engineering proceedings

Proceeding Paper

A Very Short Term Photovoltaic Power Forecasting Model using LDA Method and Deep Learning based on Multivariate Weather datasets †

Zemouri Nahed [1,*], Mezaache Hatem [2] and ChouderAissa [3]

[1] Department of electronics, University of Mohamed Boudiaf M'sila
[2] Electrical Engineering Laboratory (LGE), Department of electronics University of Mohamed Boudiaf M'sila; hatem.mezaache@univ-msila.dz
[3] Electrical Engineering Laboratory (LGE), University Mohamed Boudiaf of M'sila; aissa.chouder@univ-msila.dz
* Correspondence: nehed.zemouri@univ-msila.dz
† Presented at the 4th International Electronic Conference on Applied Sciences, 27 October–10 November 2023; Available online: https://asec2023.sciforum.net/.

Abstract: Photovoltaic (PV) system-generated solar energy has inconsistent and variable properties, which makes controlling electric power distribution and preserving grid stability extremely difficult. A photovoltaic (PV) system's performance is profoundly affected by the amount of sunlight that reaches the solar cell, the season of the year, the ambient temperature, and the humidity of the air. Every renewable energy technology, sadly, has its problems. As a result, the system is unable to function at its highest or best level. To combat the unstable and intermittent performance of solar power output, it is essential to achieve a precise PV system output power. This work introduces a new approach to enhancing accuracy and extending the time range of very-short-term solar energy forecasting (15 min step ahead) by using multivariate time series inputs in different seasons. First, Linear Discriminat Analysis (LDA) is used to select the relevant factors from the mixed meteorological input data. Secondly, two very short-term deep learning prediction models, CNN and LSTM, are used to predict PV power for a shuffled and reduced database of weather inputs. Finally, the predicted outputs from the two models are combined using classification strategy. The proposed method is applied to one year of real data collected from a solar power plant located in southern Algeria, to demonstrate that this technique can improve the forecasting accuracy compared to other techniques, as determined by statistical analysis involving normalized root mean square error (NRMSE), mean absolute error (MAE), mean bias error (MBE), and determination coefficient. (R^2).

Keywords: very-short-term; solar power forecasting; deep learning; aggregation; weather prediction

Citation: Nahed, Z.; Hatem, M.; ChouderAissa A Very Short Term Photovoltaic Power Forecasting Model using LDA Method and Deep Learning based on Multivariate Weather datasets. *Eng. Proc.* **2023**, *52*, x. https://doi.org/10.3390/xxxxx

Academic Editor: Simeone Chianese

Published: date

1. Introduction

Nowadays, solar energy has become a crucial element in electricity production systems worldwide. For the proper functioning and cost-effective integration of solar power, accurate photovoltaic power forecasting is essential [1]. Several studies have been conducted, and different methods have been suggested in the literature. There are three categories of solar energy forecasting methods: physical methods, statistical methods, and machine learning methods [2,3]. N. Zemouri et al., in [4] they proposed a novel approach using multi-models statistical ensembles to predict short-term GHI using the classification strategy. In [2] Kejun Wang et al. they proposed a hybrid deep learning model LSTM-CNN applied to photovoltaic power prediction. Elham M. Al-Ali and al in [5] they presented a Hybrid CNN-LSTM-Transformer Model to forecast solar energy production. The rest of this paper is structured as follows: the next section give in briefly description of the proposed approach. To evaluate the proposed approach we choose the statistical criteria which are presented in Section 3,the practical experimental results and comparisons for the evaluation of the proposed technique are described in Section 3. Finally, the concluding remarks of this study can be found in Section 4.

2. Methodology

Figure 1 illustrate the operational flow of the proposed approach for executing very short-term forecasting of PV output power using different stage technique. The prediction system is trained using real weather input data (solar radiation, temperature, relative humidity, pressure and wind speed) to forecasted solar PV generation. After the actual methodological data has been subjected to dimensionality reduction to acquire an ideal collection of past variables, this collection serves as an input component for every predictive model trained using deep learning methods. The fusion of these two models enables the ultimate prediction of photovoltaic solar energy output. All these steps are detailed in the following subsection.

