J. Phys. &Chem. Res. 3, Issue 1, June (2024) 1–25

Journal of Physical & Chemical Research

Journal homepage: https://ojs.univ-bba.dz

Daily Forecasting of Photovoltaic Power Generation with Multi-Technological data Using Enhanced Long Short-Term Memory Networks

Ferial El Robrini[1*], Badia Amrouche[2]

[1,2] Laboratory of electrical systems and remote control (LabSET), Renewable Energies Department, Faculty of technology, Blida 1 University.(09000), Algeria.

** Corresponding author: Tel.: +213558780269; E-mail address: elrobrinferial@univ-blida.dz/ ferialelrobrini@gmail.com*

Article history
Received May 6, 2024
Accepted for publication November 23, 2024

Abstract

The spotlight on Algeria's efforts to tap into its solar resources and enhance its photovoltaic capabilities has sparked widespread interest. With current achievements totaling 567.1 MW, the country plans for a surge to 3000 MW in the foreseeable future. This increasing reliance on intermittent solar energy underscores the importance of precise PV power forecasting for ensuring grid flexibility and reliability. Deep learning methods have demonstrated promising outcomes in handling intricate data and understanding systematic biases, surpassing conventional approaches. This study explores the effectiveness of LSTM in predicting PV power output across diverse PV technologies. Our methodology involves training the LSTM model extensively on large-scale Poly-Silicon module data and subsequently applying this pretrained model to forecast power output in regions with similar climatic conditions but different PV technologies. Specifically, the model, initially trained on extensive data from the Djelfa power plant (January 2018 to December 2019), is tested using data from the Ghardaïa PV station (July to December 2014) on four other PV technologies with different structures, including Cadmium Telluride (CdTe), Amorphous Silicon, Mono-crystalline and Poly-crystalline with fixed structures, and Mono-crystalline and Poly-crystalline equipped by a stacker system. This approach demonstrates the substantial benefits of applying a pretrained model to smaller datasets in similar climatic regions, particularly when dealing with varying PV technologies. The performance of our LSTM model, evaluated using metrics such as RMSE $\leq$ 0.2090, NRMSE< 21.36%, r $\leq$ 0.9475, and MAE $\leq$ 0.1516, confirms its robust prediction capability across different technological setups, highlighting its practical applicability in diverse PV forecasting scenarios

Keywords: Grid-Connected Photovoltaic power plant, Long Short-Term Memory, deep learning, prediction.

Nomenclature

ANN: Artificial Neural Networks
Bilstm: Bidirectional Long Short-Term Memory
BMO: Barnacle Mating Optimization
Cdte: Cadmium Telluride
CEEMDAN: Complete Ensemble Empirical Mode Decomposition With Adaptive Noise
CFNN: Cascade-Forward Neural Network
CNN: Convolutional Neural Networks
ConvLSTM: Convolutional LSTM
D: Day Number
ELM: Extrem Learning Machine
FFNN: Feed-Forward Neural Network
GA: Genetic Algorithm
GHI: Global Horizontal Irradiance (W/M2)
GNI : Direct Normal Irradiance (W/M2)
GRNN: General Regression Neural Network
GRU: Gated Regression Unit
GSA: Grid Search Algorithm
GSR: Global Solar Irradiance (W/M2)
H: Humidity
H_r: Relative Humidity (%)
IAMFN: Inception-Embedded Attention Memory Fully-Connected Network
IEDN: Inception Embedded Deep Neural Network
IEDN-RNET: Inception Embedded Deep Neural Network
IEDN-RNET: Inception Embedded Deep Neural Network with Resnet
IF: Terative Filtering Decomposition Method
IMF: Intrinsic Functions
LR: Linear Regression
LSTM: Long-Short-Term-Memory
L_T: Local Time
MAE: Mean Absolute Error
MLR: Multiple Linear Regression
MWSO: Modified White Shark Optimization Algorithm
NARX: Non-Linear Autoregressive Neural Network with Exogenous Inputs
nRMSE : Normalized Root Mean Squared Error
P_a:Atmospheric Pressure
PCA: Principal Component Analysis
PDPP: Partial Daily Pattern Prediction
PL :Power-Law
P_{PV}: Output Power (Kw)
Principal Components Analysis: PCA
PV: Photovoltaic
r: Coefficient of correlation
R^2: Coefficient of Determination
RBFN: Radial Basis Function Nerual Network
RBM: Restricted Boltzmann Machine
RBM: Restricted Boltzmann Machine
RE: Relative Error
Resnet: Residual Networks
RF: Random Forest
RMSE: Root Mean Squared Error
RNN: Recurrent Neural Network
RPL: Rational-Power Law
rRMSE: Relative Root Mean Squared Error
SAE: Stacked AutoEncoder
SVM : Support Vector Machine
Std : Standard deviation
T_a: Ambient Temperature (◦C)
TCM: Time Correlation Modification
Time2Vec: Time To Vector
TVF-EMD: Time-Varying Filter-Empirical Mode Decomposition
VAE: Variational Auto-Encoder
W_d: Wind Direction
WDCNN: Convolutional Neural Network With A Wider First Layer Kernel
Wind Speed (M/S)
Wp: Wind Output Power (Kw)
WS: White Shark
Ws: Wind Speed (M/S)
WSOA: White Shark Optimization Algorithm

Introduction

The availability of power system data – including large-scale renewable energy generation and aggregate demand – has improved in recent years. However, there are major challenges given the system's growing complexity. The surge in the adoption of batteries, heat pumps, and electric vehicles is transforming a multitude of assets associated with the transmission and distribution of electricity into a vast ecosystem teeming with millions of data points. For industries and businesses procuring their clean electricity, it is increasingly difficult to securely match real-time consumption data with information on the state of the grid, emissions intensity, and the power mix at the point of consumption – all of which are crucial for effective decarbonization strategies [1]. For utility-grid-connected renewable systems, particularly solar and wind, data availability and transparency of their generation are critical for power system suppliers.

For utility-grid-connected renewable systems, particularly solar and wind, data availability and transparency of their generation are critical for power system suppliers. Given the intermittency inherent in these sources impacting the power system's voltage, frequency, protection, harmonics, rotor angle stability, and flexibility requirement, there is an increasing need for robust and consistent data tools to process and present their information coherently, ultimately facilitating seamless integration within the power system [2]. Accurate photovoltaic (PV) forecasting is one of the proposed data tools aiming to guarantee resource adequacy within a power system. Inadequate resource planning may result in limited reserve capacity to address unexpected system conditions. Breaching these limits can jeopardize voltage stability. PV fluctuations caused by factors like cloud cover can lead to undesirable voltage fluctuations in distribution feeders. Accurate PV power forecasting enables generation companies and system operators to plan operations effectively, ensuring the power supply aligns with the load demand. As PV penetration increases, precise forecasting becomes paramount for reserve allocation and grid stability [3].

The prediction of photovoltaic generation involves estimating the future energy output of a specific PV station. This estimation is based on diverse factors, including spatial and temporal resolution [4,5], geographical location, meteorological conditions, seasonal variations, solar panel efficiency, power plant area, and other technologies used for solar energy conversion. Predictions are generated by analyzing historical data, identifying trends and patterns, specifying correlations, and extrapolating this information to create accurate projections or forecasts. In the literature, many photovoltaic power prediction models have been introduced, all of which aim to achieve better forecast accuracy with less computational cost [4]. They are classified mainly into persistence methods[6,7] , physical techniques [4], statistical techniques (empirical [8,9]machine learning [10–12]), and hybrid models [13–15].

1.1. *Related work*

In the past few years, several approaches and results related to forecasting photovoltaic generation have been published. There are fundamental differences between these methods, mainly due to their use of various input data such as PV, solar irradiation, temperature, humidity, air pressure, wind speed, and direction, among others. Additionally, these methods differ in their forecasting horizons, methodologies, and algorithms. Nowadays, hybrid methodologies that combine different types of models have proven to be effective solutions for improving prediction performance. Table 1 and 2 depicts relevant hybrid models used in this context and their appropriate analysis according to their results.

In the work of Gang et al. [16] a hybrid forecasting model is developed for photovoltaic and wind power generation. This model incorporates a Time2Vec embedding layer for data preprocessing, a Convolutional Neural Network with a wider first layer kernel (WDCNN) for feature extraction, and a Bidirectional Long Short-Term Memory (BiLSTM) network for predictive modeling. The Time2Vec layer plays a crucial role in simplifying the input data preprocessing by decomposing the time series data into both non-periodic and periodic components. One notable characteristic of the WDCNN is the utilization of a wider first convolutional kernel to achieve a larger receptive field, while smaller kernels are employed in subsequent layers to enhance network depth and expand the receptive field. Additionally, stacked BiLSTM layers are incorporated to extract temporal correlations from past and future datasets through an information-encoding mechanism. The performance of the proposed

Time2Vec-WDCNN-BiLSTM model has been compared to various other combinations, including WDCNN, BiLSTM, Time2Vec-WDCNN, Time2Vec-BiLSTM, Time2Vec-CNN-BiLSTM, Time2Vec-WDCNN-LSTM, and Time2Vec-WDCNN-GRU. The results demonstrate the superior predictive accuracy and ability to uncover complex relationships of the proposed model.

Nature-inspired meta-heuristic algorithms exhibit significant potential for addressing optimization problems. The adaptability of these optimization algorithms is closely linked to their tuning parameters. The White Shark Optimization Algorithm (WSOA) has been employed as a standard optimization technique, effectively addressing control applications without substantial modification to its tuning parameters. The white shark (WS) is a top-tier predator and a highly agile navigator, possessing a streamlined physique that enables rapid tracking of its targets. Numerous attributes underlie the excellence of WS behavior in nature as an optimization process, primarily pertaining to its ability to track, explore, and search for prey in close proximity. Mansoor et al. [17] proposed two new hybrid models using a modified white shark optimization algorithm-based General regression neural network (MWSO-GRNN) and radial basis function neural network (MWSO-RBFN) for short-term wind power forecasting. Seasonal results are compared graphically and statistically through 15 min ahead forecasting with four hybrid ANN topologies in combinations with Particle Swarm Optimization (PSO) and Barnacle Mating Optimization (BMO) stochastic optimization algorithms that are respectively: PSO-RBFN, PSO-GRNN, BMO-RBFN, and BMO-GRNN. The results show that the proposed MWSO-RBFNN model outperforms the classic models in all cases for point forecasting and interval forecasting with a higher convergence rate and lower stochastic error. The model achieves an average Nash-Sutcliffe constant score of 0.979 and exhibits superior performance with the least RMSE, RE, R^2, and MAE.

