

Received 29 June 2024, accepted 15 July 2024, date of publication 23 July 2024, date of current version 5 August 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3432574



## RESEARCH ARTICLE

# Forecasting and Performance Analysis of Energy Production in Solar Power Plants Using Long Short-Term Memory (LSTM) and Random Forest Models

KADİR OLÇAY<sup>ID1</sup>, (Member, IEEE), SAMET GİRAY TUNCA<sup>ID1</sup>, AND MUSTAFA ARİF ÖZGÜR<sup>ID2</sup>

<sup>1</sup>Dumlupınar Vocational School, Kütahya Dumlupınar University, 43030 Kütahya, Türkiye

<sup>2</sup>Mechanical Engineering Department, Kütahya Dumlupınar University, 43030 Kütahya, Türkiye

Corresponding author: Kadir Olcay (kadir.olcay@dpu.edu.tr)

**ABSTRACT** The rapid increase in energy demand and the disadvantages of using fossil fuels in electricity production have led to a greater emphasis on renewable energy sources. Consequently, research on the use of renewable resources has gained importance. Numerous factors influence the energy production of power plants that generate electricity from these sources. Power plants utilizing solar energy, one of the renewable energy sources, are significantly affected by environmental factors and meteorological variables, impacting the continuity of electrical energy production in solar power plants (SPPs). For these reasons, this study developed prediction models using two different methods based on machine learning and artificial intelligence to analyze and predict changes in the electrical energy production of SPPs due to environmental factors and meteorological changes. The data used in the study are real data collected from a 180 kWe solar power plant currently in operation. Data collection started from the day the power plant was commissioned. Using real data, the effects of pollution and environmental impacts on PV panels' energy production are demonstrated. To mitigate these effects and examine the impact of adverse conditions on production efficiency, two different analysis methods were used: Random Forest Regression (RFR) Model and Artificial Neural Networks (ANN). This allowed for a comparison of the results between the models. Long Short-Term Memory (LSTM) networks, a type of artificial neural network, were utilized. A prediction model was created for the decrease in energy production of the power plant due to pollution and environmental impacts using Random Forest (RF) regression analysis, which analyzes energy production based on non-linear independent input variables and creates a prediction model. The model estimated SPP's electrical energy production based on environmental impact measurements and pollution. A graph comparing estimated energy production amounts with actual values is shown. In another analysis phase, neural networks were trained with data from the SPP and measurement station using Long Short-Term Memory (LSTM) networks. The energy production of the power plant was estimated with the trained LSTM neural networks, and the results are shown graphically. A very large data set was used in the two different prediction models. The training data set includes hourly sunshine duration, accumulated irradiation ( $\text{Wh}/\text{m}^2$ ), hourly maximum temperature, hourly minimum temperature, humidity (%), hourly temperature, hourly total precipitation ( $\text{kg}/\text{m}^2$ ), and daily and hourly data since the power plant began operation. It consists of data on wind speed (m/s), pollution, and energy production values of the power plant. This means a total of 119,808 data points were processed in the prediction model, highlighting the detail of the analysis. The results were evaluated using four different performance measures: correlation coefficient (R), fractional gross error (FGE), mean standard error (MBE), and root mean square error (RMSE). The RF results showed a correlation coefficient of 0.8111 with the predictions. In contrast, the LSTM network predictions had an R value of 0.9759. Comparing RFR and LSTM, it is evident that LSTM provides much better results in models created with the entire data set.

The associate editor coordinating the review of this manuscript and approving it for publication was Arash Asrari<sup>ID</sup>.

**INDEX TERMS** Solar power plants, random forest, renewable energy, efficiency, pollution effects, artificial neural networks, long short-term memory, LSTM.

## I. INTRODUCTION

As technology advances, the demand for energy continues to grow. This situation gradually increases the usage of fossil fuels and exacerbates their environmental impacts. Increasing energy needs and negative environmental impacts disrupt the balance of the life cycle on earth. Additionally, energy costs are rising, leading to a global power struggle. All these reasons indicate that it is necessary to shift towards renewable energy to meet energy needs.

Although energy demand is met economically, the primary cause of greenhouse gas emissions is fossil fuels [1]. In the studies conducted, the environmental effects of solar energy systems were discussed in detail and it was emphasized that the relevant procedures should be followed [2]. Carbon emissions encourage studies in the field of renewable energy, and a study on the benefits of renewable energy in terms of emissions was conducted in [3]. Although the use of fossil fuels is considered economical, it has many harmful effects on the environment [4]. However, the generation of large amounts of greenhouse gases has a negative impact on the carbon footprint [5].

It is predicted that the world population will double in the next 25 years [6], [7]. The need for energy has increased as a result of increasing population growth and parallel development of industry in the world [8]. The fact that fossil fuels do not have sufficient reserves to meet the increasing energy need shows that renewable energy sources are needed [9]. Reducing dependence on fossil fuels encourages the transition to renewable energy technologies [10]. To prevent a climate crisis in the future, the 2015 Paris Agreement set a target of limiting global temperature rise to 1.5 to 2 degrees above pre-industrial levels [11]. Therefore, minimizing the use of fossil fuels and ensuring the use of energy obtained from clean energy sources should become the goal [12]. It is estimated that in 2040, more than one third of the energy produced from renewable energy sources will be obtained from solar and wind energy and solar energy production will reach 7200 TWh [13]. The sun has the capacity to provide  $1.7 \times 10^{22}$  J of energy in 1.5 days, which is equal to the total energy that 3 trillion barrels of oil resources can provide [14], [15].

Solar energy systems are seen as a useful energy source in regions where transmission lines are weak [3]. It has been examined that photovoltaic panels can be operated hybirdly with batteries and diesel fuel. Data obtained from solar panels on the truck were used in the study [16].

Therefore, harnessing solar energy will be a successful way to meet the world's energy needs, eliminate energy poverty in developing countries, and accelerate the transition to clean energy. The main benefits of using solar energy to generate electricity are widespread availability, easy installation, scalability, and environmental friendliness [17].

It indicates the ideal solar energy system size to be installed in different regions with the least system cost. Technical investigation of the fields has shown that there is a direct relationship between the efficiency of the system and the climate parameters [18].

The solar panel is a critical part of the solar energy system. PV technology collects solar radiation using semiconductor-based solar cells (panels) and converts it into electrical energy [19]. They perform better when there are no environmental issues and the panels are oriented and tilted to ideal angles. The accumulation of airborne dust on solar panel surfaces is one of the worst problems because it significantly reduces the efficiency of the panels [20].

PV cells produce electricity through an energy conversion process. Most often, photons or light energy strike photovoltaic cells, energizing electrons in semiconductor material atoms, resulting in an electrical voltage and current [21].

Their biggest disadvantage is that they are sensitive to weather changes. Environmental factors that affect the efficiency of solar systems include wind speed, solar radiation, ambient temperature, humidity and dust [22]. In another study, PV pollution was simulated with an aerosol model of dust particle densities typically collected from solar panels [23]. When dust accumulates on the surface of solar panels, less light reaches the solar cell and solar flux is restricted, resulting in power loss [24].

The effects of atmospheric effects on the efficiency of solar panels have been investigated. The performance effect of different particles on monocrystalline and polycrystalline panels has been studied [25]. It presents the result of a study on the impact of dust accumulation on power output in solar PV modules in the eastern region of Saudi Arabia. The study shows that solar PV modules left dirty for more than six months can experience power degradation of up to 50%. The solar tracker increases power output and helps reduce the impact of dust accumulation during off-peak times by 50% [26].

PV panel material, weather and location are just some of the many factors that affect the accumulation, adhesion and clearance processes of air particles [27]. In another study, it was stated that dust accumulating on the panel during certain periods reduces energy efficiency [28].

Soiling, the accumulation of dirt and dust on solar panels, is a major obstacle to their efficiency, leading to a significant drop in performance. This review [29] highlights how soiling has a particularly severe impact in arid and semi-arid regions, where abundant sunlight offers vast untapped potential for solar energy [29].

According to another study, air dust reduces the performance efficiency of solar energy systems by reducing the amount of sunlight reaching the surface [28]. Dust particles

cover the surface of the panels in a thin layer. In general, the size of these dust particles can be up to 10 mm. It has been determined that this situation may vary depending on geographical conditions [30]. This affects the performance of the panels [31].

Power output and module efficiency decrease as the mass of dust accumulation increases, and as particle size decreases, power output also decreases because smaller particles block more radiation and prevent it from reaching the surface of the solar system. This can cause a 6.5% loss in solar panel efficiency after two months of pollution; in desert environments, this rate of decline can be as high as 40% [21]. The pollution rate in the relevant regions determines this. The Middle East, North Africa, India and China are the regions most prone to dust accumulation, with losses ranging from 10% to 70% [13].

In a study conducted in Norway, energy losses due to pollution were examined. A high-precision method has been developed in which the weights of the cloths used to clean accumulated dust residues are evaluated [32].