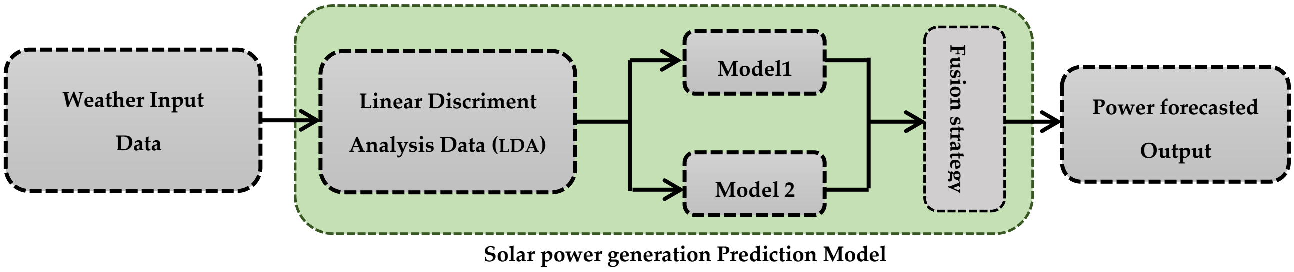


Figure 1. The proposed model to generated solar power forecasting.

2.1. Meteorological Weather Input Parameters

In the present wok, deferent meteorological weather input parameters are used to forecasted the solar power which are: temperature (T), relative humidity (RH), wind speed (WS) and solar radiation (SR)because are uncomplicated to mesure. The process of choosing input variables is a significant aspect in time series prediction. Selecting the appropriate input variables is important step for modeling time series effectively. When there is a set of independent variables, it is necessary to perform variable selection to exclude variables that have minimal impact on the forecast and ensure important variables are not overlooked. The selection of variables is influenced by factors such as data availability, quality, and correlation with the target variable [6]. In order to determine the connection between the meteorological variables collected and the output of a solar power plant, a statistical analysis is conducted. Pearson correlation is utilized as a metric to assess the linear correlation between the two variables, which must be explained in the following subsection.

2.2. Correlation between Weather Inputs Data and Power

The process of choosing input variables is a significant aspect in time series prediction. Selecting the appropriate input variables is critical for modeling time series effectively. When there is a set of independent variables, it is necessary to perform variable selection to exclude variables that have minimal impact on the forecast and ensure important variables are not overlooked. The selection of variables is influenced by factors such as data availability, quality, and correlation with the target variable [7]. In order to determine the connection between the meteorological variables collected and the output of a solar power plant, a statistical analysis is conducted. Pearson correlation is utilized as a metric to assess the linear correlation between the two variables. It calculates the strength and direction of the relationship between the variables, ranging from -1 to +1. A value of +1 indicates a perfect positive linear relationship, 0 indicates no linear relationship, and -1 indicates a perfect negative linear relationship. The equation (1) represents the mathematical expression for the Pearson correlation coefficient.

$$\boldsymbol{R} = \frac{\boldsymbol{n}\sum \boldsymbol{XY} - \sum \boldsymbol{X} \sum \boldsymbol{Y}}{\sqrt{(\boldsymbol{n}\sum \boldsymbol{X}^2 - (\sum \boldsymbol{X})^2).(\boldsymbol{n}\sum \boldsymbol{Y}^2 - (\sum \boldsymbol{Y})^2)}} \quad (1)$$

Table 1 assesses the effectiveness of the correlations between SP (Solar Power) and the different meteorological components, emphasizing the best results in bold. The impact of these relationships is evaluated using the correlation coefficient (R), which shows the degree to which the two variables are related.

Table 1. Performance comparison of solar power vs. meteorological data in terms of correlation coefficient (R) Bolded values are the best in the comparison.

