



Multi Time Horizon Ahead Solar Irradiation Prediction Using GRU, PCA, and GRID SEARCH Based on Multivariate Datasets

Wadie Bendali^{1*}, Ikram Saber¹, Bensalem Bourachdi², Omar Amri³, Mohammed Boussetta⁴, Youssef Mourad⁴

¹Laboratory of Technologies and Industrial Services, Faculty of Sciences and Technologies Fez, University Sidi Mohammed Ben Abd Allah, Fez 30000, Morocco

²Team L2MC laboratory, National Graduate School of Arts and Crafts, Moulay Ismail University Mekenes, Mekenes 50000, Morocco

³Laboratory of Industrial Technologies, Faculty of Sciences and Technologies Fez, University Sidi Mohammed Ben Abd Allah, Fez 30000, Morocco

⁴Superior School of Technology, University Sidi Mohammed Ben Abd Allah, Fez 30000, Morocco

Corresponding Author Email: wadie.bendali@usmba.ac.ma

<https://doi.org/10.18280/jesa.550102>

ABSTRACT

Received: 30 December 2021

Accepted: 26 February 2022

Keywords:

dimensionality reduction, forecasting hyperparameters, gate recurrent unit, grid-search optimization, principal component analysis

Precise forecasting of solar irradiation helps to minimize photovoltaic power plant energy wastage, avoid system damage due to the high variability of solar irradiation, and focus the integration of power output among the various power grids. To forecast solar irradiation, it is crucial to consider the multiple dimensions of historical weather data as temperature, wind speed, and different type of irradiation, in addition to the categorical time data. The high dimensionality of these data can perturb the performance and can introduce very low calculation in a forecasting model. Hybrid combination between PCA and GRU with Grid Search hyperparameters optimization, proposed in this work to predict solar irradiation in different time horizons, using multiple variable data. Firstly, PCA changes the multiple variables to a few variables named components. Secondly, the prediction model GRU trained in optimized by using Grid search method. Finally, the optimized model predicts the solar irradiation. The proposed model compared in this study with simple GRU, LSTM, MLP and RNN models. The result of experience indicates that the PCA-GRU have a good forecasting accuracy in different time horizon, and has the better performance, and faster training compared with other models.

1. INTRODUCTION

Renewable energy sources (RES) are growing in the areas of innovative energy systems and their use is very significant considering their low Greenhouse Gas (GHG) emissions, their durability and their economic efficiency compared to other energy sources [1, 2]. They are used to take the place of the conventional form of production, for example, coal-fired power generation, etc. Within RES, wind and solar energy represent the highest degree of acceptability and the most desirable because of their inherent potential and availability. As the source of solar energy, the sun serves as a kind of natural black body emitter of radiation with a 5800 K surface temperature, providing 1367 W/m² of solar energy from above the atmosphere [3, 4]. One study indicates that we obtain about 1.8 × 10¹¹ MW of energy from the sun in an instant. World energy consumption today, nonetheless, represents less than the amount of energy coming from the sun [5].

One of the most widely used renewable energy systems for producing electrical energy is photovoltaic (PV). Used principally to convert solar energy directly into electrical energy for supply systems. Thanks to the availability of low cost maintenance, higher payback of installation costs in a given time period, and the important role it plays in the provision of clean and durable energy [6], the overall contribution of PV will grow rapidly in several countries [7].

At the same time, the production of solar energy is unreliable due to natural fluctuation. This can lead to the degradation of the reliability and stability of the grid systems, in addition to the economic benefit [1].

One of the major challenges for the industry today is the integration of these new kinds of electric power plants into the mains grid. In the absence of a proper managing of this volatile energy, problems of power reserve and frequency regulation will result [8]. To address these issues, advanced prediction methods have been developed to enable us prepare enough energy supply and respond to demand simultaneously, while ensuring the reliability of the system's frequencies. Furthermore, the forecasts also apply to power plant managers, the energy trading market and grid operators. Based on the knowledge of the future development of grid power generation, financial and operational uncertainties are minimized for all parts of the market.

Much research into solar energy forecasting has been carried out around the world in recent years, aiming to achieve higher accuracy and lower complexity in the calculation. Usually, most targeted by PV prediction methods is solar irradiation or PV power. It can be used as input of model with one variable, which is historical data of solar irradiation, or multiple variables in addition to solar irradiation as temperature, wind and other parameter related to target [9]. The forecasting of solar irradiation needs, sometimes Big-Data,

which can make a model low in term of computing. In addition, the features should be chosen according to its correlation with overall solar irradiation for minimizing the degree of freedom [10]. For dealing with these issues, it is important to apply dimension reduction to Data. In literature, there are so many methods could do this function. The seven most commonly used techniques for data-dimensionality reduction, including: Ratio of missing values [11], Low variance in the column values [12], High correlation between two columns [13], Principal component analysis (PCA) [14], Candidates and split columns in a random forest [15], Backward feature elimination [16], Forward feature construction [17]. Among all of these techniques PCA has special characteristics to deal with numeric columns as solar irradiation, temperatures, etc.

The forecast of solar irradiation signifies that PV output is forecasted one or several times in advance. Forecasting assists us in developing diverse energy system applications [18, 19]. A number of applications using forecasting in function of time horizon to enhance power system scheduling and operation are shown in Figure 1.

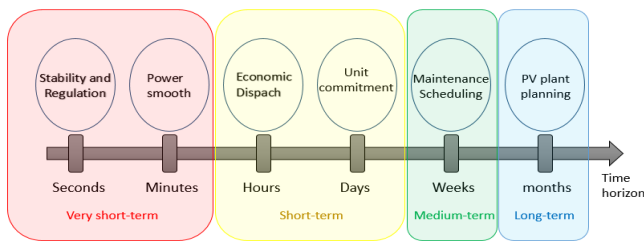


Figure 1. Required time resolution of prediction

A great deal of research has been carried out around the world in recent years on the prediction of solar irradiation. In literature, there are four types of solar irradiation or PV power generation forecasting models as indicate in Figure 2. Many statistical models for solar irradiance can be found in the publications as ARMA [20], ARIMA [21] and persistence model [22] and other methods [20]. The physical methods take into account meteorological and geological parameters through numerical weather prediction (NWP) as a result predict future parameters of interest using complex weather models [23, 24]. The accuracy of the physical forecast model is greater when the meteorological conditions are steady [25]. Artificial Intelligent (AI) models for forecasting have the ability to identify complex relationships linking the variables in the forecast with no need of complicated mathematics, for this reason, these models are much preferred to the other mentioned above [26]. Machine learning (ML) as one of AI methods has a good accuracy and performance in prediction of time series. In this case, there are many types of machine learning that it can be used for forecasting for example, Support Vector Machine (SVM), which is one of the supervised ML techniques, founded on the concept of Structural Risk Minimization (SRM). SRMs reduce a larger limit of anticipated loss. Consequently, the SVM is capable of reducing the error of the training models [27]. To speed up processing speed, the Extreme Learning Machine (ELM) boasts a training feature that is simple and resistant to blocking [28]. Artificial neural network (ANN) is a prediction method based on a big set of artificial neurons. These neurons are approximately identical with the nerve center of a biological brain. ANN applied in many fields, including signal and image processing, computer science and forecasting time series [29].

This method's successful use depends on the limitations of statistical methods in the treatment of non-linear data. They are composed of three main features: the neurons, the activation functions and the bias. The neurons can be either input, hidden or output neurons. Different ANN architectures are available, including the Multilayer Perceptron NN (MLPNN) [30], the Multilayer Feed Forward NN (MLFFNN) [31], the Radial Baseline Function NN (RBFNN) [32], the Recurrent NN (RNN) [30], and the Adaptive Neuro-Fuzzy Inference System (ANFIS) [32].

In AI, there are many developed methods, deeper than ML called Deep Learning (DL) methods that have the ability to model the data dynamics leads to successful results in a number of different applications. Gated Recurrent Unit (GRU) and Long Short-Term Memory (LSTM) [33] are an extended of RNN, these architectures could deal with the problem of vanishing gradient in RNN, and they have good performance in term of accuracy with sequence data. In addition, there are also other kinds of RNN architectures named Echo Stat Network (ESN) and deep ESN, which are coming from reservoir computing methods [34].