It has been found in many studies that the angle of inclination has a significant effect on dust accumulation. In this regard, it has been determined that the panels in Europe are more inclined than in the Ecuadorian region and the dust effect should be evaluated appropriately [33].

For example, one study showed that Iraq has the highest annual average solar energy in the world and is among the countries with the greatest potential for the installation of large photovoltaic systems. The country also has the highest political support and interest in changing the energy mix and initiatives to develop photovoltaic power plants. It is reported to have one of the radiation values [34]. However, given that Iraq is a largely desert country, the country's climate and environmental factors can pose a significant obstacle [29].

Renewable energy sources, especially solar energy, play a critical role in the transition to sustainable energy systems. Accurate estimation of energy production is important for the efficient operation of solar power plants and their integration into the energy grid. In this context, the Random Forest (RF) regression model is widely used due to its robustness and high accuracy in handling complex, nonlinear relationships in large data sets. RF is an ensemble learning method that combines multiple decision trees to improve prediction performance and reduce overfitting. Advantages of RF compared to other regression models include its ability to capture interactions between variables well, its low risk of overfitting, and its robustness to missing data or noise. These features make RF stand out in multivariate and noisy data sets such as energy production. Breiman [35] demonstrated the effectiveness of RF in various regression tasks, and subsequent studies have demonstrated the successful use of this method in environmental modeling and energy forecasting applications. In [36], hourly solar energy estimation was made using RF. It has been said that the model created using RF in predicting energy production for three different regions has high prediction

accuracy. In the same study, it was emphasized that RF is widely used with large data sets due to its high tolerance and adaptability to bad data. In another study [37], hourly global solar radiation prediction was made using RF. Here, firefly algorithm was also used to optimize the created RF model. In addition, the optimized RF model was compared with traditional artificial neural networks (ANN) and ANN predictions optimized using the firefly algorithm, and it was said that RF gave better results. In another study, the RF model was proposed to achieve compatibility with nonlinear data in time series and to make a successful prediction [38]. With the created data set, solar radiation was estimated and high performance was achieved. Measuring variable importance and the problem of overfitting brings machine learning methods to the fore and is gaining popularity in studies in the field of energy. With this incentive, another study [39] recommends the use of RF for solar radiation estimation. A prediction model was created using meteorological variables and air pollution index for three different regions in China. The results were compared with empirical models and were shown to give better results. In [40], RF was used to obtain a new data set and estimate solar radiation by using estimates from different radiation data sets.

In recent years, deep learning techniques, especially Long Short Term Memory (LSTM) networks, have emerged as powerful tools in time series forecasting. LSTM is a type of recurrent neural network (RNN) widely used in time series data due to its ability to capture long-term dependencies and temporal patterns in sequential data. The advantages of LSTM compared to other artificial neural network models include its ability to preserve past information for a longer period of time, to be resistant to the vanishing gradient problem, and to model temporal dependencies better. This makes LSTM a promising approach for solar energy production predictions, where energy output is affected by time-varying environmental factors. Previous research has shown that LSTM networks can outperform traditional statistical and machine learning models in a variety of prediction applications. In [41], ANN types used in different solar energy studies were discussed and it was said that the use of ANN in modeling not only energy production but also solar devices such as solar cookers, dryers and heaters has significant benefits due to high accuracy, generalization ability and low calculation time. ANNs are used not only for the prediction or modeling of variables in solar electricity generation systems but also for the analysis of different solar-based applications. In [42], ANN was used to model a large thermal energy system and the prediction results obtained were evaluated as satisfactory. Various hybrid models and different approaches have also been studied in different studies. Studies on modeling solar energy electricity systems, analyzing their performance, and making various predictions using artificial intelligence methods between 2009 and 2019 were examined in [43]. In the literature review, more than 90 publications were examined and the reliability of

ANN, fuzzy logic and genetic algorithm (GA) in analyzing and predicting the performance of solar radiation and PV systems was demonstrated. In [44], a literature review was conducted for the use of ANN in estimating solar radiation originating from solar energy, and it is said that ANN is more accurate in hourly and monthly solar radiation predictions. In [45], solar radiation prediction was made by comparing ANN models among themselves. In the study, predictions were made using data from a local meteorological station and deep learning methods models were created using traditional ANN recurrent neural network (RNN). The results showed that RNN prediction had higher accuracy than the performance criteria. From this, it can be seen that the suitability of deep learning algorithms for the prediction to be made within themselves is important and that the criteria for this suitability should be reviewed when determining the algorithm to be chosen.

In this research, RF and LSTM networks models were created to predict the energy production of the solar power plant using hourly data for a year. Data were collected from this field from the day an SPP was installed and started producing energy. The collected data set; solar power plant electrical energy production, hourly measurement data, measurement month, measurement time, sunshine duration, cumulative radiation ( $\text{Wh/m}^2$ ), hourly maximum temperature, hourly minimum temperature, humidity (%), hourly temperature, hourly total rainfall ( $\text{kg/m}^2$ ), wind speed (m/s) and pollution data. Using these data, how the RF and LSTM models model temporal relationships and their prediction performance were analyzed. These two methods were evaluated and compared using a comprehensive dataset containing approximately 120,000 data observations. This comparative analysis aims to provide insights into the relative strengths and weaknesses of RF and LSTM in the context of solar energy forecasting, contributing to optimizing forecasting models for renewable energy production. In addition, when looking at the literature, the scarcity of studies in this scope and by creating a large data set through a implemented SPP gives this study a special importance. In addition, RF and ANN methods have been used to estimate different variables of solar energy, as will be discussed in the literature review. In addition, the use of the LSTM and prediction model in SPP energy production is also important in terms of the contribution of prediction methods to the literature.

## II. MATERIAL AND METHODS

In this chapter; In the study, the SPP from which the data was collected is explained. The features of SPP and field information are shown in detail. Information on the efficiency losses and energy production that will occur are given here. Additionally, data collection methods from SPP and types of data collected are explained. The characteristics of the data received, their numbers and the analysis made for their effect on energy production are explained. The data set created for SPP energy production estimation and the methods of



**FIGURE 1.** Solar power plant image.

estimation with RF and LSTM are explained. Details of these methods and the reasons for their use are also given.

### A. SPP INFORMATION AND LOSS OF PRODUCTIVITY

The aim is to determine the highest efficiency operating point in solar power plants. For this reason, the main power plant equipment that needs to be examined is the PV panels used in the power plant. In this study, an analysis was made with the data obtained from the solar power plant installed on the land. The power plant was built on  $3525 \text{ m}^2$  of land. 576 PV panels and aluminum construction were used. There are also three inverters with 60 kW output power. The power plant consists of 16 stands, and each stand has a total of 36 PV panels, consisting of 4 rows placed horizontally and vertically. Total DC installed power is 190.08 kW. Figure 1 shows the visual of the power plant.

The ground of the field was covered with white gravel in the first application phase. The reason for this is to prevent mud and dust from accumulating on the panels due to rain or wind. However, it should be noted that at this stage, the arrangement and cleaning of the power plant is not done. In this case, it makes it very suitable for the analysis of power plant data and pollution within the scope of this study. SPP is connected to the electrical grid with a 250 kVA transformer at a voltage level of 34.5.

Information about the panels and inverter used in the power plant is shown in Table 1.

The solar panel data given in Table 1 was obtained as a result of analyzes performed under standard test conditions (STC). These conditions were established to compare the performance of solar panels and provide a standard reference point. STC;

- Radiation Intensity:  $1000 \text{ W/m}^2$ ,
- Cell Temperature:  $25^\circ\text{C}$ , G
- Solar Spectrum: AM 1.5G (solar spectrum under 1 air mass index)

**TABLE 1.** PV panel characteristic data.

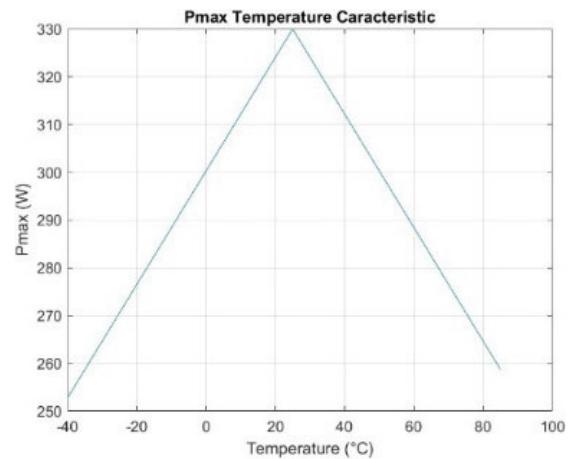
Max. Power (Pmax)	330 Wp
Module Efficiency	19,78
Maximum Power Voltage (Vmp)	33,38 V
Maximum Power Current (Imp)	9,98 A
Open Circuit Voltage (V)	40,78 V $\pm 3\%$
Open Circuit Current (A)	10,19 A $\pm 5\%$
Fuse Rating	20 A
Maximum System Voltage	DC1000V
Cell Technology	Mono-Si
Operating Temperature Range	-40~+85 °C
Temperature Coefficient (Isc )	0.042%/°C
Temperature Coefficient (Vsc)	-0.284%/°C
Temperature Coefficient (Pmax)	-0.38%/°C
Panel Size(mm)	1665x1002x35
Max. Input Voltage	1100 V (DC)
Max. Input Current	22A/22A/22A/22A/22A/22A (DC)
Isc	30A/30A/30A/30A/30A/30A
MPP Range DC	200-1000V
Out Nominal Voltage (AC)	380/400 V AC;3(N)P+E
AC Nominal Operating Frequency	50/60 Hz
Output Rated Power	60 kW (AC)
Output Max Apparent Power	66 kVA (AC)
Output Max. Current	100A;380 V AC/ 95,3A;400V AC/ 79,4A;380 V AC
Power Factor	0,8(lagging)-0,8(leading)
Operating Temperature Range	-25 - +60 °C
Pollution Degree	III
Inverter Topology	Non-Isolation
Communication	MBUS/RS485

