Article

Solar Radiation Forecasting by Pearson Correlation Using LSTM Neural Network and ANFIS Method: Application in the West-Central Jordan

Hossam Fraihat [1], Amneh A. Almbaideen [2], Abdullah Al-Odienat [2], Bassam Al-Naami [3], Roberto De Fazio [4] and Paolo Visconti [4],*

1 Department of Electrical Engineering, Al-Ahliyya Amman University, Amman 19328, Jordan; h.fraihat@ammanu.edu.jo
2 Department of Electrical Engineering, Mutah University, Al Karak 61710, Jordan; a.mbaideen@mutah.edu.jo (A.A.A.); odienat@mutah.edu.jo (A.A.-O.)
3 Department of Biomedical Engineering, Engineering Faculty, The Hashemite University, Zarqa 13133, Jordan; b.naami@hu.edu.jo
4 Department of Innovation Engineering, University of Salento, 73100 Lecce, Italy; roberto.defazio@unisalento.it
* Correspondence: paolo.visconti@unisalento.it; Tel.: +39-0832-297334

Abstract: Solar energy is one of the most important renewable energies, with many advantages over other sources. Many parameters affect the electricity generation from solar plants. This paper aims to study the influence of these parameters on predicting solar radiation and electric energy produced in the Salt-Jordan region (Middle East) using long short-term memory (LSTM) and Adaptive Network-based Fuzzy Inference System (ANFIS) models. The data relating to 24 meteorological parameters for nearly the past five years were downloaded from the MeteoBleu database. The results show that the influence of parameters on solar radiation varies according to the season. The forecasting using ANFIS provides better results when the parameter correlation with solar radiation is high (i.e., Pearson Correlation Coefficient PCC between 0.95 and 1). In comparison, the LSTM neural network shows better results when correlation is low (PCC in the range 0.5–0.8). The obtained RMSE varies from 0.04 to 0.8 depending on the season and used parameters; new meteorological parameters influencing solar radiation are also investigated.

Keywords: solar energy forecasting; Pearson Correlation Coefficient PCC; principal component analysis filter; deep learning LSTM; machine learning ANFIS

1. Introduction

Solar irradiation is the total quantity of electromagnetic irradiation emitted by the sun over a frequency range. Solar energy is one of the most abundant and adaptable renewable energy sources; it can be used directly or indirectly. However, among all the non-conventional energy sources, solar energy is the greatest option since it is both cost-effective and ecologically friendly [1–4]. The increasingly intensive use of renewable energy sources to produce clean electricity will reduce dependence on fossil fuels, allowing a strong reduction in carbon emissions [5–7].

According to the Renewable Energy Policy Network for the Twenty-First Century, solar energy will reach a total production of 8000 GWatt in 2050 [8–11]. Solar irradiation is variable and intermittent, leading to significant output-power variability; this limit represents a serious challenge for the generated photovoltaic energy (PV) that must be continuously fed into the grid [4,12,13]. Solar and wind sources are particularly suitable for mega-project investments in this field. Before starting the design of any renewable energy production plant, the factors that influence solar energy production must be studied. The parameters influencing solar radiation are well known and addressed in each region worldwide. Some researchers found that the efficient way to study the relationship between

solar radiation and environmental parameters is through the wide application of Machine Learning (ML) or Deep Learning (DL) techniques, providing an accurate prediction of the energy that will be produced. Therefore, different databases worldwide have been collected and generated to aid researchers in generating the best conditions and methods for forecasting solar radiation. Thus, the location, weather specifications, ML algorithm, and the number of selected parameters play an important role in forecasting operations. In other words, the forecasting procedure can be different from place to place, depending on which parameters are used and their total number. For this reason, project lenders and designers must rely on accurate and dependable forecasting models to protect their investments.

Jordan primarily depends on imported crude oil for electricity generation, becoming heavily dependent on imports to meet its energy needs. The local energy produced in 2020 (by natural gas, crude oil, and renewable energy) reached around 610 ktoe (thousand tons of oil equivalent), representing only 6% of the energy needed to meet demands. The cost of imported energy is as much as 19% of GDP (Gross Domestic Product). The shortage in electrical energy is covered by importing energy from Egypt through a submarine cable in the seasons of high demand for electrical energy. Jordan has been blessed with bountiful solar energy located within the Sunbelt. The intensity of direct solar irradiance in Jordan is one of the highest globally, with an annual daily average of 4–8 kWh/m^2, which corresponds to 1400–2300 kWh/m^2 per year. The sun shines on more than 300 days per year.

According to the national strategy, the Jordanian government plans to boost electricity generation capacity from renewable sources to 3.22 GW by 2025. The peak load reached 4010 MW in 2022. The production of electrical energy grew at an average rate of about 12% during the last 20 years. The Government of Jordan is keen to expand solar energy usage domestically for small- and large-scale solar power generation. It is taking important steps to make solar energy a major contributor to overall energy needs, such as the legislation, the compulsory energy efficiency code, and incentives for investing in the solar energy sector. However, compared to dispatchable plants, the variability of renewable sources, including solar energy, creates a major reliability problem for any electrical power system. Two of the most difficult aspects of integrating renewables into the Jordanian power system are unpredictability and intermittent electricity delivery. Therefore, solar power forecasting is very important for stable electric grid operation and optimal dispatch. Machine learning techniques are employed to overcome these problems and successfully integrate the produced electrical power into the national grid. Various local and global parameters influence solar energy production. The availability of solar energy's reliable information becomes essential to allow the design and construction of energy production plants with high yields in economic terms.

Solar irradiation forecasting is investigated and evaluated in this research work, with particular attention to the capability of new environmental parameters to influence solar irradiation levels. The meteorological parameters reported in the literature on the international level and in Jordan were focused on a few parameters such as solar radiation, wind, and air temperature. Moreover, the influence of these parameters was not studied per season, and researchers did not investigate them in the western region of Jordan. Other parameters that can influence solar radiation are analyzed in this article: specifically, evapotranspiration, humidity, and pressure. The obtained results showed that these parameters have a remarkable influence on solar radiation depending on the season.

Therefore, this paper aims to study twenty-four solar radiation metrological parameters (inputs to the ML or DL algorithms) instead of a few parameters as in other research works. To the best of our knowledge, this complete set of parameters (listed in Table 1) has been tested in the west-central area of Jordan for the first time. The data, relating to a time interval of almost five years, were downloaded from the international database Meteobleu and analyzed using the Long Short-Term Memory (LSTM) and ANFIS methods to build the solar radiation forecasting model. In more detail, some parameters usually not used in the literature as they are considered not useful for the solar radiation forecasting

(such as the direct and diffuse short-wave radiation, evapo-transpiration, vapor pressure deficit at 2 m, relative humidity, sunshine duration, and soil temperature), have been taken into consideration in this research work for the first time, demonstrating their effects and contribution.

Table 1. The twenty-four parameters employed in our study for solar radiation forecasting.

Number	Parameter
1	Solar radiation (sum of direct and diffuse short-wave radiation) (W/m^2)
2	Direct short-wave radiation
3	Diffuse short-wave radiation
4	Temperature (2 m above ground)
5	Vapor pressure deficit (VPD) at 2 m
6	Relative humidity (2 m above ground)
7	Growing degree days (2 m) estimates plants' growth and development, depending on the temperature variation
8	Sunshine duration
9	Soil temperature (0–10 cm under the ground level)
10	Total cloud cover (percent)
11	Low cloud cover (percent)
12	Geopotential (height 500 mb) represents the average air temperature in the vertical column
13	Evapotranspiration represents the sum of evaporation from the land surface plus transpiration from plants
14	Soil moisture (0–10 cm under the ground level)
15	Wind speed (10 m above ground)
16	Total precipitation amount (mm/m^2)
17	Medium cloud cover (percent)
18	Snowfall amount (cm/m^2)
19	Wind direction (80 m above ground)
20	High cloud cover (percent)
21	Wind gust (10 m above ground)
22	Wind speed (80 m above ground)
23	Convective available potential energy CAPE (180 mb) measures the air parcel's potential energy per kilogram of the air mass. High CAPE value means that atmosphere is unstable and would produce a strong updraft.
24	Wind Direction (10 m above ground)

These parameters have been studied case by case, starting with the historical data of the solar radiation itself (as input) to predict the solar radiation values. Afterwards, by analyzing each parameter and determining its influence, they have been divided into two groups: the first contains parameters already investigated in the literature around the world (namely, solar radiation, air temperature, wind and direction speed, cloud cover, humidity, rain and snow precipitation, soil temperature, pressure, and sunshine duration). The second group contains only the parameters already used in Jordan for solar radiation forecasting (i.e., solar radiation, wind, and air temperature).

Related Work

In the literature, significant performance differences from the different ML and DL applications in solar radiation forecasting were reported. Castangia et al. [14] used five machine learning models based on Feedforward, Echo State, 1D-Convolutional, LSTM neural networks, and Random Forest (RF) method. They used in their study six parameters: the cloud cover, air temperature, relative humidity, dew point, wind bearing, and sunshine duration. The Root Mean Square Error (RMSE) equals 6.60% for hourly forecasting. However, since some significant parameters are not considered in the study, this could impact the prediction accuracy negatively. In [15], the convolutional LSTM method is used to forecast solar irradiance on several locations simultaneously; the obtained RMSE is less

than 15%. However, in this study, the proposed DL technique could require additional training (transfer learning) if different locations are considered.

On the other hand, the DL applications have been successfully applied in the last few years. For example, the hybrid DL model proposed by Yan et al. [16] was applied for one-year forecasting solar radiation, comparing the obtained RMSE values for different time intervals related to the four seasons. However, the performed study considered only short time intervals to generate solar irradiance predictions.

In [17], the Support Vector Machines (SVM) and Random Forest (RF) models were applied to forecast solar irradiance using weather parameters, such as temperature, humidity, rainfall and wind speed. In this work, the authors utilize SVM and RF models to predict individual PV generator output and compare their performances. Convolutional Neural Network (CNN) and LSTM have been used to forecast solar power, utilizing input datasets related to temperature, wind speed, humidity, ground temperature; the obtained RMSE equals 0.0987 [18]. In Poolla et al. [19], the local and global meteorological data (air temperature and wind speed), in the time interval from December 2017 to May 2018, were employed for the forecasting process based on an auto-regressive time-series (ARTS) model, achieving an accuracy of 80%. However, in this research work, the air temperature and wind speed are the only weather variables considered in the prediction process.

