 6: Short-term PV generation forecast for a cloudy day in May 2020.

Source: Authors, (2021).

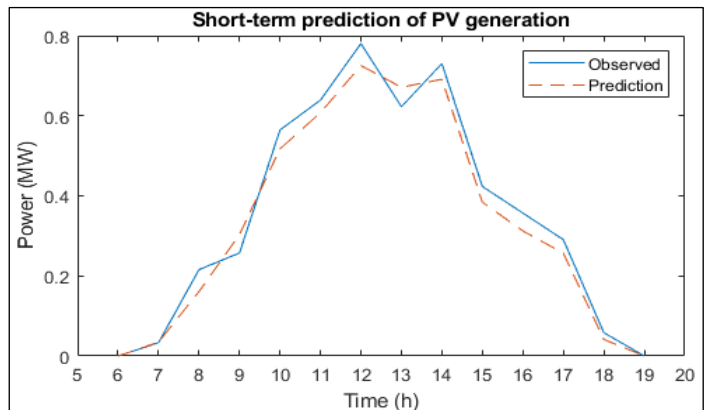


Figure 7: Short-term PV generation forecast for a partly sunny day in January 2021.

Source: Authors, (2021).

As can be seen in the previous figures, the predictions made by the LSTM deep learning model accurately reflected the PV power generation patterns on each day analyzed, even though the behavior of each of these days was different. Evidently the highest accuracy of the prediction model was obtained for the sunny day and the highest uncertainties in the prediction were presented in the cloudy day as expected.

Table 8 shows the prediction accuracy of the LSTM deep learning model for a short-term time horizon (14 hours) and for each of the analyzed day types. Three fundamental metrics were

used to determine the quality of the predictions made (RMSE, MAE and MAPE).

The lowest prediction errors were obtained for the sunny day, followed by the partly sunny day and the cloudy day. As the incident cloudiness over the PV plant increased, there was more variation and intermittency in the incident solar radiation and thus in the generated power.

The high variability made generation prediction more difficult and resulted in higher prediction errors compared to other days with more stable behavior.

Table 8: Prediction accuracy for a short-term horizon for different day types.

Type of day	Metrics used to evaluate prediction error		
	RMSE (kW)	MAE (kW)	MAPE (%)
Sunny day	25.22	16.38	5.86
Cloudy day	43.45	32.32	19.25
Partial sunny day	32.14	23.21	9.48

Source: Authors, (2021).

The model performance was also evaluated for a medium-term forecast horizon. Figure 8 shows a medium-term forecast that was performed with the LSTM deep learning model for six days in the month of February 2021 (February 5 - 10, 2021).

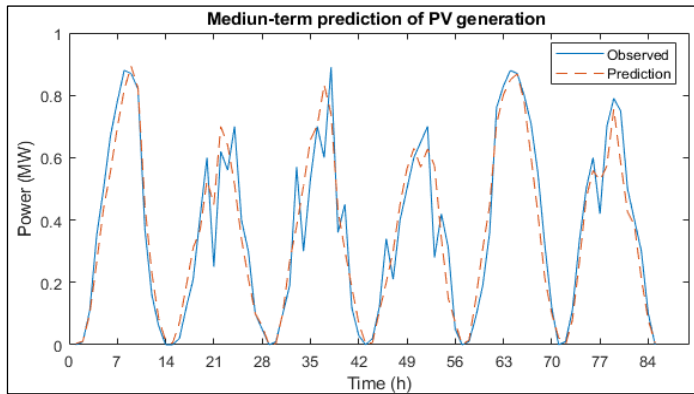


Figure 8: Medium-term forecast of PV power generation from February 5 to February 10, 2021.

Source: Authors, (2021).

The previous figure shows that the predictions of the LSTM model for a medium-term time horizon were accurate, even in the periods of time when there was a high intermittency in the generation due to the cloudiness over the photovoltaic plant, the model was able to adequately adjust to these atypical behavior patterns. Figure 9 shows the behavior of the Mean Absolute Percentage Error (MAPE) for a medium-term prediction horizon.