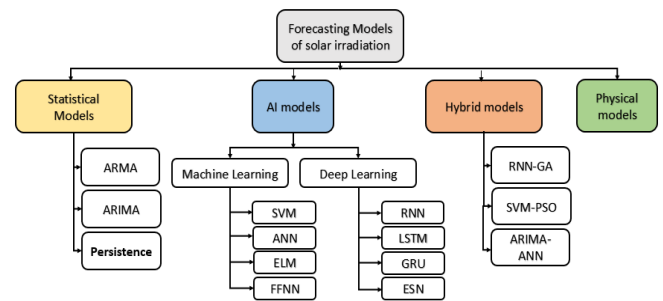


Figure 2. Different types of solar irradiation forecasting models

The combination of two or more methods called hybrid model. This mixing could improve the global forecasting accuracy of the hybrid by integrating the benefits of the different individual techniques. Some study found that one technique or method is not sufficient to meet the request to improve the precision and efficiency of solar irradiation forecasting for the system. Various methods have been combined by recent research to enhance the precision of solar irradiation forecasting, and demonstrated better performance compared to using a single method [35].

In general, in ML or DL, hyperparameters are relevant as they control the behaviour of the training algorithm directly, finding the best hyperparameters needs optimization method. In this field, there are many famous technics, among them, Grid Search [36], Random Search [37] and the Bayesian Optimization [38].

In this work, GRU is combined with PCA based on Grid-search optimization, as a new model of solar irradiation forecasting with multivariable data. GRU-PCA model tested in different time horizon and compared with different DL methods.

The rest of the paper is constructed as follows: after the introduction, description and analyses of data, including a general explanation of PCA for dimensionality reduction in section 2. Describing the DL model architectures and his mathematical equations in section 3. Description of Grid-search for hyperparameters optimization, in section 4. While in section 5, an illustration about application of models.

Section 6, discussion and results. Finally, the conclusion and the future research work.

2. DATA ENGINEERING

2.1 Time series data

A time series represent a dataset of observations measurements during a given period, describing a phenomenon [39]. The observation in this work is the global solar irradiation (GHI). Theoretically, there are two major assumptions required for a more precise analysis, firstly, the observation time interval is fixed, and secondly, the data pre-processing technique should be identical throughout the observation or measurement time.

Some definitions need to be illustrated before using time series equation. The actual value in t of the time series X is given as X_n , with t being the time interval from 1 to n , with n representing the whole number of observations. Then the number of predicted values of the time series is noted k . The prediction of the temporal series from $(n+1)$ to $(n+k)$, given the historical data from X_1 to X_n , it is named the prediction horizon (horizon 1 ... horizon k). Assuming horizon 1, the global forecasting model can be shown in Eq. (1).

$$X_{t+1} = F_n(X_t, X_{t-1}, \dots, X_{t-p+1}) + \varepsilon(t+1) \quad (1)$$

where, ε is the difference between the forecasted value and the observed value, F_n is the estimated mathematical model and t is the temporal parameter assuming the following $(n-p)$ values: $n, n-1 \dots p+1$, p ; p is the number of samples studied by the model, based on the hypothesis that: $n \gg p$.

2.2 Data scaling

The scaling of data input can greatly assist in reducing the training gap problem as well as the calculation cost by correctly training the historical pattern. Consequently, the predictive model accuracy can be greatly increased by pre-processing the row of input data. Several methods have been applied to scale the inputs of the predictive models [40]. Among these methods standardization [41], is very suitable method for scaling input data. Standardization is a scaling approach in which the values are focused at the mean using a unit of standard deviation. This signifies that the mean of the feature becomes zero and the distribution has a unit standard deviation. The formula of standardization represents as follow in Eq. (2):

$$I_{standar} = \frac{I - \mu}{\sigma} \quad (2)$$

where, μ is the mean of the variable values and σ is the standard deviation of the variable values. Unlike normalization method which scale the range values between $[0, 1]$ or $[-1, 1]$, standardization is not restricted to a certain range, it depends on the data values.

2.3 Principal Component Analyses

Principal Component Analysis or PCA is a dimensionality reduction technique frequently applied to decrease the column size of large datasets, by converting a large set of features into a shorter set containing the majority of the whole information

of the large set [14].

Reducing the number of variables of a data set naturally comes at the expense of accuracy, but the trick in dimensionality reduction is to trade a little accuracy for simplicity. Because smaller data sets are easier to explore and visualize and make analyzing data much easier and faster for ML or DL algorithms without extraneous variables to process [14].

The PCA method changes a big size matrix of linearly related variables into N_p linearly independent uncorrelated sets of PCs (dominant eigenvalues) collecting the majority of the data variance [42].

In order to apply the PCA, it is necessary first to use a sample covariance matrix Cov of features $\hat{F}_x(N \times N_v) \in R^{N \times N_v}$ that represent the vector of variables F_x divided by the mean of observations (where N are the number of observations and N_v are the number of variable) is computed as follows:

$$Cov = \frac{1}{N} \hat{F}_x^T (N_v \times N) \hat{F}_x (N \times N_v) \quad (3)$$

The goal consists in projecting F_x into $F_p \in R^{N \times P}$, where $N_v < N_p$, there by maximizing the variance contained in F_p . The information in F_x could be redundant, if the features N_v are correlated, in this case, Decomposition of the eigenvalue of the covariance matrix is carried out. To compute the eigenvectors and eigenvalues based on the correlation matrix. In this way, the eigenvalues are listed in descending order, and the N_p eigenvectors are chosen based on highest eigenvalues, which represent the number of principal components.

2.4 Data presentation

Table 1. Features description

Features	Abbreviation	Unites
Days of month (Exogenous)	-	-
Hours of Day (Exogenous)	-	-
Minute of hour (Exogenous)	-	-
Global horizontal irradiation	GHI	W/m ²
Wind speed (Exogenous)	WS	m/s
Air temperature (Exogenous)	Ta	°C
Irradiance of bean (Exogenous)	IBN	W/m ²
Height of sun (Exogenous)	Hs	Deg
Extraterrestrial radiation horizontal (Exogenous)	XH	W/m ²
Global radiation with raised horizon (Exogenous)	GRH	W/m ²
Clear sky global radiation (Exogenous)	CSG	W/m ²
Diffuse radiation horizontal (Exogenous)	DH	W/m ²

The data collected in Fez city from Morocco (33,3 °N, - 5,0 °E, Altitude=579 m), These data were collected using high-performance satellite data based on the Metronome 7 software platform, allowing users to access historical irradiation and temperature time series. The exogenous feature represents any variable that is independent of the predicted variable. In reality, the variation of global horizontal irradiance is influenced by many exogenous factors, the addition of these factors gives more information to the prediction model, which leads to more accurate and perform prediction model. In this case, the weather software gives us 9 normal characteristics in addition to the time variable, these characteristics are represented in Table 1.