Han et al. [20] considered air pressure, zenith angles, temperature, and humidity parameters. The cross-correlation coefficient between the predicted and measured solar radiation values was 0.947. Wang et al. studied day-ahead photovoltaic power forecasting by utilizing the convolutional LSTM networks [21]. The recorded data, collected continuously for two years, referred to five parameters as inputs to LSTM neural network: temperature, pressure, humidity, wind and speed direction. As result of the power forecasting, the RMSE showed a very acceptable value of 0.0865. However, in long-term forecasting, the efficacy of the proposed models is not investigated. A forecasting model based on Artificial Neural Network (ANN) has been proposed in [22]; it used the temperature, dew point, relative humidity, and wind speed as inputs. The Mean Absolute Percentage Error (MAPE) of 0.53% was achieved for 14 days of prediction. In the same domain of forecasting solar irradiation, the LSTM is widely used. In [23], four deep learning algorithms were independently trained to predict the solar irradiance of Johannesburg city: LSTM, CNN, Convolutional LSTM, and CNN-LSTM hybrid models. Based on obtained results, the Convolutional LSTM provided the best performance with a normalized RMSE of 1.62% (corresponding to a RMSE of 7.18).

In [24], the authors used an artificial neural network, CNN, bidirectional and stacked LSTM to predict solar irradiance values. The used parameters are humidity, station and ambient temperature, station altitude, sea level pressure, absolute pressure and wind speed. A dataset from September 2019 to February 2020 was used to obtain a Mean Absolute Error (MAE) of 41.738; the authors concluded that stacked LSTM is the best model for predicting solar irradiance. In [25], the data over four years were used to forecast temperature, precipitation, and wind speed parameters; the obtained MAE value was equal to 0.708. The authors evaluated the forecasting performance of a stacked bidirectional LSTM (SB-LSTM) approach for both day-ahead and week ahead load. They recommended three approaches to further improve the performance of SB-LSTM: increasing the size of the processed dataset, allowing capturing of variations not included in the limited available dataset, and implementing other architectures.

Some researchers found that analyzing moving clouds is very useful for predicting the future position of the clouds and sun occlusion using image processing algorithms [26]. The researchers proposed a framework to forecast the solar irradiance changes; by combining image processing and machine learning. CNNs were used for processing whole sky images with 6-month datasets; the obtained RMSE was equal to 6.11.

Alvarez et al. [27] used SVM, Linear Regression (LR) and Neural Network models (NNM) to forecast the solar energy by using collected data every hour from 1 January 2020 to 5 June 2020, in Aguascalientes (Mexico). The authors used weather parameters for the forecasting, such as wind velocity and direction, temperature, pressure, humidity, sunrise time, and sunset time. Different machine learning approaches were used; the multi-layer perceptron (MLP) algorithm gave the best outcome with an MSE equal to 0.2222. In [28], the backpropagation (BP) neural network is used to construct an effective forecasting model of solar radiation. As inputs to the developed model, the authors used weather parameters, specifically the rainfall, air humidity, and clear-air index; the obtained RMSE equals 0.4708. In the literature, few studies were performed in Middle Eastern regions and partially in Jordan. Alomari et al. [29] used ANN to forecast solar radiation in central Jordan using a dataset from 15 May 2015 to 30 September 2017. Instead, Al-Sbou et al. [30] used ANN to forecast solar radiation in southern Jordan, obtaining an MSE of 0.00237. Furthermore, Shboul et al. [31] used ANN to forecast solar radiation in southern, central, and north Jordan. The dataset from 1 January 1999 to 30 September 2019, was related to wind, air temperature and solar radiation. The obtained MAPE values did not exceed the 3% value.

In our work, the proposed prediction process uses two learning models, LSTM and ANFIS, with the aim to perform accurate forecasting of solar energy radiation in west-central Jordan based on meteorological data for the last five years. LSTM can memorize the data sequence and contains a set of modules where the data streams are captured and stored. In contrast, ANFIS is a hybrid model that uses numerical and linguistic knowledge. Its advantages include abilities of adaptation, nonlinearity, and rapid learning. The proposed prediction process aims to accurately forecast solar radiation in west-central Jordan based on meteorological data for the last five years. The two models are compared with each other. The Principal Component Analysis (PCA) is used to filter the input signals. New parameters not previously addressed in other studies in the literature are considered; it is concluded that some of these parameters used, such as direct short-wave radiation, diffuse short-wave radiation, and evapotranspiration, greatly influence solar radiation.

This research study is organized as follows: in the introduction, a brief analysis of the different machine learning classifiers proposed in the literature to predict solar radiation was provided, indicating the used parameters and obtained forecast results. Section 2 describes the methods used in this study; data standardization, Principal Component Analysis (PCA) for noise filtering, Pearson Correlation Coefficient (PCC) for feature selection, and the application of LSTM network and ANFIS for the prediction process. Section 3 is devoted to the obtained results related to the solar radiation forecasting for five years with the consideration of seasonal variation. Section 4 provides a comparison of results obtained using ANSIF and LSTM. Finally, Section 5 presents the work's conclusions.

2. Materials and Methods

In this research work, the data for nearly the past five years (i.e., from 1 January 2017, until 22 August 2021) were downloaded from the international database Meteobleu (https://www.meteoblue.com/en/historyplus (accessed on 2 February 2022)). This website provides meteorological data, with hourly updates relating to the twenty-four parameters shown in Table 1, starting from 1985 for worldwide locations. We have analyzed Jordan's west-central region data to forecast solar radiation as accurately as possible; the final dataset comprises 40.676 samples for each parameter. According to the Pareto principle, the dataset needs to be split up into train and test subsets with an 80:20 ratio; in other words, the learning model will use 80% of the dataset for training, while the remaining 20% (test subset) will be used for the solar radiation prediction. Figure 1 shows the main phases of the prediction process. The data are pre-processed by applying a standardization technique; afterwards, a Principal Component Analysis (PCA) noise filter is used. The most significant input variables are then selected using Pearson Correlation Coefficient (PCC) to predict solar radiation, while the remaining variables are removed from the learning set. Once the selected training data are prepared, the deep learning LSTM and machine learning

ANFIS models can be trained to predict solar radiation. Finally, the trained models are evaluated by calculating the root-mean-square error (RMSE) and compared according to their prediction performance for the total period of 5 years and for each season (summer, autumn, winter, and spring).
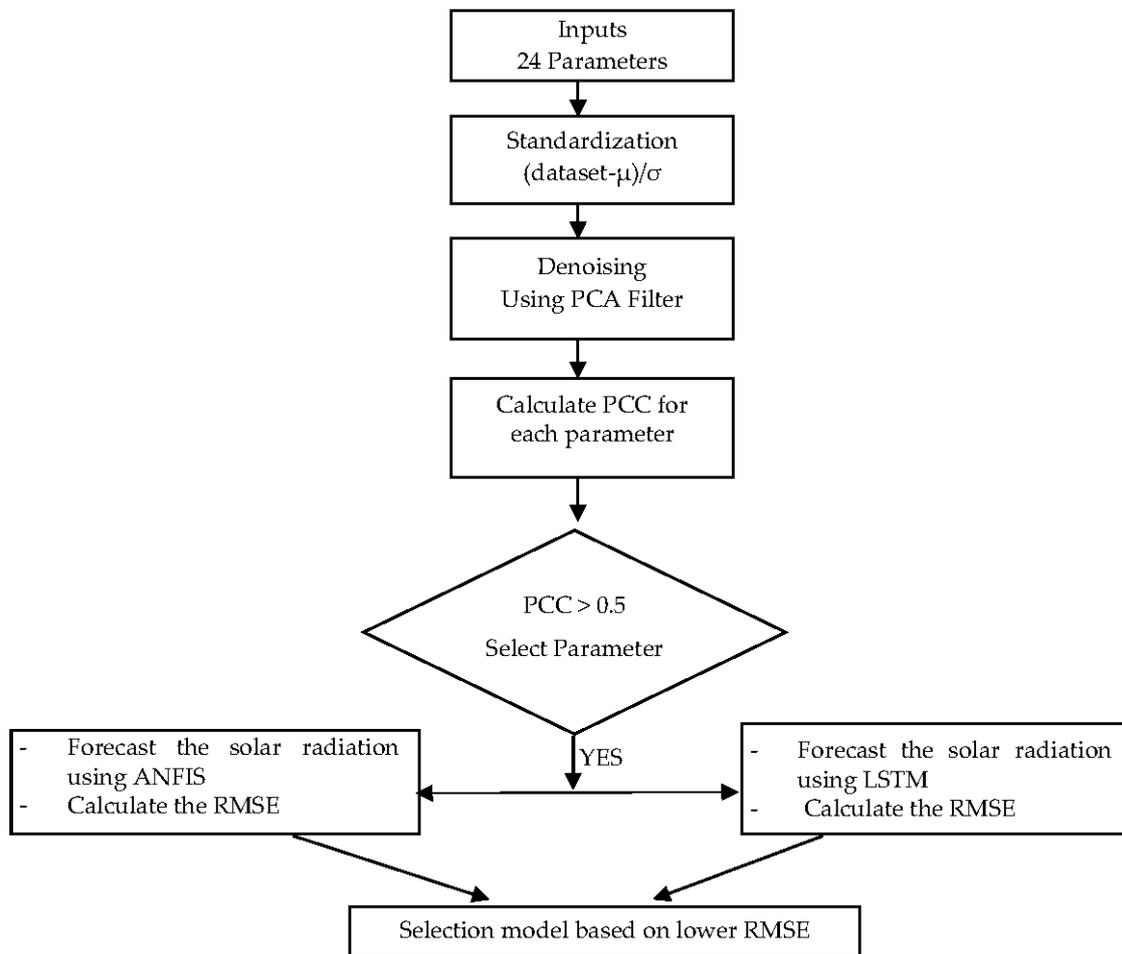

Figure 1. The flow chart of the proposed prediction process.

Figure 1 presents the flow chart of the proposed prediction process. We conducted two different sets of experiments related to meteorological parameters influencing solar radiation, the first one with the ANFIS model, the second one with LSTM; then, the outcomes were evaluated based on the obtained RMSE values.

2.1. Data Standardization

Data standardization aims to ensure the application of a common measurement scale to improve data quality. The standardization formula is given below [32,33]:

$$dataset standardized = \frac{dataset - \mu}{\sigma} \tag{1}$$

where μ is the average value and σ the standard deviation of the dataset distribution.

### 2.2. Principal Component Analysis (PCA) for Noise Filtering

Karl Pearson posited principal component analysis (PCA) for the first time in 1901. It has gained specific applications in many areas, such as chemometrics, image processing, sociology, and economics. PCA is commonly used for data clustering, filtering, extraction and classification, outlier detection, data compression, and for minimizing the correlation between variables [34–36]. PCA is a data compression approach that reduces the data dimensionality. PCA collects the covariance matrix's eigenvectors and eigenvalues to construct the data's uncorrelated principal components. Principal components are conceived of as new axes, the orientations of which reveal the most significant variance in the source data. The eigenvectors of the covariance matrix define the directions of the principal components. The eigenvalues, on the other hand, are the coefficients or weights of the principal components that reflect the amount of variation carried by each component [37,38]. Shaker et al. [39] investigated how data mining can be employed to estimate the aggregated solar power generation from a large set of solar power generation plants without continuously measuring the output of every single site by using only the measured values from a small number of representative sites. The obtained results showed that the proposed framework is capable of estimating solar generation with good accuracy; the combination of linear regression and the proposed hybrid k-means + PCA dimension reduction method gave the best results, demonstrating that PCA is very fast and computationally efficient.