It is stated as. In light of this data, it shows that the Pmax value of the solar panel will change with the temperature. This situation will also affect the efficiency of the power plant. According to the data in Table 1, the temperature and power change graph of the panel was created in MATLAB and shown in Figure 2.

### B. DATA TYPES AND COLLECTION METHODS

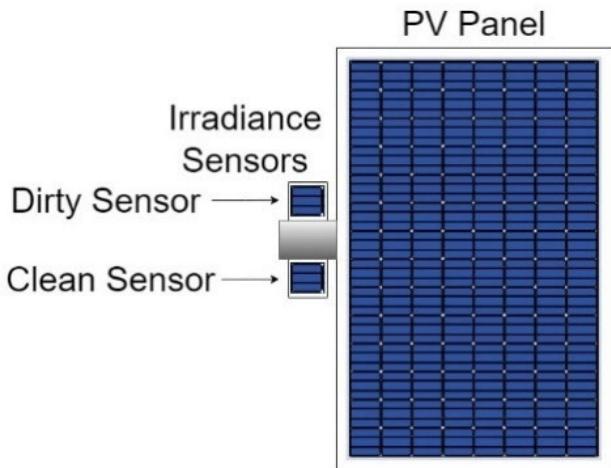
There are many factors that can affect the efficiency of energy production in SPPs. In order to determine these effects, in this study, various data were taken from the power plant whose land type was given above. First of all, data on the electrical energy produced by the power plant was collected. Electricity production data has been recorded hourly, daily and monthly since the day the power plant was produced. In addition, meteorological data were collected with a measurement station. The meteorological station where measurements were made is shown in Figure 3.

Data on sunshine duration, humidity, temperature, precipitation meter, wind speed sensor, wind direction sensor, accumulated irradiation, maximum temperature and minimum temperature are received from the meteorological measurement station shown in Figure 3. These data can be recorded at different time intervals. In this study, all these data were monitored instantly and the data used in the study were determined as hourly data. Accordingly, hourly sunshine duration, accumulated irradiation ( $\text{Wh}/\text{m}^2$ ), hourly maximum

**FIGURE 2.** Solar panel temperature-power graph.**FIGURE 3.** Meteorological measuring station.

temperature, hourly minimum temperature, humidity (%), hourly temperature, hourly total rainfall ( $\text{kg}/\text{m}^2$ ) and wind speed ( $\text{m}/\text{s}$ ) data were used. From this entire data set, data was collected hourly for 416 days and a total of 79872 measurement data sets were created as impact factors. As explained in the section above, these actual measured values have a high impact on the efficiency of the SPP and it is of great importance to take them into account when calculating or predicting the efficiency and change of the power plant. In addition, the size of the data pool consisting of real data collected for a power plant that is operating and producing energy increases the importance of the study. In addition, it shows that it will make a great contribution to SPP operators and investors.

Apart from the comprehensive data received from the measurement station for the analysis of the effects on SPP energy production, pollution that causes a decrease in energy production is also very important. These values should also be taken into consideration when examining the environmental impacts of the power plant. For this purpose, the pollution of



**FIGURE 4.** Dust measuring station.

the PV panels in the power plant was monitored and measurements were taken and recorded. Figure 4 simply shows the system installed in the field for dust measurement.

As seen in Figure 4, the pollution on the PV panels was measured through a measurement system located next to the strings in the power plant. In the system here, analysis was made by comparing two different measurements on two sensors. One of the sensors is constantly cleaned and remains clean. The other sensor is left undisturbed and pollution is determined. By measuring and comparing the radiation values with both the dirty sensor and the clean sensor, the contamination rate is determined using Equation (1).

$$SR = \left( 1 - \frac{SRV_{soiled}}{SRV_{clean}} \right) \cdot 100 \quad (1)$$

Looking at Equation (1), SR; soiling rate,  $SRV_{soiled}$ ;  $SRV_{clean}$  refers to the soiling rate value measured from the clean sensor. Here, measurements can be taken and recorded at different time intervals. In order to maintain the matrix form while analyzing the data set, measurements taken hourly were used. Referring to the IEC 61724-1 standard, which defines requirements for performance monitoring, measurement and analysis of photovoltaic (PV) systems, a daily rate for the contamination rate can be recommended. Because, as in the system here, the measurement made on radiation may mislead as a result of radiation fluctuation due to any reason. This will not be a problem since all parameters will be compared together here. Additionally, hourly measurement records will eliminate the effect of instantaneous balances. With this measurement system, not only the amount of dust but also the cell temperature in dirty cells is measured. Cell temperature also affects the efficiency of PV panels in the power plant. Therefore, it is important to include it in the analyses. The data set mentioned here consists of a total of 416 days and 19968 data in the same measurement intervals as the data set collected from the meteorological measurement station. With these, the system grows even more. Since SPP PV panels are not cleaned, the pollution of the panels increases and



**FIGURE 5.** SPP panel pollution.

the energy to be produced decreases. Figure 5 shows SPP pollution at different time intervals.

The short circuit current ( $I_{sc}$ ) of photovoltaic (PV) panels often decreases due to contamination. Contamination occurs when substances such as dust, dirt, leaves and bird droppings accumulate on the surface of PV panels. These substances reduce the amount of radiation by preventing sunlight from reaching the surface of the panels. As a result, the electrical current produced by the panels also decreases. In the IEC 61724-1 standard, it is shown that monitoring the short circuit current is appropriate among the contamination measurement methods.

Accumulated irradiation (AI) data was obtained with a measurement station installed in the power plant. Reference panel cell was used for measurement. Measurements taken with this cell were recorded via RS485 connection. The measurement made here is actually the same as the operation of solar panels. In other words, energy production begins under radiation from the solar cell. By performing this production, the short circuit current through the cell is measured. This is how AI data was collected.

Temperature measurements of SPP solar panels were made with a thermal camera. Thermal camera measurements checked whether there was any damage to the cells or connections. Because in case of physical damage, a high tendency to resist electrical conduction is expected in damaged areas. When electric current passes through high resistance points, more heat will be generated and thus can be detected with a thermal camera. Measurements were taken at weekly intervals. A few of the measurement images are shown in Figure 6.

In temperature measurements made with a thermal camera, it was observed that there was a temperature difference not only in malfunctions but also due to contamination. Polluted areas receive less sunlight and therefore less energy production. The temperature is higher in these regions than in cleaner areas.

By monitoring the current and voltage values for three inverters in the power plant, the effects of physical damage and pollution were monitored. Input current and voltage data, active power, reactive power and power factor values were recorded for each inverter on a daily basis. A daily chart created with this data is shown in Figure 7.

Figure 7 shows the active power, reactive power, power factor, input current and input voltage values of the power plant inverter, which has eleven inputs. The data in the graph

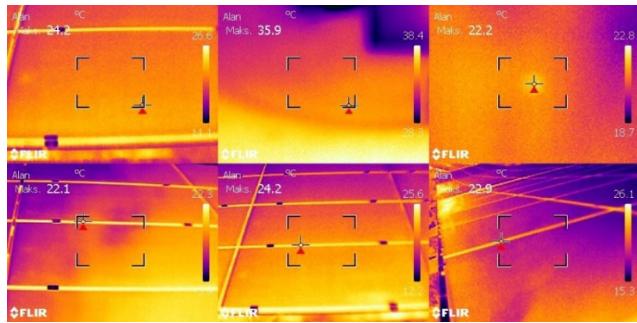


FIGURE 6. Panels thermal measurement.