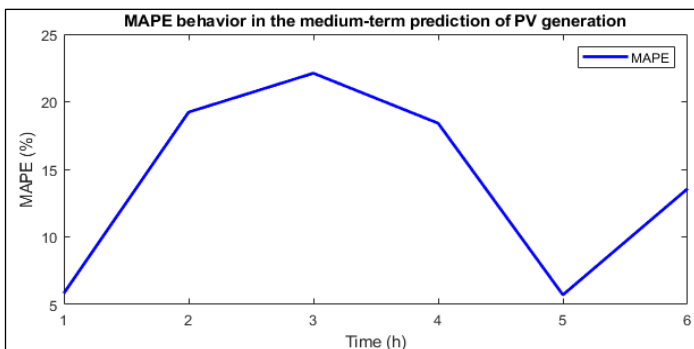


Figure 9: Behavior of the MAPE for a medium-term forecast from February 5 to February 10, 2021.

Source: Authors, (2021).

Again the largest prediction errors were obtained on cloudy days (day 2, day 3 and day 4), followed by a partly sunny day (day 6) and finally sunny days (day 1 and day 5). The LSTM model that was implemented due to its characteristics was able to learn long-term temporal dependencies in the time series and was able to adequately predict PV power generation on days with different behavior and for different time horizons.

V. CONCLUSIONS

The prediction of electric power generation in photovoltaic installations is a difficult task of great importance nowadays. In the present work, a deep learning artificial intelligence model (LSTM model) is proposed. The model is trained with historical data obtained from the UCLV photovoltaic plant, such data is statistically analyzed and prepared with the objective of achieving more accurate and faster predictions. The LSTM model is used for short-term (14 hours in the future) and medium-term (6 days) prediction of the power generated by the photovoltaic installation on days with different behavior (sunny day, cloudy day and partially sunny day), obtaining good results in all cases, with acceptable values of the prediction error.

The main contribution of the work lies in the development of a simple but effective methodology to achieve the proposed objective, which is applicable to any photovoltaic solar farm. However, other deep learning models should be analyzed and their results compared with the proposed model in order to select the most appropriate one according to the type of prediction to be presented.

VI. AUTHOR'S CONTRIBUTION

Conceptualization: Reinier Herrera Casanova.

Methodology: Reinier Herrera Casanova.

Investigation: Reinier Herrera Casanova.

Discussion of results: Reinier Herrera Casanova.

Writing – Original Draft: Reinier Herrera Casanova.

Writing – Review and Editing: Reinier Herrera Casanova.

Resources: Reinier Herrera Casanova and Lesyani León Viltres.

Supervision: Lesyani León Viltres.

Approval of the final text: Reinier Herrera Casanova and Lesyani Leon Viltres.

VII. REFERENCES

- [1] R. Ahmed., et al., "A review and evaluation of the state-of-the-art in PV solar power forecasting: Techniques and optimization", *Renew Sustain Energy Rev*, vol. 124, pp. 109792, March 2020, doi: 10.1016/j.rser.2020.109792.
- [2] S. Zhang, J. Wang, H. Liu, J. Tong & Z. Sun, "Prediction of energy photovoltaic power generation based on artificial intelligence algorithm", *Neural Computing and Applications*, vol. 33, no 3, pp. 821-835, 2021, doi:10.1007/s00521-020-05249-z.
- [3] A. Nespoli, et al., "Day-ahead photovoltaic forecasting: a comparison of the most effective techniques", *Energies*, vol. 12, no. 9, pp. 1621, 2019, doi:10.3390/en12091621.
- [4] T. Fan, T. Sun, H. Liu, X. Xie, & Z. Na, "Spatial-Temporal Genetic-Based Attention Networks for Short-Term Photovoltaic Power Forecasting", in *IEEE Access*, vol. 9, pp. 138762-138774, 2021, doi: 10.1109/ACCESS.2021.3108453.
- [5] M. N Akhter., S. Mekhilef, H. Mokhlis, & N. M Shah., "Review on forecasting of photovoltaic power generation based on machine learning and Metaheuristic techniques", *IET Renew. Power Gener.*, vol. 13, no. 7, pp. 1009–1023, May 2019, doi:10.1049/iet-rpg.2018.5649.
- [6] M. S. Mahdi, M. Longo, & F. Foidadelli, "A-Day-Ahead Photovoltaic Power Prediction Based on Long Short Term Memory Algorithm", 2020 *International*