Another characteristic of PCA is that the loading factors are sorted by their contribution to the variance of the original data, which means that the first loading factor captures most of the total variance of the original data among the other elements. The second factor then accounts for the second-largest portion of the total variance, and so on. The contribution of the final loading factors will be minimal, and they are typically used to model noise. As a result, these factors may be overlooked to suppress most of the noise and eliminate the variables' redundancy. Once the loading factors are determined, they are used in connection with the original data to compute the raw data scores. Score vectors are the projection amounts of the nth feature in the original data on the entire loading vectors. Because the scores in each data set are orthogonal, uncorrelated, and span the entire data range, the pseudoinverse of the score matrix can be computed. Thus, the raw data can be represented based on the definition of the loading vectors and scores as:

$$A = TP' \qquad (2)$$

The T rows are the score vectors, and P columns are the loading vectors. Singular value decomposition (SVD) [40], Eigenvector decomposition [41], nonlinear iterative partial least squares (NIPALS) [42], the covariance method [12], expectation-maximization method [13], and successive average orthogonalization (SAO) are all used to generate the PCA model [35]. Figure 2 shows the effect of applying the PCA-based noise filter on a set of observations representing the temperature data; from Figure 2, it is clear that the noise variance of the signal is reduced.

For comparison, we added another denoising tool, wavelet signal denoising, and we compared the results of the two techniques. In Figure 3, wavelet signal denoising is used; it decomposes the signal into different scales, significantly improving the denoising step in the prediction process.
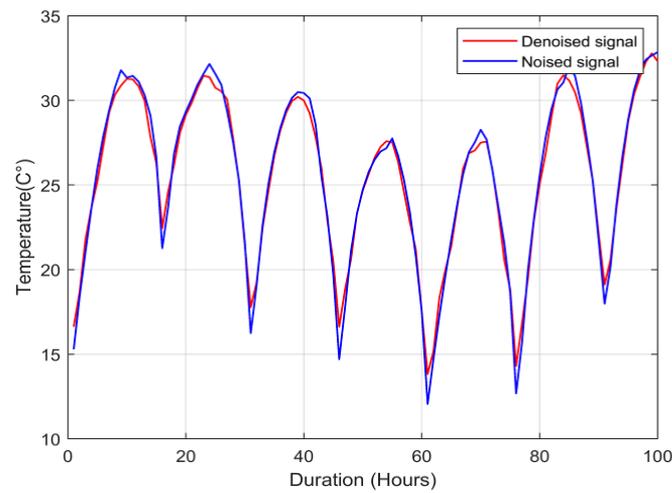
Figure 2. Application of PCA-based noise filtering to temperature data.
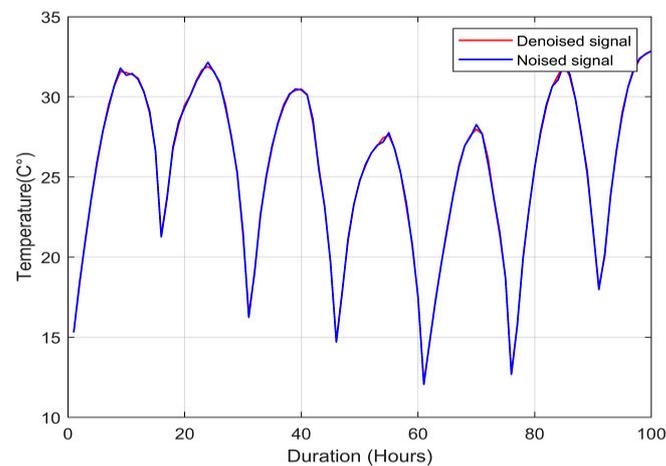
Figure 3. Application of wavelet signal denoising to temperature data.

The denoising performance of each method (PCA and wavelet signal denoising) was assessed using Signal-to-Noise Ratio (SNR) parameter. To calculate SNR, the residual noise is determined as follows:

$$\text{Resid\_Noise} = \text{Noisy\_data\_temperature} - \text{Denoised\_data\_temperature}$$

SNR is then calculated by:

$$\text{SNR} = (\text{mean}(\text{Noisy\_data\_temperature}^2)/\text{mean}(\text{Resid\_Noise}^2)$$

The SNR value is then converted to decibels (dB) using the equation:

$$\text{SNR\_dB} = 10 \times \log10(\text{SNR})$$

It is found that the SNR_dB is equal to 47.48 and 32.25 for the PCA denoising and wavelets signal denoising methods, respectively. It is clear that denoising the temperature signal is more effective using the PCA filter.

2.3. Feature Selection

Pearson correlation coefficient (PCC) measures the linear correlation between each input feature with the solar radiation. It represents the ratio between the covariance of two features and the product of their standard deviations [43,44].

$$\rho_{x,y} = \frac{\text{cov}(X,Y)}{\sigma_x \sigma_y}$$

(3)

where cov is the covariance, σ_x the standard deviation of one input feature, and σ_y the standard deviation of the solar radiation feature (output). This technique allows determining the correlation of each meteorological measurement with solar radiation [14], correlations that can be different with the season. The prediction accuracy can be statistically evaluated using PCC as a metric; a larger PCC intuitively reflects a higher linear correlation between the predicted and true values [33,45].

Figure 4 shows the PCC values ranking for all meteorological parameters with respect to solar radiation. The threshold value equal to 0.5 (red line in Figure 4) between the selected and discarded parameters was determined during the learning phase of the model. It is found that the significant parameters, with PCC values greater than or equal to 0.5, selected for the solar radiation prediction in the time interval from 1 January 2017 to 21 August 2021 are: Direct Short-wave Radiation, Diffuse Short-wave Radiation, Temperature, Vapor Pressure Deficit at 2 m, Relative Humidity at 2 m, Growing Degree Days at 2 m elevation, Sunshine Duration and Soil Temperature (0–10 cm) (Figure 4).
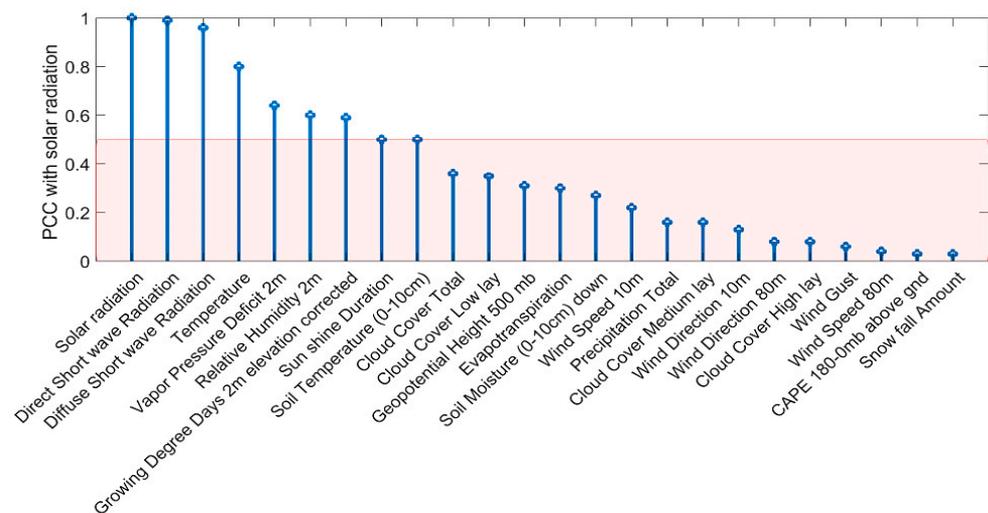
Figure 4. PCC between different meteorological parameters and solar radiation, ranked according to the PCC value (the threshold for parameters' selection corresponds to the red line).

2.4. Evaluation Measures

The most common forecasting indices are the root-mean-square error (RMSE), the mean squared error (MSE) and mean absolute error (MAE); they are used to evaluate the performance of the solar radiation prediction. The above reported error indices between the predicted X and actual Y values in the test dataset are calculated as follows:

$$\text{RMSE} = \sqrt{\left(\frac{1}{N} \cdot \sum_{n=1}^{N} (Y - X)^2\right)}$$

(4)

$$\text{MSE} = \frac{1}{N} \cdot \sum_{n=1}^{N} (Y - X)^2$$

(5)

$$\text{MAE} = \frac{1}{N} \cdot \sum_{n=1}^{N} |Y - X|$$

(6)

2.5. Long Short-Term Memory (LSTM) Network

LSTM is a recurrent neural network that remembers the problem for longer, having a chain structure to repeat the module. Other networks repeat the module whenever the input receives new information. LSTM interacts in a particular way and contains four layers of neural networks. The data transfer process is the same as standard recurrent neural networks, while the information dissemination operation is different. As information passes through, the operation decides which information needs to be further processed and which is to be discarded. The main operation consists of cells and gates (Figure 5); the former contains various activations called sigmoids, containing certain values ranging from zero to one. They help to forget and retain information. If the data is multiplied by one, the value remains the same; if the data is multiplied by zero, the value becomes zero and disappears.
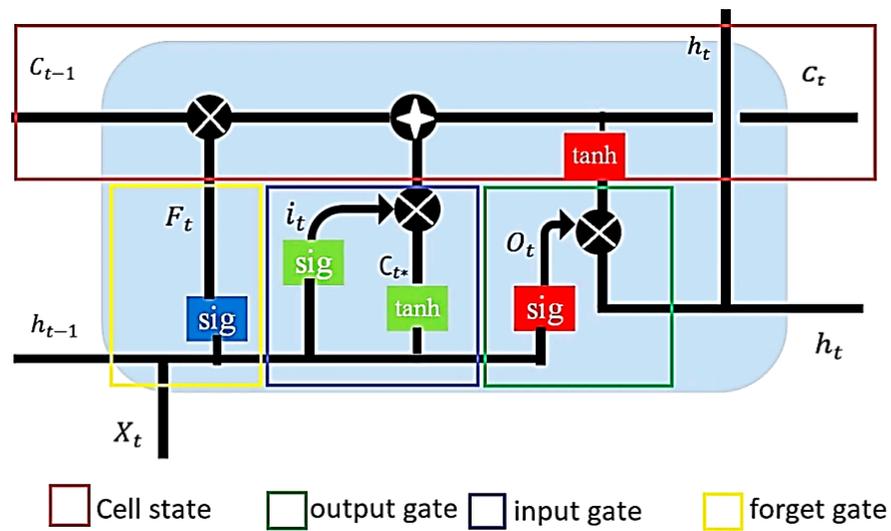

Figure 5. LSTM unit architecture.