Relationship	SP(kw)	T(C°)	SR(w/m2)	RH(%)	WS(m/s)	Pres(HPa)
SP(kw)	**1.0000**	0.9329	0.8520	−0.5015	0.1032	0.1711
T(C°)	0.9329	**1.0000**	0.8430	−0.6964	0.1220	0.0562
SR(w/m²)	0.852	0.8430	**1.0000**	−0.4595	0.0162	0.0702
RH(%)	−0.5015	−0.6964	−0.4595	**1.0000**	−0.2519	0.3060
WS(m/s)	0.1032	0.1220	0.01620	−0.2519	**1.0000**	−0.3088
Pres(HPa)	0.1711	0.0562	0.07020	0.3060	−0.3088	**1.0000**

2.3. Deep Learning Application Method to forecasted Power

In the following Section, brief introduction for the standard LSTM and 1D-CNN model which are used to forecasted the power output.

2.3.1. LSTM-RNN Predictor Model

LSTM is new kind of RNN, it is substantially distinct compared to standard ANNs which is often used to represent sequential data, such as time series or natural language. LSTMs, however, are able to recall long-term dependencies because they were specifically created to address the vanishing gradient problem [8].The LSTM comprises four basic interconnected layers (three gates and cell state): Cell state to store information, Forget gate to controls how many information is forgotten buy the cell memory, Input gate used to determines which inputs are utilized to update the memory of the cell and finally, Output gate: defines which components of the cell memory are updated to change the LSTM cell's hidden state. Figure 2a illustrates the LSTM fundamental unit architecture. The output cell$\mathbf{c_t}$ and upward output h_t are determined using all the following equations:

$$\tilde{c}_t = \tanh(W_c x_t + U_c h_{t-t} + b_c) \quad (2)$$

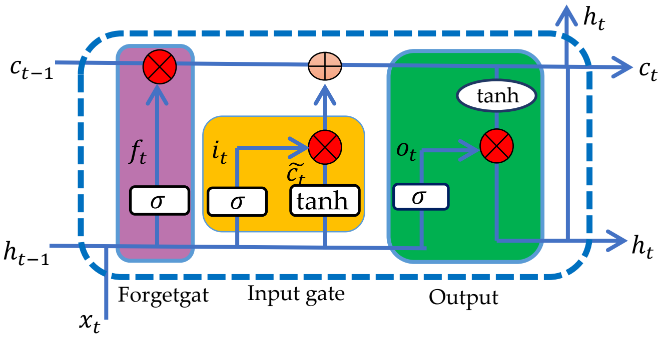
$$c_t = f_t \otimes c_{t-1} + i_t \otimes \tilde{c}_t \quad (3)$$

$$h_t = o_t \otimes \tanh(c_t) \quad (4)$$

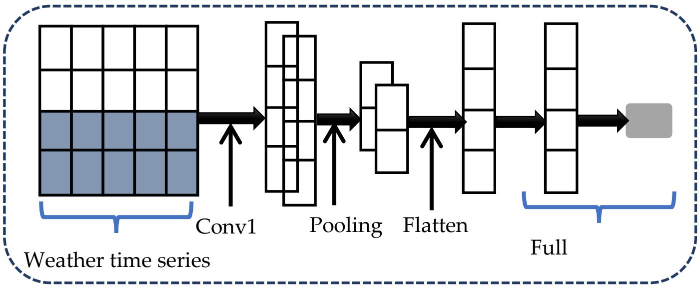
where σ,tanhrepresents sigmoid andhyperbolic tangent activation functions respectively, $\otimes$ denotes the operation of matrix multiplication. In [8] general introduction to the LSTM model is presented. The second deep learning of prediction used in our work isCNN as mentioned in the following subsection.

2.3.2. CNN Predictor for 1D Signal

In 1989, Deep ANNs of kind Convolutional Neural Networks (CNN) were first used to recognize zip codes [9]. At present, CNN mlethod has been created specifically to handle one-dimensional data called also 1D-CNN. It was very recently suggested and performs well in a number of applications, particularly for 1D signals as our case. The components that comprise into a 1D-CNN model contain are: Convolution layers, Pooling layers and Fully-connected layers or dense layers as shows in Figure 2b, the tasks of each layer are presented in [10–12] and the details of the mathematical equations is outlined in [13].



(a)



(b)

Figure 2. (a) LSTM unit architecture [14]. (b) 1D-CNNarchitecture.