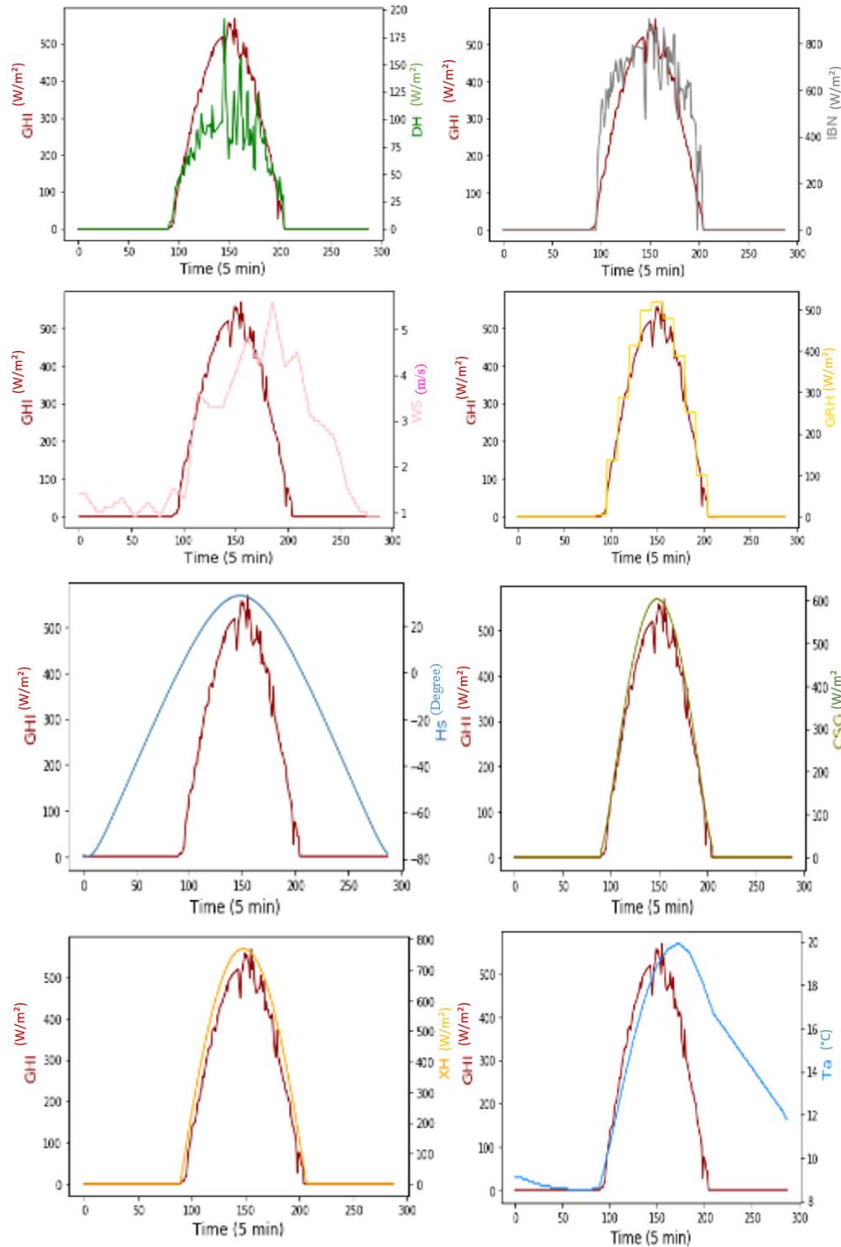


Figure 3. Different features and global horizontal irradiation pattern for a particular day

In this work, the data have time steps of 5 min, are used to predict solar irradiation in horizon of 5 min, 10 min, 20 min, 30 min and 1h, aiming to test and verify the efficiency and performance of proposed model in very short term and short term forecasting. Figure 3 illustrate relation with solar irradiation with other features in one day.

2.5 One hot encoder

One hot encoding is a technic used in ML and DL to transfer the categorical data into a 1 and 0 data [43]. The number of columns of this data represent the number of categories expected, regarding the same number of samples or rows in this data. Therefore, the one indicates in which class it is and everything else equals zero. In this work, the use of one hot encoding of time variables (minute of the hour – hour of day-day of month) as categorical features gives the model more flexibility in training and gives more information to the models [43]. However, one hot encoding product height dimensionality, in this case, to regard the effect of time variables on target, and reduce in the same time the dimension

of variables, the use of PCA will be very impotent.

2.6 Sliding windows

A noisy input can seriously disrupt the memory of prediction models. In order to give the model input greater stability, the previous time steps to the model input are included, by using sliding windows method [44] as shown in Figure 4. with $G(t)$ is a variable from the data and size means the size of sliding windows, in this work we choose windows size = 4.

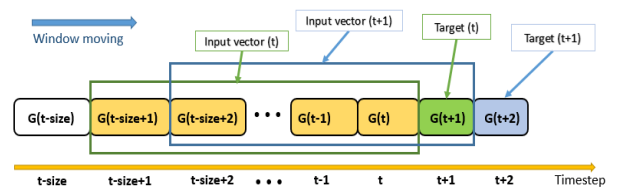


Figure 4. Sliding windows

3. DEEP LEARNING MODELS

In this part, definition, architectures and hyper-parameters of deep learning models have been illustrated, RNN models with their extension GRU and LSTM.

3.1 Recurrent neuron network

RNN, is a generation of deep learning neural network, precisely, it is a FFNN with an internal memory. By nature, the RNN is cyclic because it performs the identical activation function with any data input, whereas the output of the actual input is based on the most recent calculation. After generating the result, the output is automatically copied, and forwarded within the recurrent network. To decide, it considers the actual inputs and the output that it trained from the last input. RNNs, unlike FFNNs, are able to use their internal state (memory) to address input sequences. That makes it useful for functions like prediction time series data. All inputs in other neural networks are independent, on the other hand, there is a relation between inputs in RNN (Figure 5).

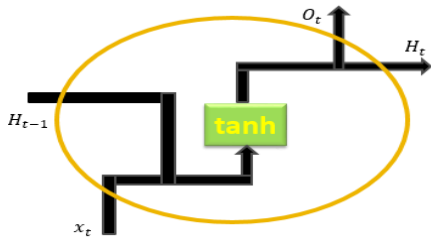


Figure 5. RNN architecture

$$H_t = \text{Tanh}(Ux_t + WH_{t-1}) \quad (4)$$

$$o_t = \text{SIG}(vH_t) \quad (5)$$

Under u is the input weight matrix, w is the past hidden state weight matrix, v is the output weight, H_t is the hidden state, and x_t is the input vector. Many works in this field illustrate that it is same disadvantage of simple RNN models, among them [45]:

- Vanishing gradient: This makes the network hard to train because supplementary layers are added to the neural networks by using certain activation functions.
- Exploding gradients: This problem Means that the gradients have a long time for training, which can make the model instable
- It is difficult to process very long sequences by using Tanh or Relu as an activation function.

3.2 Long short term memory

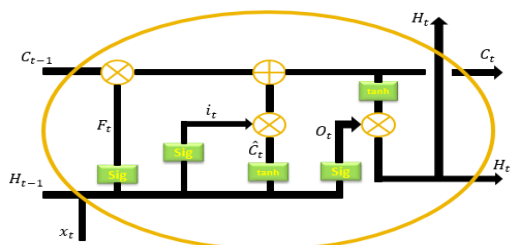


Figure 6. LSTM architecture

LSTM is a developed form of RNN, allowing easier recall of past data in memory. The problem of vanishing gradient is eliminated immediately here. LSTM is well adapted to predict time series with unknown delay times. The model is trained using back-propagation. LSTM has three gates as showed in Figure 6.

Input Gate - finds out which input value has to be used in order to change the memory. Using the Sigmoid function, it is possible to set the values from 0 to 1. Tanh function is used to give weight for variables, given a level of importance ranging from 1 to 1.

$$i_t = \text{Sig}(H_{t-1}W^i + x_tU^i) \quad (6)$$

Forget gate - find out what need to be removed from the block. Decisions are made by the sigmoid function. It examines preceding hidden state h_{t-1} and current content entry x_t to produce output of a number between 0 (omit this) and 1 (keep this) for every values in the C_{t-1} cell state.

$$f_t = \text{Sig}(H_{t-1}W^f + x_tU^f) \quad (7)$$

Output gate - inputs and block memory are needed to make decisions about the output. After this, a decision is made by Sigmoid function to make values between 0 and 1. Like input gate, output gate has Tanh function to know the level of importance in the range -1 to 1.

$$o_t = \text{Sig}(H_{t-1}W^o + x_tU^o) \quad (8)$$

The memory storing is according to the cell status or internal memory, having two stages to update, first stage, gate sigmoid helps to identify the values available to be updated, whereas the TanH equation creates a new candidate \tilde{C}_t , ended up being given to the State. Secondly, it relates the input layer with the new value vector to make an update of the state, based on the equations below:

$$\tilde{C}_t = \text{Tanh}(H_{t-1}W^g + x_tU^g) \quad (9)$$

$$C_t = \tilde{C}_t \times i_t + C_{t-1} \times f_t \quad (10)$$

After to the cell status C_t update, he used this equation to achieve this next step's hidden state:

$$H_t = \text{Tanh}(C_t) \times O_t \quad (11)$$

3.3 Gate recurrent unit

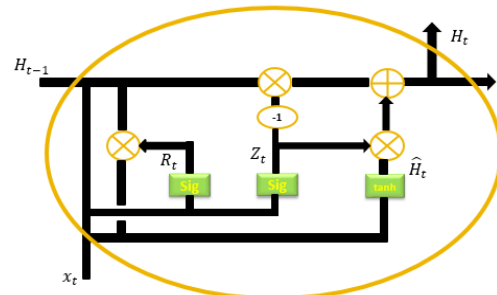


Figure 7. GRU architecture

GRU is another kind of RNN, which is identical to an LSTM with same difference between them. GRU discarded

the cells state and used the hidden state to forward information (Figure 7). Furthermore, there are only two available gates in the GRU, a reset and update gate:

- Update Gate: this gate functions in the same way as LSTM forget and input gates. Basically It makes the decision to throw away and add information.
- Reset Gate: it represents an important gate that is used to make decisions about how much information should be forgotten.

