

A Proposed Model to Forecast Hourly Global Solar Irradiation Based on Satellite Derived Data, Deep Learning and Machine Learning Approaches

Badr Benamrou^{1*}, Mustapha Ouardouz¹, Imane Allaouzi², Mohamed Ben Ahmed²

¹ Department of Mechanical Engineering, Faculty of Sciences and Techniques, University Abdelmalek Essaadi, Tangier, Morocco

² Department of Computer Science, List laboratory, Faculty of Sciences and Techniques, University Abdelmalek Essaadi, Tangier, Morocco

* Corresponding author's email: badr.benamrou@gmail.com

ABSTRACT

An accurate short-term global solar irradiation (GHI) forecast is essential for integrating the photovoltaic systems into the electricity grid by reducing some of the problems caused by the intermittency of solar energy, including rapid fluctuations in energy, management storage, and the high costs of electricity. In this paper, the authors proposed a new hybrid approach to forecast hourly GHI for the Al-Hoceima city, Morocco. For this purpose, a deep long short-term memory network is trained on a combination of the hourly GHI ground measurements from the meteorological station of Al-Hoceima and the satellite-derived GHI from the neighbouring pixels of the point of interest. Xgboost, Random Forest, and Recursive Feature Elimination with cross-validation were used to select the most relevant features, the lagged satellite-derived GHI around the point of interest, as input to the proposed model where the best forecasting model is selected using the Grid Search algorithm. The simulation and results showed that the proposed approach gives high performance and outperformed other benchmark approaches.

Keywords: solar energy, forecast of hourly global solar irradiation, satellite derived GHI, deep learning

INTRODUCTION

Photovoltaic (PV) solar energy, like other renewable energies (such as wind energy) has traditionally been considered an unreliable source of energy due to its dependence on the weather conditions. The integration of photovoltaic solar energy into electrical networks requires a prior knowledge of the output power of photovoltaic systems, which is mainly related to the global solar radiation received on the module plan.

The incident radiation predictions can be grouped into two distinct types according to the following prediction horizons [Pavlovski A., Kostylev V. 2011]: (1) short-term predictions with a prediction horizon of one to three days. These predictions are related to the programming of power plants and the planning operations on the

electricity market. (2) Very short-term predictions with a horizon ranging from one to six hours aiming at contributing to the stability of the electricity grid by limiting the very fast fluctuations in the energy generated by a photovoltaic system.

Nowadays, there is a need to develop effective solar radiation forecasting methods for different horizons to address the problems resulting from the intermittency of solar energy such as: voltage fluctuation, grid management, continuity of the supply/demand balance, and costs of electricity production [Voyant C., Notton G., Kalogirou S., Nivet M-L., Paoli C., Motte F., Fouilloy A. 2016] in order to ensure network stability, local power quality, effective power system planning, and optimal storage management.

In this work, the authors presented their contribution to the task of a very short-term

forecasting of hourly global solar irradiance, especially with a one-hour horizon. For this purpose, a high-precision forecasting model based on deep short-term memory network (LSTM) and machine learning techniques using GHI ground measurements as well as satellite-derived data has been developed.

Therefore, to achieve this objective, the following aspects were considered:

- Development of a LSTM based model for one-hour ahead forecasting of GHI using the past measured data (endogenous data).
- Creation of a spatio-temporal grid of pixels surrounding the point of interest. Each pixel (i,j) corresponds to the time series of the satellite-derived GHI.
- Combination of the GHI ground measurements and the satellite-derived GHI around the point of interest.
- Identification of the most relevant surrounding pixels for different time lags using the Recursive Feature Elimination (RFE) approach.
- Optimization of the proposed deep Long Short-Term Memory network and other Machine Learning techniques with the Grid Search algorithm.
- Evaluation and comparison of the proposed model against the state-of-the-art models from the literature.

RELATED WORKS

The GHI time series predictions consist of the future GHI prediction based on historical data. These data can be measured at ground level, or extracted from satellite or sky images.

In recent years, a very large number of solar radiation prediction techniques have been developed and proposed in the literature. Machine learning and especially deep learning algorithms are the most widely adopted techniques, as they have proven their efficiency by achieving state-of-the-art results in most GHI forecasting tasks: In [Alzahrani A., Shamsi P., Dagli C., Ferdowsi M. 2017], the authors used a deep Recurrent Neural Network (RNN) to forecast the short-term solar irradiation in Ontario, Canada. The authors in [Xiangyun Q. and Yugang N. 2018] proposed a LSTM algorithm to predict hourly day-ahead solar irradiation. Their experimental results demonstrated that the proposed LSTM model outperformed multi-layered feedforward neural

networks (FFNN), linear least square regression, and persistence algorithm in the island of Santiago, Cape Verde. In [Yu Y., Cao J. and Zhu J. 2019], the study showed that the LSTM model has better prediction accuracy than the Support Vector Machine-Regression (SVR), ARIMA, and FFNN models under different weather conditions in New York and Atlanta. In [Loutfi H., Bernatchou A., Tadili R. 2017], the authors presented a comparison between a FFNN and a neural autoregressive with exogenous inputs (NARX) for generating the hourly global solar radiation in the city of Fes, Morocco. The results showed that the NARX with external inputs gives the best performances. In [Huang X., Shi J., Gao B., Tai Y., Chen Z. and Zhang J. 2019], the authors proposed a method that combines Wavelet Transform (WT) and Elman Neural Network (ENN) to predict the hourly solar irradiance in Kuning (China) and Denver (USA). The proposed method achieved better prediction accuracy compared with other methods such as SVR, BPNN, and persistence. In [Urraca, R., Antonanzas, J., Martinez, M.A., Martinez-de-Pison, F.J., Torres, F.A. 2016], the GHI is predicted one hour ahead in the south of Spain by using SVR and other machine learning techniques, the experiment results showed that the SVR gives a higher accuracy than Random Forest (RF) and k-nearest neighbors (KNN). In [Benali L., Notton G., Fouilloy A., Voyant C., and Dizene R. 2019], a comparative study of three methods in forecasting the hourly solar radiation for time horizons from h+1 to h+6 on the site of Odeillo, France, was discussed. The results demonstrated that RF outperformed smart persistence and FFNN. The authors in [Pan C. and Tan J., 2019] proposed a method based on cluster analysis and ensemble model to predict day-ahead hourly solar generation with high accuracy in Australia. In [Ji W. and Chan, C.K. 2011], the authors proposed a hybrid model that combines ARMA and Time Delay Neural Network (TDNN) to predict the hourly solar radiation. In [Crisosto C. Hofmann M., Mubarak R., Seckmeyer G., 2018], the authors developed a method which combines the advantages of using all-sky images and Levenberg–Marquardt Artificial Neural Network (LV-ANN) in order to predict GHI one hour ahead in Hannover, Germany. In [Ameen B., Balzter H., Jarvis C. and Wheeler J.S. 2019], the satellite derived dataset such as clear sky, top atmosphere (TOA), and other observed data are used as inputs to the FFNN model to accurately forecast