2.4. Supervised Dimensionality Reduction Based in LDA Technique

Linear Discriminate Analysis (LDA) is one famous supervised learning technique used in statistics and another fields, to find a linear combination of features that characterizes or separates two or more classes of data [15]. It can achieved in three principal steps: (1) calculate the separabolity between different classes caled as between-class variance (Sb), (2) calculate the distance between the mean and sample of each class which is called the withinclass variance (SW), (3) construct the lower dimentional space wich maximizes between class variance (Sb) and minimizes the within class variance (SW) as mentioned in Figure 3.

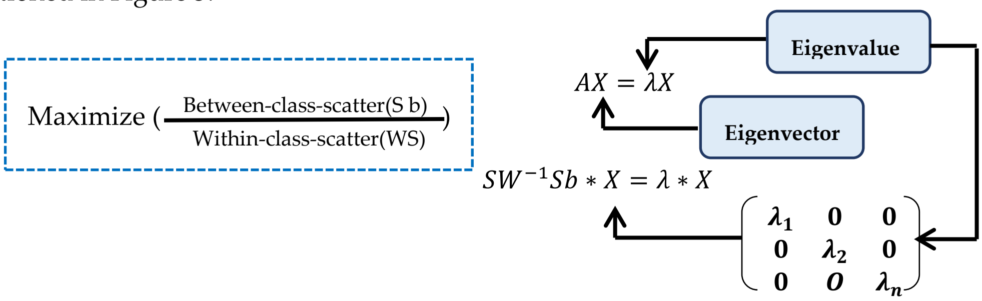


Figure 3. Linear Discriminate Analysis Principe.

2.5. Classification Aggregation Strategy

A combination strategy combines various techniques in order to enhance the performance of global forecasting. It presents a conscious decision to increase forecasting accuracy by using a practical approach, in present work two model of predion was combined using classification strategy which is based on the hypothesis that several highperformance predictors may be associated with various regions of the input time series variable space [4] the best predictor is the one that minimizes error throughout the whole variable space wee used the RMSE error to suggested our proposed strategy.

3. Results and Discussion

The data that was used in this study are real data were collected with a 15 min, during a period from January 1, 2019 to December 31, 2019 from photovoltaic solar power plant with a capacity of 20 megawatts, located in city situated in East of Algeria in North Africa. Meteorological data includes are taken as the proposed model inputs. The collected dataset was divided into three subsets, this division was realized with the aim is to obtained acceptable and accurate forecasting data. The first subset includes as train set 70%and the second were regarded as test set with 30% of the total data. All this data are used to forecasted the power solar to evaluated our approach proposed in this work deferent criteria are used: mean square error (NRMSE), mean absolute percentage error (MAPE), mean bias error (MBE), and coefficient of determination (R^2) there equation are mentioned in [4]. The characters of the two method of prediction used in this work are mentioned in Table 2.

Table 2. Setting Parameters of the prediction models.

Model	Parametres	Value
LSTM	Optimizer, Mini Batch Size, Number of Hidden Units, learning rate	ADAM, 80, 107, 0.0005
CNN	Convolutional layer, Max-pooling layer, Dropout layer Initial learning rate, Mini batch size	3, 5, 0.005 0.002, 2

The prediction accuracy of the proposed model is compared in Table 3, which show that the used strategy based in classification aggregating is performs significantly better than the other individual mode with LDA and without LDA in term of R^2 with 0.9337 witches indicated that the classification-based strategy outperforms the other ones by a large margin. Performance samples from the prediction of solar power with LSTM and CNN after using LDA reduction is compared with their combination as shown in Figure 4.