The study of Feroz Mirza et al. [18]introduced a hybrid inception embedded deep neural network with ResNet architecture termed IEDN-RNET, combining Inception modules with various kernel sizes for capturing diverse abstraction levels, ResNet blocks for addressing gradient vanishing issues and capturing local and global patterns, Bidirectional weighted LSTM and Bidirectional weighted GRU layers for handling sequential data's long-term dependencies and dynamics from historical and forthcoming information simultaneously, and Time2Vec method for capturing periodic patterns. Comparative analysis against IAMFN, CNN-RNN, and CNN-BiLSTM shows IEDN-RNET outperforming others with 12% lower mean absolute error, 13% lower root mean square error, 19% lower normalized mean absolute error, 20% lower normalized root mean square error, higher R-Square, and correlation coefficients. Despite its accuracy, the model's training time is relatively high (3805 s), suggesting room for future research in optimizing its architecture for efficiency.

Kedouda et al. [19]explored the utilization of a Feed-Forward Neural Network (FFNN) along with two regression models: the Rational-Power Law (RPL) and Power-Law (PL). The objective was to predict the power output of a 160 W photovoltaic (PV) panel in El-Oued, Algeria, based on a dataset encompassing six days of experimental data (172,800 × 7 data points). The study identified solar irradiation, ambient temperature, and module temperature as key factors strongly correlated with PV power generation. Notably, the Levenberg-Marquardt algorithm delivered the best results for training the ANN model. Results demonstrated that both the ANN and the RPL and PL models achieved a remarkable level of precision, with R2 values of 0.997, 0.998, and 0.996 and Mean Absolute Error (MAE) values of 1.998, 1.156, and 1.242, respectively. It is important to note that while the models from this study exhibit considerable accuracy and robust predictive ability, further investigations are advisable, especially in scenarios involving significant changes in climatic conditions, which should be considered.

Wang et al. [20] developed an independent day-ahead PV power forecasting model based on a Long Short-Term Memory Recurrent Neural Network (LSTM-RNN) to address the issue of PV power fluctuations using data from the previous day. Subsequently, the Time Correlation Modification (TCM) principle is applied to adjust the output of the LSTM-RNN model based on trends and regularities observed in historical data from previous years. Since the proportions of periodic and random components of PV power can vary with different weather conditions, a Partial Daily Pattern Prediction (PDPP) model is introduced to predict the patterns of forecasting days. This allows for the selection of optimal parameters for the TCM process, further enhancing the accuracy of the proposed day-ahead PV power forecasting model. Simulation results based on actual data confirmed the performance of the LSTM-RNN model and the TCM method, as well as the effectiveness of the PDPP framework in improving accuracy, particularly for days with partially accurate patterns.

Sahin et al. [21] presented a comparative analysis between Feed-Forward Neural Networks (FFNN) and multiple linear regression (MLR) to investigate the forecasted energy production of a 500 kWp photovoltaic (PV) solar power plant in the Igdir province. The performance evaluation demonstrates the effectiveness of artificial neural networks in capturing the complex relationships between features and efficiency, even in cases of limited data availability. Principal Component Analysis (PCA) was used to reduce feature dimensions, and the results show that accurate efficiency prediction remains achievable even with a reduced set of features. The findings indicate that the system performed well despite limited data availability. Among a total of seven detailed features used, only three parameters; solar irradiation, module power, and module temperature had the most significant impact on the efficiency of the PV generation.

Matera et al. [22] developed a network of Artificial Neural Networks (ANNs) to forecast the hourly worldwide electrical power produced by eight PV modules with different electrical characteristics. They created six different ANNs based on six PV modules, using hourly temperatures and hourly solar radiation data from 24 different localities worldwide obtained through TRNSYS simulation. The validation and generalization performances were assessed by considering the six PV modules in an additional 24 localities and by including two more PV modules in all 48 localities. The excellent results in terms of accuracy metrics confirm that the network of ANNs is a reliable, simple, and accurate tool that can be used to predict the hourly performance of any PV module in any location worldwide.

Khelifi et al. [23] analyzed the implementation of a time-varying filter-empirical mode decomposition (TVF-EMD) and an extreme learning machine (ELM) model. The suggested TVF-EMD-ELM approach has been established to a maximum horizon of 30 minutes and has been assessed and verified on four separate Algerian PV power datasets with varying climate conditions. The use of decomposition algorithms allows the identification and separation of different components in time series data, such as trend, seasonality, and noise. The combination of EMD-TVF is used to enhance the performance of addressing unexpected events or changes, such as sudden changes in weather conditions or equipment failures, and to maximize the hyperparameter tuning. In all the regions examined, the TVF-EMD-ELM model generates less than 4% error in terms of normalized root mean square error (nRMSE).

Melit [24] utilized a recurrent neural network (RNN) to forecast daily electricity generation in a PV system located in Tahifet, Algeria's southern region. The RNN effectively interpolated solar PV output and key parameters, showing strong performance even with unusual cases.

Variational AutoEncoder (VAEs) are powerful unsupervised generative techniques known for automatically extracting information from data. They excel at dimensionality reduction, compressing high-dimensional data effectively. VAEs also approximate complex data distributions efficiently through stochastic gradient descent. They mitigate overfitting issues through built-in regularization during training, making them effective for diverse applications involving complex data. Dairi et al. [25] provided a Variational AutoEncoder (VAE) for single- and multi-step-ahead forecasting of a 9 MW grid-connected PV power plant in Timimoune. They compared the VAE-based method's forecasting outputs with seven deep learning approaches (RNN, LSTM, BiLSTM, ConvLSTM, GRU, SAE, RBM) and two traditional machine learning methods (LR and SVM). The results highlight the strong performance of deep learning techniques in solar power forecasting, with the VAE consistently outperforming other methods. This underscores the VAE's ability to learn high-level features that enhance forecasting accuracy.

Bouchouicha et al. [26] conducted a comparative forecasting analysis of ANN (Artificial Neural Network) models and MLR (Multiple Linear Regression) models in a 20 MW grid-connected PV plant. The performance analysis demonstrates that all the ANN-based models outperform the MLR models in terms of prediction accuracy and stability. Among these ANN models, the Cascade Forward Neural Network-based models (CFNN) yield the most accurate results.

Hassan et al. [27], implemented a hybrid model based on a non-linear autoregressive neural network with exogenous inputs (NARX) and utilized a genetic algorithm for gradient-free training (GA) to forecast the power output of PV systems. Through an evaluation of the NARX-GA models at various locations and time horizons in Algeria and Australia, the study found that these models provide highly accurate estimates, with relative RMSE ranging between 10% and 20%. Moreover, the introduction of exogenous models improved the forecasting accuracy of corresponding endogenous models by up to 19% when considering only day number and local time as external variables. When additional external parameters, including ambient temperature, relative humidity, wind speed, and global horizontal irradiance, were incorporated, the performance of endogenous NAR-GA models increased by up to 22.3%. Across the different forecasting horizons considered (ranging from 5 to 60 minutes), the NARX-GA models consistently outperformed persistent models by up to 58.41%.

Guermoui et al [28]conceived a new integrated model based on the Recursive Intrinsic Functions decomposition technique (Recursive-IF) and the Extreme Learning Machine (ELM). The methodology is adapted for a maximum forecasting horizon of 60 minutes. Time series PV power was decomposed into various IMFs functions through the IF method, from high- to low-frequency sequences. Then, the decomposed IMFs were used as inputs for the ELM forecasting model to generate the desired PV power output. The effectiveness of the proposed methodology was validated on three different PV plants worldwide, each with distinct technology, capacity, climate conditions, and forecasting horizons. The Recursive IF-ELM approach shows promising results, significantly improving upon direct IF-ELM and outperforming the Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN).

Ziane et al.[29], explored the relationship between the meteorological variables and the output of the grid-connected PV station of Zawiet Kounta (Adrar) in terms of performance assessment and production estimation. Feature selection and Principal Components Analysis (PCA) analysis were used as dimensional reduction techniques and pre-processing input data for training random forest models. The results revealed that pre-processing reduced the computing duration from 6,12 s to 2,65 s for the feature selection and 3,99 s for the (PCA).

1.2. *Motivation and main contribution*

In examining the challenges of improving daily and intraday PV power forecasting models, several key research gaps have been identified. Firstly, there is an over-reliance on data from polysilicon technology, with a lack of model testing and validation for other PV technologies. Secondly, there's a notable absence of data from PV stations with tracker systems, as most studies focus on fixed-structure stations, overlooking the influence of tracking systems on power generation forecasts. Thirdly, existing research is predominantly limited to small-scale PV systems, disregarding the increasing importance of large-scale PV plant data in the renewable energy sector. Finally, most studies focus on very short-term (less than 1 hour) or medium-term (daily) forecasting, neglecting the crucial 2 to 4-hour timeframe required for conventional power plants to initiate. This timeframe is essential to understand for effective photovoltaic energy compensation during peak hours.

This work is intended to explore the forecasting ability of deep learning techniques across different PV data. We applied a Long Short-Term Memory (LSTM) to forecast intraday PV power. The distinctive contributions of this paper include:

1. **Technology and structure diversity:** The scope of our analysis was expanded beyond the confines of a single technology by incorporating data from six different photovoltaic technologies and configurations: polysilicon, mono-silicon, CdTe, and amorphous which are installed on fixed structures, and Poly and Mono-silicon that are equipped with tracker systems.