There are three types of gates [14,46–50]:

- Forget gate: its function is to decide whether to keep or forget the information. Only information from previously hidden layers and current input remain with the sigmoid function. Any value closer to one will remain, while values closer to zero will disappear:

$$f_t = \sigma\left(x_t U^f + h_{t-1} W^f\right) \tag{7}$$

where x_t is the input vector, h_{t-1} are the output of the previous block, W and U the weight matrices of the hidden state and input respectively for each gate, σ is the sigmoid activation function.

1. Input Gate: the front door helps to update the cell condition. Current input and previous state information go through the sigmoid function, which updates the value by multiplying it by zero and one. Likewise, for network regulation, data also pass through the tanh function (Equation (9)); i_t is the input gate vector.

$$i_t = \sigma\left(x_t U^i + h_{t-1} W^i\right) \tag{8}$$

$$C_{t*} = \tanh\left(x_t U^g + h_{t-1} W^g\right) \tag{9}$$

The cell state vector aggregates the two components (old memory via the forget gate and new memory via the input gate)

$$C_t = \sigma\left(f_t * C_{t-1} + i_t * C_{t*}\right) \tag{10}$$

C_{t-1} is a memory from the previous block, C_t is defined as a memory from the current block; the "$*$" operator is the Hadamard product.

- Output Gate: the next hidden state is set in the output gate. The sigmoid output has to be multiplied by the tanh function; the result of this multiplication decides which information the hidden state h_t should carry. This hidden state is used for the prediction. After, the new hidden state and cell state will move on to the next step:

$$o_t = \sigma \left(x_t U^{\circ} + h_{t-1} W^{\circ} \right) \tag{11}$$

$$h_t = \tanh{(C_t)} * o_t \tag{12}$$

where o_t is the output gate vector, and h_t the current block output. Table 2 illustrates the used training hyper-parameters for the LSTM neural network.

Table 2. Training hyper-parameters for LSTM.

Parameters	Value
Optimizer	Adam
Epoch	250
Learning rate	0.0001
Hidden units	200
Gradient threshold	0.01
Layers	Regression
Input size	1
Output response size	1

2.6. Adaptive Neuro-Fuzzy Inference System (ANFIS)

Neuro-Flous systems combine the advantages of two complementary techniques (Fuzzy system and neural network). Fuzzy systems provide a good representation of knowledge. The integration of neural networks within these systems improves their performance because of the learning capacity of neural networks.

ANFIS is an optimization method for Takagi and Sugeno's-type fuzzy inference system, based on the use of multilayer networks. ANFIS uses least squares estimation (LSE) combined with the gradient descent backpropagation methods to model a training data set (Figure 6) [51–54].
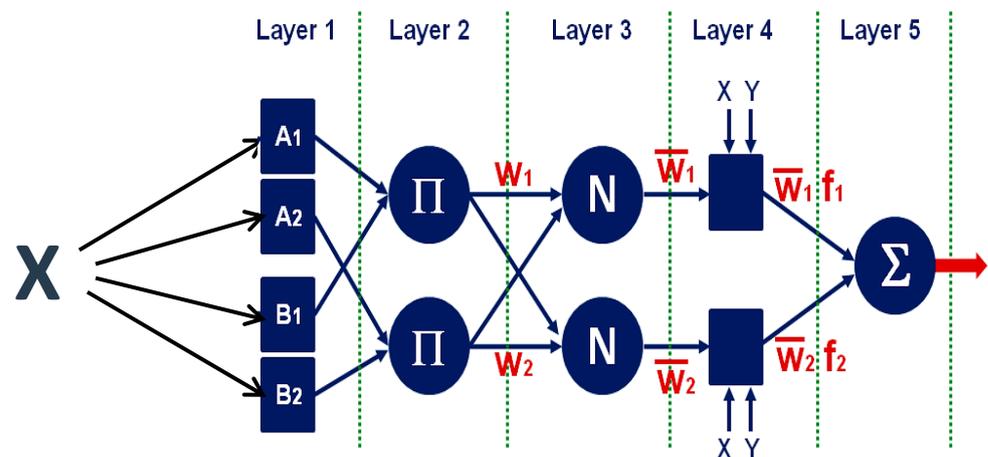
Figure 6. ANFIS structure for one input variable.

The rule base contains one fuzzy if-then-rule of Takagi and Sugeno's type (with X the input and f the output data).

Rule: If x is A1, then $f_1 = pl\ x + rl$

where f_i is the fuzzy inference according to the desired output:

$$\mu_{Ai}(x) = e^{-\left(\frac{x-c_i}{a_i}\right)^2} \tag{13}$$

where $\{a_i, c_i\}$ is the parameter set. ANFIS uses a 5-layer MLP (Multilayer perceptron) as following described:

Layer 1: Generating degree of membership:

$$O_{1,i} = \mu_{Ai}(x) \tag{14}$$

The first layer of an ANFIS-type architecture comprises as many neurons as there are fuzzy subsets in the inference system represented. Each neuron calculates the degree of truth of a particular fuzzy subset by its transfer function. The activation function of neurons i of the first layer is $O_{1,i}$; x is the input to neuron i, and Ai is a fuzzy subset corresponding to x. $O_{1,i}$ is the membership function of Ai and indicates the degree to which a given x satisfies the quantifier Ai. We choose $\mu_{Ai}(x)$ to be in the Gaussian form.

Layer 2: Fuzzy intersection:

The outputs of this layer are the weights w_i of the rules; they are obtained by a simple multiplication of the inputs in each cell. The neurons receive as input the truth degree of the different fuzzy subsets making up this premise and are responsible for calculating their truth degree. The activation functions used for these neurons depend on the operators present in the rules (AND or OR).

The activation function of neurons i of this layer is the following:

$$O_{2,i} = w_i = \mu_{Ai}(x) \cdot \mu_{Bi}(x), \ i = 1, 2 \tag{15}$$

Layer 3: Normalization:

This layer corresponds to the normalization of the weights of the rules. It calculates the ratio between the weights w_i of the rule and the sum of all the weights of the rules.

$$O_{3,i} = \overline{w_i} = \frac{w_i}{w_1 + w_2} \ i = 1, 2 \tag{16}$$

Layer 4: Defuzzification:

Each node i in this layer is calculated as reported in Equation (17); $\overline{w_i}$ are the outputs of layer 3, whereas $p_i x + r_i$ are the consequent parameters of the output function.

$$O_{4,i} = \overline{w_i} f_i = \overline{w_i}(p_i x + r_i) \tag{17}$$

Layer 5: The output layer:

The cell represents the sum of all the input signals and therefore returns, at the output, the approximate value of the desired function.

$$O_{5,i} = \sum \overline{w_i} f_i \tag{18}$$

Table 3 illustrates our training parameter of the ANFIS software.

Table 3. Training parameters for ANFIS software.

Name	FIS
Type	Sugeno
And-Method	Prod:
Or-Method	Probor
DefuzzMethod	Wtaver (Weighted average of all rule outputs)
ImpMethod	Prod
AggMethod	Sum
Input Size	1
Output Response size	1
Rules	7
Epoch	250
Ranges of influence	0.4

3. Results

In this study, LSTM and ANFIS learning models are used to predict the amount of solar radiation available in the west region of Jordan. Twenty-four meteorological parameters are considered in the prediction process, selected for five different scenarios (i.e., referring to the five-year time interval or specific seasons, autumn, summer, spring and winter). The results show that the degree of influence of these parameters depends on seasonal variation. The forecast RMSE related to the five-year dataset is calculated for the different scenarios. The meteorological parameters are ranked based on PCC values using LSTM and ANFIS and properly selected for the different seasons.

As a first result of the research work, Figure 7 depicts the forecasted solar radiation based on temperature parameter in the summer season using the LSTM model. When the temperature changes, the solar radiation changes in the same way with a prediction error of the solar radiation according to the temperature variation. The obtained RMSE value is 0.14, proving that the model accurately forecasts the solar radiation.
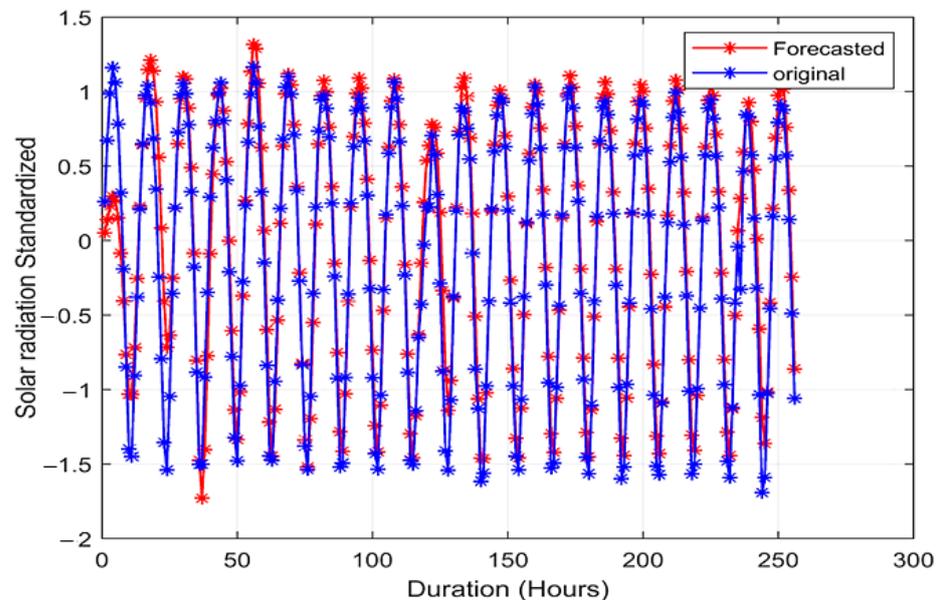
Figure 7. Solar radiation forecasted in the summer season (from 21 June to 20 September 2020).

