A Digital Twin Design Applied to a Photovoltaic Conversion System

Touileb Mohammed Islam
Dept. Génie Electrique et Informatique Industrielle
Ecole Nationale Supérieure des Technologies Avancées
Alger, Algérie
m_touileb@ensta.edu.dz

YAICHE Fahem
Dept. Génie Electrique et Informatique Industrielle
Ecole Nationale Supérieure des Technologies Avancées
Alger, Algérie
f_yaiche@ensta.edu.dz

ZAOUI Abdelhalim
Dept. Génie Industriel et Maintenance
Ecole Nationale Supérieure des Technologies Avancées
Alger, Algérie
abdelhalim.zaoui@ensta.edu.dz

AKEL Fethi
Unité de dévlopment énergie solaire
Tipaza Algérie
akel.fethielt@gmail.com

***Abstract*—This article presents the creation of a digital twin for a photovoltaic conversion chain and proposes a predictive model for the electrical parameters of this chain using the LSTM (Long Short-Term Memory) model. The digital twin enables accurate, real-time simulation of the photovoltaic system, integrating short-term forecasting to optimize performance. The LSTM model is employed to predict critical electrical parameters such as DC current, DC voltage, AC current, and AC voltage, based on environmental and electrical data collected. The study demonstrates that the LSTM model can effectively capture temporal dependencies and provide accurate predictions, contributing to better management and optimization of the photovoltaic system. The results show excellent model performance with minimal errors, and the article discusses the challenges related to the observed discrepancies as well as opportunities to improve forecast accuracy.**

***Mots-Clé*—Digital Twin, Photovoltaic Conversion Chain, LSTM Model, Short-Term Forecasting, Performance Optimization, Electrical Parameters**

I. Introduction

In the rapidly evolving field of renewable energy, solar energy systems, while offering promising prospects for a sustainable future, present operational challenges such as sunlight variability, equipment degradation, and maintenance complexity. Managing these complexities becomes crucial to optimizing energy yields and ensuring a solid return on investment. Traditional monitoring and maintenance methods, which are primarily reactive, are often insufficient to meet the demands of these sophisticated systems. This highlights the importance of integrating advanced technologies to optimize performance and ensure sustainability.

One of the most promising innovations in this field is the development of a digital twin for photovoltaic (PV) conversion chains. The concept of a digital twin bridges the gap between the physical and digital worlds by establishing a symbiotic relationship between virtual and real environments. It provides a real-time virtual replica of a physical system, equipped with predictive analysis capabilities. By analyzing real-time data from solar installations, digital twins offer deep operational insights, enabling preventive actions against inefficiencies or potential failures. This real-time connection allows organizations to make data-driven decisions, optimize operations, and predict future behaviors. [1]

Deep learning (DL) and reinforcement learning (RL) are among the most promising advanced methodologies applied to digital twins in the solar energy sector. Deep learning algorithms leverage time series data to enable accurate energy production forecasts. Simultaneously, reinforcement learning agents provide digital twins with the ability to dynamically adjust system parameters in real-time, in response to changes in environmental factors and equipment performance. Together, these methods contribute to a more adaptive, resilient, and efficient solar energy infrastructure. [2]

In this study, we used a machine learning method to create a digital twin of the photovoltaic conversion chain. LSTM (Long Short-Term Memory) networks, designed to avoid the long-term dependency problem, are capable of capturing abstract concepts in the sequences of the photovoltaic (PV) chain, which improves the prediction of various parameters of this conversion chain. This paper is organized as follows: Section 2 presents the literature review. Section 3 explains the proposed method. Section 4 presents and discusses the experimental results. Conclusions and some directions for future work are provided in Section 5.

II. Literature Review:

Several studies have demonstrated the effectiveness of machine learning in predicting energy production from photovoltaic systems using various forecasting methods. The primary objective is to achieve higher accuracy while minimizing complexity and computational cost. A review of photovoltaic power forecasting using machine learning [3] It has been demonstrated that methods such as Bagging models, deep learning, genetic algorithms, random forests, gradient boosting, and artificial neural networks, as well as hybrid models combining different algorithms, can produce good results and improve the forecasting of photovoltaic solar power.