2. **Applying a Pretrained Large-Scale PV Model to a Small-Scale Plant in Similar Climatic Conditions**: Our approach involves the application of a pretrained model, initially developed using data from a large-scale 53MW PV system, to a smaller-scale PV plant with

a capacity of 1.1 MW. Both PV stations, despite their size difference, share similar climate conditions. This strategy allows for an effective evaluation of the model's scalability and adaptability across PV systems of varying capacities within the same climatic environment.

3. **Extended Forecasting horizon:** To provide a comprehensive understanding of photovoltaic generation potential, we employed a forecasting horizon of three hours (3h). This extended timeframe offers valuable insights, especially in the context of energy compensation during peak hours, considering the startup durations of conventional power plants.

By addressing these research gaps and employing a comprehensive approach, this study aims to significantly enhance the accuracy and applicability of PV power forecasting models, contributing to the advancement of photovoltaic energy integration in the broader renewable energy field. The rest of the paper is arranged as follows: Section 2 describes the methodology architecture, including an explanation of the LSTM model, the PV station's characteristics, and the main data processing steps. In Section 3, the model evaluation is processed and discussed. Finally, Section 4 provides the concluding results and perspective research.

Table 1: Some related work from different localities around the world

Location	**Forecasting Horizon**	**Input variables**	**Methods**	**Performance coefficients**	**Comment**
China [16]	1h	GHI, DNI T_a,H_r, W_s, P_a, PV, W_p	Time2Vec-WDCNN-BiLSTM	NMAE : $1.8308.10^{-2}$ NMSE : $1.6348.10^{-3}$ NRMSE : $4.0432.10^{-2}$	(+) Accurate prediction for sunny days (-) Lower precision for rainy days (-) Low interpretability
Malaysia [17]	15 min	GHI,T_a, H,P_a, W_s,W_d,W_p	MWSO-RBFNN	RMSE: $5{,}73.10^{-4}$ MAE: $2{,}01.10^{-8}$ R^2 :0,99731	(+) Higher predictability (+) Good stability (+) Higher convergence rate (+) Lower stochastic error (-) Overhead costs:
Turkey [17]				RMSE: $6.42.10^{-5}$ MAE: $2{,}01.10^{-8}$ R^2: 0.99731	
China [18]	15min	GHI, DNI, GSR, H_r,T_a,P_a, P_{pv}	IEDN-RNET	MAE: 0.723 RMSE: 1.914 NRMSE :0.0451 R^2 :0.98052	(+) Accurate prediction (+) Enhanced Feature Extraction (-) Complex architecture (-) Time consuming (3805 sec)
Nevada, USA [20]	24h	$P_{pv-1,}$ $P_{pv-2,}$ $P_{pv-3,}$	PDPP-TCM-LSTM- RNN	RMSE: 5,68 MAE: 2,35 R^2: 97,76	(+) Accurate prediction (+) Captures the trend and regularity reflected by historical data.
Marroco [21]	Short therm	GSR, T_a,H_r,W_s P_a, T_{pv} P_{pv}	FFNN	R^2 : 0,9628 RMSE :23,89 MAE :25,09	(+) Successful performance (-) Limited data availability
48 localities arround the world[22]	1h	GHI, T_a,	FFNN	Localities validations $0.8975 < R^2 < 0.9971$ $0.013 < RMSE < 0.075$ $0.008 < MAE < 0.039$	(+) High accuracy (+) Reliability (+) Simple network (+) Generalization capacity

Table 2: Related work from different localities in Algeria

Location	Forecasting Horizon	Input variables	Methods	Performance coefficients	Comment
ElOued, [19]	3 sec	GSR, T_a,H_r, W_s,W_d, T_{pv}, P_{pv}	FFNN	R^2: 0,997 MAE:1,998 RMSE :4,10	(+) High accuracy (-) Period of study too short, doesn't consider climatic conditions among the seasons
Djelfa, Ghardaia, Laghouat, Sidi BelAbbes [23]	30min	P_{pv}	TVF-EMD-ELM	$99,80 < R^2$ (%)$< 99,94$ 0,509 < RMSE < 585,31 2,27 < nRMSE < 3,64 1,54 < nMAE < 2,70	(+) High precision (+) Easy to build (+) Fast convergence
Tahifet [24]	daily	GHI, T_a	RNN	r: 0.97 RMSE: 0.07087	(+) High accuracy
Timimoune [25]	15 min - 30 min 45 min - 60 min	P_{pv}	VAE	0.977 < R2 < 0.995 199.645 <RMSE < 420.03 99.838 < MAE < 193.157	(+) Ability to learn higher-level features (+) High accuracy
Adrar [26]	15 min	W_s, T_a, H_r, G, P_a	CFNN	RMSE: 9,546%, MAE: 6,37% r: 0,984	(+) Accurate results (+) Stable precision among the seasons
Adrar [27]	15 min - 30 min 45 min - 60 min	D, L_T, W_s, T_a, H_r , G, P_{pv}	NARX - GA	$0.952 < R^2 < 0.956$ 10.033 < rRMSE < 11.595	(+) High accuracy (+) Generalization capacity
Adrar [28]	15 min - 30 min 45 min - 60 min	P_{pv}	Recursive IF-ELM	12,44<RMSE(kW)<3,77.103 0.339< nMAE<4,144 0,9792< r (%) <0,9999	(+)Simple implementation (+) Speed convergence (+) Very high forecasting score
Zaouiet Kunta [29]	Not mentioned	W, T_a, H_r , GSR, T_a	RF FS+ RF PCA+RF	R2 = 0.9965, t= 06.12 s R2 = 0.9959, t= 02.65 s R2 = 0.9807, t= 03.99 s	(+) Fast computing time (-) Black Box

2. Material and methods

2.1. *The forecasting architecture*

The objective of this study is to investigate the effectiveness of data mining techniques in deep learning algorithms, particularly Long Short-Term Memory (LSTM) models, for intraday PV power forecasting over an extended three-hour horizon. We utilized datasets representing various PV technologies, each with its distinct characteristics, to evaluate the predictive capability of the LSTM model for different technological setups. The research question we address is whether a deep learning model, once developed, can reliably forecast PV output across these diverse technologies. As depicted in Figure 1, our methodology involves collecting the data from multiple PV technologies, pre-processing this data to ensure its suitability for analysis, selecting representative data samples, training the LSTM algorithm on these samples, and finally, validating the algorithm's performance across different technology types.

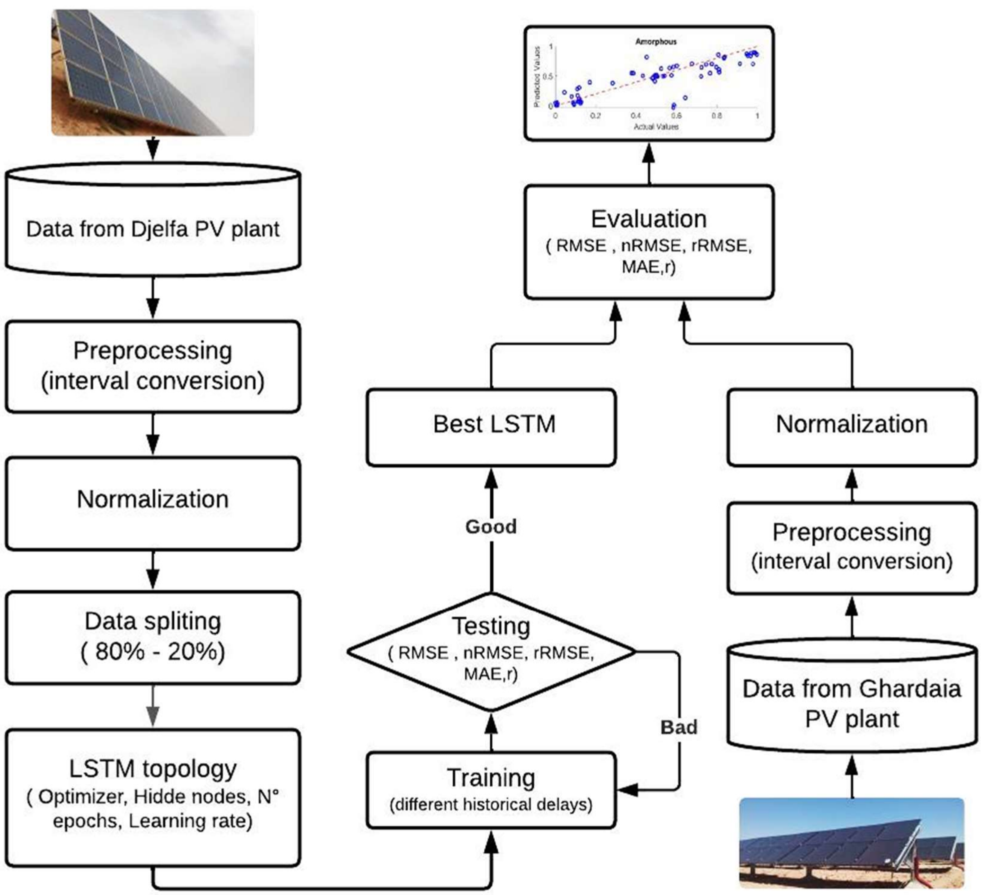

Figure 1: Flowchart of the investigation methodology

This comprehensive approach aims to ascertain the adaptability and accuracy of the LSTM model in forecasting PV output under varying technological conditions. The detailed steps of the proposed study's methodology are the following:

1. **Data Collection and Enhancement:** The initial phase of the study is focused on detailed data collection and comprehensive processing. Essential activities in this stage encompass rectifying missing values, identifying and removing anomalies, and excluding non-essential nighttime data.