These two equations form the remainder and the update gate:

$$R_t = \text{Sig}(H_{t-1}W^r + x_tU^r) \quad (12)$$

$$Z_t = \text{Sig}(H_{t-1}W^z + x_tU^z) \quad (13)$$

To update the hidden state of GRU, two steps are necessary, first, it is important to calculate the hidden state of the reset gate that noted \widetilde{H}_t , and then use it to get H_t by producing with z_t as they are presented in Eq. (14) and Eq. (15):

$$\widetilde{H}_t = \text{Tan}((H_{t-1} \times r_t)W^h + x_tU^h) \quad (14)$$

$$H_t = (1 - z_t) \times H_{t-1} + z_t \times \widetilde{H}_t \quad (15)$$

3.4 Hyperparameters

In general, in DL, the parameters are the variables that are used to train the model on own during training, by adjusting the data to obtain the expected results. On the other hand, hyper-parameters have the characteristic that regulate the whole training process. They comprise variables that determine the organization structure of the network (as number of unites, number of layers, and number of sliding windows), and the variables which determine how the network is trained (for example, Learning Rate). Here is an illustration of some hyperparameters:

- Learning rate (LR):

The model will be of long duration to take a good state, because of too much small value of learning rate, or it would move beyond the best state, because of larger values of learning rate compared with optimal values. As a result, it is very important to choose your learning rate carefully, because the model would have several parameters, having each its own error curve, and the learning rate manages them all. LR has a small positive value, often between 0 and 1 [46].

- Mini-Batch Size:

The gradient descent is carried out with a loss function, which is generated by summing all single losses. The gradient

of individual losses can be computed in the same time, whereas it must be computed step by step sequentially in the case of stochastic gradient descent. It is preferable to use a constant number of examples (for example 16, 32 or 128) to construct the loss function than a single example or the whole data set, in order to create what is known as a mini-batch. The choose of adequate mini-batch size performs a gradient descent to achieve adequate stochasticity avoiding local minima [47].

- Dropout rate

Dropout is a regulation method to address the problem of overfitting of RNN models, based on choosing certain neurons at random and ignoring them during training. The dropout rate is therefore the probability of skipping a neuron in each weight adjustment run [48].

4. GRID SEARCH OPTIMIZATION

In general, the selection of best hyperparameter influence on the performance of DL algorithms [49]. Even though there are some rules of the road in the research community regarding the adequate values of these hyperparameters [50], It is necessary to use an optimization algorithm because the most appropriate values would be according to the nature of the data employed in the comparisons, the particular data sets used, the performance requirements, and several other factors. Grid Search is one such algorithm, allowing you to choose the optimal hyperparameters for your optimization problem based on a list of hyperparameters options that you supply [51]. As shown in Table, hyperparameters, which are optimized, are learning rate, dropout rate, Bache-size as shown Table 2.

Table 2. Hyperparameters ranges

Hyperparameter	Search Range
Learning rate	0.1, 0.01, 0.005, 0.001
Dropout rate	0.1, 0.2, 0.3, 0.4
Bache size	32,64,128,256

I. Application

DL models are built in the Python environment utilising TensorFlow with version 1.8.0 and KERAS 2.6.1 framework, PCA and grid search are built using Sklearn framework. All the simulation results were made with a personal computer with the configuration and environment: Intel Core(TM) i7-7700HQ CPU @ 2.8 GHZ, NVIDIA, GeForce GTX1050Ti graphics card.

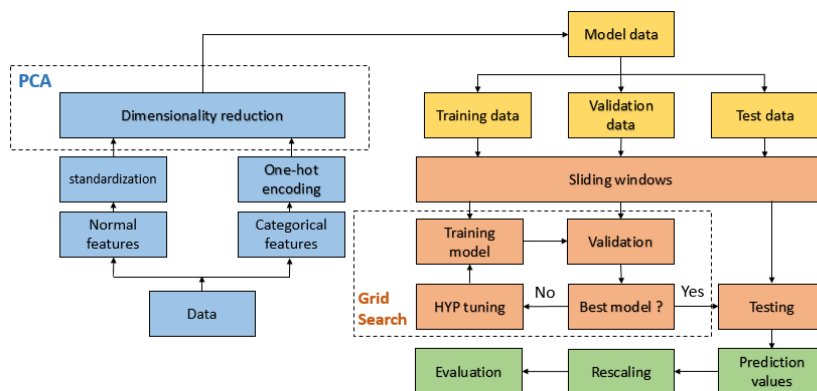


Figure 8. Flowchart of proposed method

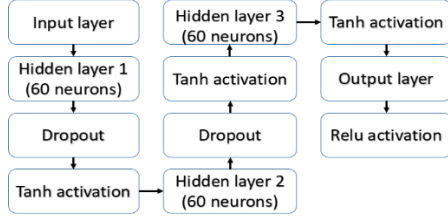


Figure 9. Proposed DL architectures

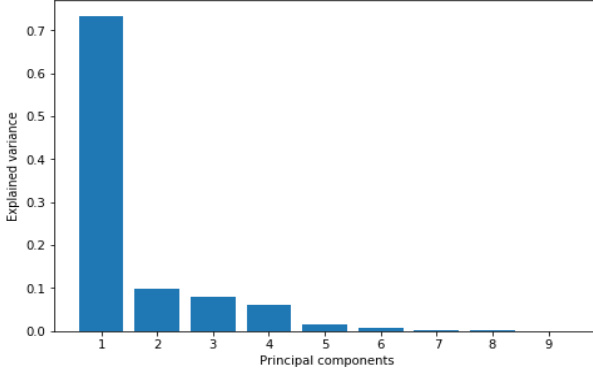


Figure 10. Explaining variance for each component

Three steps are used in the application of proposed models, preprocessing, processing and post processing (Figure 8).

During preprocessing, features of data devised into two types normal data, which have nine features of weather, and categorical features, which represent days of the month, hours of the day, and minutes of the hour. The normal data are scaled by using standardization, meanwhile, the categorical data are encoded by using one-hot-encoding, which give 68 variables. After that, the normal and the categorical data are concatenated for obtaining the new data of 77 variables, then the dimension of this data are reduced using the PCA method, which transforms the vectors or variables into one eigenvector that has the highest eigenvalue of the covariance matrix (Figure 10).

In processing, the Sliding Windows method is applied to the data, for obtaining the input and the target of the model, which divided into 60 % for training, 20 % for validation, and 20 %

for testing. At first, the model is trained and compared under different drop-out, Bach-size and learning rate values by using Grid-Search optimization. Then, the best models obtained in function of the comparison between validate data and out-put of the model, using MSE metrics equation (15). Concerning the models, which are compared with GRU-PCA model, are respectively MLP, RNN, LSTM, GRU as DL methods. The architectures of the models (Figure 9) contain three hidden layers, each one of them has 60 neurons and Tanh as activation functions, and output layer, which has one neuron and Relu activation function, the models are trained in 50 epochs using Bach gradient descent.

In post-processing step, forecast data are rescaled to the origin values and compared with test values. The performance of the models can be evaluated by using metric errors that measure the accuracy of the forecasting. In literature, there are several considerations about choosing a metric of evaluation error as indicated in [52]. The mean square error (MSE) and mean absolute error (MAE) are, however, extensively used in the research literature. Calculation equations for MSE and MAE are given in equations (16) and (17) respectively.

$$MSE = \frac{1}{N} \sum_{i=1}^N (I_{forecast}(t) - I_{true}(t))^2 \quad (16)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |I_{forecast}(t) - I_{true}(t)| \quad (17)$$

With N is the number of samples, $I_{forecast}(t)$ is the forecast solar irradiation at time t, and $I_{true}(t)$ is the true solar irradiation at time t.