Furthermore, forecasting techniques can be divided into three main categories [4]. Long-term forecasting, as discussed in article [5], explores the potential to predict energy demand up to 10,000 hours into the future. Researchers have considered using the Random Forest model for such long-term forecasts. This category of forecasting can greatly assist in planning and organizing the production, transmission, and distribution of energy by anticipating future energy demand.

Another category of forecasting is medium-term forecasting, which is used for the efficient operation and maintenance of the electrical system by predicting the future availability of electrical energy.

Short-term forecasting of photovoltaic power is the most widely studied area, with numerous articles and studies dedicated to it, such as those in [6]–[8], as it is essential for real-time grid management and the efficient use of solar energy.

Forecasting photovoltaic power using neural networks for short- and medium-term dependencies [9] has shown that the LSTM model exhibits good accuracy in predicting photovoltaic power.

III. METHODOLOGY:

In this methodology, we focus on the integration of short-term forecasting for the design of a digital twin of a photovoltaic conversion chain using a Long Short-Term Memory (LSTM) model. The integration of short-term forecasting offers significant advantages, as it enables real-time optimization and control, which are essential for maximizing efficiency and responding quickly to dynamic environmental conditions. This approach utilizes real-time data from sensors and weather stations to provide accurate and immediate insights into the performance of the photovoltaic system. As a result, the digital twin can dynamically adjust operational parameters, ensuring optimal energy production and load balancing. This real-time adaptability enhances the reliability and efficiency of the photovoltaic conversion process and allows for rapid responses to unexpected changes, thereby reducing downtime and maintenance costs.

The emphasis on short-term forecasting addresses the need for immediate and actionable information in modern photovoltaic systems, promoting both operational excellence and economic benefits.

A. DATA COLLECTION

In this work, we use a dataset comprising environmental information collected from sensors installed in an internal weather station, as well as electrical data from a photovoltaic inverter. The data, sourced from Kaggle, covers a period of 30 days with 1-minute intervals between each data point. The dataset includes the time (hour and minute), environmental variables (irradiance, temperature, wind speed, wind direction, and precipitation), as well as electrical variables, such as the direct current (DC) current and voltage generated by the photovoltaic panel, and the alternating current (AC) current and voltage fed into the electrical grid.

B. DATA EXPLORATION

The predictive model developed in this study is a data-driven approach. In this type of approach, the input data and its quality play a crucial role in the accuracy of predictions. Therefore, exploring the input data is one of the main steps in data preprocessing and feature selection [10]. In [11] , researchers used correlation values to identify and select input features containing relevant information. This method demonstrated high accuracy in predictions. Table 1 presents the correlations between weather variables and electrical measurements. The correlation between our target variable and the two parameters, irradiance and temperature, was particularly high, leading to these two parameters being considered the only inputs. In contrast, the correlations with other parameters (time, wind speed, and wind direction) were much weaker.

	Tension DC	Current DC	Tension AC	Current AC
Irradiance	0.543915	0.967080	0.671383	0.967151
Temperature	0.536676	0.791488	0.436498	0.787360
Time	-0.226386	-0.352723	-0.305055	-0.259540
Wind Speed	0.217650	0.314070	0.233604	0.318541
Wind Direrction	0.047837	0.032502	0.109597	0.032128

TABLE I: The correlation between electrical values and weather values.

C. Model Development:

We use the LSTM-RNN model to predict direct current (DC), direct voltage (DC), alternating current (AC), and alternating voltage (AC) in photovoltaic systems. LSTM models are capable of capturing temporal variations in the data, thereby improving forecast accuracy. In the following subsections, we briefly describe the RNN model, explain the structure of the proposed LSTM model, and detail the selection of hyperparameters.

1) Basic RNN Model : During the learning phase, traditional neural networks cannot leverage the information learned from previous time steps when modeling current data, which is a significant limitation. Recurrent Neural Networks (RNNs) address this issue by incorporating loops that transfer information from one step of the network to subsequent steps, as illustrated in Figure 1, allowing information to persist. In other words, RNNs connect past information to the present task. Utilizing samples from previous sequences can help better understand the current sample.