the hourly GHI in Iraq. In [Mazorra Aguiar L., Pereira B. Lauret, P. Díaz, F. David, M. 2016], the researchers proposed a FFNN trained on a combination of ground measurements, satellite derived data, and Numerical Weather Prediction (NWP) to improve the prediction of the intraday solar radiation in Gran Canaria, Spain. In [Benmouiza K., and Cheknane A. 2013], the authors combined an unsupervised k-means clustering algorithm and FFNN in order to improve the results on the forecasting the hourly global solar radiation task.

METHODOLOGY

In this section, the authors presented the methodology employed for building the proposed model for the task of One-Hour Forecasting of the Global Solar Irradiance. As shown in Figure 1, this methodology is based mainly on the following steps:

1. Data preparation.
2. Features extraction and selection (Recursive feature elimination).
3. LSTM optimization.
4. The choice of the best one-hour ahead GHI forecasting Model.

Data preparation

Dataset

In this study, the authors selected Al-Hoceima as the area of interest. This study area is geographically located in north-central Morocco on the Mediterranean coast. Its climate is Mediterranean, the summers are very hot and dry, and the winters are rainy and cold.

Ground measured data

The hourly GHI ground data was measured from a meteorological station located in Al-Hoceima, Morocco, between 2015 and 2017. In the study, the authors were only interested in the GHI measured during the day, so the values measured at night were excluded. In order to obtain a high quality of data, the authors also removed inconsistent, missing, noisy data, and the data format was standardized.

Satellite derived data set

The satellite-derived dataset of the solar irradiation has become an important source of information for various solar energy applications.

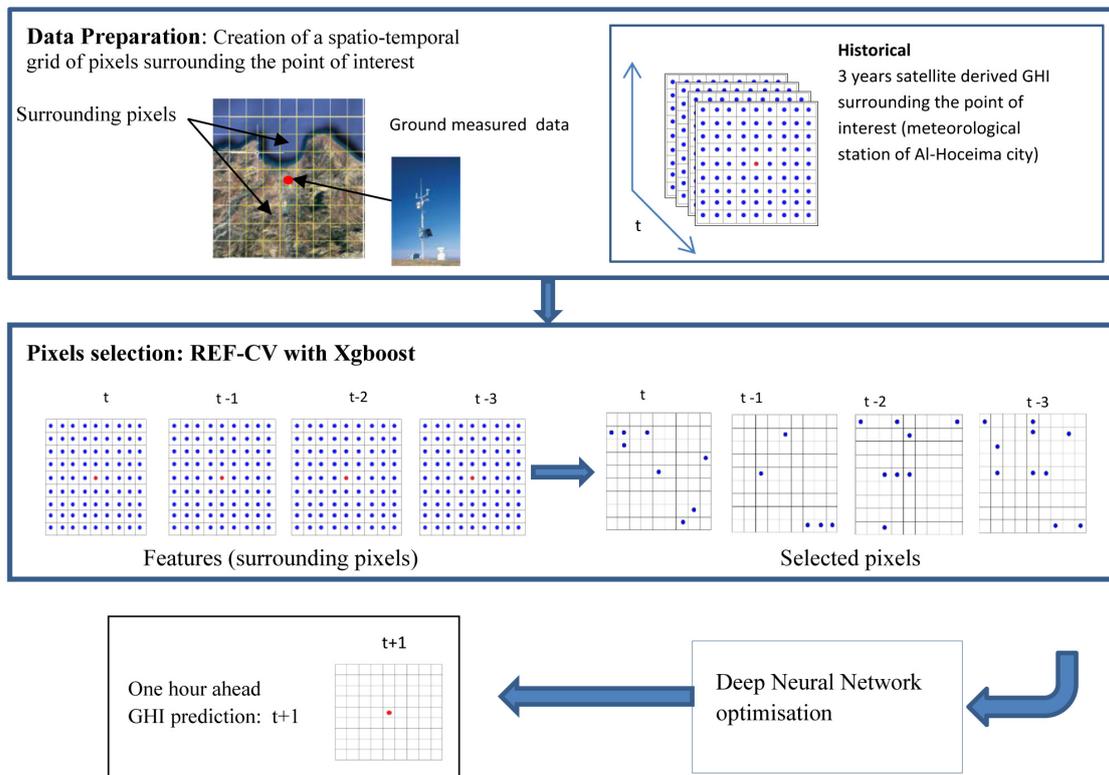


Figure 1. The proposed methodology

There are several datasets and services which provide the GHI data with various spatial and temporal resolutions. For example, Copernicus Atmosphere Monitoring Service (CAMS) provides time series of global horizontal irradiation derived from satellite images (Meteosat) using the new heliostat-4 method [Qu Z. et al. 2017].

In this study, CAMS was used to collect time series of hourly GHI for pixel points around the meteorological station of the Al-Hoceima city. For this purpose, an area located from 35.03 to 35.32 of latitude and -4.02 to -3.66 of longitude, resulting in a spatio-temporal grid of 9x9 pixels (see Figure 2) was selected. The located area consists of pixels surrounding the point of interest, where each pixel (i,j,t) represents a spatial resolution of $5 \times 5 \text{ km}^2$ with a temporal resolution of one hour. The distance from the station location to the surrounding pixels ranges from 5 km to 28 km, each pixel (i,j) corresponds to the time series of the satellite-derived GHI.