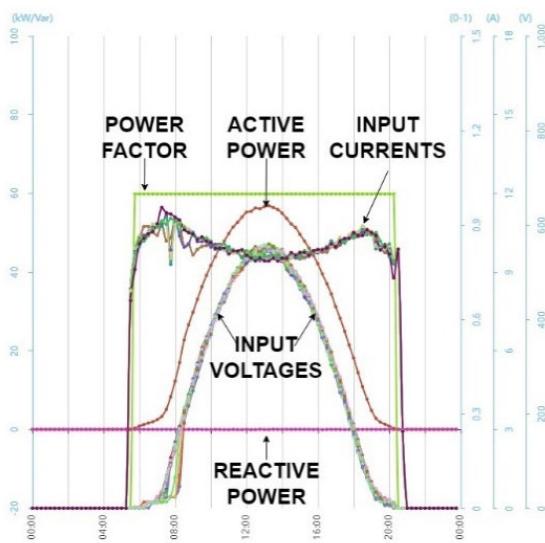


FIGURE 7. Inverter-1 data.

are the values recorded within a full day and are included in the analyzes in the study, confirming the analysis of the effects of pollution on the power plant.

A performance rate (PR) was created based on these data. PR is shown in Equation (2).

$$PR = E_T / AI_{av} \quad (2)$$

In Equation 1, PR reflects the efficiency of energy production to solar radiation per unit area.  $E_T$  shows the total amount of energy produced in a certain period of time and  $AI_{av}$  shows the average AI value in a certain period. Production and irradiation data of the power plant are shown in Figure 8.

Equation (3) is given to analyze the efficiency of the power plant:

$$\eta = \frac{PR_{clean} - PR_{soiled}}{PR_{clean}} \quad (3)$$

$\eta$  in Equation (3); yield value,  $PR_{clean}$ ; When the panels are clean in the power plant, the performance rate value and  $PR_{soiled}$ ; It shows the performance rate value when the panels in the power plant are dirty.

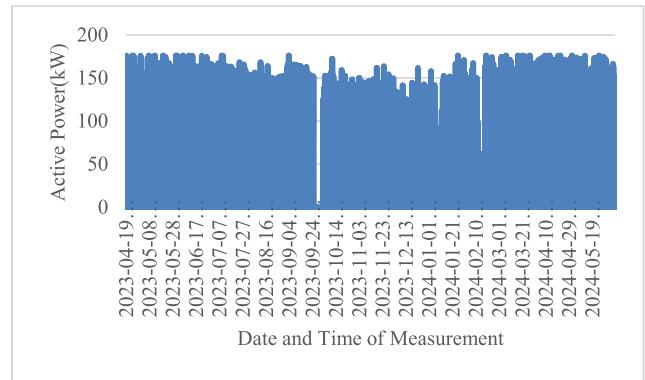


FIGURE 8. SPP energy yield-irradiation change chart according to dates.

### C. FACTORS AFFECTING SPP YIELD LOSS AND RANDOM FOREST REGRESSION ANALYSIS

The efficiency of solar panels that produce energy in SPPs begins to decrease from the moment of first installation. For this situation, degradation curves of PV panel manufacturers are shared. Although these values generally decrease by 1.5% to 2.5% in the first two years of the panel, it is expected to decrease by a total of 10% in the next ten years.

When we look at the literature, it can be seen that the highest effects causing efficiency loss in SPP energy production are caused by contamination. Experimental and theoretical studies have been carried out to determine this decrease. In [45], pollution refers to a situation that occurs as a result of particulate matter collected on the windshield surface of the PV module, preventing the interaction of sunlight on the module. Studies have shown that if panels are not cleaned for 3-6 months, there can be up to a 30% loss of efficiency. Additionally, it has been determined that excessive contamination can lead to a cumulative power loss of more than 1% per day. In another study [26], it is said that the efficiency loss in SPPs exceeds 50% when cleaning intervals are more than 6 months. In order to determine the effects of dust accumulation on the energy production of PV modules, approaches were made in experimental studies according to the IEC 61724-1 standard. The PV modules installed in [46] were divided into two parts and the short circuit currents ( $I_{sc}$ ) of the modules left dirty and the cleaned modules were measured and the pollution rates of the panels were determined. It has been said that the contamination rate is the ratio of the measured  $I_{scs}$ .

Random Forest (RF) Regression Analysis is a machine learning method and was proposed by Breiman [35]. Random Forest regression is an algorithm that aims to make more accurate and reliable predictions by using multiple decision trees. Each decision tree is trained on a random subset of the data set and the results of these trees are combined. This process increases the generalization ability of the model and reduces the problem of overlearning. Due to these advantages, it has been used in prediction processes for different applications. In [47], a short-term load forecast was made by creating a hybrid model using the RF and mean generating

function (MGF) model together. In another study, RF was used for solar power forecasting, again in the field of renewable energy resources [48]. In another similar study [49], RF was used to estimate the amount of solar energy radiation in an area. Sunshine duration was used when making the prediction. Installation of solar power plants on roofs and in the field is quite common. PV potential estimates are very important in the feasibility of these power plants. Many studies have examined the solar energy potential in small and large-scale rooftop solar power plants [50], [51], [52], [53], [54], [55], [56]. By performing solar energy potential analysis for rooftop SPP using RF, high accuracy was achieved for this large-scale rooftop power plant [57].

In the working principle of RF regression, N bootstrap samples are first created from the given training data set. Bootstrap sampling is randomly selected subsets of data (which can be selected recursively) from the data set. This process creates different training datasets for each decision tree. The training dataset and bootstrap sampling are shown in Equation (4) and Equation (5).

$$D = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\} \quad (4)$$

$$D^* = \{(x_1^*, y_1^*), (x_2^*, y_2^*), \dots, (x_N^*, y_N^*)\} \quad (5)$$

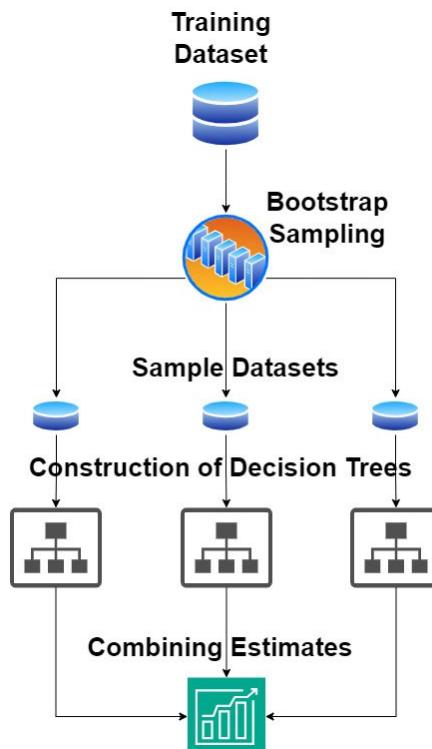
In Equation (4), D represents the training data set. In Equation (5),  $x_i^*$ ,  $y_i^*$  values represent the data points selected from bootstrap sampling. A decision tree is built on each bootstrap instance. In each branch of the decision tree, the best split point is selected among  $m$  randomly selected variables (features). The split point is made on the j-th variable in the feature space X:

$$\operatorname{argmin}_{j,s} \sum_{i:x_i^j \leq s} (y_i - \hat{y}_{left})^2 + \sum_{i:x_i^j > s} (y_i - \hat{y}_{right})^2 \quad (6)$$

In Equation (6),  $\hat{y}_{left}$  and  $\hat{y}_{right}$  are the average target values of the left and right subnodes, respectively. The other value in the equation is; j indicates which feature will be divided, the division point, that is, the value at which the property values will be divided. This process continues splitting the branches (nodes) of the tree until a certain depth or until certain stopping criteria are met. The predictions given by each decision tree are taken based on the test data. In the case of regression, the final prediction is obtained by averaging the predictions from all trees. The predictions made by each decision tree  $T_b$  are denoted as  $\hat{f}_b(x)$ . The final prediction for B decision trees is shown in Equation (7):

$$\hat{f}(x) = \frac{1}{B} \sum_{b=1}^B \hat{f}_b(x) \quad (7)$$

Equation (7) is the simple arithmetic average of the predictions of all trees. The predictions given by each decision tree are taken based on the test data. In the case of regression, the final prediction is obtained by averaging the predictions from all trees. Figure 9 shows the working principle of the RF regression model.



**FIGURE 9.** RF regression working principle diagram.

In this study, RF regression analysis was first used to estimate the electrical energy production of SPP with environmental factors and pollution. In RF regression, the training data set includes hourly sunshine duration, accumulated irradiation ( $\text{Wh/m}^2$ ), hourly maximum temperature, hourly minimum temperature, humidity (%), hourly temperature, hourly total precipitation ( $\text{kg/m}^2$ ), wind speed (m/s). and pollution data. With these data, the electrical energy production of SPP will be estimated using RF regression. All these data consist of 109,824 separate data in total [58]. This means that the data pool is quite large and sufficient.