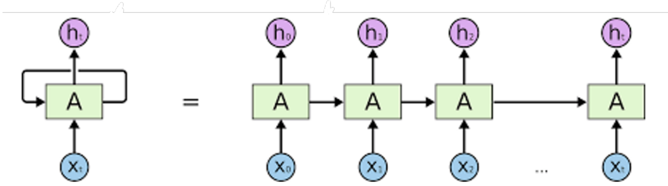

Fig. 1: Structure of the RNN Model

2) LSTM Structure : LSTMs (Long Short-Term Memory) are a special type of RNN (Recurrent Neural Networks) capable of learning both short-term and long-term dependencies [12]. Unlike RNNs, LSTMs are designed to avoid the problem of long-term dependencies. The LSTM network is trained using backpropagation through time, overcoming the vanishing gradient problem. Traditional neural networks have neurons, while LSTM networks consist of memory blocks connected across successive layers. Each block contains gates that manage the block's state and output. In the LSTM unit, there are three types of gates: the forget gate, the input gate, and the output gate. The function of each gate can be summarized as follows:

- The forget gate : determines which information should be discarded from the block based on certain conditions.
- The input gate : determines which values from the input should be used to update the memory state based on certain conditions.
- The output gate : determines what to produce as output based on the input and the block's memory, also according to certain conditions.

As shown in Figure 2, an LSTM block receives an input sequence, and each gate uses activation units to decide whether to activate or not. The commonly used activation functions in an LSTM (Long Short-Term Memory) model are as follows:

1.Sigmoid Function (σ): Used for the gates in the LSTM (forget gate, input gate, and output gate). The sigmoid function compresses the input into a range of 0 to 1, which is useful for deciding how much of a certain value should be retained or forgotten.

$$\sigma(x) = 1/(1 + e^{-x})$$

2. Tanh Function (tanh): Used in updating the cell state and often in the hidden state. The tanh function compresses the input into a range of -1 to 1, which helps keep values within a controlled range while allowing for both positive and negative values.

$$tanhx = (e^x - e^{-x})/(e^x + e^{-x})$$

La structure d'un LSTM comprend généralement ces fonctions pour contrôler le flux d'information et maintenir les dépendances à long terme.

This operation makes the state change and information addition through the block conditional. The gates have weights that can be learned during the training phase. Indeed, the gates make LSTM blocks smarter than classical neurons and enable them to remember recent sequences.

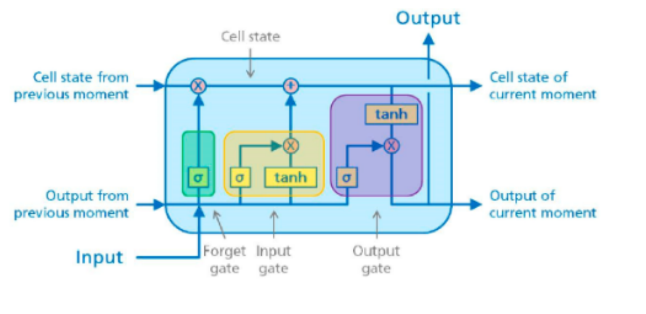

Fig. 2: Representation of an LSTM Cell (Long Short-Term Memory)

3) Mathematical Representation of the LSTM Model : like we said before Each LSTM unit has three key gates: the input gate i_t , forget gate f_t , and output gate o_t . These gates control how information flows through the LSTM, along with the cell state and hidden state.

For our LSTM model, we use two input features, irradiance and temperature , Let $x_t = [I_t, T_t]$ be the input vector at time step t , where I_t and T_t are irradiance and temperature respectively.

the two input feature processed through a two-layer architecture with 20 LSTM units in each hidden layer.

For each LSTM unit l_1 , l_2 in the first layer and the seconde layer , at time step t:

- Forget gate : $f_t = \sigma(W_f.[h_{t-1}, x_t] + b_f)$
- Input gate : $i_t = \sigma(W_i.[h_{t-1}, x_t] + b_i)$
- Cell candidate : $\hat{c}_t = tanh(W_c.[h_{t-1}, x_t] + b_c)$
- Cell state : $c_t = f_t.c_{t-1} + i_t.\hat{c}_t$
- Output gate = $o_t = \sigma(W_o.[h_{t-1}, x_t] + b_o)$
- Hidden state : $h_t = o_t.tanh(c_t)$

The final hidden state from the second LSTM layer h_t is passed to a dense (fully connected) layer to generate the 4 outputs ;

$$y_t = W_{out}.h_t + b_{out}$$

where :

- W_{out} is the weight matrix for the output layer, mapping the hidden state to the 4 output parameters.
- b_{out} is the bias for the output layer .
- $y_t = [V_{DC}, I_{DC}, V_{AC}, I_{AC}]$ the predicted outputs for AC voltage, DC voltage, AC current, and DC current.