2. **Data splitting Strategy:** The dataset is strategically divided into training and testing segments. For the Djelfa station, 80% of the data is allocated for training the LSTM model, while the remaining 20% is reserved for testing its performance. Conversely, at the Ghardaïa station, the entire dataset, spanning a six-month period, is exclusively used as a testing ground for the trained LSTM model, providing a unique evaluation environment.
3. **Optimizing the LSTM Model:** In this stage, the deep LSTM model undergoes a rigorous tuning process through grid search methodology. This process is aimed at identifying the most effective hyper-parameters, ensuring the model's ability to deliver precise and reliable PV power forecasts.
4. **Model Evaluation and Validation across Stations:** The optimized LSTM model is then subjected to a thorough evaluation and validation process. This is done using the test sets from both Djelfa and Ghardaïa stations, allowing for a comprehensive assessment of the model's forecasting accuracy and generalizability across different datasets.
5. **Comparative Analysis and Performance Metrics Assessment:** Additional steps include conducting a comparative analysis of the model's forecasts against actual data, and evaluating its performance using a range of metrics such as RMSE, MAE, and others. This analysis is crucial in understanding the model's strengths and limitations in real-world scenarios.
6. **Scalability and Transferability Testing:** The study also tests the model's scalability and transferability by applying it to datasets of varying sizes and characteristics. This helps in determining the model's effectiveness in diverse operational contexts.
7. **Technological Diversity Consideration**: Finally, the study takes into account the diversity of PV technologies. It examines how the LSTM model performs across different PV technology types, thereby evaluating its versatility and applicability in a broader spectrum of photovoltaic systems.

2.2. *Long Short-Term Memory Theory*

The LSTM model, a form of recurrent neural network, incorporates distinctive features such as weighted connections, memory, and feedback functions [30]. A pivotal element within LSTM is the memory cell (MC), which functions as enduring storage throughout the computational process. The MC facilitates information transfer across the entire sequence, regulating the flow based on decisions made by gate mechanisms. In contrast to traditional RNNs, LSTM excels in efficiently managing valuable information over extended durations, thereby mitigating the vanishing issues associated with conventional RNNs [31].

In its detailed version, LSTM involves adding three gate structures: input, output, and forget gates, as represented in Figure 2. Forget Gate helps to forget the redundant information and save only the relative information to proceed with prediction [32]. The input gate (i_t) is responsible for controlling the flow of new information and specifies whether and, if so, to what extent new information should be used in the current state cell (c_t). The output gate (o_t) determines how much of the information from the previous time step is transferred to the next along with the information from the current time step [33]. The calculated values are in the 0 to 1 forget gate range. When the f_t is close to 1 and the i_t is close to 0, LSTM can achieve the long-term memory function; otherwise, it can realize the short-term memory function [30,31]. The mathematical formulations used in the LSTM network are the following [31,33–35]:

To calculate the gate units

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \tag{1}$$

$$f_t = \sigma\left(W_f x_t + U_f h_{t-1} + b_f\right) \tag{2}$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \tag{3}$$

To update the memory unit

$$\bar{c}_t = tanh(W_c x_t + U_c h_{t-1}) \tag{4}$$

$$c_t = f_c * c_{t-1} + i_t * \bar{c}_{t-1} \tag{5}$$

To calculate the output of the LSTM unit:

$$h_t = o_t * tanh(c_t) \quad (6)$$

The output of the MC is denoted c_t , and the candidate MC is expressed as $\bar{c}_t$, where c_{t-1} represents the cell state at time t-1. Additionally, x_t refers to the input components, and h_t corresponds to the hidden nodes. W_i, W_f, W_o, and Wc are the weights for the i_t, f_t, o_t, and c_t, respectively. U_i, U_f , U_o , and U_c represent the weight matrices for the hidden layers. Additionally, b_i, b_f, and b_o stand for the bias vectors associated with the three gates. The activation function utilized includes the sigmoid function (σ) and the hyperbolic tangent function (tanh).

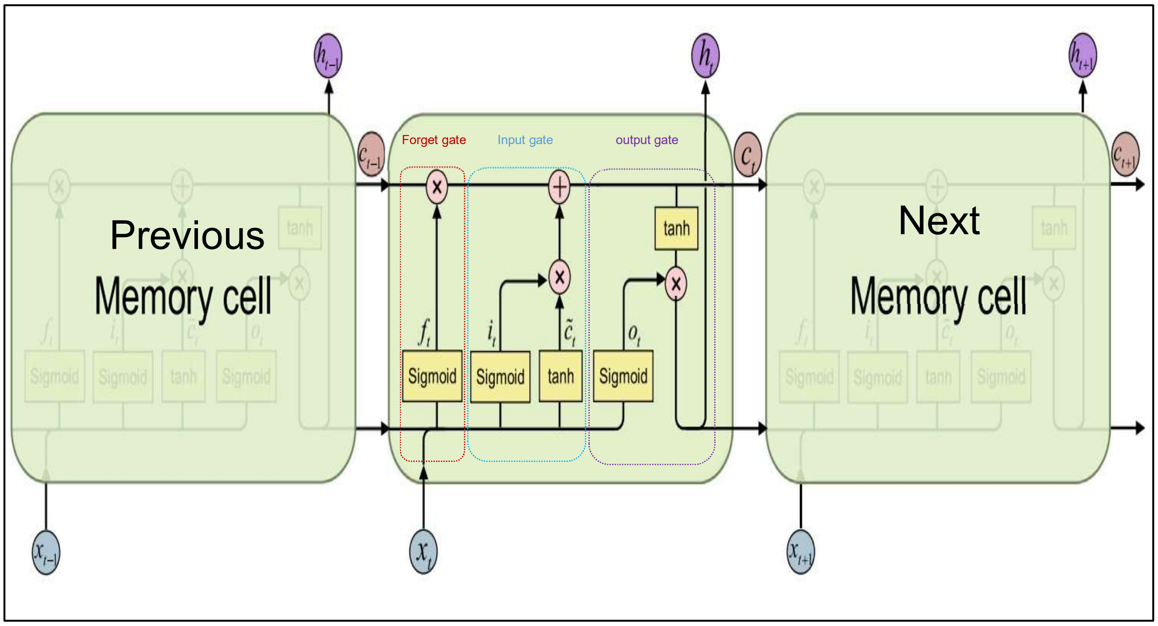

Figure 2: The structure of a LSTM unit [31,35–37]

2.3. *PV station description*

In this study, a dataset comprising two grid-connected photovoltaic (PV) stations located in distinct regions within Algeria was utilized, with each station employing diverse technologies. The first station (Figure 3), situated in Djelfa province (34°20'42"N 3°09'49"E) and characterized by considerable photovoltaic potential (Figure 5), utilizes polycrystalline silicon panels, generating a total capacity of 53MW and covering an area of 120 hectares. The second PV plant (Figure 4), located in Ghardaïa (32°36'02"N 3°41'58"E) characterized by a high photovoltaic potential (Figure 5), has two implemented structures. The fixed structure incorporates four different technologies: amorphous (a-Si), Cadmium Telluride-based thin film (Cd-Te), polycrystalline silicon (Poly-Si), and monocrystalline silicon (Mono-Si). Additionally, the motorized structure (tracker system) is employed for the Poly-Si and Mono-Si technologies. More features of the two power stations are listed in Table 3, while their related PV power statistical indicators are represented in Table 4.

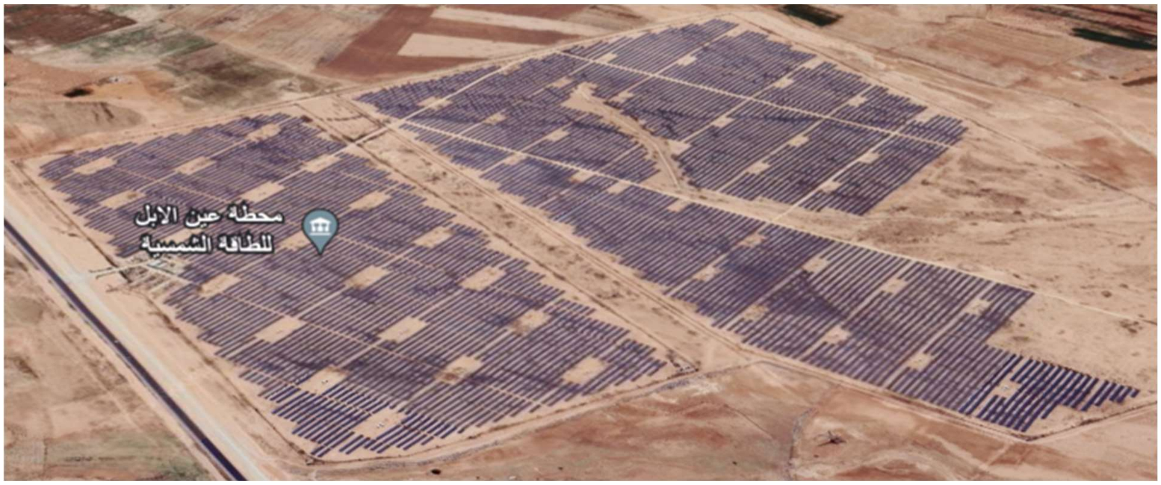

Figure 3: Djelfa solar PV plant [38]

Figure 4: Ghardaïa Solar PV plant [38]

Table 3: Main features of the studied PV plant

PV plant	Sub-field	Photovoltaic technology	Structure Type	Capacity (MW)		No. Panels	Tilt angel
Djelfa I + II		Crystalline poly (YL250P-29b)	Fixed	53		212212	33°
Ghardaïa	01	Mono-Si (SOLARIA S6M-2G)	Tracker	0.102		420	17° ± 55°
	02	Poly-Si (ATERSAA-235P)	Tracker	0.098		420	17° ± 55°
	03	CdTe (FIRST SOLAR FS-380)	Fixed	0.108		1260	30°
	04	Amorphous (SCHOTT ASI 103)	Fixed	0.100	1.13	972	30°
	05	Mono-Si (SOLARIA S6M-2G)	Fixed	0.103		420	30°
	06	Poly-Si (ATERSAA-235P)	Fixed	0.113		480	30°
	07	Mono-Si (SOLARIA S6M-2G)	Fixed	0.249		1020	30°
	08	Poly-Si (ATERSAA-235P)	Fixed	0.256		1100	30°