Regression analysis codes were written in MATLAB. Independent variables are normalized when finding the equation model. Z-score normalization was used for this and is shown in Equation (8).

$$z = \frac{x - \alpha}{\sigma} \quad (8)$$

In this equation, z; z-score normalized value,  $\alpha$ ; the mean of the data set for each of the independent variables, and  $\sigma$ ; It shows the standard deviation of the data set for each of the independent variables.

Hyperparameters have been optimized to improve performance when performing RFR. In RF regression analysis, hyperparameter optimization and training the model with these hyperparameters are important steps to maximize the performance of the model. Hyperparameters are parameters that must be adjusted during the training process of machine learning algorithms and determine the configuration of the

model. These directly affect the model's learning process and performance but are not learned by the model itself; instead, they are determined and optimized externally before the model is trained. The hyperparameters in the RF model are number of trees, minimum leaf size, maximum depth, maximum number of features and bootstrap. The number of trees indicates how many decision trees will be used in the model. It determines the minimum number of observations that should be present in the leaf nodes (last nodes) of each decision tree. A small leaf size creates more detailed and complex trees, which can lead to overfitting. A large leaf size creates more general trees and reduces the risk of overlearning. Maximum depth determines the maximum depth of a tree. As depth increases, the tree becomes more complex. Deep trees may show overlearning, while shallow trees may show underfitting. The maximum number of features determines the maximum number of features to be considered for splitting at each node. Using random subsets of features reduces the correlation between trees and increases the generalization ability of the model. These variables are adjusted at certain intervals during the optimization process and used to find the best values. Bayesian optimization was used to find the best combination within the hyperparameter ranges determined for optimization. This method takes into account the results of past attempts, ensuring that subsequent attempts are more productive. This optimization is done to maximize the performance of the model and minimize overlearning or under learning problems. After the best hyperparameters were determined during the Bayesian optimization process, the Random Forest model was trained using these hyperparameters.

Four different performance evaluation criteria were made for the result obtained: correlation coefficient (R), fractional gross error (FGE), mean standard error (MBE) and root mean square error (RMSE). Equations (9), (10), (11) and (12) were used for these performance evaluation measures.

$$R = \frac{\sum (Y_{\text{observed}} - \bar{Y}_{\text{observed}}) \times (Y_{\text{predicted}} - \bar{Y}_{\text{predicted}})}{\sqrt{\sum (Y_{\text{observed}} - \bar{Y}_{\text{observed}})^2} \times \sqrt{\sum (Y_{\text{predicted}} - \bar{Y}_{\text{predicted}})^2}} \quad (9)$$

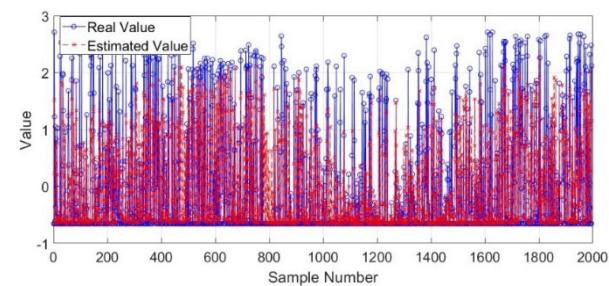
$$FGE = \sum_{i=1}^n \left| \frac{Y_{\text{observed},i} - Y_{\text{predicted},i}}{Y_{\text{observed},i} + Y_{\text{predicted},i}} \right| \quad (10)$$

$$MBE = \frac{\sum_{i=1}^n (Y_{\text{observed},i} - Y_{\text{predicted},i})}{n} \quad (11)$$

$$RMSE = \sqrt{\sum_{i=1}^n (Y_{\text{observed},i} - Y_{\text{predicted},i})^2} \quad (12)$$

In the equations given here,  $Y_{\text{observed}}$ ; observed values,  $Y_{\text{predicted}}$ ; predicted values,  $\bar{Y}_{\text{observed}}$ ; average of measured values,  $\bar{Y}_{\text{predicted}}$ ; the average of the predicted values, n; It represents the number of measured and predicted values.

The data collected for RF regression analysis were transferred to the algorithm and predictions were made. The predicted values and actual values in the prediction results are shown in Figure 10.



**FIGURE 10.** Actual-predicted values as a result of RF regression analysis.

**TABLE 2.** RF regression analysis results performance metrics.

Root Mean Squared Error (RMSE)	0.5865
Mean Bias Error (MBE)	0.0001
Fractional Gross Error (FGE)	0.4356
Correlation Coefficient (R)	0.8111

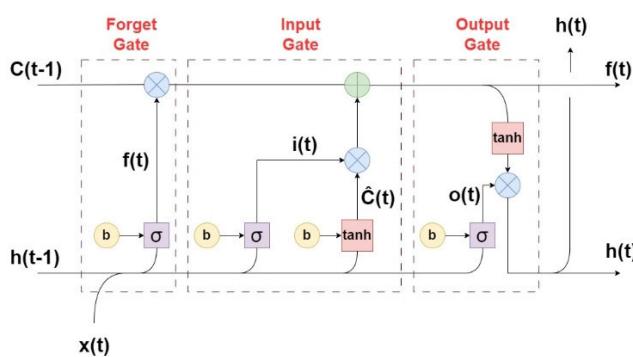
Performance evaluation criteria were also determined for the values obtained as a result of the regression analysis. The results of the performance measures calculated according to Equations (9), (10), (11) and (12) are shown in Table 2.

#### D. FACTORS AFFECTING SPP EFFICIENCY LOSS AND ARTIFICIAL NEURAL NETWORKS

Artificial Neural Networks (ANN) are mathematical models that aim to mimic the information processing capabilities of the human brain. ANNs were developed inspired by the working principles of biological neural networks and are especially effective in learning complex relationships and patterns in large data sets. Its basic structure consists of layers of interconnected artificial neurons, and each neuron produces an output by multiplying the input information it receives with certain weights. This output is usually processed through an activation function and passed to the next layer.

However, classical artificial neural networks face some limitations in processing sequential and dependent data such as time series data. In such data, where information from previous time steps is important for future predictions, traditional ANN cannot perform adequately. In this context, Recurrent Neural Networks (RNN) come into play. RNNs can model dependencies of data over time. However, RNNs can also have difficulty learning long-term dependencies, leading to the “long-term dependency” problem.

Long Short-Term Memory (LSTM) networks have been developed to solve this problem. LSTMs are a type of RNNs and are particularly good at learning long-term dependencies. LSTM cells have a special cell structure that allows information to be stored for longer periods of time without being lost. This cell structure contains components such as the forget gate, entry gate, and exit gate. While the forget gate determines what information is forgotten, the entry gate controls how new information is added to the cell. The output gate



**FIGURE 11.** LSTM architecture.

determines which information will be transferred from the cell state to the new hidden state. Thanks to these gates and cell state, LSTMs effectively learn long-term dependencies in time series data and make predictions.

ANN applications and predictions are popular in engineering studies and are often preferred for different purposes [59], [60]. These studies provide effective results. In a study examining solar radiation prediction using ANN, 24 different articles were selected and examined among 373 articles that obtained results using ANN with different input data, and it was said that the ANN study gave accurate predictions in different climatic conditions [44].

LSTM artificial neural networks also have great potential in solar power plant production forecasting. Solar energy production may vary depending on various weather conditions and environmental factors. Therefore, accurately analyzing historical data and trends is critical to predicting future energy production. LSTMs can be used as an effective tool for production forecasting of solar power plants, thanks to their ability to process such complex and time-dependent data. In this way, energy management and planning can be made more efficient. The architectural structure showing the working principle of LSTM networks is shown in Figure 11 [61].

In the working structure of LSTM networks, time series data is entered at the input gate. The gate time series data in this study are shown in Equation (13).

$$x_t = \begin{bmatrix} x_{t,1}, x_{t,2}, x_{t,3}, x_{t,4}, x_{t,5}, x_{t,6}, x_{t,7}, x_{t,8}, \\ x_{t,9}, x_{t,10}, x_{t,11}, x_{t,12} \end{bmatrix}^T \quad (13)$$

The values of the vector elements containing the input information defined in Equation (13) are as follows:

- $x_{t,1}$ : Month in which the measurement was taken,
- $x_{t,2}$ : Day of the month in which the measurement was taken,
- $x_{t,3}$ : Time of day of the month in which the measurement was taken,
- $x_{t,4}$ : Hourly sunshine duration,
- $x_{t,5}$ : Accumulated irradiation ( $\text{Wh}/\text{m}^2$ ),
- $x_{t,6}$ : Hourly maximum temperature,
- $x_{t,7}$ : Hourly minimum temperature,

- $x_{t,8}$ : Humidity,
- $x_{t,9}$ : Hourly temperature,
- $x_{t,10}$ : Total hourly rainfall ( $\text{kg}/\text{m}^2$ ),
- $x_{t,11}$ : Wind speed (m/s),
- $x_{t,12}$ : Pollution.