Feature extraction using Recursive Feature Elimination with Xgboost

Finding the optimal features to use for a learning model can sometimes be a difficult task to accomplish. In order to address this problem, Recursive Feature Elimination (RFE) can be employed. The RFE technique consists of recursively removing the weakest features, creating a model using the remaining features, and then evaluating its performance using the cross-validation (CV) technique. CV is a technique that aims at evaluating the machine learning models by training several models on the subsets of the available input data and evaluating them on the complementary subset of the data. Thus, the optimal and most important subset of features is then selected.

In this work, Extreme Gradient Boosting (Xgboost) [Chen T. and Guestrin C. 2016], Random

Forest [Breiman L. 2001], and Support Vector Regressor SVR were used as the basic machine learning models for the RFE method because they are among the most efficient algorithms used in feature engineering.

Xgboost is one of the successful and powerful machine learning algorithms that provide a fast and high performance results for most time series prediction problems. It is based on the gradient boosting algorithm [Freund Y. and Schapire R. 1997, Friedman J. 2001.] that consists mainly in training weak learners in parallel.

Random Forest is an ensemble method that uses the same fundamental principal as bagging and decision tree methods. It consists of a number of individual decision trees that operate as an ensemble. The output is the average output of individual trees, while SVR is the application of Support Vector Machine (SVM) [Vapnik V.N., 1995, Smola A.J., Scholkopf B., 2004] for the case of regression. SVR is a very powerful machine learning technique that solves complex models with high dimensionality and noisy data by applying the kernel functions.

The procedure of the RFE method is illustrated in Figure 3. As can be seen, RFE first trains all the surrounding pixels for time lags 1,2,3, and 4 hours ((GHI(i,j,t), GHI(i,j,t-1), GHI(i,j,t-2), GHI(i,j,t-3)) using the RF, SVR or Xgboost algorithm, the features are then sorted from high to low according to their importance and their forecasting contribution. Finally, the least important features are eliminated and the ML algorithm is retrained and evaluated on the new feature subset. This recursive procedure is repeated until the feature set is empty, which gives a list of the performance measurement values corresponding to each subset. On the basis of this list, the best subset of features to use in order to address the considered problem was selected.

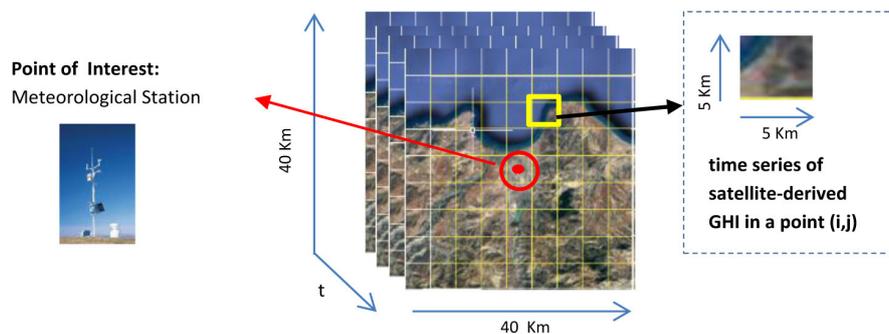


Figure 2. The spatio-temporal satellite-derived GHI around the meteorological station

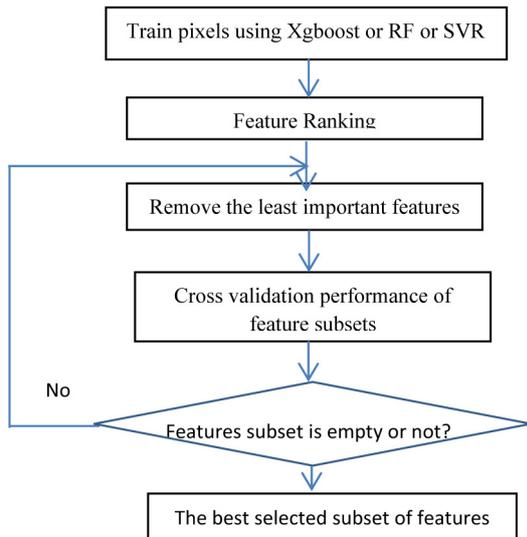


Figure 3. The RFE process

Forecasting the GHI one hour ahead

Forecasting the GHI one hour ahead is a sequential prediction problem that involves using a historical sequence of information to predict the next value in the sequence. For this purpose, a short-term memory network was used [Hochreiter S., Schmidhuber J., 1997], which is one of the most widely neural networks in time series prediction tasks because it allows remembering the information for long periods and facilitates the task of future prediction using periods of historical records.

Unlike traditional neural networks, instead of having neurons, the LSTM networks have memory blocks. These blocks facilitate the task of remembering values for long or short periods of time by taking into account the current input data, the short-term memory from the previous cell, and the long-term memory, usually known as the cell state.

Long-term short-term networks (LSTM) are an extension for recurrent neural networks, which extend their memory. Therefore, it is well-suited for learning important experiences that have very long delays between the two.

LSTMs allow RNNs to remember their inputs over a long period of time. This is because LSTMs store their information in a memory, which is very similar to the memory of a computer because the LSTM can read, write and delete the information from its memory.

This memory can be seen as a gated cell, where gated means that the cell decides to store or delete the information (for example if it opens

the gates or not), depending on the importance it assigns to the information. The attribution of the importance is done through weights, which are also learned by the algorithm. It just means that it learns over time what information is important and which is not.

An LSTM consists of a memory and three gates. The forget gate (forget) controls which part of the previous cell will be forgotten. The i (input) gate chooses the relevant information that will be transmitted to the memory. The output o (output) controls which part of the cell state will be exposed as a hidden state.