5. RESULT AND DISCUSSION

5.1 Dimensionality reducing results

Figure 10, illustrate the correlation between different components, from this illustration, the best component in term of covariance used as multivariate input data. As a result, the number of principal components k is determined to be greater than or equal to 70%. According to Figure 10, the number of principal components is one.

5.2 Hyperparameters optimization results

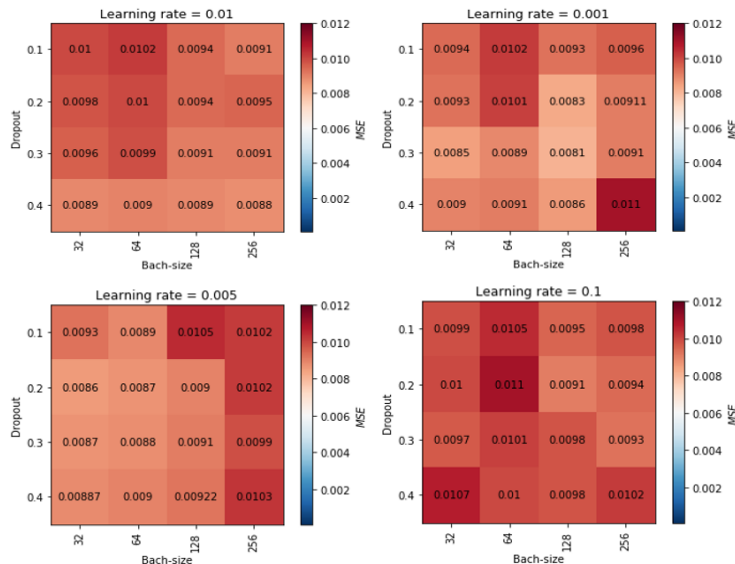


Figure 11. Grid search results for PCA-GRU

Table 3. Model hyperparameters selected for each DL model

	Time horizon	MLP	RNN	LSTM	GRU	PCA-GRU
Time training (s)	5 min	121	243	501	495	405

From Figure 11, it can be seen that the grid search find the best hyperparameters values of PCA-GRU, which are 0.3 for dropout, 128 for mini Bach size and 0.001 for learning rate, related with minimum values of MSE which is equal to 0.0081. Grid search is applied also to optimize other DL models, Table

3 describe the best hyperparameters of MLP, RNN, LSTM, GRU, PCA-GRU. Based on these results, the models are tested in evaluated.

5.3 Forecasting results

Table 4. Result of MSE and MAE in deferent time horizon for compared DL models Time training in time horizons of 5 min for compared DL models

		RNN	LSTM	MLP	GRU	GRU-PCA
MSE	5	0.0097	0.0061	0.0065	0.0051	0.0053
	10 min	0.00976	0.00671	0.0068	0.0059	0.0062
	20 min	0.00991	0.00765	0.0076	0.00762	0.0078
	30 min	0.012	0.0098	0.0099	0.0095	0.0097
MAE	5 min	0.048	0.036	0.041	0.0301	0.0304
	10 min	0.049	0.039	0.044	0.0341	0.0345
	20 min	0.052	0.042	0.0461	0.039	0.0411
	30 min	0.061	0.045	0.0483	0.039	0.0452

Table 5. Time training in time horizons of 5 min for compared DL models

	RNN	LSTM	MLP	GRU	GRU-PCA
Hidden units/layer	60	-	-	-	-
Number of layers	3	-	-	-	-
Optimization solver	Adam	-	-	-	-
Number of epochs	50	-	-	-	-
Drop-out	0.3	0.3	0.1	0.3	0.3
Learning-Rate	0.01	0.005	0.001	0.005	0.001
Bach-size	32	32	64	32	128
MSE Validation	0.012	0.0082	0.0089	0.0079	0.0081

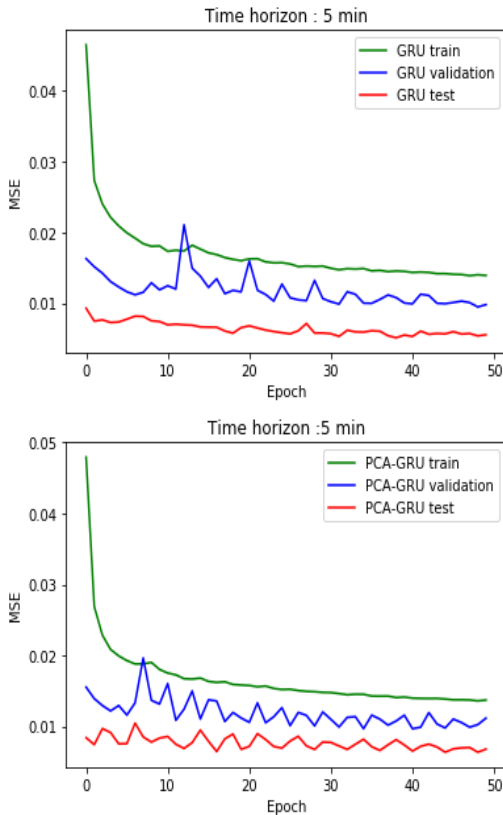


Figure 12. Training, validation and testing MSE of GRU and PCA-GRU according to the epoch for time horizon of 5 min

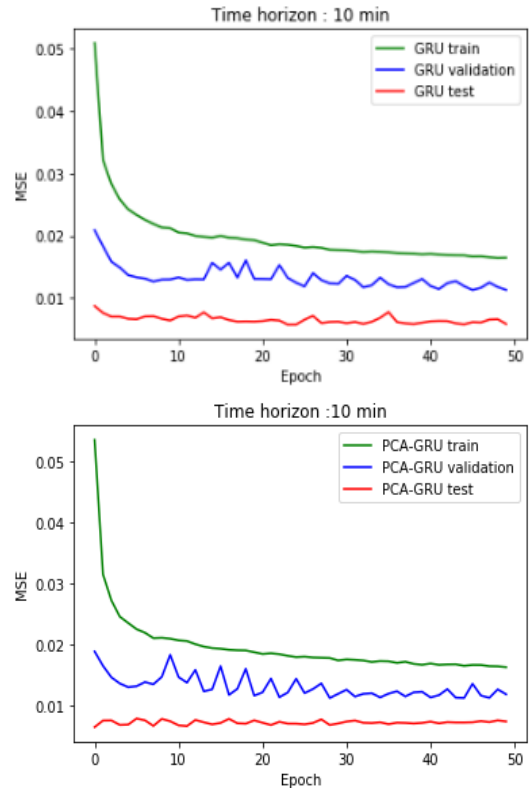


Figure 13. Training, validation and testing MSE of GRU and PCA-GRU according to the epoch for time horizon of 10 min

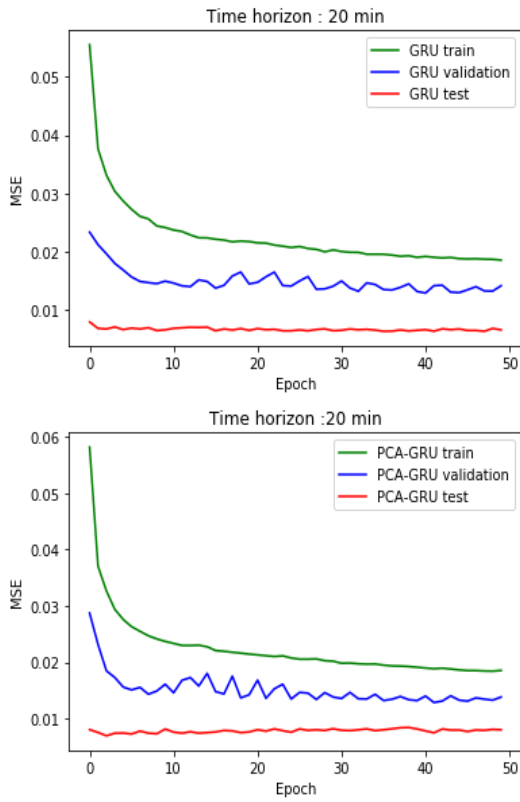


Figure 14. Training, validation and testing MSE of GRU and PCA-GRU according to the epoch for time horizon of 20 min.