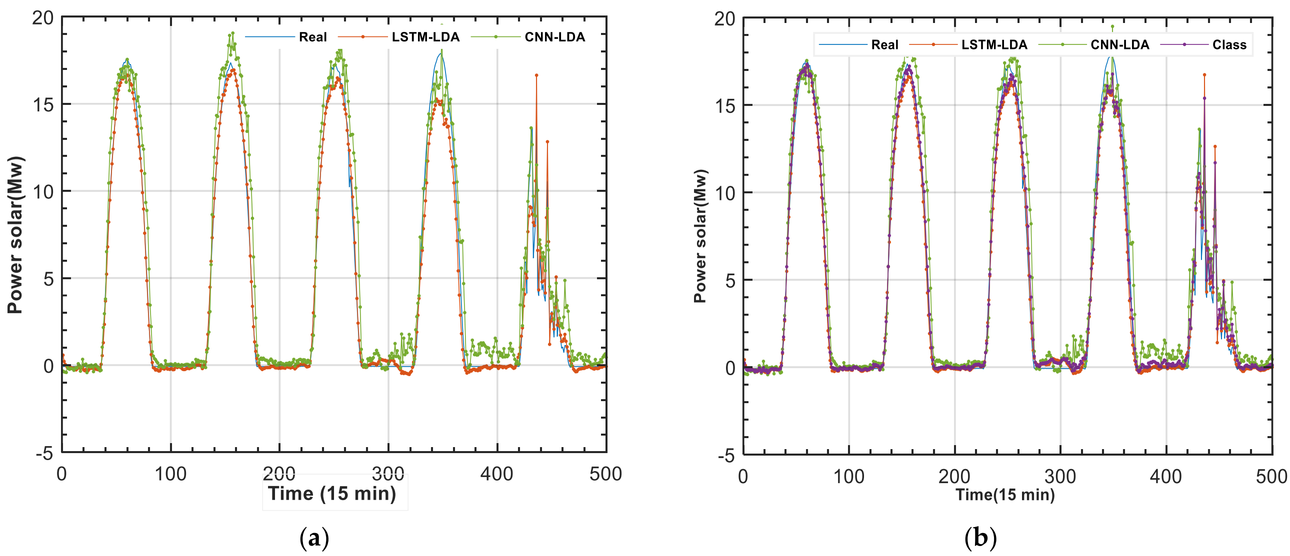


Figure 4. 15-min ahead solar power forecasting for different strategy deep learnig method compared with their combination.

Table 3. Prediction accuracy for different proposed methods. The best results are highlighted in bold.

Methods	CNN	CNN-LDA	LSTM	LSTM-LDA	Class
NRMSE	0.1697	0.1480	0.0019	0.0016	**0.0014**
MAPE	43.6935	35.6423	17.7424	16.055	**16.012**
MBE	1.5077	1.0137	0.0580	0.0225	**0.0183**
R^2	0.8701	0.8828	0.9234	0.9319	**0.9337**

4. Conclusions

Aggregating deep laerning techniques with a dimensionality reduction framework, called the Linear Discriminate Analysis (LDA) method, is used to forecast solar power. Their combination based on classification strategy is proposed in this paper. The suggested approach demonstrates experimental results in forecasting tasks involving solar power, indicating that the combination strategy gives the best results compared with individual models. The good performance of the proposed models demonstrated here should make them more attractive for a large variety of forecasting problems.

Future Research

Under feasible forecasting conditions, the proposed approach has achieved good forecasting accuracy and it can be further improved, and as research work that can be carried out later we can try to propose other reduction methods dimensionality such as PCA or MRMR add a block of decomposition of the signal of the weather time series using for example EMD, CEEMD or VMD whose objective and always is the improvement of the precision of the forecasts of the model of proposed forecast.

Author Contributions:

Funding:

Institutional Review Board Statement:

Informed Consent Statement:

Data Availability Statement:

Conflicts of Interest:

Nomenclature

b_c, c_t	Bias term for generating candidate, Output Cell
c_{t-1}, $\widetilde{c_t}$	Cell state from previous time step, Candidate cell state
f_t	Forget gate
h_t, h_{t-1}	Upward Output, Previous hidden state
i_t	Input value of weather time series
LDA	Linear Discriminate Analysis
o_t	Output gate
R, R^2	Pearson correlation coefficient, Coefficient of determination
RH,SR, SP	Relative Humidity, Solar Radiation, Solar Power
WS, T	Wind Speed, Temperature
W_c	Weight matrix for generating candidate
x_t	Input value of weather time series

References

1. Abdel-Nasser, M.; Mahmoud, K. Accurate photovoltaic power forecasting models using deep LSTM-RNN. *Neural Comput. Appl.* **2019**, *31*, 2727–2740. https://doi.org/10.1007/s00521-017-3225-z.
2. Wang, K.; Qi, X.; Liu, H. Photovoltaic power forecasting based LSTM-Convolutional Network. *Energy* **2019**, *189*, 116225. https://doi.org/10.1016/j.energy.2019.116225.