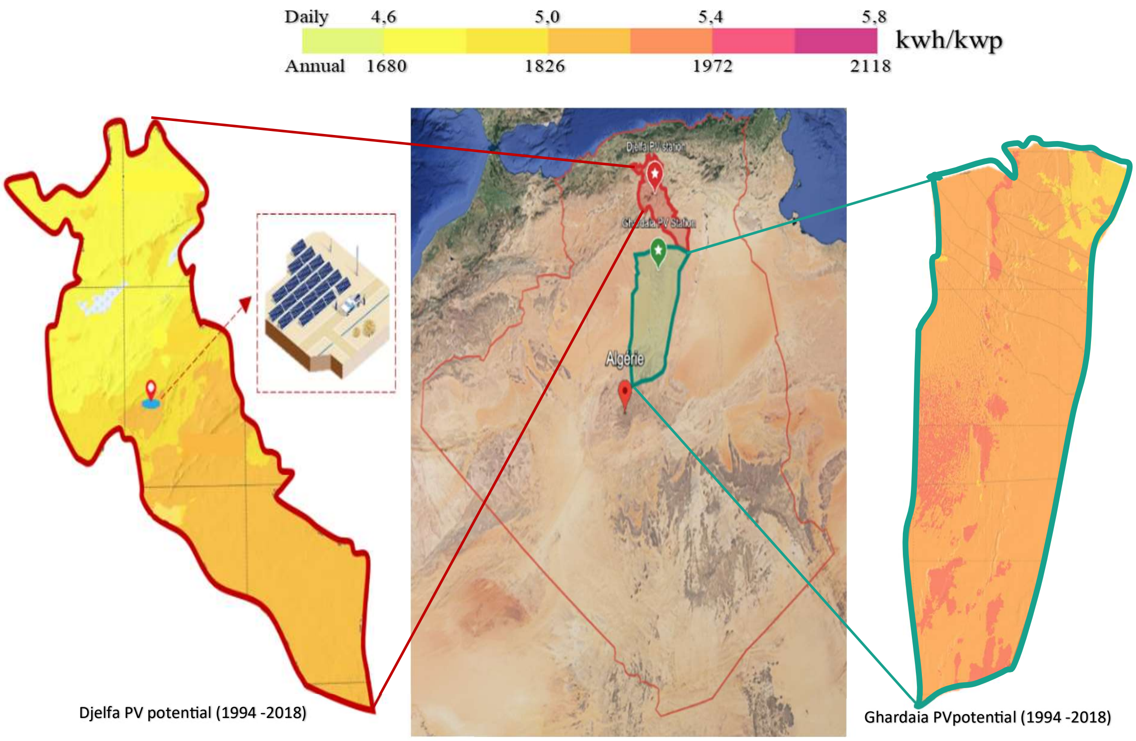

Figure 5: PV potential for the studied regions calculated from 25 recent years historical data (1994 -2018) [39]

Table 4: Statistical parameters of the PV generation

Region	Mean	Mean	Max.	Std	Skewness	Kurtosis
Djelfa (MW)	Djelfa	22.23	58.34	14.82	-0.03	1.58
Ghardaia (kW)	Motorized Mono-Si (kW)	51.52	87.36	27.65	-0.67	1.90
	Motorized Poly-Si	39.11	78.39	23.91	-0.33	1.53
	CdTe	56.60	91.47	25.07	-0.50	1.97
	Amorphous	53.83	93.25	29.20	-0.28	1.71
	Mono-Si	44.88	81.36	24.33	-0.20	1.67
	Poly-Si	41.10	84.60	24.66	-0.24	1.61

2.4. *Data collection, preprocessing, and splitting*

This study is founded on 21,046 measurements of photovoltaic (PV) generation data recorded between 2018 and 2019 at the Djelfa PV plant. The recordings span from early morning (6 a.m.) to evening (20:00) at 30-minute intervals. For the Ghardaïa power plant, a total of 52,706 measurements for each technology, were recorded at 5-minute intervals, covering the entire day from 00:00 to 23:55, over the period from July to December 2014. Hence, to standardize the data interval, we took only the data from (6 a.m.) to evening (20:00) at 30-minute intervals. The recording process at the Ghardaïa power plant encountered numerous disturbances, resulting in a substantial portion of the recorded data points being excluded. Instead of removing this data, which would introduce discontinuity and data offset in the input layer, we opted to retain one clean and continuous interval for each technology. While this shortens the dataset for each technology, it preserves the integrity of data mining processes. Despite the robust support of power supply systems for the intraday forecast horizon, we adjusted the temporal step from 30 minutes to 3 hours, as depicted in Figure 6. This modification is deemed more practical for assessing the intraday potential of the Djelfa PV plant and, consequently, for grid scheduling.

In constructing the model, 80% of the entire dataset associated with the Djelfa Power Plant was allocated for the training phase, while the remaining 20% was reserved for testing purposes. To evaluate the predictive capabilities of the developed model, the Ghardaïa dataset was independently introduced to the model and subjected to testing. More details are given in Table 5.

Table 5: Train and testing dataset

Location	Technology	Test/train	Period		Total data
Djelfa	Poly-Si	Train :80%	01/01/2018	07/08/2019	(585 days x 5)
		Test 1 :20%	08/08/2019	31/12/2019	(146 days x 5)
Ghardaïa	Mono-Si (track.)	Test 2	25/08/2014	25/09/2014	(32 days x 5)
	Poly-Si (track.)	Test 3	01/09/2014	30/09/2014	(30 days x 5)
	CdTe	Test 4	01/10/2014	24/10/2014	(24 days x 5)
	Amorphous	Test 5	01/10/2014	16/10/2014	(16 days x 5)
	Mon-Si (Fixed)	Test 6	28/09/2014	29/10/2014	(32 days x 5)
	Poly-Si (Fixed)	Test 7	03/08/2014	25/12/2014	(145 days x 5)

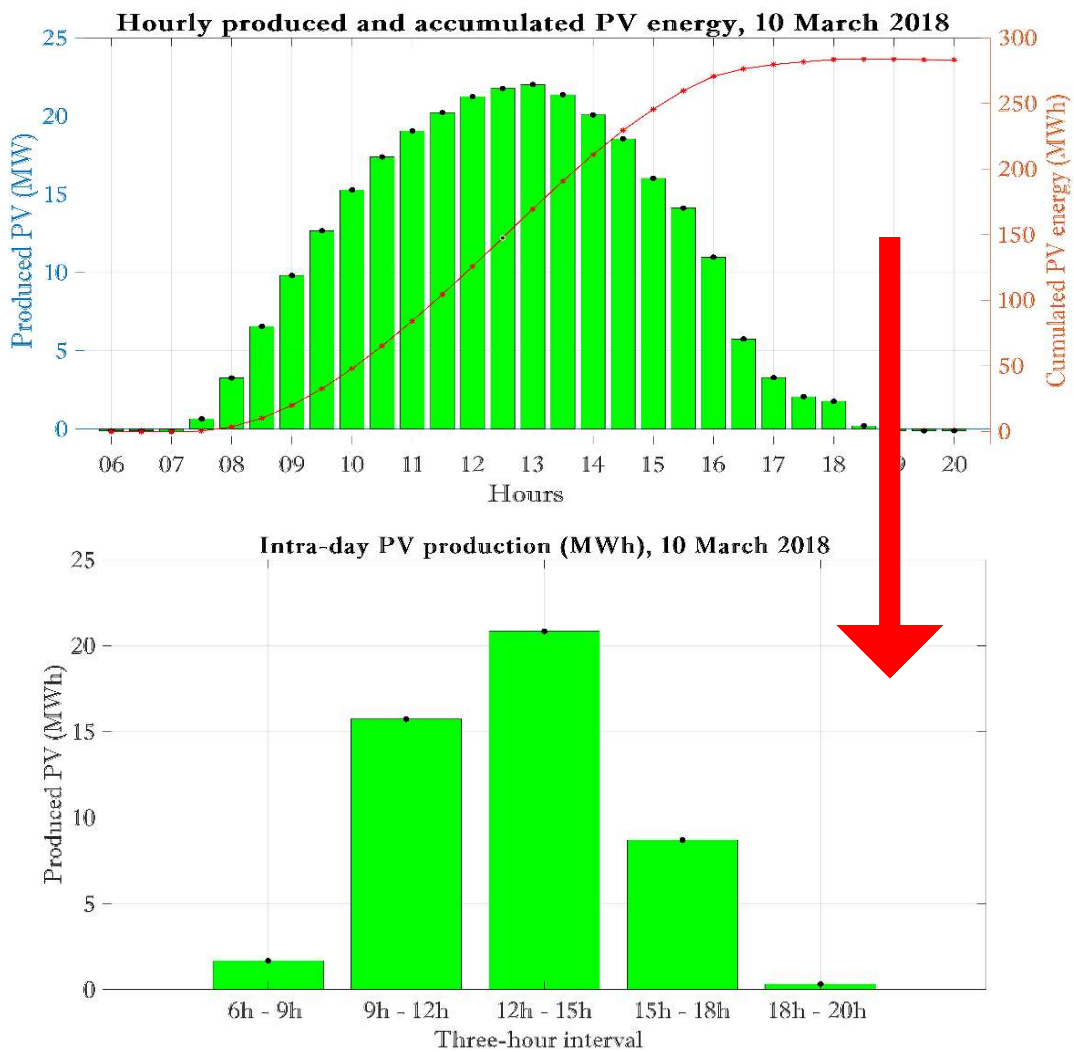

Figure 6: Time horizon conversion

After all, Bias may happen in the developed model because the input data has a different variation scale; the maximal range for each technology is different. Besides, outliers may occur within the same dataset. By normalizing the data, all features are brought to a similar scale, and the impact of outliers data is reduced by bringing them closer to the range of other data points. Different normalization types exist. In our work, the Min-Max normalization is used and represented in equation (7) as follows [37]:

$$\bar{X}_i = \frac{X_i - X_{min}}{X_{min} - X_{max}} \quad (7)$$

$\bar{X}_i$: Is the normalized data

X_i : Is the original data

X_{min} , X_{max} : are minimal and maximal values contained in the dataset.

3. Results and discussions

3.1. *Hardware and software requirements.*

The development of the model is carried out in the MATLAB R2018b platform, and all the simulations were conducted on a computer with a 64-bit operating system, 16.00 GB of RAM, and an Intel(R) Core (TM) i7-9850H CPU @ 2.60GHz.