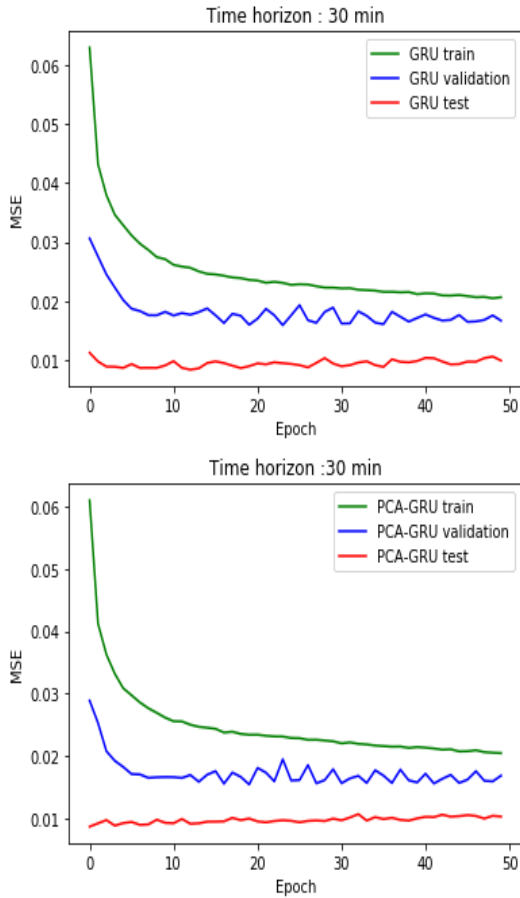


Figure 15. Training, validation and testing MSE of GRU and PCA-GRU according to the epoch for time horizon of 30 min

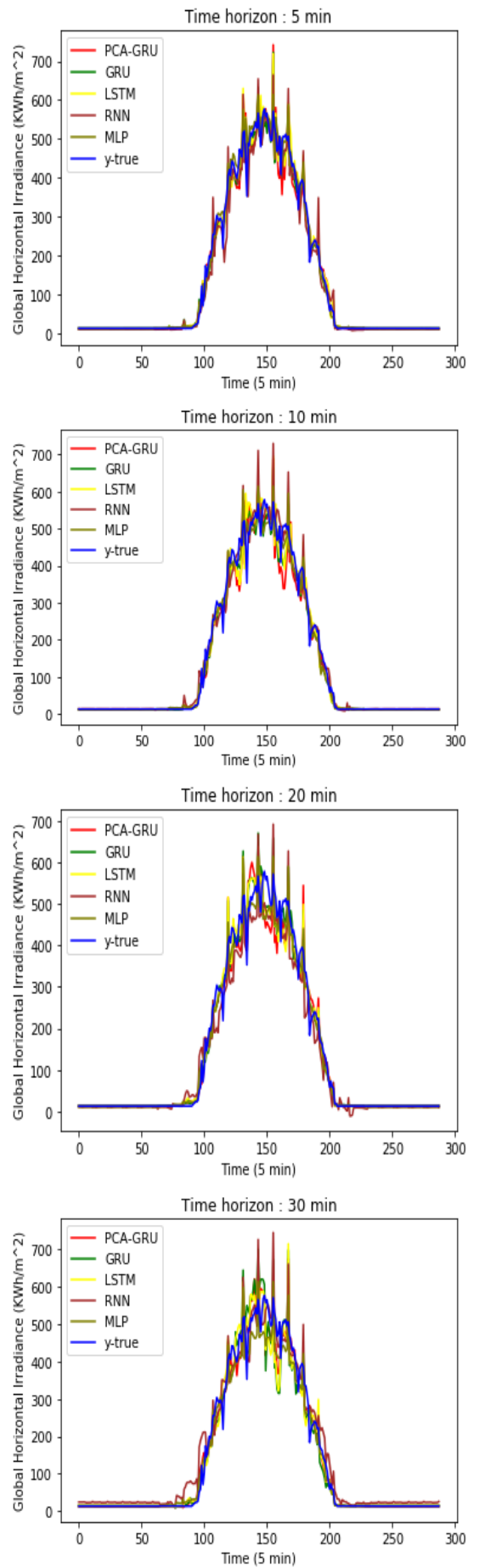


Figure 16. True values vs. forecasting values of PCA-GRU, GRU, LSTM, RNN, and MLP in different time horizon

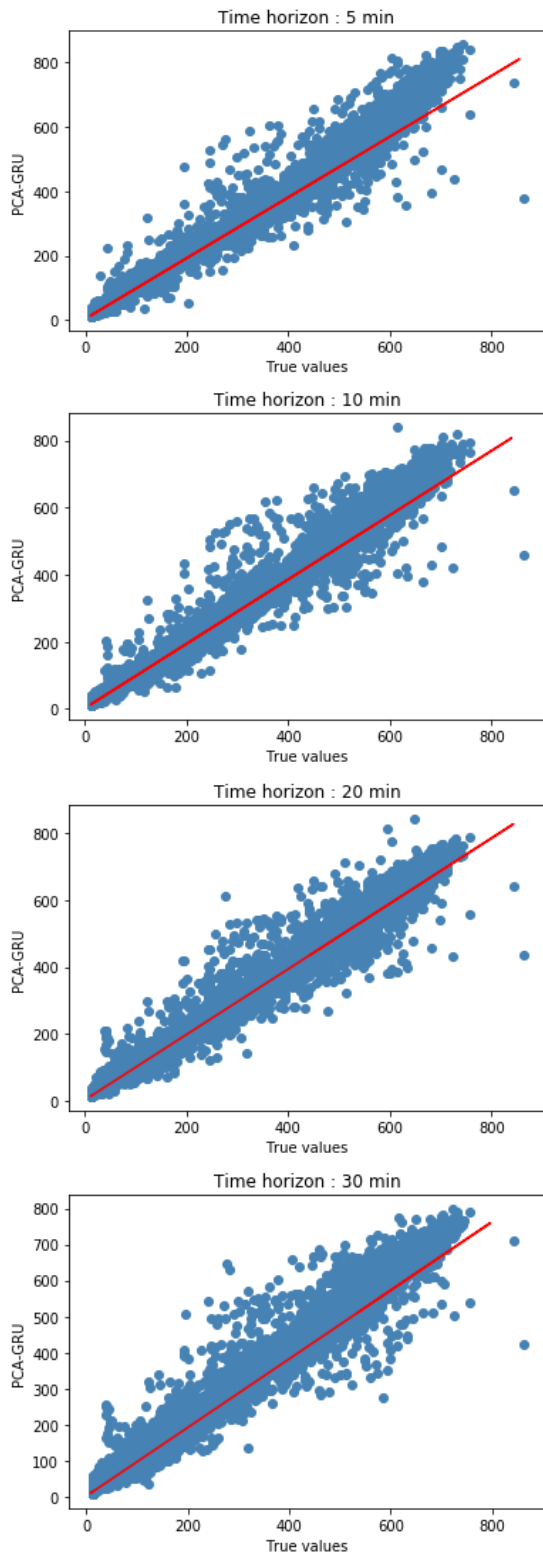


Figure 17. Scatter plots of the True values and predicted solar irradiance for the PCA-GRU in different time horizon

In order to evaluate proposed hybrid forecasting model in the different time horizon, experiments have been carried out in the following time horizons: 5 min, 10 min, 20 min and 30 min, accordingly, to forecast solar irradiation. The forecast accuracy was mostly greater for 5 minutes through the five comparative methods, and the accuracy progressively reduced from time horizon of 5 minutes, to 30 minutes. Generally, the proposed GRU-PCA model achieves the best forecasting compared with RNN, LSTM, MLP, and GRU models as it

illustrated in Figures 16, 17 and Table 5. Table 5 presents the specific evaluation error metrics of MAE and MAE for the five models in different time range. The experimental results in Figures 12, 13, 14, and 15 are shown that there is a small difference value between training, testing and validation MSE during the number of epochs, these figures can be explaining the advantage of PCA in front of overfitting. In Table 4, it can be seen that the dimensionality reducing of PCA allows the GRU model very fast in term of training.