In order to compute the gates and the final LSTM output, the following equations were used:

$$f_t = \text{Sig}(W_f[h_{t-1}, x_t] + b_f)$$

$$\check{C}_t = \tanh(W_c[h_{t-1}, x_t] + b_c)$$

$$C_t = f_t \cdot C_{t-1} + \check{C}_t \tag{1}$$

$$i_t = \text{Sig}(W_i[h_{t-1}, x_t] + b_i)$$

$$O_t = \text{Sig}(W_o[h_{t-1}, x_t] + b_o)$$

$$h_t = O_t \cdot \tanh(C_t)$$

where: i_t represents the input gate, f_t represents the forget gate, O_t represents the output gate, C_t represents cell state, Sig represents sigmoid function.

W_x and b_x are respectively the weight and bias of a respective gate x.

h_t represents the output of the current LSTM block.

For the implementation of the LSTM, the Keras deep learning library was used [Schapire R.E., 1990]. Keras provides the tools for building and training high-level neural networks using as backend frameworks such as TensorFlow, CNTK, or Theano.

Evaluation of the performance

In this paper, the performance of the proposed models was evaluated on the basis of statistical parameters such as: Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and the coefficient of the determination (R^2).

Their formulas are given in the following equations:

Mean Absolute Error:

$$MAE = \frac{1}{N} \times \sum_{i=1}^N |\hat{y}(i) - y(i)|$$

Root Mean Square Error:

$$RMSE = \sqrt{\frac{1}{N} \times \sum_{i=1}^N (\hat{y}(i) - y(i))^2}$$

Coefficient of Determination:

$$R^2 = 1 - \frac{\sum_{i=1}^N (y(i) - \hat{y}(i))^2}{\sum_{i=1}^N (y(i) - \bar{y})^2}$$

where: \hat{y}_i is the predicted value and y_i is the actual value.

SIMULATION AND RESULTS

In this section, the authors discuss and present in detail the simulations and results obtained by the proposed one hour ahead GHI forecasting model.

The proposed one hour ahead GHI forecasting model is an LSTM network. For this model, different tuning parameter combinations were created and evaluated to obtain good estimates of how well each candidate performs. The best combination can be determined in various ways

but the most common approach is to pick the candidate with the empirically best results. Once the results are calculated, the best tuning parameter combination is chosen and the final model is fit to the entire training set with this value.

In this work, three scenarios were conducted to find the optimal forecasting model:

Scenario 1

The proposed model is trained and evaluated using the hourly GHI measured by the meteorological station of Al-Hoceima city between 2015–2017. To this end, an LSTM model is tuned to discover the hyperparameters that result in the most skilful predictions. As shown in Table 1, R^2 varies significantly between the five best combinations of parameters where the best setting had a corresponding R^2 of 0.848 (see Figure 4).

Scenario 2

For this scenario, a spatio-temporal grid of pixels surrounding the point of interest is created where each pixel corresponds to time series of the hourly satellite-derived GHI. Then, a LSTM model is trained and evaluated on the past ground

Table 1. The best five LSTM structures of the scenario 1

Model architecture		Hyperparameters					Time lag	R ²
Hidden Layer	Neurons in each hidden layer	Optimizer	Activation Function	Batch size	Learning rate	Epoch		
2	100–20	adam	Tanh	100	0.01	500	5	0.848
2	20–20	adam	Tanh	100	0.01	500	4	0.835
1	80	adam	Relu	100	0.01	500	3	0.824
2	30–60	adam	Tanh	20	0.01	500	4	0.828
2	20	adam	Relu	100	0.01	500	3	0.807

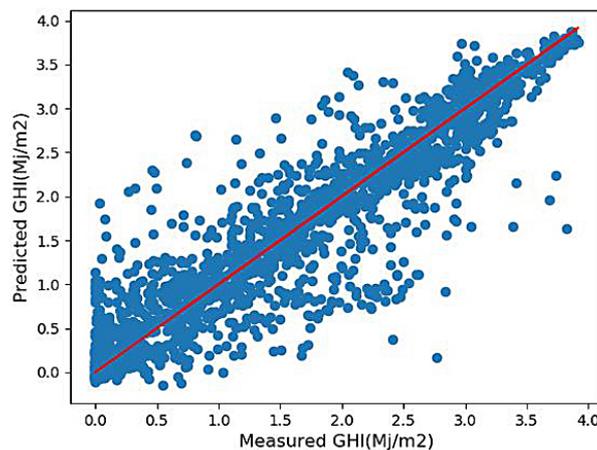


Figure 4. The measured and predicted GHI obtained by the best model of the scenario 1

measured GHI and surrounding pixel data located at a distance of 20 km in each direction from the station location for lags 1, 2, 3, and 4 hours, resulting in a total of 324 input features.

The table 2 shows the five best combinations of hyperparameters in terms of R^2 where the best model achieves a R^2 of 0.874 (see Figure 5).

Scenario 3

In this case, an LSTM model is trained and evaluated using the past ground measured GHI and surrounding pixels of the point of interest for lags 1, 2, 3, and 4 hours but this time instead of using all features as input, the RFE-CV technique is adopted to remove the weakest features and select the optimal features for which the model accuracy is the best.

In order to find the best basic regressor model of the RFE-CV, three ensemble machine learning algorithms that enable to automatically provide the estimates of feature importance from a trained predictive model are evaluated.

RFE-CV with Xgboost

A recursive Xgboost procedure is trained with 3-fold cross-validation to determine the initial importance scores as well as to remove the lowest importance features. As shown in Figure 6, for all of the surrounding pixels in different time lags

GHI (i,j,t) (324 features) a maximum CV score is reached at 32 pixels. It should also be noted that there does not seem to be much improvement in the CV score of the model after around 10 pixels.

The 32 optimal satellite-derived pixels selected by RFE-CV-Xgboost are distributed as follows: 8 pixels for time lag t, 5 pixels for time lag t-1, 8 pixels for time lag t-2, and finally 10 pixels for time lag t-3 as shown in Figure 8.

The selected pixels are employed as inputs to the network while the hyperparameter tuning is used to find the best LSTM model. Therefore, the five best combinations of hyperparameters in terms of R^2 are presented in the table 3 where the most accurate model achieves R^2 of 0.916 (see Figure 7).