The output of LSTM networks and the electrical energy production in the power plant that will correspond to this data. The neural network here is the electrical energy production of SPP as a result of the relevant values at the same time, and the network is trained with this. At the output, it also makes predictions as a result of this training. In the forget gate, the previous state  $C_{t-1}$  and the previous cell state  $h_{t-1}$  are taken. Using the sigmoid activation function, it is decided which information will be forgotten. The forget gate is calculated in Equation (14).

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (14)$$

In this equation:

- $f_t$ : Output of the forget gate,
- $\sigma$ : Sigmoid activation function,
- $W$ : Weight matrix for the forget gate,
- $h_{t-1}$ : Previous hidden state,
- $x_t$ : Current entry,
- $b_f$ : Bias term for forget gate.

By taking the current input ( $x_t$ ) and previous cell state ( $h_{t-1}$ ) at the input gate, Sigmoid and tanh activation functions are used to decide which new information will be added to the cell state. The entrance door is shown in Equation (15).

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (15)$$

According to equation 15:

- $i_t$ : Output of the input gate,
- $W_i$ : Weight matrix for the input gate,
- $b_i$ : Bias term for the input gate.

It then calculates the candidate cell state using the tanh activation function for new information that is a candidate for cell disruption. These calculation equations are found in (16).

$$\hat{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (16)$$

In the Equation (16) above:

- $\hat{C}_t$ : Candidate cell state,
- $W_C$ : Candidate cell state weight matrix,
- $b_C$ : Candidate cell state bias term.

After this stage, the cell state is updated. In other words, the information coming from the Forget gate is multiplied by the previous cell state. The information coming from the input gate is multiplied by the candidate new cell state and the results are summed to obtain the new cell state. This is shown in Equation (17).

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \hat{C}_t \quad (17)$$

**TABLE 3.** LSTM neural networks analysis results performance metrics.

Root Mean Squared Error (RMSE)	0.2183
Mean Bias Error (MBE)	0.0041
Fractional Gross Error (FGE)	0.1360
Correlation Coefficient (R)	0.9759

In this equation:

- $C_t$ : New cell state,
- $C_{t-1}$ : Previous cell state,
- $f_t.C_{t-1}$ : The part of the previous cell state passing through the forget gate,
- $i_t.\hat{C}_t$ : The part of the candidate cell state passing through the input gate.

Then, the current input and previous cell state are received at the output gate. The output port determines what information is transferred from the cell state to the hidden state. This process is also calculated using the sigmoid activation function. It is shown in Equation (18).

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (18)$$

In this equation:

- $o_t$ : Output of the exit gate,
- $W_o$ : Weight matrix for the output gate,
- $b_o$ : It is the bias value for the output gate.

Output information is obtained by using sigmoid and tanh activation functions. In the LSTM output layer, the output information obtained from the LSTM cell is used for the final prediction. The new hidden state is calculated using the tanh activation function by combining the new cell state and the output gate. It is shown in Equation (19).

$$h_t = o_t \cdot \tanh(C_t) \quad (19)$$

In the Equation (19) above:

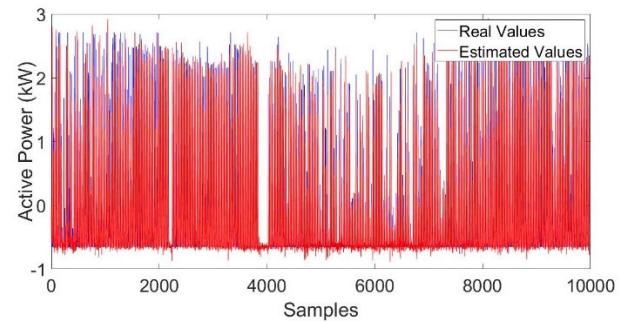
- $h_t$ : New hidden status,
- $o_t$ : Output of the exit gate,
- $C_t$ : New cell status

The new hidden state  $h_t$  is transferred to a fully connected layer or another neural network layer. This layer is used to make the final prediction. In this process, it is usually processed with a weight matrix and activation function. The final estimate is defined in Equation (20).

$$\hat{y}_t = W_{out} \cdot h_t + b_{out} \quad (20)$$

where:

- $\hat{y}_t$ : Final estimate,
- $W_{out}$ : Weight matrix for final prediction,
- $b_{out}$ : Final prediction bias value.

**FIGURE 12.** Actual and predicted values as a result of LSTM networks analysis.

In this process,  $h_t$  is the output of the LSTM cell and is an important intermediate step used in making the final prediction. Using the fully connected layer,  $h_t$ , the final prediction  $\hat{y}_t$  is calculated.

The energy production of the power plant was estimated using LSTM networks with the data set created by long-term measurements made in the SPP field. The LSTM model is trained with input and output data. To increase the performance of the model, 500 epochs were used, the number of neurons was increased, and the learning rate was adjusted with the optimization algorithm. Predictions were made with the trained model and the results were visualized by calculating performance evaluation metrics. These steps were taken to increase the accuracy of the model and make better predictions. The result obtained is shown in Figure 12.

Performance evaluation criteria were also determined for the values obtained as a result of LSTM networks analysis. The results of the performance measures calculated according to Equations (9), (10), (11) and (12) are shown in Table 3.

### III. CONCLUSION

In this study, the efficiency loss of environmental factors and pollution in SPPs was made by creating a linear regression analysis model. Prediction models were created using SPP data, which produces 180 kWe energy in Dumlupinar District of Kutahya province of Turkey. It is aimed to analyze the effects of environmental factors and meteorological variables on energy production and to predict the energy production of the power plant based on these effects. Prediction models were developed using two different machine learning methods, Random Forest Regression (RFR) Model and Long Short-Term Memory (LSTM) artificial neural networks. The data used in the study is based on real data collected from a 180 kWe solar power plant in operation. These data were collected from the day the power plant was put into operation and a total of 119,808 data sets were included in the analysis. The results obtained were evaluated with R, RMSE, MBE and FGE performance criteria. Predictions made using the RF model were used to analyze how the power plant's energy production was affected by environmental impacts and pollution.

The obtained R, RMSE, MBE and FGE results are 0.8111, 0.5865, 0.0001 and 0.4356 respectively. The results of the RF model show that the model generally exhibits good prediction performance. The correlation coefficient is 0.8111, which shows that the model's predictions are quite compatible with the real values. A low RMSE value indicates that the prediction errors are generally small, while a MBE value very close to zero indicates that the model does not have a systematic error. The fact that the FGE value is 0.4356 shows that the model has an acceptable error rate in its predictions.

LSTM networks were used as the secondary method. Predictions made using the LSTM network aimed to obtain higher accuracy results by learning more complex and non-linear dependencies. The obtained R, RMSE, MBE and FGE results are 0.9759, 0.2183, 0.0041 and 0.1360, respectively. The results of the LSTM model show that the prediction performance of this model is much higher than that of the RF model. The correlation coefficient is 0.9759, indicating that the predictions of the LSTM model are in almost perfect agreement with the actual values. A very low RMSE value indicates that the prediction errors of the model are very small, while a low MBE value indicates that the model does not have a systematic error. The fact that the FGE value is 0.1360 shows that the model has a very low error rate in its predictions.

When the results of the two different prediction models used in the study are compared, it is seen that LSTM artificial neural networks perform much better than the RF model. This is due to the fact that LSTM networks have a better capacity to learn time series data and complex dependencies. The ability of the LSTM model to make high-accurate predictions allows solar power plants to predict future energy production amounts more reliably. This is important to ensure continuity in energy production and to minimize the impact of adverse environmental conditions.

## ACRONYMS

Acronyms used in this paper are tabulated here.

AI	Accumulated Irradiation
ANN	Artificial Neural Networks
FGE	Fractional Gross Error
LSTM	Long Short-Term Memory
MBE	Mean Standard Error
MGF	Mean Generating Function
PR	Performance Rate
PV	Photovoltaic
R	Correlation Coefficient
RF	Random Forest
RFR	Random Forest Regression
RMSE	Root Mean Square Error
RNN	Recurrent Neural Network
SPP	Solar Power Plant
SR	Soiling Rate
STC	Standard Test Conditions