6. CONCLUSIONS

The effective use of solar energy in smart-grids, gives an importance of solar irradiation forecasting in these systems. In this work, based on Keras frameworks, which has functionality to create the DL model, and Sc learn framework, which has data analyses and processing functions. The new hybrid DL model proposed to forecast solar irradiation namely, PCA-GRU using multivariable data, based on Grid Search for hyperparameters optimization. Most of the study focuses on the encoding of the categorical data by using the One-Hot-encoder, dimensionality reduction of the data based on PCA, DL models, and hyperparameters optimization by using Grid Search method. In particular, a PCA-GRU has a remarkable accuracy in dealing with multivariable data, surmounts the challenge of the high dimension of the original data, makes the training faster, and gives an accurate prediction result for solar irradiance in different time horizon, especially very short term and short term time horizon as illustrated in the results. The combination of DL prediction model with data mining technique substantially enhanced the accuracy of the energy production forecasting, making the grid more stable and reliable. Given that the model proposed is firstly applied to solar irradiation prediction, which means that it has achieved an unprecedented level of accuracy in solar irradiation forecasting with multivariable data. Furthermore, as future work, we will try to develop the model by extending the time horizon for making medium- and long-term solar irradiation predictions.

REFERENCES

- [1] Boussetta, M., El Bachtiri, R., Khanfara, M., El Hammoumi, K. (2017). Assessing the potential of hybrid PV-Wind systems to cover public facilities loads under different Moroccan climate conditions. *Sustainable Energy Technologies and Assessments*, 22: 74-82. <https://doi.org/10.1016/j.seta.2017.07.005>
- [2] Das, U.K., Tey, K.S., Seyedmahmoudian, M., Mekhilef, S., Idris, M.Y.I., Van Deventer, W., Stojcevski, A. (2018). Forecasting of photovoltaic power generation and model optimization: A review. *Renewable and Sustainable Energy Reviews*, 81: 912-928. <https://doi.org/10.1016/j.rser.2017.08.017>
- [3] Gueymard, C.A. (2004). The sun's total and spectral irradiance for solar energy applications and solar radiation models. *Solar Energy*, 76(4): 423-453. <https://doi.org/10.1016/j.solener.2003.08.039>
- [4] Saraçoğlu, N., Gündüz, G. (2009). Wood pellets—tomorrow's fuel for Europe. *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, 31(19): 1708-1718.

- <https://doi.org/10.1080/15567030802459677>
- [5] Sawin, J.L., Martinot, E., Sonntag-O'Brien, V., McCrone, A., Roussel, J., Barnes, D., Flavin, C. (2010). Renewables 2010 global status report. Chapter 1. Global Market Overview. 15-27.
- [6] Chaibi, Y., Allouhi, A., Salhi, M., El-jouni, A. (2019). Annual performance analysis of different maximum power point tracking techniques used in photovoltaic systems. *Protection and Control of Modern Power Systems*, 4(1): 1-10. <https://doi.org/10.1186/s41601-019-0129-1>
- [7] Motahhir, S., Chouder, A., El Hammoumi, A., Benyoucef, A.S., El Ghzizal, A., Kichou, S., Silvestre, S. (2020). Optimal energy harvesting from a multistrings PV generator based on artificial bee colony algorithm. *IEEE Systems Journal*, 15(3): 4137-4144. <https://doi.org/10.1109/JSYST.2020.2997744>.
- [8] Zhong, C., Zhou, Y., Yan, G. (2021). Power reserve control with real-time iterative estimation for PV system participation in frequency regulation. *International Journal of Electrical Power & Energy Systems*, 124: 106367. <https://doi.org/10.1016/j.ijepes.2020.106367>
- [9] Wang, L., Lu, Y., Zou, L., Feng, L., Wei, J., Qin, W., Niu, Z. (2019). Prediction of diffuse solar radiation based on multiple variables in China. *Renewable and Sustainable Energy Reviews*, 103: 151-216. <https://doi.org/10.1016/j.rser.2018.12.029>
- [10] Lin, P., Peng, Z., Lai, Y., Cheng, S., Chen, Z., Wu, L. (2018). Short-term power prediction for photovoltaic power plants using a hybrid improved Kmeans-GRA-Elman model based on multivariate meteorological factors and historical power datasets. *Energy Conversion and Management*, 177: 704-717. <https://doi.org/10.1016/j.enconman.2018.10.015>
- [11] Sim, J., Lee, J.S., Kwon, O. (2015). Missing values and optimal selection of an imputation method and classification algorithm to improve the accuracy of ubiquitous computing applications. *Mathematical Problems in Engineering*, 2015: Article ID 538613. <https://doi.org/10.1155/2015/538613>
- [12] Hu, Z., Bhatnagar, R. (2010). Algorithm for discovering low-variance 3-clusters from real-valued datasets. In 2010 IEEE International Conference on Data Mining, 236-245. <https://doi.org/10.1109/ICDM.2010.77>
- [13] Thangavel, K., Pethalakshmi, A. (2009). Dimensionality reduction based on rough set theory: A review. *Applied Soft Computing*, 9(1): 1-12. <https://doi.org/10.1016/j.asoc.2008.05.006>
- [14] Jolliffe, I.T., Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065): 20150202. <https://doi.org/10.1098/rsta.2015.0202>
- [15] Gong, H., Sun, Y., Shu, X., Huang, B. (2018). Use of random forests regression for predicting IRI of asphalt pavements. *Construction and Building Materials*, 189: 890-897. <https://doi.org/10.1016/j.conbuildmat.2018.09.017>
- [16] Maldonado, S., Weber, R. (2009). A wrapper method for feature selection using support vector machines. *Information Sciences*, 179(13): 2208-2217. <https://doi.org/10.1016/j.ins.2009.02.014>
- [17] Mahapatra, B., Patnaik, S. (2015). Data reduction in manets using forward feature construction technique. In 2015 International Conference on Man and Machine Interfacing (MAMI), pp. 1-3. <https://doi.org/10.1109/MAMI.2015.7456620>.
- [18] Lipperheide, M., Bosch, J.L., Kleissl, J. (2015). Embedded nowcasting method using cloud speed persistence for a photovoltaic power plant. *Solar Energy*, 112: 232-238. <https://doi.org/10.1016/j.solener.2014.11.013>
- [19] Lonij, V.P., Brooks, A.E., Cronin, A.D., Leuthold, M., Koch, K. (2013). Intra-hour forecasts of solar power production using measurements from a network of irradiance sensors. *Solar Energy*, 97: 58-66. <https://doi.org/10.1016/j.solener.2013.08.002>
- [20] Pappas, S.S., Ekonomou, L., Karamousantas, D.C., Chatzarakis, G.E., Katsikas, S.K., Liatsis, P. (2008). Electricity demand loads modeling using Autoregressive Moving Average (ARMA) models. *Energy*, 33(9): 1353-1360. <https://doi.org/10.1016/j.energy.2008.05.008>
- [21] Benvenuto, D., Giovanetti, M., Vassallo, L., Angeletti, S., Ciccozzi, M. (2020). Application of the ARIMA model on the COVID-2019 epidemic dataset. *Data in Brief*, 29: 105340. <https://doi.org/10.1016/j.dib.2020.105340>
- [22] Yang, X., Ren, J., Yue, H. (2016). Photovoltaic power forecasting with a rough set combination method. In 2016 UKACC 11th International Conference on Control (CONTROL), pp. 1-6. <https://doi.org/10.1109/CONTROL.2016.7737652>
- [23] Lara-Fanego, V., Ruiz-Arias, J.A., Pozo-Vázquez, A.D., Gueymard, C.A., Tovar-Pescador, J. (2012). Evaluation of DNI forecast based on the WRF mesoscale atmospheric model for CPV applications. In AIP Conference Proceedings, 1477(1): 317-322. <https://doi.org/10.1063/1.4753895>
- [24] Ma, L., Luan, S., Jiang, C., Liu, H., Zhang, Y. (2009). A review on the forecasting of wind speed and generated power. *Renewable and Sustainable Energy Reviews*, 13(4): 915-920. <https://doi.org/10.1016/j.rser.2008.02.002>
- [25] Soman, S.S., Zareipour, H., Malik, O., Mandal, P. (2010). A review of wind power and wind speed forecasting methods with different time horizons. In North American Power Symposium 2010, pp. 1-8. <https://doi.org/10.1109/NAPS.2010.5619586>
- [26] Lawan, S.M., Abidin, W.A.W.Z., Chai, W.Y., Baharun, A., Masri, T. (2014). Different models of wind speed prediction; a comprehensive review. *International Journal of Scientific & Engineering Research*, 5(1): 1760-1768.
- [27] Fan, J., Wang, X., Wu, L., Zhou, H., Zhang, F., Yu, X., Xiang, Y. (2018). Comparison of support vector machine and extreme gradient boosting for predicting daily global solar radiation using temperature and precipitation in humid subtropical climates: A case study in China. *Energy Conversion and Management*, 164: 102-111. <https://doi.org/10.1016/j.enconman.2018.02.087>
- [28] Tang, P., Chen, D., Hou, Y. (2016). Entropy method combined with extreme learning machine method for the short-term photovoltaic power generation forecasting. *Chaos, Solitons & Fractals*, 89: 243-248. <https://doi.org/10.1016/j.chaos.2015.11.008>
- [29] Abiodun, O.I., Jantan, A., Omolara, A.E., Dada, K.V., Mohamed, N.A., Arshad, H. (2018). State-of-the-art in