RFECV with RF

In this case, REF-CV is combined with Random Forest in order to select the optimal features. Figure 9 shows the cross validation score of the different pixel subset sizes where the maximum cross validation score is reached at 54 features, which are distributed as follows: 28 pixels for time lag t, 12 pixels for time lag t-1, 7 pixels for time lag t-2 and finally 7 pixels for time lag t-4 (see Figure 11). It should be noted that the dimension of pixels selected by RFECV-RF is greater than that determined by RFECV-Xgboost.

Table 2. The best five LSTM models of the scenario 2

Model architecture		Hyperparameters					Time lag	R^2
Hidden Layer	Neurons in each hidden layer	Optimizer	Activation Function	Batch size	Learning rate	Epoch		
3	20–25–60	adam	relu	50	0.01	500	3	0.874
3	20–25–60	adam	relu	50	0.01	1000	5	0.858
2	20–50	adam	tanh	500	0.01	500	4	0.84
2	20–25	adam	tanh	50	0.01	1000	4	0.82
3	20–60	adam	tanh	20	0.01	500	5	0.801

Table 3. The best five LSTM models that use as inputs GHI ground measurements and surrounding pixels selected by RFE-CV-Xgboost

Model architecture		Hyperparameters					Time lag	R^2
Hidden Layer	Neurons in each hidden layer	Optimizer	Activation Function	Batch size	Learning rate	Epoch		
3	10–100–10	Adam	Relu	100	0.01	200	4	0.916
2	10–50	Adam	Relu	50	0.01	500	4	0.905
3	10–10–40	Adam	Relu	100	0.01	1000	4	0.896
2	20–30	Adam	tanh	100	0.01	1000	4	0.872
2	10–30	Adam	Tanh	50	0.01	1000	4	0.86

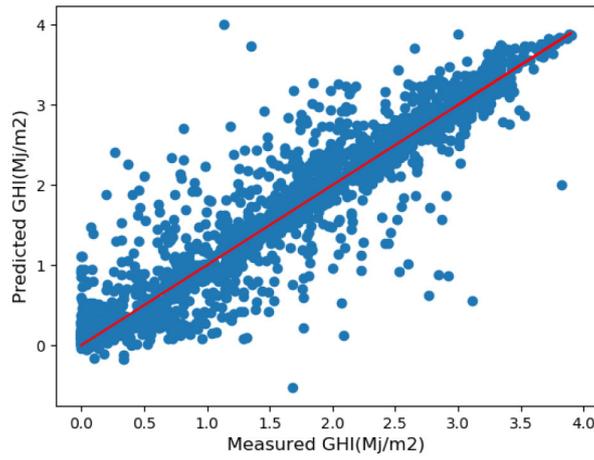


Figure 5. The measured and predicted GHI obtained by the best model of the scenario Scenario 3

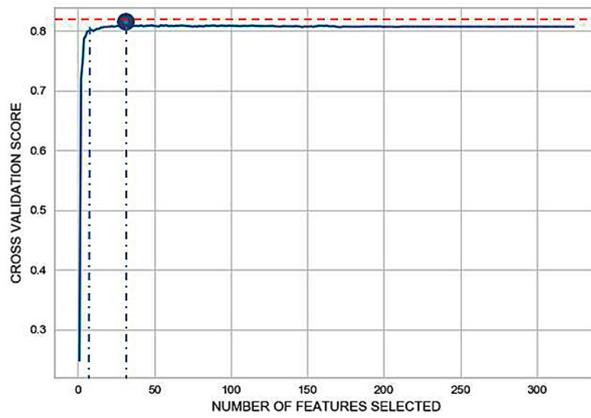


Figure 6. Feature selection using RFE-CV-Xgboost

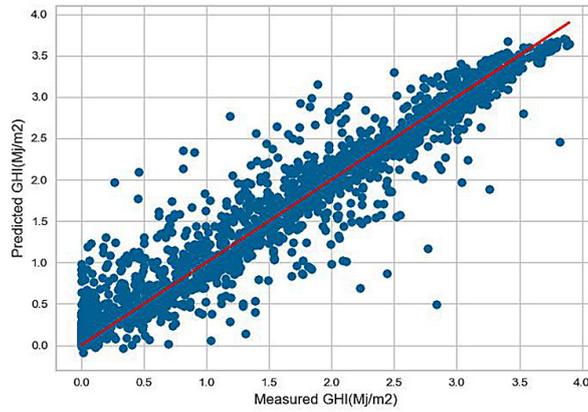


Figure 7. The measured and predicted GHI using the selected features from RFE-CV-xgboost

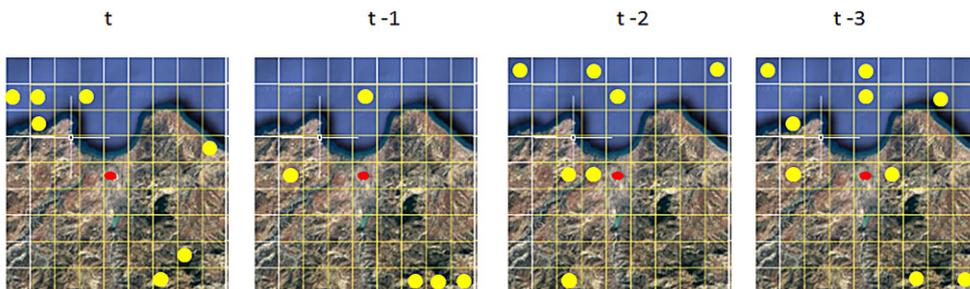


Figure 8. Distribution of the optimal selected features around the meteorological station for time lag t, t-1, t-2, and t-3