## REFERENCES

- [1] T. Lehtola and A. Zahedi, "Solar energy and wind power supply supported by storage technology: A review," *Sustain. Energy Technol. Assessments*, vol. 35, pp. 25–31, Oct. 2019.
- [2] M. Rabaia, M. Abdelkareem, E. Sayed, K. Elsaied, K. J. Chae, T. Wilberforce, and A. Olabi, "Environmental impact of solar energy systems: A review," *Sci. Total Environ.*, vol. 754, Feb. 2021, Art. no. 141989.
- [3] A. Shahsavari and M. Akbari, "Potential of solar energy in developing countries for reducing energy-related emissions," *Renew. Sustain. Energy Rev.*, vol. 90, pp. 275–291, Jul. 2018.
- [4] M. L. Adekanbi, "Optimization and digitization of wind farms using Internet of Things : A review," *Int. J. Energy Res.*, vol. 45, no. 11, pp. 15832–15838, Sep. 2021.
- [5] A. Sharif, M. S. Meo, M. A. F. Chowdhury, and K. Sohag, "Role of solar energy in reducing ecological footprints: An empirical analysis," *J. Cleaner Prod.*, vol. 292, Apr. 2021, Art. no. 126028.
- [6] D. Gu, K. Andreev, and M. Dupre, "Major trends in population growth around the world," *China CDC Weekly*, vol. 3, no. 28, pp. 604–613, 2021.
- [7] R. Sadigov, "Rapid growth of the world population and its socioeconomic results," *Sci. World J.*, vol. 2022, pp. 1–8, Mar. 2022.
- [8] T.-Y. Kim and S.-B. Cho, "Predicting residential energy consumption using CNN-LSTM neural networks," *Energy*, vol. 182, pp. 72–81, Sep. 2019.
- [9] M. Zoghi, A. H. Ehsani, M. Sadat, M. J. Amiri, and S. Karimi, "Optimization solar site selection by fuzzy logic model and weighted linear combination method in arid and semi-arid region: A case study Isfahan-IRAN," *Renew. Sustain. Energy Rev.*, vol. 68, pp. 986–996, Feb. 2017.
- [10] C. Harsito, T. Triyono, and E. Roviyanto, "Analysis of heat potential in solar panels for thermoelectric generators using ANSYS software," *Civil Eng. J.*, vol. 8, no. 7, pp. 1328–1338, Jul. 2022.
- [11] W. Lamb, T. Wiedmann, J. Pongratz, R. Andrew, M. Crippa, J. Olivier, and J. House, "A review of trends and drivers of greenhouse gas emissions by sector from 1990 to 2018," *Environ. Res. Lett.*, vol. 16, Jun. 2021, Art. no. 073005.
- [12] S. Eshiemogie, J. Ighalo, M. Adekanbi, T. Banji, S. Eshiemogie, R. Okoh, and K. Dulta, "Current effect and projected implications of climate change on Nigeria's sustainable development plan," in *Climate Change Impacts on Nigeria*. Cham, Switzerland: Springer, 2023.
- [13] A. Ghosh, "Soiling losses: A barrier for India's energy security dependency from photovoltaic power," *Challenges*, vol. 11, no. 1, p. 9, May 2020.
- [14] M. B. Hayat, D. Ali, K. C. Monyake, L. Alagha, and N. Ahmed, "Solar energy—A look into power generation, challenges, and a solar-powered future," *Int. J. Energy Res.*, vol. 43, no. 3, pp. 1049–1067, Mar. 2019.
- [15] E. Kabir, P. Kumar, S. Kumar, A. A. Adelodun, and K.-H. Kim, "Solar energy: Potential and future prospects," *Renew. Sustain. Energy Rev.*, vol. 82, pp. 894–900, Feb. 2018.
- [16] A. Laabid, A. Saad, and M. Mazouz, "Integration of renewable energies in mobile employment promotion units for rural populations," *Civil Eng. J.*, vol. 8, no. 7, pp. 1406–1434, Jul. 2022.
- [17] S. Dixit, "Solar technologies and their implementations: A review," *Mater. Today. Proc.*, vol. 28, no. 4, pp. 2137–2148, 2020.
- [18] G. Panayiotou, S. Kalogirou, and S. Tassou, "Design and simulation of a PV and a PV-wind standalone energy system to power a household application," *Renew. Energy*, vol. 37, no. 1, pp. 355–363, Jan. 2012.
- [19] D. Jordan, J. Wohlgemuth, and S. Kurtz, "Technology and climate trends in PV module degradation: Preprint," in *Proc. 27th Eur. Photovoltaic Solar Energy Conf. Exhib.*, Frankfurt, Germany, Sep. 2012, pp. 24–28.
- [20] S. Toth, M. Hannigan, M. Vance, and M. Deceglie, "Enhanced photovoltaic soiling in an urban environment," in *Proc. IEEE 46th Photovoltaic Specialists Conf. (PVSC)*, Jun. 2019, pp. 2904–2907.
- [21] A. Hussain, A. Batra, and R. Pachauri, "An experimental study on effect of dust on power loss in solar photovoltaic module," *Renewables, Wind, Water, Sol.*, vol. 4, no. 1, p. 9, Dec. 2017.