Conference on Smart Energy Systems and Technologies (SEST). IEEE, pp. 1-6, 2020, doi: 10.1109/SEST48500.2020.9203481.

[7] B. Sivaneasan, C. Y. Yu & K. P. Goh., “Solar forecasting using ANN with fuzzy logic pre- processing”, *Energy Procedia*, vol. 143, pp. 727–32, 2017, doi: 10.1016/j.egypro.2017.12.75.

[8] M. Q. Raza, N. Mithulananthan, J. Li., K. Y. Lee, & H. B. Gooi, “An ensemble framework for day-ahead forecast of PV output power in smart grids”, *IEEE Transactions on Industrial Informatics*, vol. 15, no. 8, pp. 4624–4634, Aug. 2019.

[9] B. Wolff, J. Kühnert, E. Lorenz, O. Kramer, D. Heinemann, “Comparing support vector regression for PV power forecasting to a physical modeling approach using measurement, numerical weather prediction, and cloud motion data”, *Sol Energy*, vol. 135, pp. 197–208, 2016, doi: 10.1016/j.solener.2016.05.05.

[10] Z. Chen, Y. Chen, L. Wu, S. Cheng, P. Lin, L. You, “Accurate modeling of photovoltaic modules using a 1-d deep residual network based on iv characteristics”, *Energy Conversion and Management*, vol. 186, pp. 168-187, 2019, doi: 10.1016/j.enconman.2019.02.032.

[11] F. Harrou, F. Kadri, & Y. Sun, “Forecasting of Photovoltaic Solar Power Production Using LSTM Approach”, In *Advanced Statistical Modeling, Forecasting, and Fault Detection in Renewable Energy Systems*, London, UK: IntechOpen, 2020.

[12] J. Zhang, R. Verschae, S. Nobuhara, J-F. Lalonde, “Deep photovoltaic nowcasting”, *Sol Energy*, vol. 176, pp. 267–276, 2018, doi.org/10.1016/j.solener.2018.10.024.

[13] L. Wen, K. Zhou, S. Yang, X. Lu, “Optimal load dispatch of community microgrid with deep learning based solar power and load forecasting”, *Energy*, vol. 171, pp. 1053–1065, 2019, doi: 10.1016/j.energy.2019.01.075.

[14] R. Zhang, M. Feng, W. Zhang, S. Lu, F. Wang, “Forecast of solar energy production - a deep learning approach”, In: 2018 IEEE international conference on big knowledge. ICBK), pp. 73–82, 2018.

[15] M. S. Hossain and H. Mahmood, “Short-Term Photovoltaic Power Forecasting Using an LSTM Neural Network and Synthetic Weather Forecast”, in *IEEE Access*, vol. 8, pp. 172524-172533, 2020, doi: 10.1109/ACCESS.2020.3024901.

[16] Z. Pang, F. Niu, Z. O’NEILL, “Solar radiation prediction using recurrent neural network and artificial neural network: A case study with comparisons”, *Renewable Energy*, vol. 156, pp. 279-289, 2020, doi: 10.1016/j.renene.2020.04.04.

[17] A. Sherstinsky, “Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network”, *Phys. D: Nonlinear Phenomena*, vol. 404, Art. no. 132306, 2020, doi: 10.1016/j.physd.2019.132306.

[18] Y. Ding, Y. Zhu, J. Feng, P. Zhang & Z. Cheng, “Interpretable spatio-temporal attention LSTM model for flood forecasting”, *Neurocomputing*, vol. 403, pp. 348–359, 2020, doi: 10.1016/j.neucom.2020.04.110.