4) Hyperparameter Selection : In machine learning, hyperparameters are parameters that are not learned during training but control the model's structure or the training process. For our LSTM model with a two-layer hidden architecture with 20 LSTM units in each layer, processing 50 previous time steps of sequential data to make predictions. The chosen loss function is the Mean Squared Error (MSE), while the 'Adam' optimizer is used to adjust the model weights. By default, the activation functions used in the LSTM model are the sigmoid function for the gates and the tanh function for the cell state and hidden state. An indefinite number of epochs was allowed, with early stopping implemented to prevent overfitting. To evaluate the forecasting performance of the models, we used the Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE), defined as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - y_i^*)^2}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i^* - y_i|$$

To clarify the internal workings of our LSTM model, the following figure provides a visual representation of its structure.

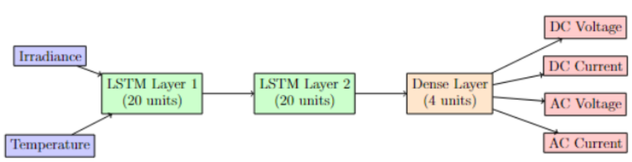

Fig. 3: The developed model structure

IV. RESULTS AND DISCUSSION:

From the collected data, we identified three distinct days with varying weather conditions to test the model: one very cloudy day, one clear day, and one slightly cloudy day. The remaining data was used for training the model. Figure 4 and 5 illustrates the variation of input features (irradiance and temperature) in the test dataset.

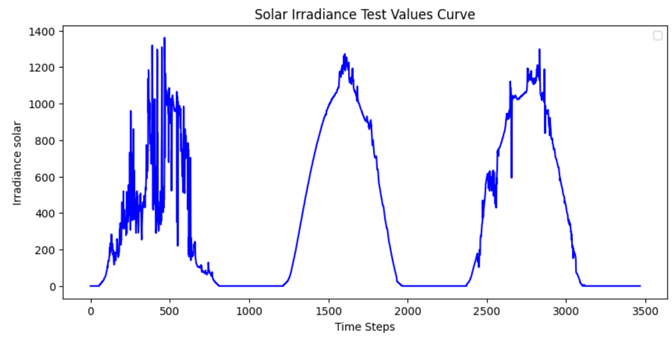

Fig. 4: Irradiance Test Value Variation

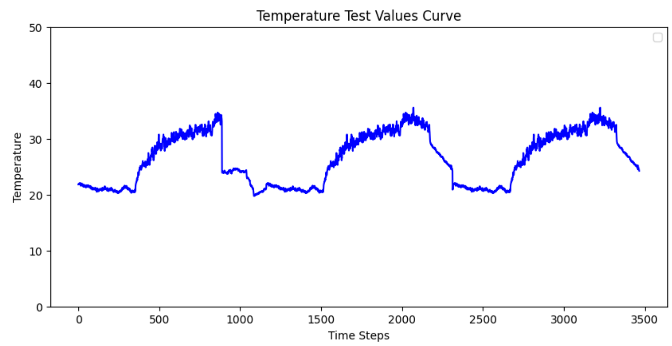

Fig. 5: Temperature Test Value Variation

As shown in the results above, we can see that the LSTM model fit the test sets very well, with a Root Mean Squared Error (RMSE) of 0.0943 and a Mean Absolute Error (MAE) of 0.0738 .

Based on the graphs shown in Figure 5, the prediction model demonstrates a reasonable ability to capture the general trend of the actual values. The model's predictions generally follow the pattern of the real data, suggesting that it has learned some underlying characteristics of the system.