3. Antonanzas, J.; Osorio, N.; Escobar, R.; Urraca, R.; Martinez-de-Pison, F.J.; Antonanzas-Torres, F. Review of photovoltaic power forecasting. *Sol. Energy* **2016**, *136*, 78–111. https://doi.org/10.1016/j.solener.2016.06.069.
4. Zemouri, N.; Bouzgou, H.; Gueymard, C.A. Multimodel ensemble approach for hourly global solar irradiation forecasting. *Eur. Phys. J. Plus* **2019**, *134*, 594. https://doi.org/10.1140/epjp/i2019-12966-5.
5. Al-Ali, E.M.; Hajji, Y.; Said, Y.; Hleili, M.; Alanzi, A.M.; Laatar, A.H.; Atri, M. Solar Energy Production Forecasting Based on a Hybrid CNN-LSTM-Transformer Model. *Mathematics* **2023**, *11*, 676. https://doi.org/10.3390/math11030676.
6. Al-Shamisi, M.H.; Assi, A.H.; Hejase, H.A.N. Artificial neural networks for predicting global solar radiation in Al Ain City—UAE. *Int. J. Green Energy* **2013**, *10*, 443–456. https://doi.org/10.1080/15435075.2011.641187.
7. Al Shamisi, M.H.; H., A.; N. Hejase, H.A. Using MATLAB to Develop Artificial Neural Network Models for Predicting Global Solar Radiation in Al Ain City—UAE. *Eng. Educ. Res. Using MATLAB* **2011**. https://doi.org/10.5772/25213.
8. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735.
9. Gu, J.; Wang, Z.; Kuen, J.; Ma, L.; Shahroudy, A.; Shuai, B.; Liu, T.; Wang, X.; Wang, G.; Cai, J.; et al. Recent advances in convolutional neural networks. *Pattern Recognit.* **2018**, *77*, 354–377. https://doi.org/10.1016/j.patcog.2017.10.013.
10. Ghimire, S.; Nguyen-Huy, T.; Deo, R.C.; Casillas-Pérez, D.; Salcedo-Sanz, S. Efficient daily solar radiation prediction with deep learning 4-phase convolutional neural network, dual stage stacked regression and support vector machine CNN-REGST hybrid model. *Sustain. Mater. Technol.* **2022**, *32*, e00429. https://doi.org/10.1016/j.susmat.2022.e00429.
11. Wang, F.; Yu, Y.; Zhang, Z.; Li, J.; Zhen, Z.; Li, K. Wavelet decomposition and convolutional LSTM networks based improved deep learning model for solar irradiance forecasting. *Appl. Sci.* **2018**, *8*, 1286. https://doi.org/10.3390/app8081286.
12. Jalali, S.M.J.; Ahmadian, S.; Kavousi-Fard, A.; Khosravi, A.; Nahavandi, S. Automated Deep CNN-LSTM Architecture Design for Solar Irradiance Forecasting. *IEEE Trans. Syst. Man, Cybern. Syst.* **2022**, *52*, 54–65. https://doi.org/10.1109/TSMC.2021.3093519.
13. Lemke, C.; Gabrys, B. Meta-learning for time series forecasting and forecast combination. *Neurocomputing* **2010**, *73*, 2006–2016. https://doi.org/10.1016/j.neucom.2009.09.020.
14. Bouktif, S.; Fiaz, A.; Ouni, A.; Serhani, M.A. Single and multi-sequence deep learning models for short and medium term electric load forecasting. *Energies* **2019**, *12*, 149. https://doi.org/10.3390/en12010149.
15. ArchanaH., T.; Sachin, D. Dimensionality Reduction and Classification through PCA and LDA. *Int. J. Comput. Appl.* **2015**, *122*, 4–8. https://doi.org/10.5120/21790-5104.