Figure 8 provides the RMSE values related to forecasted solar radiation for the five-year dataset by employing LSTM (red line) and ANFIS (blue line) models. The RMSE values are lower for meteorological parameters with greater PCC values (close to 1) relative to solar radiation (i.e., direct short-wave radiation, diffuse short-wave radiation, and temperature). The pink area (below the red line related to the 0.5 PCC threshold) indicates the parameters

set with PCC values lower than 0.5 (not significant for the forecasting process), which have therefore been excluded from the learning phase [55]. The solar radiation forecast by the ANFIS model gives a better result than the LSTM one when the PCC is high (>0.95); instead, for (0.5 ≤ PCC < 0.95), the LSTM provides better results. It can be concluded that the proposed methodology gives good results, with an RMSE equal to 0.12 (by ANFIS model) for direct short-wave radiation (PCC = 0.98) up to an RMSE value of 0.32 for temperature parameter (PCC = 0.80) provided by the LSTM model.
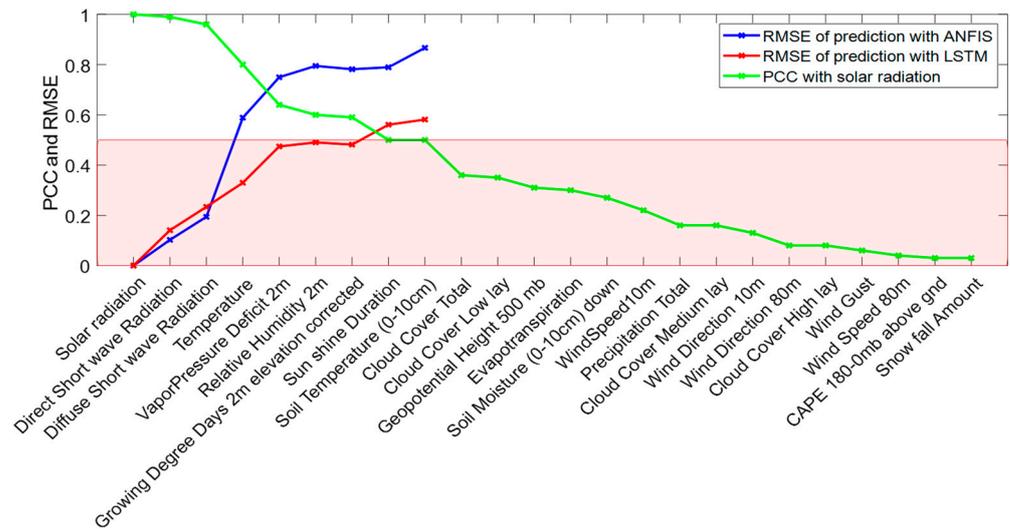
Figure 8. Meteorological parameters' ranking based on Pearson Correlation Coefficients and Comparative performance of LSTM and ANFIS models with significant parameters selection (PCC value ≥ 0.5) (employed dataset from 1 January 2017, until 22 August 2021).

Figure 9 shows the meteorological parameters' ranking for the 2020 summer season. The selected parameters with PCC ≥ 0.5 are listed below with decreasing PCC values: the direct and diffuse short-wave radiation, temperature, sunshine duration, growing degree days at 2 m elevation corrected, vapor pressure deficit at 2 m, and relative humidity. As explained above, the other parameters with a PCC value less than 0.5 were not used in the forecasting models, but discarded.
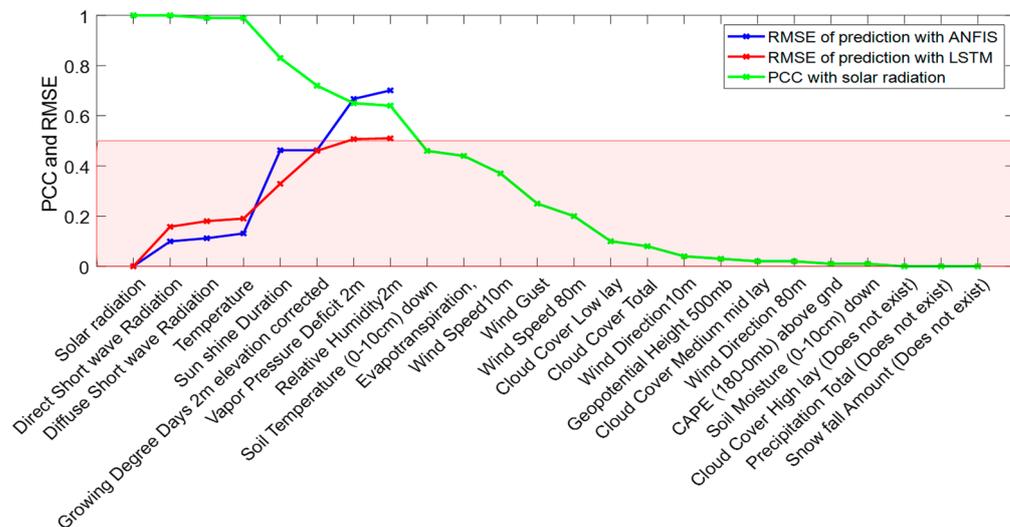
Figure 9. Summer season ranking between the different meteorological parameters based on Pearson Correlation Coefficients and comparative performance of LSTM and ANFIS models with significant parameters selection (PCC ≥ 0.5) (dataset from 21 June 2020 to 20 September 2020).

Figure 10 shows the meteorological parameters' ranking for the 2020 autumn season. The selected parameters with PCC $\geq$ 0.5 are listed below with decreasing PCC values: the solar radiation, direct and diffuse short-wave radiation, temperature, vapor pressure deficit at 2 m, growing degree days at 2 m elevation corrected, relative humidity, sunshine duration and evapotranspiration.
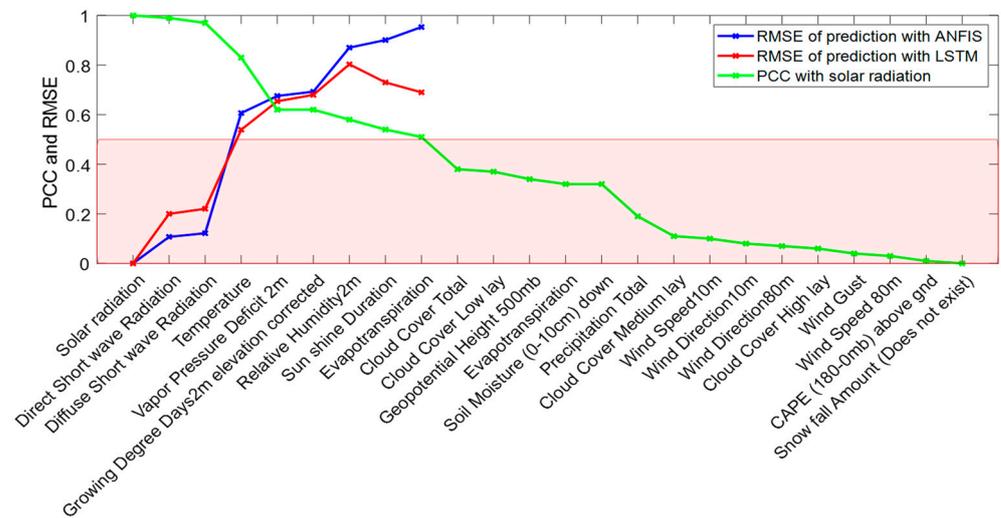

Figure 10. Autumn season ranking between different meteorological parameters based on Pearson Correlation Coefficients and comparative performance of LSTM and ANFIS models with parameters selection (PCC $\geq$ 0.5) (employed dataset from 21 September to 20 December 2020).

Figure 11 shows the meteorological parameters' ranking for the 2020 winter season (employed dataset from 22 December 2020 to 20 March 2021). The parameters with PCC $\geq$ 0.5 (significant for the forecasting process) are the solar radiation, direct and diffuse short-wave radiation, temperature, evapotranspiration, vapor pressure deficit, relative humidity, growing degree days at 2 m elevation corrected, and sunshine duration.

Figure 11. Winter season ranking between different meteorological parameters based on Pearson Correlation Coefficients and comparative performance of LSTM and ANFIS models with significant parameters selection (PCC $\geq$ 0.5) (dataset from 22 December 2020 to 20 March 2021).

Figure 12 shows the meteorological parameters' ranking for the 2021 spring season. The parameters significant for the forecasting process (i.e., with PCC ≥ 0.5) are listed below with decreasing PCC values: the solar radiation, direct and diffuse short-wave radiation, temperature, vapor pressure deficit at 2 m, growing degree days at 2 m elevation corrected, and relative humidity at 2 m.
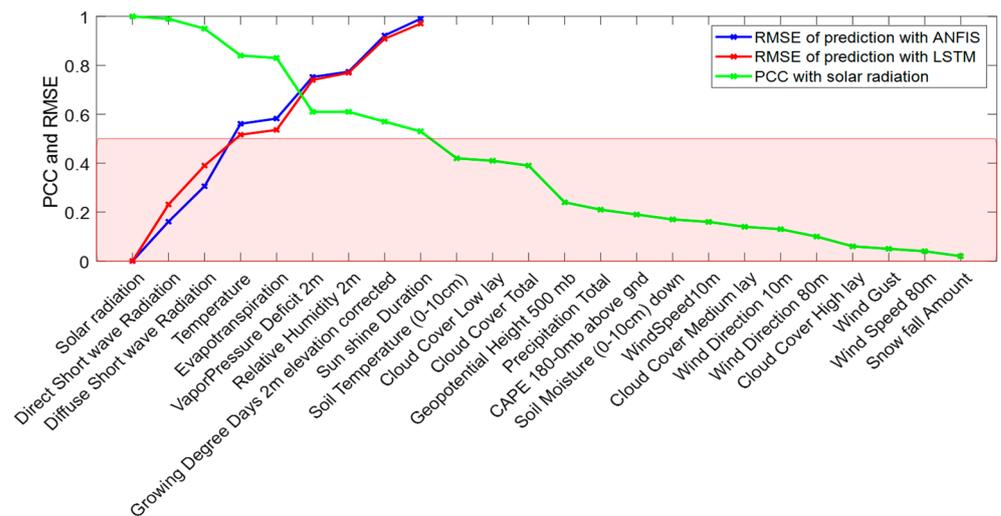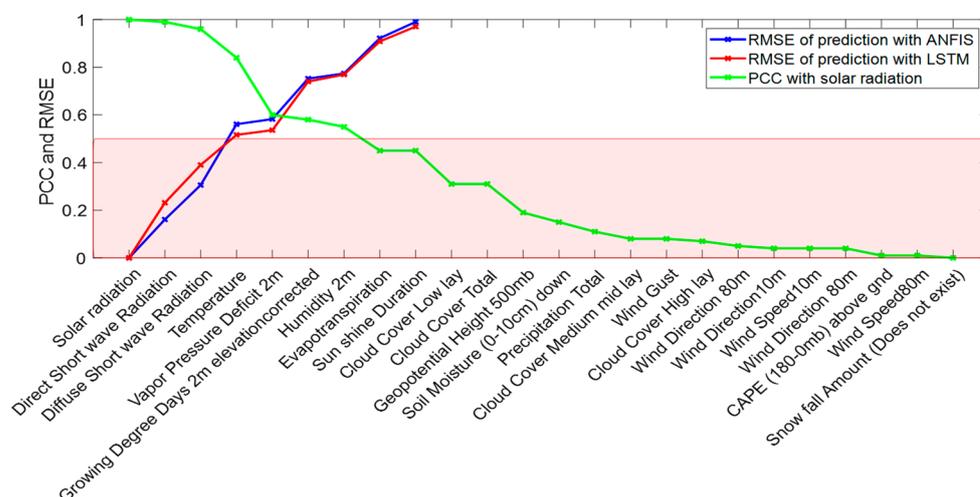

Figure 12. Spring season ranking between different meteorological parameters based on Pearson Correlation Coefficients and comparative performance of LSTM and ANFIS models with significant parameters selection (PCC ≥ 0.5) (dataset from 21 March 2021 to 20 June 2021).