However, noticeable discrepancies exist between the predicted and actual values, particularly during certain fluctuations. These discrepancies could be attributed to several factors:

- Noise and measurement errors: The actual data exhibits fluctuations, likely due to measurement noise. This inherent can impact the model's prediction accuracy. .

- Model complexity: The current model architecture may not sufficiently capture the nuances of output behavior. A more advanced recurrent neural network could enhance performance.

- Data quality and quantity: The quality and quantity of training data greatly affect model accuracy. Insufficient or noisy data can impede effective learning.

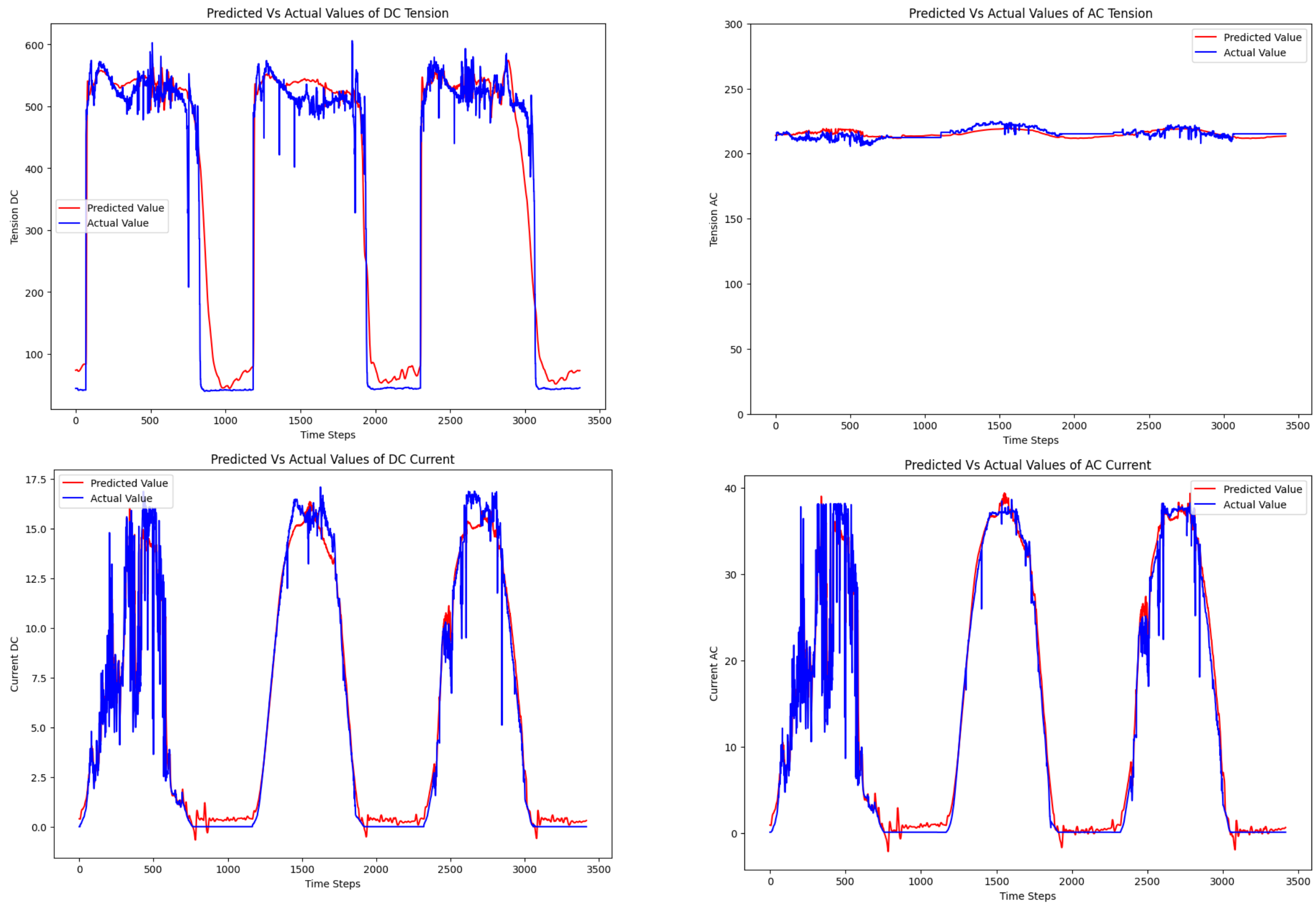

Fig. 6: Electrical parameter prediction using the LSTM model

Despite these discrepancies, the model can still be considered a very good predictive model, as indicated by the values of the Root Mean Square Error (RMSE) and Mean Absolute Error (MAE), which suggest overall strong performance.