3.2. *The model parameters*

The accuracy of the LSTM model is influenced by various factors, including the volume of training data, the network architecture, hyperparameters, and optimizers employed for weight and bias optimization [34]. However, these parameters were kept constant among all the processes as listed in Table 6, to value the data diversity impact on the accuracy of the developed model.

Table 6: Main parameters of the LSTM model.

Activation function	Hyperbolic tangent
Optimization algorithm	Adam
Number of hidden nodes	200
Maximum number of training epochs	100
Mini-batches used during training	64
The initial learning rate	0.01

3.3. *Evaluation metrics*

The performance validation of the LSTM model is presented in this section. To assess the results, several metrics are employed, including the Mean Absolute Error (MAE), Root Mean Square Error (RMSE), relative and normalized Root Mean Square Error (r$RMSE$) end (nRMSE), and the coefficient of correlation (r), whose mathematical expressions are given by:

$$\mathbf{r} = \frac{\sum_{i=1}^{n}\left(\left(I_{i,predicted}-\hat{I}_{i,predicted}\right)\left(I_{i,measured}-\hat{I}_{i,measured}\right)\right)}{\sqrt{\sum_{i=1}^{n}\left(I_{i,predicted}-\hat{I}_{i,predicted}\right)^{2}}\sqrt{\sum_{i=1}^{n}\left(I_{i,measured}-\hat{I}_{i,measured}\right)^{2}}} \quad (8)$$

$$\mathbf{MAE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left|\mathbf{predicted_i}-\mathbf{measured_i}\right|} \quad (9)$$

$$\mathbf{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}\left(I_{i,predicted}-I_{i,measured}\right)^{2}}{N}} \quad (10)$$

$$\mathrm{nRMSE} = \frac{\mathbf{RMSE}}{\mathrm{Max}(I_{measured})-\mathrm{Min}(I_{measured})}\,x100 \quad (11)$$

$$\mathrm{rRMSE} = \frac{\mathbf{RMSE}}{\mathrm{Mean}(I_{measured})}\times 100 \quad (12)$$

3.4. *Results and discussion*

The current study seeks to assess the effectiveness of the Long Short-Term Memory (LSTM) model in forecasting photovoltaic (PV) generation across various technologies. This evaluation leverages data from two sizable PV stations, each characterized by differing maximum PV capacities, geographical locations, and module technologies. A stand-alone LSTM model and a Grid Search Algorithm based LSTM were first developed and assessed using the dataset from the initial power plant (Djelfa).

Subsequently, a comprehensive reevaluation was conducted utilizing the best of the two models by applying the dataset from the second station (Ghardaïa), as detailed in Section 2.4. The assessment involved the analysis of evaluation metrics, yielding the subsequent results.

Table 7 displays the performance parameter details of the LSTM model through the building phase. Varying historical delays (from 1 to 15) were tested. The MAE occupies a range from 0.0732 (D = 14) to 0.0893 (D = 2), indicating enhanced predictive accuracy. The RMSE varies from 0.1056 (D = 14) to 0.1304 (D = 2). Importantly, the correlation coefficient (r) values consistently maintained a high level across the experiment and varied from 0.9081 (D = 2) to 0.9342 (D = 14). The NRMSE values span from 11.84% (D14) to 14.62% (D=2). The rRMSE went from 31.59% (D = 14) to 38.97% (D = 2). According to these statistical results, the 14Th historical delay is selected as the best forecasting delay, explained by lower error values and a higher correlation coefficient. On the other side, Table 8 showcases impressive results related to the Grid Search algorithm integration. For instance, the Mean Absolute Error is between 0.0731 (D = 8) and 0.0905 (D = 1), indicating consistently low absolute errors across various scenarios. The RMSE varies from 0.1045 (D = 8) to 0.1280 (D = 1), demonstrating the model's robust performance with relatively small root mean square errors. The correlation coefficient (r) values maintained a better level across the experiment and varied from 0.9073 (D = 2) to 0.9350 (D = 8), reflecting a consistently strong linear relationship between predicted and actual values. The normalized RMSE values are between 11.40% (D4) and 14.35% (D=1). The relative RMSE went from 31.27% (D = 8) to 38.20% (D = 1). Moreover, it can be inferred that the 8th historical delay exhibits the highest correlation coefficient with lower error values.

As per the statistical findings, the two LSTM-based approaches demonstrate a strong ability to forecast the PV generation within the Djelfa power plan based on short (D=1) or long (D=15) historical delays. However, through a refined analysis and comparison, the Grid Search Algorithm yields superior results and permits the transition to a shorter historical delay.

Table 7: Results for Djelfa power plant testing (Stand-alone LSTM)

	MAE	**r**	**RMSE**	**NRMSE(%)**	**rRMSE (%)**
D=15	0.0736	0.9310	0.1079	12.09	32.23
D=14	**0.0732**	**0.9342**	**0.1056**	**11.84**	**31.59**
D=13	0.0777	0.9223	0.1141	12.79	34.13
D= 12	0.0768	0.9282	0.1125	12.61	33.58
D= 11	0.0756	0.9271	0.1113	12.47	33.19
D= 10	0.0745	0.9302	0.1080	12.10	32.24
D=9	0.0767	0.9243	0.1140	12.78	34.07
D=8	0.0784	0.9224	0.1136	12.74	34.00
D=7	0.0804	0.9184	0.1179	13.21	35.22
D=6	0.0801	0.9198	0.1168	13.10	34.88
D=5	0.0803	0.9191	0.1178	13.21	35.20
D=4	0.0809	0.9220	0.1164	13.05	34.83
D=3	0.0844	0.9163	0.1197	13.41	35.80
D=2	0.0893	0.9081	0.1304	14.62	38.97
D=1	0.0848	0.9164	0.1222	13.70	36.48

Table 8: Results for Djelfa power plant testing (Grid Search)

	MAE	r	RMSE	NRMSE(%)	rRMSE (%)
D=15	0.0752	0.9324	0.1064	11.92	31.76
D=14	0.0750	0.9298	0.1082	12.12	32.34
D=13	0.0765	0.9278	0.1100	12.33	32.89
D= 12	0.0767	0.9272	0.1110	12.44	33.12
D= 11	0.0749	0.9304	0.1084	12.14	32.31
D= 10	0.0772	0.9309	0.1075	12.04	32.07
D=9	0.0786	0.9268	0.1099	12.31	32.83
D=8	**0.0729**	**0.9350**	**0.1045**	**11.70**	**31.21**
D=7	0.0782	0.9210	0.1158	12.98	34.62
D=6	0.0771	0.9292	0.1089	12.20	32.51
D=5	0.0806	0.9252	0.1110	12.45	33.18
D=4	0.0772	0.9253	0.1140	12.77	34.07
D=3	0.0760	0.9319	0.1062	11.89	31.74
D=2	0.0846	0.9223	0.1156	12.96	34.55
D=1	0.0905	0.9073	0.1280	14.35	38.20

Figure 7 and 8 depict a representative sample of results obtained over three consecutive days, along with a scatter plot comparing stand-alone LSTM and Grid Search LSTM. It is interesting to note that Grid Search Algorithm based LSTM, charactereized by its systematic exploration of hyperparameter combinations, displays the closest alignment between predicted and actual values.

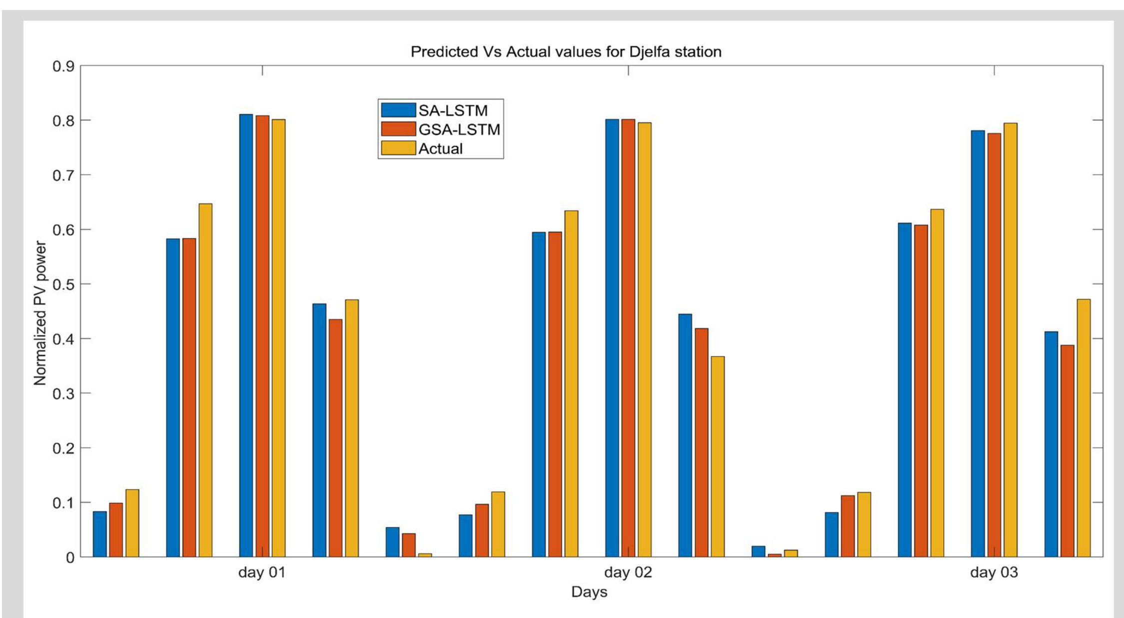

Figure 7: Comparative results between Stand-alone LSTM and Grid search LSTM

Moving to the evaluation across different technologies, the database from Table 4 (Ghardaïa Dataset) was used. The Grid search-based LSTM is used for the rest of the experimentations and the optimized historical delay found in the first building phase is considered and fed into the model for each technology (D=8). The results display varying effectiveness in predicting photovoltaic (PV) generation across different technologies, as indicated by the key performance metrics listed in Table 9. For instance, the comprehensive assessment of Grid Search-enhanced LSTM through all technologies reveals a spectrum of Root Mean Squared Errors (RMSE), ranging from 0.1250 (Fixed-Poly-Si) to 0.2090 (Motorized Poly-Si), exhibiting varying but good levels of accuracy for each technology. Similarly, the relative RMSE (rRMSE) fluctuates from 26.97% (CdTe) to 49.21 % (Motorized Poly-Si), emphasizing the model's relative predictive precision. Mean Absolute Error (MAE) demonstrates consistency, varying from 0.0832 (Fixed Poly-Si) to 0.1516 (Motorized Poly-Si). The correlation coefficient spans from 0.8171 (Motorized Poly-Si) to 0.9467 (CdTe), elucidating the robust predictive capabilities of the model. It is noteworthy that CdTe, Amorphous, Fixed Poly-crystalline, and Fixed Mono-Crystalline exhibit exceptional performance as predicted technologies, giving results to lower errors and higher correlation coefficients.