- [22] M. Jaszczur, J. Teneta, K. Styszko, Q. Hassan, P. Burzyńska, E. Marcinek, and N. Łopian, "The field experiments and model of the natural dust deposition effects on photovoltaic module efficiency," *Environ. Sci. Pollut. Res.*, vol. 26, no. 9, pp. 8402–8417, Mar. 2019.
- [23] C. Fountoukis, B. Figgis, L. Ackermann, and M. A. Ayoub, "Effects of atmospheric dust deposition on solar PV energy production in a desert environment," *Sol. Energy*, vol. 164, pp. 94–100, Apr. 2018.
- [24] P. Pandian, S. Saravanan, T. Chinnadurai, T. Ramji, N. Prabaharan, and S. Umashankar, "Mitigation techniques for removal of dust on solar photovoltaic system," in *Electrical and Electronic Devices, Circuits and Materials*. Hoboken, NJ, USA: Wiley, 2023.
- [25] A. Sameera, M. Tariq, and M. Rihan, "Investigating the impact of atmospheric factors on solar PV panel performance," in *Proc. Int. Conf. Power, Instrum., Energy Control (PIECON)*, Feb. 2023, pp. 1–6.
- [26] M. J. Adinoyi and S. A. M. Said, "Effect of dust accumulation on the power outputs of solar photovoltaic modules," *Renew. Energy*, vol. 60, pp. 633–636, Dec. 2013.
- [27] A. K. Yadav and S. S. Chandel, "Tilt angle optimization to maximize incident solar radiation: A review," *Renew. Sustain. Energy Rev.*, vol. 23, pp. 503–513, Jul. 2013.
- [28] Z. Darwish, H. Kazem, K. Sopian, M. Alghoul, and M. Chaichan, "Impact of some environmental variables with dust on solar photovoltaic (PV) performance: Review and research status," *Int. J. Energy Environ.*, vol. 7, no. 4, pp. 152–159, Jan. 2013.
- [29] M. L. Adekanbi, E. S. Alaba, T. J. John, T. D. Tundelao, and T. I. Banji, "Soiling loss in solar systems: A review of its effect on solar energy efficiency and mitigation techniques," *Cleaner Energy Syst.*, vol. 7, Apr. 2024, Art. no. 100094.
- [30] S. Mekhilef, R. Saidur, and M. Kamalisarvestani, "Effect of dust, humidity and air velocity on efficiency of photovoltaic cells," *Renew. Sustain. Energy Rev.*, vol. 16, no. 5, pp. 2920–2925, Jun. 2012.
- [31] T. L. S. D. Souza, R. L. F. A. Lima, and C. D. Lima Júnior, "Dirt on photovoltaic modules and efficient energy generation in the Brazilian semiarid," *Revista Brasileira de Engenharia Agrícola e Ambiental*, vol. 26, no. 5, pp. 321–326, May 2022.
- [32] H. Pedersen, J. Strauss, and J. Selj, "Effect of soiling on photovoltaic modules in Norway," *Energy Procedia*, vol. 92, pp. 585–589, Aug. 2016.
- [33] M. Abderrezek and M. Fathi, "Experimental study of the dust effect on photovoltaic panels' energy yield," *Sol. Energy*, vol. 142, pp. 308–320, Jan. 2017.
- [34] M. Saidan, A. G. Albaali, E. Alasis, and J. K. Kalldellis, "Experimental study on the effect of dust deposition on solar photovoltaic panels in desert environment," *Renew. Energy*, vol. 92, pp. 499–505, Jul. 2016.
- [35] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, pp. 5–32, Oct. 2001.
- [36] D. Liu and K. Sun, "Random forest solar power forecast based on classification optimization," *Energy*, vol. 187, Nov. 2019, Art. no. 115940.
- [37] I. A. Ibrahim and T. Khatab, "A novel hybrid model for hourly global solar radiation prediction using random forests technique and firefly algorithm," *Energy Convers. Manage.*, vol. 138, pp. 413–425, Apr. 2017.
- [38] S. Karasu and A. Altan, "Recognition model for solar radiation time series based on random forest with feature selection approach," in *Proc. 11th Int. Conf. Electr. Electron. Eng. (ELECO)*, Bursa, Turkey, Nov. 2019, pp. 8–11.
- [39] H. Sun, D. Gui, B. Yan, Y. Liu, W. Liao, Y. Zhu, C. Lu, and N. Zhao, "Assessing the potential of random forest method for estimating solar radiation using air pollution index," *Energy Convers. Manage.*, vol. 119, pp. 121–129, Jul. 2016.
- [40] B. Babar, L. T. Luppino, T. Boström, and S. N. Anfinsen, "Random forest regression for improved mapping of solar irradiance at high latitudes," *Sol. Energy*, vol. 198, pp. 81–92, Mar. 2020.
- [41] A. H. Elsheikh, S. W. Sharshir, M. A. Elaziz, A. E. Kabeel, W. Guilan, and Z. Haiou, "Modeling of solar energy systems using artificial neural network: A comprehensive review," *Sol. Energy*, vol. 180, pp. 622–639, Mar. 2019.
- [42] S. A. Kalogirou, E. Mathioulakis, and V. Belessiotis, "Artificial neural networks for the performance prediction of large solar systems," *Renew. Energy*, vol. 63, pp. 90–97, Mar. 2014.
- [43] K. S. Garud, S. Jayaraj, and M. Lee, "A review on modeling of solar photovoltaic systems using artificial neural networks, fuzzy logic, genetic algorithm and hybrid models," *Int. J. Energy Res.*, vol. 45, no. 1, pp. 6–35, Jan. 2021.
- [44] A. Qazi, H. Fayaz, A. Wadi, R. G. Raj, N. A. Rahim, and W. A. Khan, "The artificial neural network for solar radiation prediction and designing solar systems: A systematic literature review," *J. Cleaner Prod.*, vol. 104, pp. 1–12, Oct. 2015.
- [45] H. Jiang, L. Lu, and K. Sun, "Experimental investigation of the impact of airborne dust deposition on the performance of solar photovoltaic (PV) modules," *Atmos. Environ.*, vol. 45, no. 25, pp. 4299–4304, Aug. 2011.
- [46] M. Abraim, M. El Ydrissi, H. Ghennoui, A. Ghennoui, N. Hanrieder, S. Wilbert, O. El Alan, M. Boujoudar, and A. Azouzoute, "PVSMS: A system for quantifying soiling effects and optimizing cleaning schedule in PV solar plants," *Energy Convers. Manage.*, vol. 284, May 2023, Art. no. 116978.
- [47] G.-F. Fan, L.-Z. Zhang, M. Yu, W.-C. Hong, and S.-Q. Dong, "Applications of random forest in multivariable response surface for short-term load forecasting," *Int. J. Electr. Power Energy Syst.*, vol. 139, Jul. 2022, Art. no. 108073.
- [48] M. Abuella and B. Chowdhury, "Random forest ensemble of support vector regression models for solar power forecasting," in *Proc. IEEE Power Energy Soc. Innov. Smart Grid Technol. Conf. (ISGT)*, Washington, DC, USA, Apr. 2017, pp. 1–5.
- [49] C. Villegas-Mier, J. Rodríguez-Resendiz, J. Álvarez-Alvarado, H. Jiménez-Hernández, and Á. Odry, "Optimized random forest for solar radiation prediction using sunshine hours," *Micromachines*, vol. 13, no. 9, p. 1406, Aug. 2022.
- [50] L. Bergamasco and P. Asinari, "Scalable methodology for the photovoltaic solar energy potential assessment based on available roof surface area: Application to Piedmont region (Italy)," *Sol. Energy*, vol. 85, no. 5, pp. 1041–1055, May 2011.
- [51] L. K. Wiginton, H. T. Nguyen, and J. M. Pearce, "Quantifying rooftop solar photovoltaic potential for regional renewable energy policy," *Comput., Environ. Urban Syst.*, vol. 34, no. 4, pp. 345–357, Jul. 2010.
- [52] J. Ordóñez, E. Jadraque, J. Alegre, and G. Martínez, "Analysis of the photovoltaic solar energy capacity of residential rooftops in Andalusia (Spain)," *Renew. Sustain. Energy Rev.*, vol. 14, no. 7, pp. 2122–2130, Sep. 2010.
- [53] P. P. Patankar, M. M. Munshi, R. R. Deshmukh, and M. S. Ballal, "A modified control method for grid connected multiple rooftop solar power plants," *IEEE Trans. Ind. Appl.*, vol. 57, no. 4, pp. 3306–3316, Jul./Aug. 2021.
- [54] L. R. Rodríguez, E. Duminil, J. S. Ramos, and U. Eicker, "Assessment of the photovoltaic potential at urban level based on 3D city models: A case study and new methodological approach," *Sol. Energy*, vol. 146, pp. 264–275, Apr. 2017.
- [55] J. Peng and L. Lu, "Investigation on the development potential of rooftop PV system in Hong Kong and its environmental benefits," *Renew. Sustain. Energy Rev.*, vol. 27, pp. 149–162, Nov. 2013.
- [56] J. Byrne, J. Tamini, L. Kurdegashvili, and K. N. Kim, "A review of the solar city concept and methods to assess rooftop solar electric potential, with an illustrative application to the city of Seoul," *Renew. Sustain. Energy Rev.*, vol. 41, pp. 830–844, Jan. 2015.
- [57] J. Hofierka and J. Kaňuk, "Assessment of photovoltaic potential in urban areas using open-source solar radiation tools," *Renew. Energy*, vol. 34, no. 10, pp. 2206–2214, Oct. 2009.
- [58] D. Assouline, N. Mohajeri, and J.-L. Scartezzini, "Large-scale rooftop solar photovoltaic technical potential estimation using random forests," *Appl. Energy*, vol. 217, pp. 189–211, May 2018.
- [59] (2024). *Turkish State Meteorological Service, Türkiye*. Accessed: Jun. 1, 2024. [Online]. Available: <https://www.mgm.gov.tr/>
- [60] K. Olcay and N. Çetinkaya, "Cost analysis of electric vehicle charging stations and estimation of payback periods with artificial neural networks," in *Proc. 4th Int. Conf. Commun., Inf., Electron. Energy Syst. (CIEES)*, Nov. 2023, pp. 1–9.
- [61] K. Olcay and N. Çetinkaya, "Analysis of the electric vehicle charging stations effects on the electricity network with artificial neural network," *Energies*, vol. 16, no. 3, p. 1282, Jan. 2023.
- [62] Y. Yu, X. Si, C. Hu, and J. Zhang, "A review of recurrent neural networks: LSTM cells and network architectures," *Neural Comput.*, vol. 31, no. 7, pp. 1235–1270, Jul. 2019.



**KADİR OLÇAY** (Member, IEEE) received the B.Sc. degree from the Electrical and Electronics Engineering Department, Selçuk University, Konya, Türkiye, in 2015, the M.Sc. degree from the Department of Electrical and Electronics Engineering, Institute of Science and Technology, Konya Technical University, in 2018, and the Ph.D. degree from the Department of Electrical and Electronics Engineering, Konya Technical University, in March 2024, with a focus on integration of electric vehicles into the smart grid and the analysis of their impacts on the grid using artificial intelligence methods. From 2016 to 2019, he was a Project Engineer and the Manager in a company which operates in the field of electrical power systems, electrical transmission and distribution lines, and renewable energy systems. He founded his own energy company in 2019. Subsequently in 2023, he established a new company named Powerage Energy. He commenced his academic career at Kütahya Dumlupınar University, in 2020, and has been serving in this capacity since. He continues to work on electric vehicle charging systems and solar energy systems. His current research interests include electric vehicles, charging systems, deep learning, electric power systems, renewable energy systems, and optimization algorithms.



**MUSTAFA ARIF ÖZGÜR** received the degree from the Faculty of Engineering, Department of Mechanical Engineering, Mechanical Engineering Program, Kütahya Dumlupınar University, in July, 2000, the master's degree from the Institute of Science and Technology, Department of Mechanical Engineering, Kütahya Dumlupınar University, in July 2002, with a thesis on finding electrical energy production potential with wind data in a selected location in Kütahya, and the Ph.D. degree from the Institute of Science and Technology, Department of Mechanical Engineering, Eskisehir Osmangazi University, in June 2006, with a focus on a statistical analysis of Kütahya wind characteristics and its applicability to electricity production.

He was a Research Assistant with the Faculty of Engineering, Kütahya Dumlupınar University, from 2000 to 2002. He was a Research Assistant with the Faculty of Engineering and Architecture, Eskisehir Osmangazi University, from 2002 to 2006.

• • •



**SAMET GİRAY TUNCA** received degree from the Department of Mechanical Engineering, Kütahya Dumlupınar University, Türkiye, in 2010, and the M.Sc. degree from the Institute of Science and Technology, Kütahya Dumlupınar University, in 2015. He is currently pursuing the Ph.D. degree in fluid mechanics and wing aerodynamics. From 2011 to 2014, he was with Kütahya Magne-site Industry Inc., in the mining sector, initially as a Quality Engineer and later as an Investment Engineer. From 2015 to 2016, he was a Planning and Production Engineer with Wam Eurasia-Wam Group, Kütahya. In 2019, he embarked on his academic career with Kütahya Dumlupınar University. He is currently a Lecturer with the Alternative Energy Sources Technology Program, Dumlupınar Vocational School, where he also holds the administrative position of the Director. His academic research focuses on alternative energy sources, experimental aerodynamic studies, and energy-related fields.