4. Discussion

In this research work, twenty-four meteorological parameters have been processed to investigate their influence on solar radiation in west-central Jordan. We used the PCC to select the most significant parameters to facilitate solar radiation prediction. After selecting the parameters, ANFIS and LSTM methods are used to forecast the solar radiation and calculate the prediction RMSE for each selected parameter according to solar radiation (i.e., with PCC value ≥ 0.5). The selected parameters have been then treated to study their influence on solar radiation with the changing of the seasons.

Figure 8 shows the first attempt to study solar radiation forecasting in west-central Jordan, based on a five-year database relative to twenty-four meteorological parameters (from 1 January 2017, until 22 August 2021). Only parameters with a PCC value greater than 0.5 (listed in the previous section) were selected for the forecasting process by LSTM and ANFIS to obtain low and acceptable RMSE values. The parameters that strongly correlate with the solar radiation (PCC in the range 0.98 ÷ 1) are the solar radiation itself, direct short-wave radiation, and diffuse short-wave radiation. As for the solar radiation, ANFIS provides a low RMSE of 0.04 and an LSTM of 0.07. As regards parameters with a PCC between 0.5 and 0.8, the LSTM method certainly performs better, providing lower and acceptable RMSE values; for example, for the temperature parameter (PCC = 0.8), LSTM gives a low RMSE of 0.35, while ANFIS gives a much higher value, equal to 0.6. The sunshine duration, soil temperature and cloud cover have a low influence on solar radiation (PCC ≤ 0.5) because Jordan is poor in rain and cloud.

As for 2020 summer forecasting (Figure 9), the parameters that strongly correlate with solar radiation having PCC values between 0.98 and 1, are the solar radiation itself, direct and diffuse short-wave radiation, and temperature. Other parameters have an average correlation (PCC between 0.5 and 0.8) with solar radiation, such as the sunshine duration, growing degree days at 2 m elevation, vapor pressure deficit, and relative humidity. In particular, the sunshine duration has a remarkable influence in summer, stronger than other parameters, whereas it has no noticeable influence in other seasons.

As for the 2020 autumn season, the parameters with PCC $\geq$ 0.5 selected for LSTM and ANFIS analysis are shown in Figure 10. In more detail, parameters that strongly correlate with the solar radiation (i.e., PCC values between 0.98 and 1) are the solar radiation itself, and direct and diffuse short-wave radiation. Other parameters with significant correlation (PCC between 0.5 and 0.8) are temperature, vapor pressure deficit, growing degree days at 2 m elevation, relative humidity, sunshine duration, and evapotranspiration. Notably, the temperature and sunshine duration parameters have less influence in the autumn (PCC equal to 0.8 and 0.55 respectively) than the summer season with a PCC of 0.98 and 0.7, respectively. In comparison, evapotranspiration has more influence in autumn than summer and significantly influences solar radiation.

As for the 2020 winter season, the parameters with the highest correlation (PCC in the range 0.95 $\div$ 1) are the solar radiation, direct and diffuse short-wave radiation. Other parameters with average correlation (PCC between 0.5 and 0.8) are the temperature, evapotranspiration, vapor pressure deficit 2 m, relative humidity, growing degree days at 2 m elevation, and sunshine duration, listed by decreasing PCC values (Figure 11). Notably, the temperature has a lower influence (PCC = 0.8) than the summer season, with a PCC of 0.98. In contrast, the evapotranspiration parameter has a greater influence on solar radiation forecasting (PCC = 0.8) than in summer (PCC = 0.45).

As for the 2021 spring season, the parameters that strongly correlate with the solar radiation (PCC between 0.95 and 1) are solar radiation itself, direct short-wave radiation, and diffuse short-wave radiation, while other parameters have PCC values between 0.5 and 0.8 (Figure 12). In this season, the temperature has less influence than in the summer season, with a PCC of 0.8 compared to 0.98 in summer. Evapotranspiration and sunshine duration were not selected for the forecasting process because their PCC is less than 0.5 (0.48 for both parameters in the spring season compared to PCC values of 0.80 in winter and 0.7 in summer for the two parameters).

Based on the previous analysis of solar radiation forecasting in the different seasons, the significant parameters have been grouped into two classes depending on the determined PCC value (Table 4). From the results reported in Section 3 (shown in Figures 8–12), as for first-class parameters (i.e., likely linear correlation) as inputs to the solar radiation forecasting process, the ANFIS model gives low RMSE values compared to LSTM. Instead, concerning second class parameters (PCC ranges from 0.5 to 0.8, i.e., unlikely linear correlation), the LSTM performs better than the ANFIS method based on the RMSE values for the forecasting process performed in all four seasons.

Table 4. Parameters that influence solar radiation in the different seasons (classified into two different classes according to PCC value) and used for the solar radiation forecasting process by LSTM and ANFIS methods.

Season	Parameters	PCC Value	Parameters	PCC Value
Summer	Solar radiation	0.98 ÷ 1	Sunshine duration	0.5 ÷ 0.8
	Direct short-wave radiation		Growing degree days 2 m elevation	
	Diffuse short-wave radiation		Vapor pressure deficit at 2 m	
	Temperature		Relative humidity at 2 m	
Autumn	Solar radiation	0.98 ÷ 1	Temperature	0.5 ÷ 0.8
	Direct short-wave radiation		Vapor pressure deficit at 2 m	
	Diffuse short-wave radiation		Growing degree days at 2 m elevation	
			Relative humidity at 2 m	
			Sunshine duration	
			Evapotranspiration	
Winter	Solar radiation	0.95 ÷ 1	Temperature	0.5 ÷ 0.8
	Direct short-wave radiation		Evapotranspiration	
	Diffuse short-wave radiation		Vapor pressure deficit at 2 m	
			Relative humidity at 2 m	
			Growing degree days at 2 m elevation	
			Sunshine duration	
Spring	Solar radiation	0.98 ÷ 1	Temperature	0.5 ÷ 0.8
	Direct short-wave radiation		Vapor pressure deficit at 2 m	
	Diffuse short-wave radiation		Growing degree days at 2 m elevation	
			Relative humidity at 2 m	

This work provides the following contributions: firstly, the parameters of solar radiation, direct short-wave radiation, diffuse short-wave radiation, and temperature always have a very high degree of influence on solar radiation forecasting based on results obtained with both the complete five-year dataset as well as the seasonal ones. Secondly, evapotranspiration, sunshine duration and humidity showed a remarkable influence in west-central Jordan; instead, other parameters like cloud cover, snowfall amount, wind speed, and total precipitation amount have no influence in Jordan on the solar radiation prediction. Due to Jordan's geographical location with relatively high values of daily solar irradiance, the average sunshine duration is approximately 300 days a year, with average daily sunshine of 9.07 h. The PCC values' ranking of the different parameters with respect to solar radiation can change with the season. For example, regarding the temperature, the obtained PCC values with respect to solar radiation are 0.99 in summer, 0.85 in spring, 0.83 in autumn, and 0.81 in winter; this means that in summer, the temperature has a higher correlation with solar radiation than other seasons. The RMSE of forecasting by LSTM equals 0.14 in summer and only 0.5 in winter, as the PCC between the temperature and solar radiation decreases.

Table 5 shows the five criteria researchers must consider in order to build a reliable so-lar radiation forecasting model; namely, test location, time duration of the study, employed parameters as inputs, machine learning models, and evaluation criteria. The experimental data were collected worldwide, whereas the last three studies were conducted in Jor-dan [29–31]. For the published research articles, the time-length of studied parameters in the analysis that influence the solar radiation prediction ranged from 5 months to 20 years. The employed parameters varied in number and type.

In the presented research work, twenty-four parameters have been involved for the first time, a significantly higher number than in the scientific literature to date. The achieved RMSE ranges from 0.04 to 0.8, which is very competitive compared to other experimental results obtained in the same region. As for the research works carried out in North America and Hawaii [15,16,19], the analysis time length ranged from 6 to 20 months, obtaining RMSE values equal to 6.11 and 0.086, respectively. This means that the short-term forecast (only six months) provides a lower outcome prediction with respect to the longer ones (up to 20 months). In these studies, the DL LSTM method was employed in [15] and [16]; up to five forecasting parameters were studied, including wind, clouds, longitude, and latitude. Compared to these published results, our research work presents some advantages, such as the longer time length (5 years), significantly lower RMSE values (in the range 0.04–0.86), and a particularly high number of studied parameters (up to 24), some of which were analyzed for the first time, to our knowledge.

In [56], the authors proposed a new short-term load forecasting model that integrates different machine learning methods, such as support vector regression (SVR), grey catas-trophe, and RF modeling. The developed model, focusing on characteristics of electric load sequence as stability and flexibility sequence, can help systems to balance power supply and demand, to avoid possible catastrophes, to rationally allocate resources, and to capture trends in power system loads. In studies conducted in East-Asian countries with a time length from 6 months to 3 years, the obtained RMSE values varied from 0.086 to 1.39 [17,18,21,22]. The number of forecasting parameters was five, including the dew point and wind speed. The best RMSE (i.e., 0.086) was achieved using the LSTM model and a three-year analysis. A six-month short-term study for solar irradiation forecasting was car-ried out through ML methods in Mexico [27]. The best MSE values, by processing acquired data related to six ambient parameters, were obtained by MLP (Multi-layer perceptron) and RF (Random Forest) algorithms, respectively 0.222 and 0. In the study presented in [25] and conducted in Scotland for a four-year period, the determined MAE was equal to 0.525 for day-ahead and 0.708 for week-ahead forecasting by using only three parameters as inputs to the stacked bidirectional LSTM neural network. In Al-Sbou et al. [30], the minimum MSE value obtained relative to the solar radiation prediction was equal to 0.00237.

Compared to these reported performances, in this research work, we obtained better results, as shown in the following Table 5 (data shown in the last line). Based on the results reported in Section 3 in relation to the solar radiation prediction, we have obtained RMSE values in the range 0.04–0.8, MSE values in the range 0.0016–0.64 and MAE values between 0.034 and 0.86.

Table 5. Comparison analysis of the research studies reported in the literature with the proposed one in this work related to solar radiation forecasting carried out in the Jordanian territory and worldwide.