Table 9: Evaluation of LSTM with grid search through all technologies.

	Ghardaïa (LSTM with Grid search)					
	Motorized Mono-Si	**Motorized Poly-Si**	**CdTe**	**Amorphous**	**Fixed Mono-Si**	**Fixed Poly-Si**
RMSE	0.1792	0.2090	0.1406	0.1883	0.1528	0.1250
rRMSE (%)	36.34	49.21	26.97	34.83	27.41	32.26
NRMSE (%)	17.92	21.36	14.06	18.97	15.28	12.50
MAE	0.1371	0.1516	0.1161	0.1467	0.1318	0.0832
r	0.8831	0.8171	0.9475	0.8869	0.9467	0.9276

Visibly, Figure 9 elucidates the model's performance over three consecutive days, offering a clear depiction of its forecasting prowess along with the scatter plot of the predictive model for each technology. The observations underscore the deep learning models' proficiency in forecasting photovoltaic generation, particularly for specific technologies within the diverse array considered.

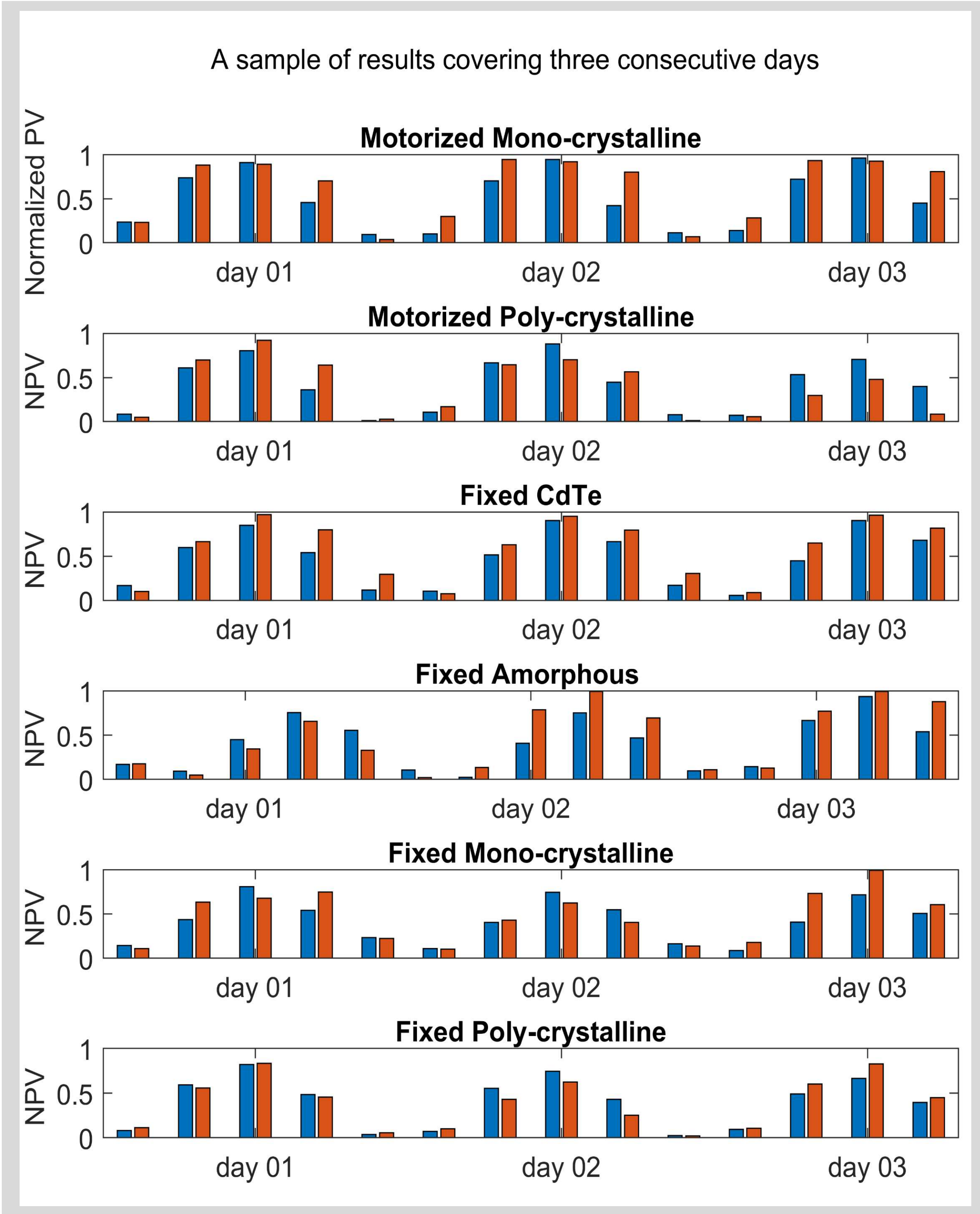

Figure 8: A sample of results covering three consecutive days for each technology.
The blue color represents predicted values, and the red represents actual values.

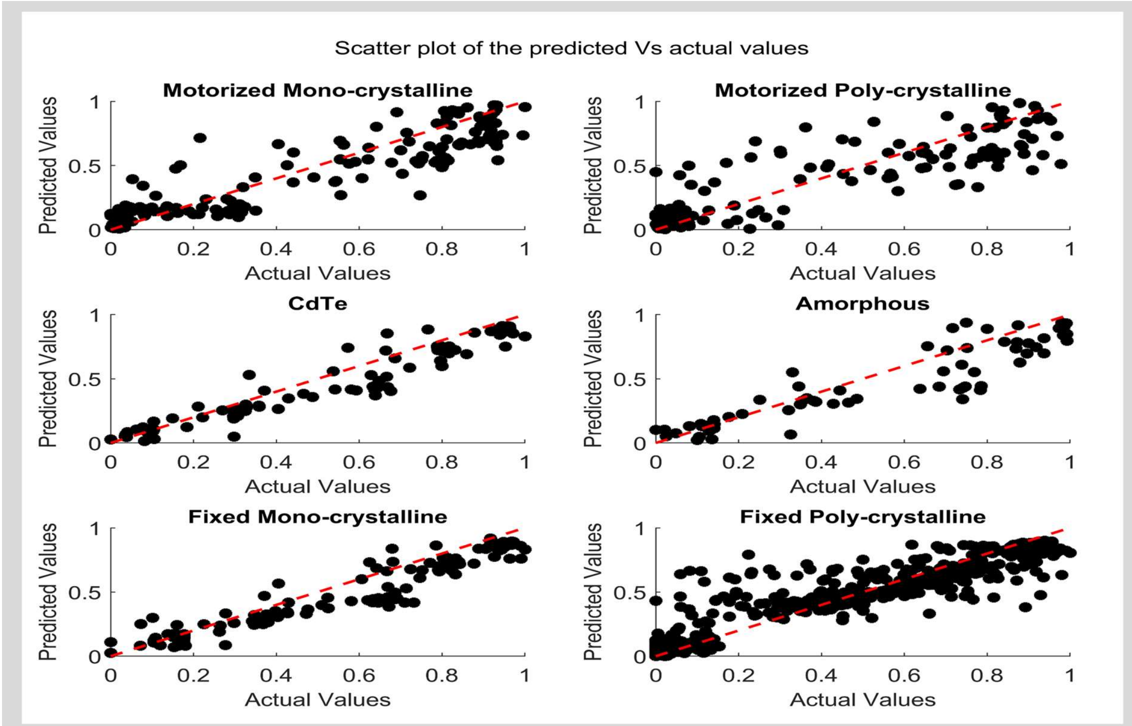

Figure 9: Scatter plot of the predicted Vs actual values for each technology within Ghardaïa power plant

The findings presented in the paper demonstrate that Long Short-Term Memory-based models are a highly efficient method for predicting PV power, as evidenced by the literature. However, while most studies have found relevant results by testing the model on the same data technology, the developed LSTM-based model in this study was built on polycrystalline technology with fixed structure data recorded from the Djelfa power plant and tested for a second PV plant from another region located in Ghardaïa. The accuracy is very satisfactory for PV power forecasting even when the testing data belong to types such as Cadmium Telluride (CdTe), amorphous silicon, Mono-crystalline and poly-crystalline silicon with motorized and fixed structures. The following table summarizes the particular features of this study:

Table 10: The particular features for this study

Training data technology	Testing data technology	Technique	horizon	Results
Fixed Poly-crystalline	Fixed Poly-crystalline	LSTM (Grid-Search)	3h	MAE $\leq$ 0.1516 0.8171$\leq$ r $\leq$ 0.9475 RMSE$\leq$ 0.2090 NRMSE (%) $\leq$ 21.36
	Fixed Mono-crystalline			
	Fixed Amorphous			
	Fixed CdTe			
	Motorized Poly-Si			
	Motorized Mono-Si			

4. Conclusion

In light of the increasing adoption of photovoltaic stations, the demand for precise photovoltaic forecasting has significantly intensified. This is due to the potential of photovoltaic power energy forecasting to assist plant owners in proactively avoiding penalties, thereby resulting in a net profit for the plant. For this purpose, deep learning techniques are being employed to investigate enhanced precision. This article aimed to examine the LSTM efficiency in predicting PV power from different technologies based on one data type. For this purpose, our prediction methodology was developed and examined. Firstly, the dataset from two stations with different module technologies was collected and processed. Secondly, the model was built and optimized according to the data from the first PV plant. Thirdly, the developed LSTM was tested for all the technologies of the second PV plant, respectively: Mono-crystalline and Poly-crystalline silicon with motorized structures, Cadmium Telluride (CdTe), Amorphous Silicon, Mono, and Poly crystalline silicon with fixed structures. Finally, the evaluation of the performance metrics was realized, and the effectiveness of the LSTM model is strongly approved with the following metrics: RMSE $\leq$ 0.2090, NRMSE< 21.36%, r $\leq$ 0.9475, and MAE $\leq$ 0.1516

In summary, the main contributions of this paper can be summarized as follows: (i) the technology diversity: four different photovoltaic technologies were used for testing, broadening the scope of analysis beyond the confines of a single technology. (ii) Structure variability: we examined and compared data from two distinct PV station structures: fixed and tracker. (iii) Large-Scale System Analysis: Our models were developed using data from a large-scale PV system of 53 MW total capacity and tested for a second-small scale PV system of 1.1 MW under similar climatic conditions. (iv) Extended forecasting horizon: We employed a forecasting horizon of three hours (3 h); extending the timeframe offers valuable insights, especially in the context of energy compensation during peak hours, considering the startup durations of conventional power plants. Future work will be guided by exploring the impact of exogenous data such as temperature and solar radiation and boosting the accuracy of the model by utilizing hybrid deep learning algorithms.