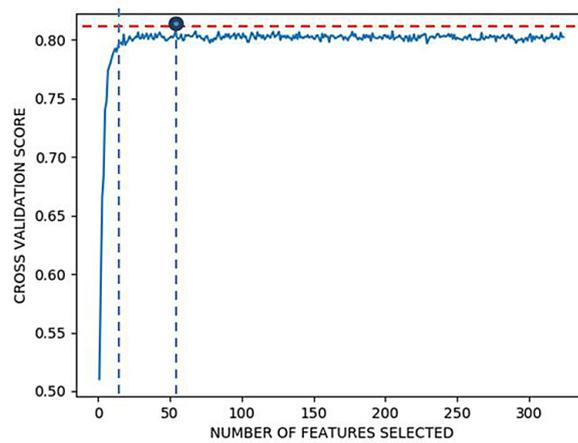


Figure 9. Results of Feature selection using RFE-CV-RF

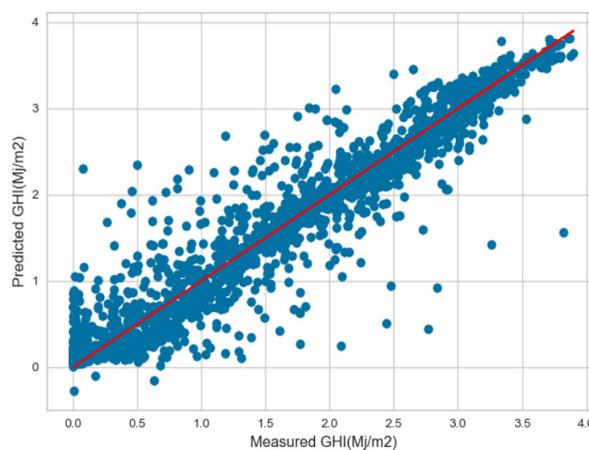


Figure 10. The measured and predicted GHI using selected features from RFE-CV-xgboost

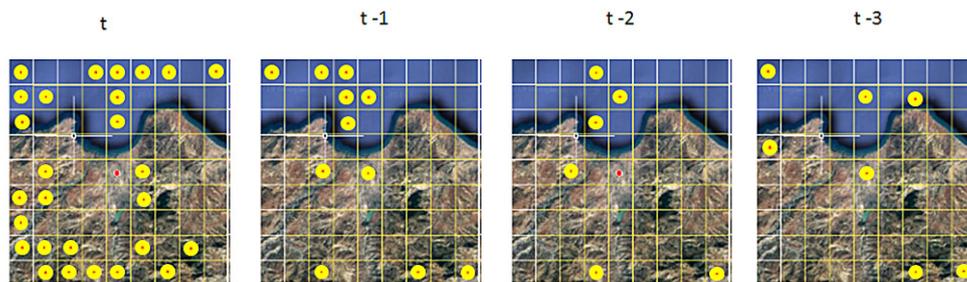


Figure 11. Distribution of the optimal selected features by RFE-CV-RF for time lag t , $t-1$, $t-2$ and $t-3$

Table 4. The best five LSTM models that use as inputs surrounding pixels and GHI ground measurement selected by RFE-CV-Random Forest

Model architecture		Hyperparameters					Time lags	R ²
Hidden Layer	Neurons in each hidden layer	Optimizer	Activation Function	Batch size	Learning rate	Epoch		
3	10-10-50	adam	Relu	100	0.01	500	4	0.908
3	10-10-10	adam	Relu	100	0.01	1000	4	0.894
2	10-10	adam	Relu	100	0.01	500	4	0.875
2	10-50	adam	Relu	50	0.01	500	4	0.85
2	30-10	adam	Relu	50	0.01	500	4	0.83

As shown in Figure 11, the GHI ground measurements and the satellite derived GHI around the point of interest are trained by LSTM. The Table 4 shows the five best combinations of hyper-parameters in terms of R^2 where the best model achieves a R^2 of 0.908 (Figure 10).

RFECV with SVR

In this section, the authors selected the optimal number of pixels with REF-CV combined with

Support Machine SVR. As shown in Figure 12, the best score is achieved with 36 features which are distributed as follows: 7 pixels for time lag t , 9 pixels for time lag $t-1$, 8 pixels for time lag $t-2$, and 12 pixels for time lag $t-3$ (see Figure 14).

After selecting the 36 optimal features, the LSTM network was optimized using hyper parameter tuning. The five best models with the optimal combinations of parameters are shown in Table 5, where the best result in terms of R^2 is 0.907 (Figure 13).

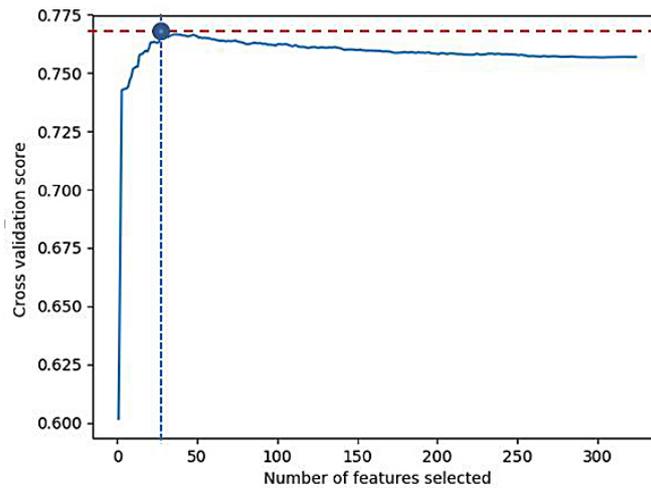


Figure 12. Feature selection using RFE-CV-SVR

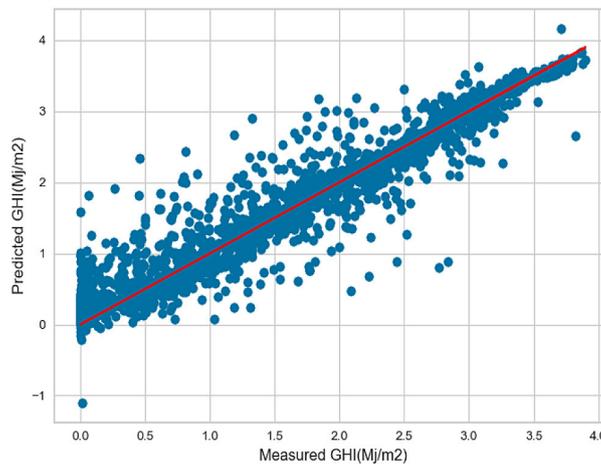


Figure 13. The measured and predicted GHI using selected features from RFE-CV-SVR