Reference	Test Location	Time Duration of the Study	Employed Parameters as Inputs	Machine Learning Models	Evaluation Criteria
Prado-Rujas et al. [15]	Oahu island (Hawaii)	Twenty months	Global Horizontal Irradiance (GHI), wind, longitude, latitude	RNN, LSTM, BiLSTM	RMSE less than 15%
Yan et al. [16]	Nevada desert, USA	One year (seasonal analysis)	Sun position, temperature, wind speed, and cloud movement.	Gated recurrent unit (GRU) NM, LSTM,	Best RMSE = 11.44 (in winter)
Yen et al. [17]	Southern Taiwan	Seventeen months	Temperature, humidity, rainfall, and wind speed.	SVM, RF	Best RMSE = 1.3912
Lee et al. [18]	South Korea	Three years	Temperature, wind speed, humidity, and ground temperature	CNN, LSTM	Best RMSE = 0.0987
Poolla et al. [19]	USA (California)	Six months	Solar irradiance, temperature, and windspeed spanning	Autoregressive ARX model	Best RMSE = 1.63 (wind)
Wang et al. [21]	India	One year	Historical power, solar irradiance, panel temperature.	LSTM, Conv-LSTM A-S	Best RMSE = 0.12 (Conv-LSTM-S)
Munir et al. [22]	Pakistan	One year	Temperature, dew point, relative humidity, and wind speed	Artificial Neural Network (ANN)	Average MAPE = 14.33%
Obiora et al. [23]	Johannesburg (South Africa)	Five years	Temperature, relative humidity, solar radiation, and sunshine duration	LSTM, CNN, ConvLSTM, and hybrid CNN-LSTM	Best RMSE = 7.18 (ConvLSTM)
de Guia et al. [24]	Morong, (Philippines)	Six month	Humidity, station temperature, ambient temperature, station altitude, sea level, absolute pressure and wind speed	ANN, CNN, bidirectional and stacked LSTM.	Best MAE = 41.738
Zou et al. [25]	Scotland	Five years	Temperature, precipitation, and wind speed	Bidirectional LSTM	MAE = 0.525 (Day-ahead) 0.708 (Week-ahead)
Tiwari et al. [26]	Johannesburg (South Africa)	Five years	Temperature, relative humidity, solar radiation, sunshine duration	Convolutional LSTM	NRMSE = 1.62%.
Alvarez et al. [27]	Aguascalientes (Mexico)	Six months	Wind velocity and direction, irradiance, temperature, humidity, pressure	SVM, Linear Regression (LR) and NNMs	Mean Squared Error (MSE) 0.2222
Alomari et al. [29]	Center Jordan	Thirty months	Solar radiation	ANN	RMSE = 0.0721
Al-Sbou et al. [30]	South Jordan	One year	Solar radiation	ANN	MSE = 0.00237
Shboul et al. [31]	North, south, center Jordan	Twenty years	Wind, air temperature, solar radiation	ANN	MAPE values < 3%
This research work	West-central Jordan	Five years	The 24 parameters listed in Table 1	ANFIS, LSTM	RMSE in the range 0.04–0.8 MSE in the range 0.0016–0.64 MAE in the range 0.034–0.86

5. Conclusions

This work presents two learning models, LSTM and ANFIS, to forecast solar energy radiation in west-central Jordan. The proposed ML models process meteorological data for the last five years, downloaded from the Meteobleu site (Table 1). Many new parameters, not yet studied before in the literature, were considered for solar radiation forecasting in our study. A PCC algorithm is used to indicate the most influencing parameters correlated with solar radiation to facilitate the training process with LSTM and ANFIS.

An important result of the proposed work is that new parameters greatly influence solar radiation which have not been previously investigated in other studies everywhere, such as direct short-wave radiation, diffuse short-wave radiation, temperature, sunshine duration and evapotranspiration. According to obtained results, these parameters remarkably influence solar radiation, differently depending on the seasons. Also, our study affirms that the LSTM is the best model for solar radiation forecasting when the PCC is not high (i.e., in the range 0.5–0.8). In contrast, the ANFIS model gives lower RMSE values concerning first-class parameters with a high correlation with the solar radiation (i.e., PCC values between 0.95 and 1) (Table 4). In total 24 meteorological parameters have been analyzed, a very large set; the results showed that the influence of each parameter varies significantly according to the season; altogether this we believe is an important result not yet reported in the literature. Summarizing the experimental results reported in Section 3 for the LSTM and ANFIS models, we obtained RMSE values in the range 0.04–0.8, MSE in the range 0.0016–0.64 and MAE between 0.034 and 0.86, very competitive values compared to the existing literature as reported in the comparative Table 5. We believe the model can be improved by building a local weather station, which provides meteorological records every 10–15 min. In addition, we plan to apply the methodology by exploiting information from another region, the city of El Kerak in the south of Jordan, with specific meteorological features different from west-central Jordan.

Author Contributions: Conceptualization, H.F., A.A.A., A.A.-O.; methodology, H.F., A.A.A., A.A.-O.; software, H.F., R.D.F.; validation, H.F., P.V.; formal analysis, B.A.-N., P.V.; investigation, H.F., A.A.-O., B.A.-N., R.D.F.; data curation, H.F., R.D.F.; writing—original draft preparation, A.A.-O., B.A.-N., R.D.F., P.V.; writing—review and editing, A.A.-O., B.A.-N., R.D.F., P.V.; supervision, A.A.-O., B.A.-N., P.V.; funding acquisition, H.F., P.V. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all involved subjects.

Data Availability Statement: Data of our study are available upon request.

Abbreviations

Acronyms	Extended Meaning of the Acronym
LSTM	Long short-term memory
ANFIS	Adaptive neuro fuzzy inference system
PCC	Pearson correlation coefficient
PCA	Principal Component Analysis
PV	Photovoltaic
ML	Machine learning
DL	Deep learning
GDP	Gross Domestic Product
VPD	Vapor pressure deficit
CAPE	Convective available potential energy

RMSE	Root-Mean-Square Error
SVM	Support Vector Machines
RF	Random Forest
CNN	Convolutional Neural Network
ARTS	Auto-regressive time-series
ANN	Artificial Neural Network
MAE	Mean Absolute Error
NNM	Neural network models
SVD	Singular value decomposition
NIPALS	Nonlinear iterative partial least squares
SAO	Successive average orthogonalization
MLP	Multilayer perceptron
RNN	Recurrent neural network
GRU	Gated recurrent unit
ARX	Autoregressive-exogenous
LR	Linear regression

References

1. Sayed, E.T.; Wilberforce, T.; Elsaid, K.; Rabaia, M.K.H.; Abdelkareem, M.A.; Chae, K.-J.; Olabi, A.G. A critical review on environmental impacts of renewable energy systems and mitigation strategies: Wind, hydro, biomass and geothermal. *Sci. Total Environ.* **2021**, *766*, 144505. [CrossRef] [PubMed]

2. de Araujo, J.M.S. Improvement of Coding for Solar Radiation Forecasting in Dili Timor Leste—A WRF Case Study. *J. Power Energy Eng.* **2021**, *9*, 7–20. [CrossRef]

3. Ziane, A.; Necaibia, A.; Sahouane, N.; Dabou, R.; Mostefaoui, M.; Bouraiou, A.; Khelifi, S.; Rouabhia, A.; Blal, M. Photovoltaic output power performance assessment and forecasting: Impact of meteorological variables. *Sol. Energy* **2021**, *220*, 745–757. [CrossRef]

4. Alawasa, K.M.; Al-Odienat, A.I. Power quality characteristics of residential grid-connected inverter ofphotovoltaic solar system. In Proceedings of the 2017 IEEE 6th International Conference on Renewable Energy Research and Applications (ICRERA), San Diego, CA, USA, 5–8 November 2017; pp. 1097–1101. [CrossRef]

5. Gielen, D.; Boshell, F.; Saygin, D.; Bazilian, M.D.; Wagner, N.; Gorini, R. The role of renewable energy in the global energy transformation. *Energy Strategy Rev.* **2019**, *24*, 38–50. [CrossRef]

6. Strielkowski, W.; Civin, L.; Tarkhanova, E.; Tvaronaviciene, M.; Petrenko, Y. Renewable Energy in the Sustainable Development of Electrical Power Sector: A Review unit. *Energies* **2021**, *14*, 8240. [CrossRef]

7. Lay-Ekuakille, A.; Ciaccioli, A.; Griffo, G.; Visconti, P.; Andria, G. Effects of Dust on Photovoltaic Measurements: A Comparative Study. *Measurement* **2018**, *113*, 181–188. [CrossRef]

8. McGee, T.G.; Mori, K. The Management of Urbanization, Development, and Environmental Change in the Megacities of Asia in the Twenty-First Century. In *Living in the Megacity: Towards Sustainable Urban Environments*; Springer: Tokyo, Japan, 2021; Volume 17, Chapter 2; ISBN 978-4-431-56899-5. [CrossRef]

9. Wilson, G.A.; Bryant, R.L. *Environmental Management: New Directions for the Twenty-First Century*; Routledge: London, UK, 2021; ISBN 9780203974988. [CrossRef]

10. Ismail, A.M.; Ramirez-Iniguez, R.; Asif, M.; Munir, A.B.; Muhammad-Sukki, F. Progress of solar photovoltaic in ASEAN countries: A review. *Renew. Sustain. Energy Rev.* **2015**, *48*, 399–412. [CrossRef]

11. Al-Odienat, A.; Al-Maitah, K. A modified Active Frequency Drift Method for Islanding Detection. In Proceedings of the 2021 12th International Renewable Engineering Conference (IREC), Amman, Jordan, 14–15 April 2021; pp. 1–6. [CrossRef]

12. Srivastava, R.; Tiwari, A.N.; Giri, V.K. Prediction of Electricity Generation using Solar Radiation Forecasting Data. In Proceedings of the 2020 International Conference on Electrical and Electronics Engineering (ICE3), Gorakhpur, India, 14–15 February 2020. [CrossRef]

13. Alawasa, K.M.; Al-Odienat, A.I. Power Quality Investigation of Single Phase Grid-connected Inverter of Photovoltaic System. *J. Eng. Technol. Sci.* **2019**, *51*, 597–614. [CrossRef]

14. Castangia, M.; Aliberti, A.; Bottaccioli, L.; Macii, E.; Patti, E. A compound of feature selection techniques to improve solar radiation forecasting. *Expert Syst. Appl.* **2021**, *178*, 114979. [CrossRef]

15. Prado-Rujas, I.I.; García-Dopico, A.; Serrano, E.; Pérez, M.S. A Flexible and Robust Deep Learning-Based System for Solar Irradiance Forecasting. *IEEE Access* **2021**, *9*, 12348–12361. [CrossRef]

16. Yan, K.; Shen, H.; Wang, L.; Zhou, H.; Xu, M.; Mo, Y. Short-term solar irradiance forecasting based on a hybrid deep learning methodology. *Information* **2020**, *11*, 32. [CrossRef]

17. Yen, C.F.; Hsieh, H.-Y.; Su, K.-W.; Yu, M.-C.; Leu, J.-S. Solar Power Prediction via Support Vector Machine and Random Forest. *E3S Web Conf.* **2018**, *69*, 01004. [CrossRef]

18. Lee, W.; Kim, K.; Park, J.; Kim, J.; Kim, Y. Forecasting solar power using long-short term memory and convolutional neural networks. *IEEE Access* **2018**, *6*, 73068–73080. [CrossRef]