- artificial neural network applications: A survey. *Heliyon*, 4(11): e00938. [10.1016/j.heliyon.2018.e00938](https://doi.org/10.1016/j.heliyon.2018.e00938)
- [30] Das, U.K., Tey, K.S., Seyedmahmoudian, M., Mekhilef, S., Idris, M. Y.I., Van Deventer, W., Horan, B., Stojcevski, A. (2018). Forecasting of photovoltaic power generation and model optimization: A review. *Renewable and Sustainable Energy Reviews*, 81: 912-928. <https://doi.org/10.1016/j.rser.2017.08>
- [31] Ahmadi, N., Akbarizadeh, G. (2018). Hybrid robust iris recognition approach using iris image pre-processing, two-dimensional Gabor features and multi-layer perceptron neural network/PSO. *IET Biometrics*, 7(2): 153-162. <https://doi.org/10.1049/iet-bmt.2017.0041>
- [32] Zaman, M.H.M., Mustafa, M.M., Hussain, A. (2018). Estimation of voltage regulator stable region using radial basis function neural network. *Journal of Telecommunication, Electronic and Computer Engineering (JTREC)*, 10(2-8): 63-66.
- [33] Bendali, W., Saber, I., Boussetta, M., Mourad, Y., Bourachdi, B., Bossoufi, B. (2020, June). Deep learning for very short term solar irradiation forecasting. In *2020 5th International Conference on Renewable Energies for Developing Countries (REDEC)*, pp. 1-6. <https://doi.org/10.1109/REDEC49234.2020.9163897>.
- [34] Chembo, Y.K. (2020). Machine learning based on reservoir computing with time-delayed optoelectronic and photonic systems. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 30(1): 013111. <https://doi.org/10.1063/1.5120788>
- [35] Guermoui, M., Melgani, F., Gairaa, K., Mekhalfi, M.L. (2020). A comprehensive review of hybrid models for solar radiation forecasting. *Journal of Cleaner Production*, 258: 120357. <https://doi.org/10.1016/j.jclepro.2020.120357>
- [36] Kaur, S., Aggarwal, H., Rani, R. (2020). Hyperparameter optimization of deep learning model for prediction of Parkinson's disease. *Machine Vision and Applications*, 31(5): 1-15. <https://doi.org/10.1007/s00138-020-01078-1>
- [37] Marti, K. (2020). *Optimization Under Stochastic Uncertainty*. International Series in Operations Research & Management Science. Springer; 1st ed. 2020 edition.
- [38] Sun, D., Wen, H., Wang, D., Xu, J. (2020). A random forest model of landslide susceptibility mapping based on hyperparameter optimization using Bayes algorithm. *Geomorphology*, 107201. <https://doi.org/10.1016/j.geomorph.2020.107201>
- [39] Bourbonnais, R., Terraza M. (2008). *Analyse de séries temporelles en économie*. Paris, Dunod, 318 pages.
- [40] Schulz, M.A., Yeo, B.T., Vogelstein, J.T., Mourao-Miranada, J., Kather, J.N., Kording, K., Bzdok, D. (2020). Different scaling of linear models and deep learning in UKBiobank brain images versus machine-learning datasets. *Nature Communications*, 11(1): 1-15. <https://doi.org/10.1038/s41467-020-18037-z>
- [41] Han, J.H., Chi, S.Y. (2016). Consideration of manufacturing data to apply machine learning methods for predictive manufacturing. In *2016 Eighth International Conference on Ubiquitous and Future Networks (ICUFN)*, pp. 109-113. <https://doi.org/10.1109/ICUFN.2016.7536995>
- [42] Jolliffe, I.T. (2002). *Principal Components Analysis*; Springer: New York, NY, USA, 2002.
- [43] Yang, H., Pasupa, K., Leung, A.C.S., Kwok, J.T., Chan, J.H., King, I. (2020). Neural information processing. *Communications in Computer and Information Science*. <https://doi.org/10.1007/978-3-030-63823-8>
- [44] Lian, R., Huang, L. (2020). DeepWindow: Sliding window based on deep learning for road extraction from remote sensing images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13: 1905-1916. <https://doi.org/10.1109/JSTARS.2020.2983788>.
- [45] Gen, M., Cheng, R. (1997). *Genetic Algorithm and Engineering Design*. John & Wiley Sons, New York.
- [46] Attoh-Okine, N.O. (1999). Analysis of learning rate and momentum term in backpropagation neural network algorithm trained to predict pavement performance. *Advances in Engineering Software*, 30(4): 291-302. [https://doi.org/10.1016/s0965-9978\(98\)00071-4](https://doi.org/10.1016/s0965-9978(98)00071-4)
- [47] Vedaldi, A., Bischof, H., Brox, T., Frahm, J.M. (2020). *Computer Vision-ECCV 2020*. Lecture Notes in Computer Science. <https://doi.org/10.1007/978-3-030-58610-2>
- [48] Park, J.J., Loia, V., Pan, Y., Sung, Y. (2021). *Advanced multimedia and ubiquitous engineering*. Lecture Notes in Electrical Engineering. <https://doi.org/10.1007/978-981-15-9309-3>
- [49] Tsai, C.W., Hsia, C.H., Yang, S.J., Liu, S.J., Fang, Z.Y. (2020). Optimizing hyperparameters of deep learning in predicting bus passengers based on simulated annealing. *Applied Soft Computing*, 88: 106068. <https://doi.org/10.1016/j.asoc.2020.106068>
- [50] Khalid, R., Javaid, N. (2020). A survey on hyperparameters optimization algorithms of forecasting models in smart grid. *Sustainable Cities and Society*, 61: 102275. <https://doi.org/10.1016/j.scs.2020.102275>
- [51] Pontes, F.J., Amorim, G.F., Balestrassi, P.P., Paiva, A.P., Ferreira, J.R. (2016). Design of experiments and focused grid search for neural network parameter optimization. *Neurocomputing*, 186: 22-34. <https://doi.org/10.1016/j.neucom.2015.12.061>
- [52] Spüler, M., Sarasola-Sanz, A., Birbaumer, N., Rosenstiel, W., Ramos-Murguialday, A. (2015). Comparing metrics to evaluate performance of regression methods for decoding of neural signals. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 1083-1086. <https://doi.org/10.1109/EMBC.2015.7318553>

NOMENCLATURE

AI	Artificial Intelligent
ANN	Artificial Neuron Network
ANFIS	Adaptive Neuro-Fuzzy Inference System
ARMA	Autoregressive moving average
ARIMA	Autoregressive integrated moving average
PV	Photovoltaic
DL	Deep Learning
ML	Machine Learning
MLPNN	Multi layer perceptron neuron network
MLFFNN	Multi Layer Feed Forward Neuron Network
GHI	Wind speed
WS	Air temperature
Ta	Irradiance of bean
IBN	Height of sun
Hs	Extraterrestrial radiation horizontal

XH	Global radiation with raised horizon	ESN	Principal component analysis
GRH	Clear sky global radiation	PCA	Deep Echo state network
CSG	Diffuse radiation horizontal	DeepESN	Support vector machine
DH	Recurrent neuron network	SVM	Extreme learning machine
RNN	long-short term memory	ELM	Radial Baseline Function neuron network
LSTM	Gate recurrent unit	RBFNN	Recurrent neuron network
GRU	Multi-layer perceptron	RNN	Renewable energy sources
MLP	Feed forward neuron network	RES	Numerical Weather Prediction
FFNN	Echo State network	NWP	Global horizontal irradiation