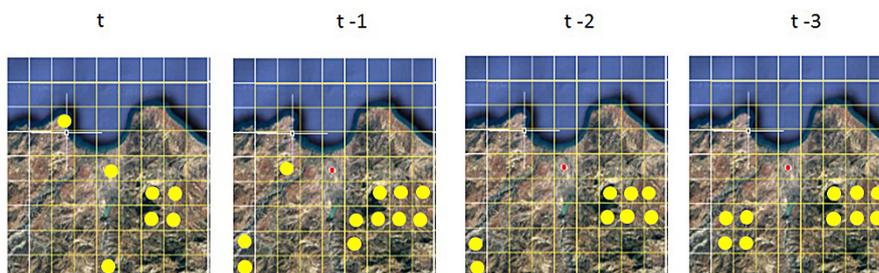


Figure 14. Distribution of the optimal selected features by RFE-CV-SVR for time lag t , $t-1$, $t-2$ and $t-3$

The best forecasting results

The comparison of the best results of each scenario (see Figure 15) shows the GHI measured by the meteorological station and the GHI predicted by the best models obtained in each scenario.

Table 6 summarizes the best results obtained from all scenarios conducted in this work to find the most accurate hourly GHI forecasting model for the considered case of study. These results are presented in terms of forecasting performance metrics including R^2 , RMSE, and MAE, where R^2 values range from 0.848 to 0.916. The RMSE values range from 0.424 to 0.29. and MAE from 0.17 to 0.28.

It is clear from the table that the lowest results were obtained using as inputs only endogenous GHI with the lowest R^2 of 0.848. the high errors RMSE of 0.42. and MAE of 0.28 . In turn, the results of scenario 2 prove that adding the satellite-derived GHI to the GHI ground measurements improved the prediction accuracy with a rate of 0.03%. However, the best results were obtained from scenario 3 with R^2 of 0.916 using the RFECV-Xgboost technique which selects only 32 optimal pixels around the target pixel for time lags 1.2.3. and 4 hours instead of all surrounding pixels, as in scenario 2.

After the comparison of all performance results. the authors concluded that the accuracy of

forecasting GHI one hour ahead model highly depends on:

1. Dataset: the forecasting model trained on the dataset composed of a combination of the GHI ground measurements and the satellite derived GHI provides better results than the one trained only with the GHI ground measurements.
2. Feature selection: the use of RFE-CV as a feature selection technique has significantly improved the performance results from a R^2 of 0.877 to 0.916 by eliminating the 292 less important features and selecting just 32 as the most relevant one. However. The choice of the best basic machine learning method for RFE is crucial because it impacts the performance results where the RFE based on Random Forest has attained a R^2 of 0.89 while the RFE based on Xgboost has achieved the best results with an R^2 of 0.916.
3. Forecasting method optimization: after the application of RECV-Xgboost, the selected pixels are trained with LSTM to predict the GHI one hour ahead. In order to optimize the deep LSTM network, a hyper parameter tuning is applied. This involves finding the best combination of hyperparameters (number of hidden layer. number of neurons in each hidden layer. optimizer. learning rate. activation function. epochs. and batch size) that provides a model with the highest prediction accuracy.

Table 5. The best five LSTM models that use as inputs surrounding pixels and GHI ground measurement selected by RFE-CV-SVR

Model architecture		Hyperparameters					Time lags	R^2
Hidden Layer	Neurons in each hidden layer	Optimizer	Activation Function	Batch size	Learning rate	Epoch		
3	10-40-40	adam	relu	100	0.01	500	4	0.907
3	10-10-60	adam	tanh	200	0.01	200	4	0.9
1	10	adam	relu	20	0.01	1000	4	0.892
1	10	adam	relu	50	0.01	200	4	0.896
1	15	adam	relu	200	0.01	100	4	0.884

Table 6. Performance metrics of the best model of each Scenario

Scenario	Model Input	RMSE	Mae	R^2
Scenario 1	Endogenous data	0.424	0.28	0.848
Scenario 2	Combining Endogenous and all surrounding pixels (324 features)	0.381	0.24	0.877
Scenario 3	Endogenous data and surrounding pixels selected by RFECV-RF (54 features)	0.324	0.20	0.908
	Endogenous and surrounding pixels selected by RFECV-Xgboost (32 features)	0.29	0.17	0.916
	Endogenous and pixels selected by RFECV-SVR (36 features)	0.327	0.21	0.907

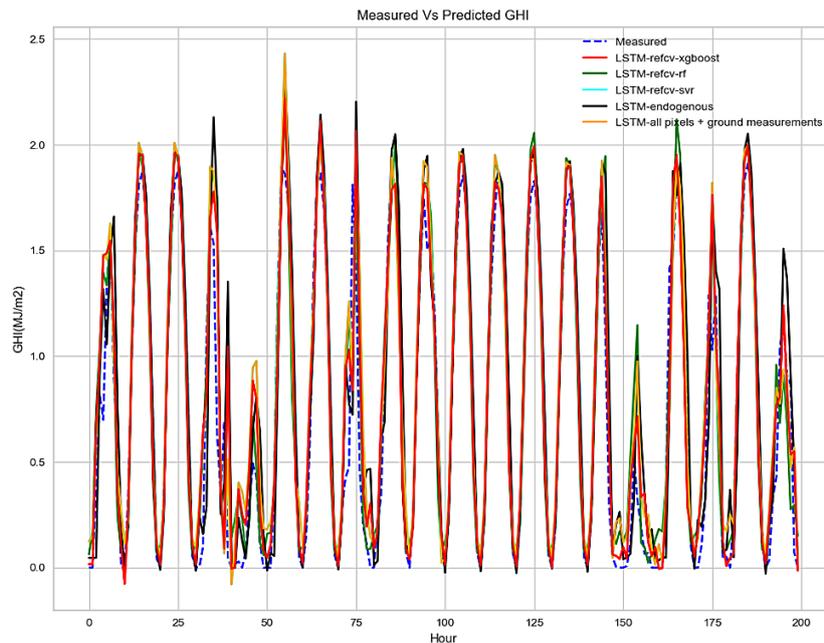


Figure 15. Measured and predicted GHI by the best models in each scenario

CONCLUSION

In this paper, the authors proposed an accurate model to forecast the Global Horizontal Irradiation one hour ahead in Alhoceima, north east of Morocco. For this purpose, different scenarios based on different Deep LSTM models and different inputs were realized. The most accurate model is obtained by combining the ground measurement and the most relevant surrounding satellite derived GHI for time lags 1, 2, 3 and 4 hours. RFE-CV that combined with Xgboost method was used to remove the weakest features and choose the most relevant one. Hyperparameter tuning was used to enhance the deep LSTM model in order to give the best results in the prediction accuracy. After analysing the results of the scenarios presented in this work, it was concluded that:

- The addition of surrounding satellite-derived GHI to the GHI ground measurements improved the prediction accuracy from R^2 0.848 to 0.916.
- The features selection methods such as RFE-CV combined with Xgboost, RFE-CV combined with RF, and RFE-CV combined with SVR are suitable and efficient for eliminating the weakest satellite derived GHI and selecting the most relevant ones in order to forecast GHI one hour ahead with high accuracy.
- Optimized deep LSTM model is very promising to forecast GHI one hour ahead.

- The surrounding pixels for time lags 1, 2, 3, and 4 hours selected by the RFE-CV with Xgboost have reached the highest prediction accuracy with R^2 of 0.916. RMSE with 0.28, and MAE of 0.17.

REFERENCES

1. Alzahrani A., Shamsi P., Dagli C., Ferdowsi M. 2017. Solar irradiance forecasting using deep neural networks. *Procedia Computer Science*. 114. 304–313.
2. Ameen B., Balzter H., Jarvis C. and Wheeler J.S. 2019. Modelling Hourly Global Horizontal Irradiance from Satellite-Derived Datasets and Climate Variables as New Inputs with Artificial Neural Networks. *Energies*. 12(1). 148.
3. Benali L., Notton G., Foulloy A., Voyant C., and Dizene R. 2019. Solar radiation forecasting using artificial neural network and random forest methods: Application to normal beam, horizontal diffuse and global components. *Renewable Energy*. 132. 871–884.
4. 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*. 75. 561–569.
5. Breiman L. 2001. Random forests. *Machine learning*. 45. 5–32.
6. Chen T. and Guestrin C. 2016. XGBoost: A scalable tree boosting system. In: *Proceedings of the*

- 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '16. New York, NY, USA. ACM. 785–794.
7. Crisosto C., Hofmann M., Mubarak R., Seckmeyer G., 2018. One-Hour Prediction of the Global Solar Irradiance from All-Sky Images Using Artificial Neural Networks. *Energies*. 11. 2906.
 8. Friedman J. 2001. Greedy boosting approximation: a gradient boosting machine. *Ann. Stat.* 29. 1189–1232.
 9. Freund Y. and Schapire R. 1997. A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. Syst. Sci.* 55. 119–139.
 10. Hochreiter S., Schmidhuber J., 1997. Long Short Memory. *Neural Computation*. 9. 1735–1780.
 11. Huang X., Shi J., Gao B., Tai Y., Chen Z. and Zhang J. 2019. Forecasting Hourly Solar Irradiance Using Hybrid Wavelet Transformation and Elman Model in Smart Grid. *IEEE Access*. 7. 139909–139923.
 12. Ji W. and Chan. C. K. 2011. Prediction of hourly solar radiation using a novel hybrid model of ARMA and TDNN. *Solar Energy*. 85(5). 808–817.
 13. Loutfi H. , Bernatchou A., Tadili R. 2017. Generation of Horizontal Hourly Global Solar Radiation From Exogenous Variables Using an Artificial Neural Network in Fes (Morocco). *international journal of renewable energy research*. 7. No.3. 1097–1107.
 14. Mazorra Aguiar L., Pereira B. Lauret. P. Díaz. F. David. M. 2016. Combining solar irradiance measurements, satellite-derived data and a numerical weather prediction model to improve intra-day solar forecasting. *Renewable Energy*. 97. 599–610.
 15. Pan C. and Tan J., 2019. Day-Ahead Hourly Forecasting of Solar Generation Based on Cluster Analysis and Ensemble Model. *IEEE Access*. 7. 112921–112930.
 16. Pavlovski A., Kostylev V. 2011. Solar power forecasting performance towards industry standards. 1st International Workshop on the Integration of Solar Power into Power Systems Aarhus, Denmark.
 17. Qu Z., Oumbe A., Blanc P., Espinar B., Gesell G., Gschwind B., Klüser L., Lefèvre M., Saboret L., Schroedter-Homscheidt M., and Wald. L. 2017. Fast radiative transfer parameterisation for assessing the surface solar irradiance: The Heliosat-4 method. *Meteorol. Z.* 26. 33–57.
 18. Schapire R.E., 1990. The strength of weak learnability. *Machine Learning*. 2. 5. 197–227.
 19. Smola A.J., Scholkopf B., 2004. A tutorial on support vector regression. *Statistics and computing* 14. 199–222.
 20. Urraca. R., Antonanzas. J., Martinez. M.A., Martinez-de-Pison. F.J., Torres. F.A. 2016. Smart baseline models for solar irradiation forecasting. *Energy Convers. Manag.* 108. 539–548.
 21. Vapnik V.N., 1995. *The nature of statistical learning theory*. New York: Springer.
 22. Voyant C., Notton G., Kalogirou S., Nivet M-L., Paoli C., Motte F., Fouilloy A. 2016. Machine Learning methods for solar radiation forecasting: a review. *Renew. Energy*. 105. 569–582.
 23. Xiangyun Q. and Yugang N. 2018. Hourly day-ahead solar irradiance prediction using weather forecasts by lstm. *Energy*. 148. 461–468.
 24. Yu Y., Cao J. and Zhu J. 2019. An LSTM Short-Term Solar Irradiance Forecasting Under Complicated Weather Conditions. *IEEE Access*. 7. 145651–145666.