19. Poolla, C.; Ishihara, A.K. Localized solar power prediction based on weather data from local history and global forecasts. In Proceedings of the 2018 IEEE 7th World Conference on Photovoltaic Energy Conversion (WCPEC) (A Joint Conference of 45th IEEE PVSC, 28th PVSEC, 34th EU PVSEC), Waikoloa, HI, USA, 10–15 June 2018; pp. 2341–2345. [CrossRef]

20. Han, J.; Park, W.-K. A Solar Radiation Prediction Model Using Weather Forecast Data and Regional Atmospheric Data. In Proceedings of the 2018 IEEE 7th World Conference on Photovoltaic Energy Conversion (WCPEC) (A Joint Conference of 45th IEEE PVSC, 28th PVSEC & 34th EU PVSEC), Waikoloa, HI, USA, 10–15 June 2018; pp. 2313–2316. [CrossRef]

21. Wang, Y.; Chen, Y.; Liu, H.; Ma, X.; Su, X.; Liu, Q. Day-Ahead Photovoltaic Power Forcasting Using Convolutional-LSTM Networks. In Proceedings of the 3rd Asia Energy and Electrical Engineering Symposium (AEEES), Chengdu, China, 26–29 March 2021; pp. 917–921. [CrossRef]

22. Munir, M.A.; Khattak, A.; Imran, K.; Ulasyar, A.; Khan, A. Solar PV Generation Forecast Model Based on the Most Effective Weather Parameters. In Proceedings of the 2019 International Conference on Electrical, Communication, and Computer Engineering (ICECCE), Swat, Pakistan, 24–25 July 2019; pp. 1–5. [CrossRef]

23. Obiora, C.N.; Ali, A.; Hasan, A.N. Estimation of Hourly Global Solar Radiation Using Deep Learning Algorithms. In Proceedings of the 2020 11th International Renewable Energy Congress (IREC), Hammamet, Tunisia, 29–31 October 2020; pp. 1–6. [CrossRef]

24. de Guia, J.D.; Concepcion, R.S.; Calinao, H.A.; Alejandrino, J.; Dadios, E.P.; Sybingco, E. Using Stacked Long Short Term Memory with Principal Component Analysis for Short Term Prediction of Solar Irradiance based on Weather Patterns. In Proceedings of the 2020 IEEE Region 10 Conference (TENCON), Osaka, Japan, 16–19 November 2020. [CrossRef]

25. Zou, M.; Fang, D.; Harrison, G.; Djokic, S. Weather Based Day-Ahead and Week-Ahead Load Forecasting using Deep Recurrent Neural Network. In Proceedings of the 2019 IEEE 5th International Forum on Research and Technology for Society and Industry (RTSI), Florence, Italy, 9–12 September 2019; pp. 341–346. [CrossRef]

26. Tiwari, S.; Sabzehgar, R.; Rasouli, M. Short term solar irradiance forecast based on image processing and cloud motion detection. In Proceedings of the 2019 IEEE Texas Power and Energy Conference (TPEC), College Station, TX, USA, 7–8 February 2019; pp. 1–6. [CrossRef]

27. Alvarez, L.F.J.; González, S.R.; López, A.D.; Delgado, D.A.H.; Espinosa, R.; Gutiérrez, S. Renewable Energy Prediction through Machine Learning Algorithms. In Proceedings of the 2020 IEEE ANDESCON, Quito, Ecuador, 13–16 October 2020; pp. 1–6. [CrossRef]

28. Huang, C.-J.; Ma, Y.; Chen, Y.-H. Solar Radiation Forecasting based on Neural Network in Guangzhou. In Proceedings of the 2020 International Automatic Control Conference (CACS), Hsinchu, Taiwan, 4–7 November 2020; pp. 1–5. [CrossRef]

29. Alomari, M.H.; Adeeb, J.; Younis, O. Solar photovoltaic power forecasting in jordan using artificial neural networks. *Int. J. Electr. Comput. Eng.* **2018**, *8*, 497–504. [CrossRef]

30. Al-Sbou, Y.A.; Alawasa, K.M. Nonlinear autoregressive recurrent neural network model for solar radiation prediction. *Int. J. Appl. Eng. Res.* **2017**, *12*, 4518–4527.

31. Shboul, B.; Ismail, A.-A.; Michailos, S.; Ingham, D.; Ma, L.; Hughes, K.J.; Pourkashanian, M. A new ANN model for hourly solar radiation and wind speed prediction: A case study over the north & south of the Arabian Peninsula. *Sustain. Energy Technol. Assess.* **2021**, *46*, 101248. [CrossRef]

32. Kassambara, A. *Practical Guide to Cluster Analysis in R: Unsupervised Machine Learning*; Statistical Tools for High-Throughput Data Analysis STHDA: Marseille, France, 2017; Volume 1, ISBN 978-1542462709.

33. Huang, H.; Jia, R.; Shi, X.; Liang, J.; Dang, J. Feature selection and hyper parameters optimization for short-term wind power forecast. *Appl. Intell.* **2021**, *51*, 6752–6770. [CrossRef]

34. Al-Odienat, A.; Gulrez, T. Inverse covariance principal component analysis for power system stability studies. *Turk. J. Electr. Eng. Comput. Sci.* **2014**, *22*, 57–65. [CrossRef]

35. Gulrez, T.; Al-Odienat, A. A New Perspective on Principal Component Analysis using Inverse Covariance. *Int. Arab J. Inf. Technol.* **2015**, *12*, 104–109.

36. Hu, C.; He, S.; Wang, Y. A classification method to detect faults in a rotating machinery based on kernelled support tensor machine and multilinear principal component analysis. *Appl. Intell.* **2021**, *51*, 2609–2621. [CrossRef]

37. Mukherjee, A.; Kundu, P.K.; Das, A. A supervised principal component analysis-based approach of fault localization in transmission lines for single line to ground faults. *Electr. Eng.* **2021**, *103*, 2113–2126. [CrossRef]

38. Guo, Y.; Zhou, Y.; Zhang, Z. Fault diagnosis of multi-channel data by the CNN with the multilinear principal component analysis. *Measurement* **2021**, *171*, 108513. [CrossRef]

39. Shaker, H.; Zareipour, H.; Wood, D. A Data-driven Approach for Estimating the PowerGeneration of Invisible Solar Sites. *IEEE Trans. Smart Grid* **2016**, *7*, 2466–2476. [CrossRef]

40. Zavareh, M.; Maggioni, V.; Sokolov, V. Investigating water quality data using principal component analysis and granger causality. *Water* **2021**, *13*, 343. [CrossRef]

41. Wang, L.; Shi, J. A Comprehensive Application of Machine Learning Techniques for Short-Term Solar Radiation Prediction. *Appl. Sci.* **2021**, *11*, 5808. [CrossRef]

42. Zhan, J.; Shi, H.; Wang, Y.; Yao, Y. Complex Principal Component Analysis of Antarctic Ice Sheet Mass Balance. *Remote Sens.* **2021**, *13*, 480. [CrossRef]

43. Benesty, J.; Chen, J.; Huang, Y. On the importance of the Pearson correlation coefficient in noise reduction. *IEEE Trans. Audio Speech Lang. Process.* **2008**, *16*, 757–765. [CrossRef]

44. Edelmann, D.; Móri, T.F.; Székely, G.J. On relationships between the Pearson and the distance correlation coefficients. *Stat. Probab. Lett.* **2021**, *169*, 108960. [CrossRef]

45. Granados-López, D.; Suárez-García, A.; Díez-Mediavilla, M.; Alonso-Tristán, C. Feature selection for CIE standard sky classification. *Sol. Energy* **2021**, *218*, 95–107. [CrossRef]

46. Gensler, A.; Henze, J.; Sick, B.; Raabe, N. Deep Learning for solar power forecasting—An approach using AutoEncoder and LSTM Neural Networks. In Proceedings of the 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Budapest, Hungary, 9–12 October 2016; pp. 002858–002865. [CrossRef]

47. Chandola, D.; Gupta, H.; Tikkiwal, V.A.; Bohra, M.K. Multi-step ahead forecasting of global solar radiation for arid zones using deep learning. *Procedia Comput. Sci.* **2020**, *167*, 626–635. [CrossRef]

48. Huynh, A.N.-L.; Deo, R.C.; An-Vo, D.-A.; Ali, M.; Raj, N.; Abdulla, S. Near real-time global solar radiation forecasting at multiple time-step horizons using the long short-term memory network. *Energies* **2020**, *13*, 3517. [CrossRef]

49. Yang, H.-T.; Huang, C.-M.; Huang, Y.-C.; Pai, Y.-S. A weather-based hybrid method for 1-day ahead hourly forecasting of PV power output. *IEEE Trans. Sustain. Energy* **2014**, *5*, 917–926. [CrossRef]

50. Zhu, T.; Guo, Y.; Li, Z.; Wang, C. Solar Radiation Prediction Based on Convolution Neural Network and Long Short-Term Memory. *Energies* **2021**, *14*, 8498. [CrossRef]

51. Al-Naami, B.; Fraihat, H.; Gharaibeh, N.Y.; Al-Hinnawi, A.-R.M. A Framework Classification of Heart Sound Signals in PhysioNet Challenge 2016 Using High Order Statistics and Adaptive Neuro-Fuzzy Inference System. *IEEE Access* **2020**, *8*, 224852–224859. [CrossRef]

52. Fraihat, H.; Madani, K.; Sabourin, C. Learning-based distance evaluation in robot vision: A comparison of ANFIS, MLP, SVR and bilinear interpolation models. In Proceedings of the 2015 7th International Joint Conference on Computational Intelligence (IJCCI), Lisbon, Portugal, 12–14 November 2015; pp. 168–173.

53. Gbémou, S.; Eynard, J.; Thil, S.; Guillot, E.; Grieu, S. A Comparative Study of Machine Learning-Based Methods for Global Horizontal Irradiance Forecasting. *Energies* **2021**, *14*, 3192. [CrossRef]

54. Rushdi, M.A.; Yoshida, S.; Watanabe, K.; Ohya, Y. Machine Learning Approaches for Thermal Updraft Prediction in Wind Solar Tower Systems. *Renew. Energy* **2021**, *177*, 1001–1013. [CrossRef]

55. Visconti, P.; De Fazio, R.; Cafagna, D.; Velazquez, R.; Lay-Ekuakille, A. A Survey on Ageing Mechanisms in II and III-Generation PV Modules: Accurate Matrix-Method Based Energy Prediction Through Short-Term Performance Measures. *Int. J. Renew. Energy Res.* **2021**, *11*, 178–194.

56. Fan, G.-F.; Yu, M.; Dong, S.-Q.; Yeh, Y.-H.; Hong, W.-C. Forecasting short-term electricity load using hybrid support vector regression with grey catastrophe and random forest modeling. *Util. Policy* **2021**, *73*, 101294. [CrossRef]