V. CONCLUSION AND FUTURE WORK:

In this work, we utilized the LSTM model to predict the voltage and current parameters generated by solar panels, as well as the voltage and current at the inverter's output, based solely on irradiance and temperature inputs. This approach yielded excellent results in terms of predicting the aforementioned electrical parameters. However, to create a complete digital twin of a photovoltaic conversion chain, it is necessary to include battery parameters, which entails adding other inputs, such as time, that influence the quality of the predictions for the entire conversion chain. Integrating the battery component requires the development of more complex artificial intelligence models, as well as improving the quality and quantity of data for more effective training.

Another important aspect to consider is the optimization of existing models to reduce computational complexity, which would enable real-time implementation on embedded systems. Lastly, it would be interesting to study the impact of component degradation in the photovoltaic system on prediction accuracy, as this could pave the way for predictive maintenance strategies based on the digital twin.

REFERENCES

[1] “P. augustine, ”the industry use cases for the digital twin idea,” *advances in computers*, elsevier, 2020.”

[2] “M. lamagna, d. groppi, m. majidi nezhad, and g. piras, ”a comprehensive review on digital twins for smart energy management system,” international journal of energy production and management, vol. 6, pp. 323-334, 2021, doi: 10.2495/eq-v6-n4-323-334.”

[3] “Here is the reference formatted in ieee style: J. gaboitaolelwe, a. m. zungeru, a. yahya, c. k. lebekwe, d. n. vinod, and a. o. salau, ”machine learning based solar photovoltaic power forecasting: A review and comparison,” *ieee access*, vol. 11, pp. 40820-40845, 2023, doi: 10.1109/access.2023.3270041.”

[4] “M. konstantinou, s. peratikou, and a. g. charalambides, ”solar photovoltaic forecasting of power output using lstm networks,” atmosphere, vol. 12, p. 124, 2021, doi: 10.3390/atmos12010124.”

[5] “A. sedai, r. dhakal, s. gautam, a. dhamala, a. bilbao, q. wang, a. wigington, and s. pol, ”performance analysis of statistical, machine learning and deep learning models in long-term forecasting of solar power production,” forecasting, vol. 5, pp. 256-284, 2023.”

[6] “L. liu, y. zhao, d. chang, j. xie, z. ma, q. sun, h. yin, and r. wennersten, ”prediction of short-term pv power output and uncertainty analysis,” applied energy, vol. 228, pp. 700-711, 2018, doi: 10.1016/j.apenergy.2018.06.112.”

[7] “T. trong, h. vu, h. dinh, h. takano, and t. nguyen duc, ”short-term pv power forecast using hybrid deep learning model and variational mode decomposition,” energy reports, vol. 9, pp. 712-717, 2023, doi: 10.1016/j.egyr.2023.05.154.”

[8] “M. ali, r. prasad, m. jamei, a. malik, y. xiang, s. abdulla, r. deo, a. farooque, and a. labban, ”short-term wave power forecasting with hybrid multivariate variational mode decomposition model,” 2023.”

[9] “R. kabir, a. elmouatamid, h. elkhoukhi, and p. w. t. pong, ”photovoltaic power forecasting using neural networks for short and medium-term dependencies,” in 2024 ieee texas power and energy conference (tpec), college station, tx, usa, 2024, pp. 1-6, doi: 10.1.”

[10] “J. d. kelleher, b. m. namee, and a. d’arcy, fundamentals of machine learning for predictive data analytics.”

[11] “N. maitanova, j.-s. telle, b. hanke, m. grottke, t. schmidt, k. v. maydell, and c. agert, ”a machine learning approach to low-cost photovoltaic power prediction based on publicly available weather reports,” energies, vol. 13, p. 735, 2020, doi: 10.3390/e.”

[12] “Hochreiter s, schmidhuber j (1997) long short-term memory..”