CRediT authorship contribution statement

Ferial EL ROBRINI: Conceptualization; Investigation; Methodology; Writing - original draft; Software, Formal analysis; Visualization; Writing - review & editing. **Badia AMROUCHE:** Conceptualization, Project administration; Supervision; Methodology; Validation.

Declaration of Competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Tavarez, E.G. et al. (2023). https://www.iea.org/commentaries/large-consumers-can-help-drive-the-clean-energy-transition-but-they-need-better-data.

[2] Gandhi, O. et al. (2020). Review of power system impacts at high PV penetration Part I: Factors limiting PV penetration. *Solar Energy*. DOI: 10.1016/j.solener.2020.06.097.

[3] Sampath Kumar, D. et al. (2020). Review of power system impacts at high PV penetration Part II: Potential solutions and the way forward. *Solar Energy*. DOI: 10.1016/j.solener.2020.08.047.

[4] Raza, M.Q. et al. (2016). On recent advances in PV output power forecast. *Solar Energy*. DOI: 10.1016/j.solener.2016.06.073.

[5] Shab, G. (2022). Prévision multi-horizon de l'éclairement global horizontal pour la gestion intelligente du réseau électrique de distribution en région Occitanie.

[6] Yang, D. (2019). Making reference solar forecasts with climatology, persistence, and their optimal convex combination. *Solar Energy*. DOI: 10.1016/j.solener.2019.10.006.

[7] Ssekulima, E.B. et al. (2016). Wind speed and solar irradiance forecasting techniques for enhanced renewable energy integration with the grid: A review. *IET Renewable Power Generation*. DOI: 10.1049/iet-rpg.2015.0477.

[8] Cryer, J.D. and Chan, K.-S. Springer Texts in Statistics Time Series Analysis With Applications in R Second Edition.

[9] Huang, J. et al. (2013). Forecasting solar radiation on an hourly time scale using a Coupled AutoRegressive and Dynamical System (CARDS) model. *Solar Energy*. DOI: 10.1016/j.solener.2012.10.012.

[10] Benmouiza, K. and Cheknane, A. (2019). Clustered ANFIS network using fuzzy c-means, subtractive clustering, and grid partitioning for hourly solar radiation forecasting. *Theoretical and Applied Climatology*. DOI: 10.1007/s00704-018-2576-4.

[11] Khelifi, R. et al. (2020). Multi-step-ahead forecasting of daily solar radiation components in the Saharan climate. *International Journal of Ambient Energy*. DOI: 10.1080/01430750.2018.1490349.

[12] Benmouiza, K. and Cheknane, A. (2013). Forecasting hourly global solar radiation using hybrid k-means and nonlinear autoregressive neural network models. *Energy Conversion and Management*. DOI: 10.1016/j.enconman.2013.07.003.

[13] Rabehi, A. et al. (2020). Hybrid models for global solar radiation prediction: a case study. *International Journal of Ambient Energy*. DOI: 10.1080/01430750.2018.1443498.

[14] Kushwaha, V. and Pindoriya, N.M. (2019). A SARIMA-RVFL hybrid model assisted by wavelet decomposition for very short-term solar PV power generation forecast. *Renewable Energy*. DOI: 10.1016/j.renene.2019.03.020.

[15] Guermoui, M. et al. (2021). A Novel Hybrid Model for Solar Radiation Forecasting Using Support Vector Machine and Bee Colony Optimization Algorithm: Review and Case Study. *Journal of Solar Energy Engineering, Transactions of the ASME*. DOI: 10.1115/1.4047852.

[16] Geng, D. et al. (2023). A hybrid photovoltaic/wind power prediction model based on Time2Vec, WDCNN and BiLSTM. *Energy Conversion and Management*. DOI: 10.1016/j.enconman.2023.117342.

[17] Mansoor, M. et al. (2023). Hybrid forecasting models for wind-PV systems in diverse geographical locations: Performance and power potential analysis. *Energy Conversion and Management*. DOI: 10.1016/j.enconman.2023.117080.

[18] Feroz Mirza, A. et al. (2023). Hybrid Inception-embedded deep neural network ResNet for short and medium-term PV-Wind forecasting. *Energy Conversion and Management*. DOI: 10.1016/j.enconman.2023.117574.

[19] Keddouda, A. et al. (2023). Solar photovoltaic power prediction using artificial neural network and multiple regression considering ambient and operating conditions. *Energy Conversion and Management*. DOI: 10.1016/j.enconman.2023.117186.

[20] Wang, F. et al. (2020). A day-ahead PV power forecasting method based on LSTM-RNN model and time correlation modification under partial daily pattern prediction framework. *Energy Conversion and Management*. DOI: 10.1016/j.enconman.2020.112766.

[21] Sahin, G. et al. (2023). Predictive modeling of PV solar power plant efficiency considering weather conditions: A comparative analysis of artificial neural networks and multiple linear regression. *Energy Reports*. DOI: 10.1016/j.egyr.2023.09.097.

[22] Matera, N. et al. (2023). Hourly forecasting of the photovoltaic electricity at any latitude using a network of artificial neural networks. *Sustainable Energy Technologies and Assessments*. DOI: 10.1016/j.seta.2023.103197.

[23] Khelifi, R. et al. (2023). Short-Term PV Power Forecasting Using a Hybrid TVF-EMD-ELM Strategy. *International Transactions on Electrical Energy Systems*. DOI: 10.1155/2023/6413716.

[24] Mellit, A. Recurrent neural network-based forecasting of the daily electricity generation of a photovoltaic power system.

[25] Dairi, A. et al. (2020). Short-term forecasting of photovoltaic solar power production using variational auto-encoder driven deep learning approach. *Applied Sciences (Switzerland)*. DOI: 10.3390/app10238400.

[26] Bouchouicha, K. et al. (2021). Comparison of artificial intelligence and empirical models for energy production estimation of 20 MWp solar photovoltaic plant at the Saharan Medium of Algeria. *International Journal of Energy Sector Management*. DOI: 10.1108/IJESM-12-2019-0017.

[27] Hassan, M.A. et al. (2021). Ultra-short-term exogenous forecasting of photovoltaic power production using genetically optimized non-linear auto-regressive recurrent neural networks. *Renewable Energy*. DOI: 10.1016/j.renene.2021.02.103.

[28] Guermoui, M. et al. (2021). Forecasting intra-hour variance of photovoltaic power using a new integrated model. *Energy Conversion and Management*. DOI: 10.1016/j.enconman.2021.114569.

[29] Ziane, A. et al. (2021). Photovoltaic output power performance assessment and forecasting: Impact of meteorological variables. *Solar Energy*. DOI: 10.1016/j.solener.2021.04.004.

[30] Agga, A. et al. (2021). Short-term self consumption PV plant power production forecasts based on hybrid CNN-LSTM, ConvLSTM models. *Renewable Energy*. DOI: 10.1016/j.renene.2021.05.095.

[31] Cao, Y. et al. (2023). Multi-timescale photovoltaic power forecasting using an improved Stacking ensemble algorithm based LSTM-Informer model. *Energy*. DOI: 10.1016/j.energy.2023.128669.

[32] Liu, Q. et al. (2023). A cloud-based Bi-directional LSTM approach to grid-connected solar PV energy forecasting for multi-energy systems. *Sustainable Computing: Informatics and Systems*. DOI: 10.1016/j.suscom.2023.100892.

[33] Dhaked, D.K. et al. (2023). Power output forecasting of solar photovoltaic plant using LSTM. *Green Energy and Intelligent Transportation*. DOI: 10.1016/j.geits.2023.100113.

[34] Al-Ja'afreh, M.A.A. et al. (2023). An enhanced CNN-LSTM based multi-stage framework for PV and load short-term forecasting: DSO scenarios. *Energy Reports*. DOI: 10.1016/j.egyr.2023.08.003.

[35] Sadeghi, D. et al. (2023). Improving PV power plant forecast accuracy: A hybrid deep learning approach compared across short, medium, and long-term horizons. *Renewable Energy Focus*. DOI: 10.1016/j.ref.2023.04.010.

[36] Limouni, T. et al. (2023). Accurate one step and multistep forecasting of very short-term PV power using LSTM-TCN model. *Renewable Energy*. DOI: 10.1016/j.renene.2023.01.118.

[37] Abou Houran, M. et al. (2023). COA-CNN-LSTM: Coati optimization algorithm-based hybrid deep learning model for PV/wind power forecasting in smart grid applications. *Applied Energy*. DOI: 10.1016/j.apenergy.2023.121638.

[38] https://earth.google.com/web/@32.60044686,3.7014045,566.44499362a,420.37578165d,35y,30.85967317h,44.73348261t,0r/data=OgMKATA.

[39] https://globalsolaratlas.info/.