



# Enhancing and Optimising Solar Power Forecasting in Dhar District of India using Machine Learning

Prabhakar Sharma<sup>1,2</sup> · Ritesh Kumar Mishra<sup>2</sup> · Parveen Bhola<sup>1</sup> · Sachin Sharma<sup>3</sup> · Gulshan Sharma<sup>4</sup> · Ramesh C. Bansal<sup>5,6</sup> 

Received: 18 September 2023 / Accepted: 8 February 2024  
© The Author(s), under exclusive licence to Springer Nature Singapore Pte Ltd. 2024

## Abstract

Power and energy systems around the world are expanding and evolving in tandem with technological advancement. In the current scenario, energy is a crucial requirement for the development for any country. Machine learning (ML) is used as a technology to address the requirement for quicker and more accurate analyses that would support the control and operation of modern power systems. In this paper, analysis is performed using Machine Learning and Deep Learning (DL) models to predict power estimation at a photovoltaic (PV) solar site with the capacity of 79.95 kW, installed in Dhar district, Madhya Pradesh (MP), India. The model's accuracy is evaluated using various statistical parameters,  $R^2$  score, Mean Square Error (MAE), Root Means Square Error (RMSE), Mean Square Error (MSE), Mean Absolute Error (MAE), Mean Absolute Percent Error (MAPE) and variance. The proposed method Linear Regression (LR) algorithm shows a maximum  $R^2$  score of 0.99994, a small error metric of MAE 0.0091, and an RMSE of 0.121, which indicate the highest accuracy model as compared to other algorithms. Accurate prediction of solar power without irradiance, season-wise (five seasons in India) and month-wise, is also predicted with high accuracy using ten different models of machine learning and one deep learning method, with comparison of its results with the existing work.

**Keywords** Artificial Intelligence (AI) · Machine learning (ML) · Deep Learning (DL) · Photovoltaic (PV) · Root Mean Square Error (RMSE) · Mean Absolute Error (MAE)

## Abbreviations

AI	Artificial Intelligence	DAC	Data Acquisition Card
ANN	Artificial Neural Networks	DT	Decision Tree
ARIMA	Autoregressive Integrated Moving Average	DL	Deep Learning
ARMA	Autoregressive Moving Average	DNN	Deep Neural Network
CNN	Convolutional Neural Network	DNI	Direct Normal Irradiation
		DA	Dragonfly Algorithm
		EN	Elastic Net
		FFNN	Feed Forward Neural Networks
		GAN	Generative Adversarial Networks
		GBC	Gradient Boosting Classifier
		GNN	Graph Neural Network
		GBSSL	Graph-based Semi-supervised Learning
		IR	Infrared
		IRENA	International Renewable Energy Agency
		KNN	K-Nearest Neighbour
		LASSO	Least Absolute Shrinkage and Selection Operator
		LSR	Least Squares Regression
		LR	Linear Regression
		LSTM	Long Short-Term Memory
		ML	Machine Learning

✉ Ramesh C. Bansal  
rcbansal@ieee.org

<sup>1</sup> Department of ECE, ITSEC- Greater Noida, Greater Noida, India

<sup>2</sup> Department of ECE, National Institute of Technology-Patna, Patna, India

<sup>3</sup> Department of Electrical Engineering, Graphic Era Deemed to Be University, Dehradun, India

<sup>4</sup> Department of Electrical Engineering Technology, University of Johannesburg, Johannesburg 2006, South Africa

<sup>5</sup> Department of Electrical Engineering, University of Sharjah, Sharjah, United Arab Emirates

<sup>6</sup> Department of Electric, Electronic and Computer Engineering, University of Pretoria, Pretoria, South Africa

MPPT	Maximum Power Point Tracking
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MSE	Mean Squared Error
MNRE	Ministry of New and Renewable Energy
NWP	Numeric Weather Prediction
PM	Particulate Matters
PML	Perceptron type with Multiple Layers
PV	Photovoltaic
QRA	Quantile Regression Averaging
RBF	Radial Base Functions
RF	Random Forest
ReLU	Rectified Linear Unit
RMSE	Root Mean Square Error
SVR	Support Vector Regression
TWh	Terawatt hours
TSF	Time Series Forecasting
VST	Very Short-Term
WTP	Wavelet Transform Package

### Introduction

Due to limited conventional resources and harmful impact on the environment, renewable energy sources like solar energy are the most attractive alternative replacement due to abundant availability in most places in the world. Solar energy on the earth is sufficient to fulfil all the energy requirements of human beings. Total 173,000 terawatt hours (TWh) amount of energy strikes the Earth in one hour and the world electricity consumption by any means (including industrial, household, vehicle, cooling, heating purpose etc.) in 2022 is 25,500 TWh [1]. Due to the unpredictability of Photovoltaic (PV) cell solar

energy, it is still unattractive to some consumers. To meet the demand for quicker and more accurate assessments that would support the operation and control of contemporary power systems, predictive analysis of a power system is required.

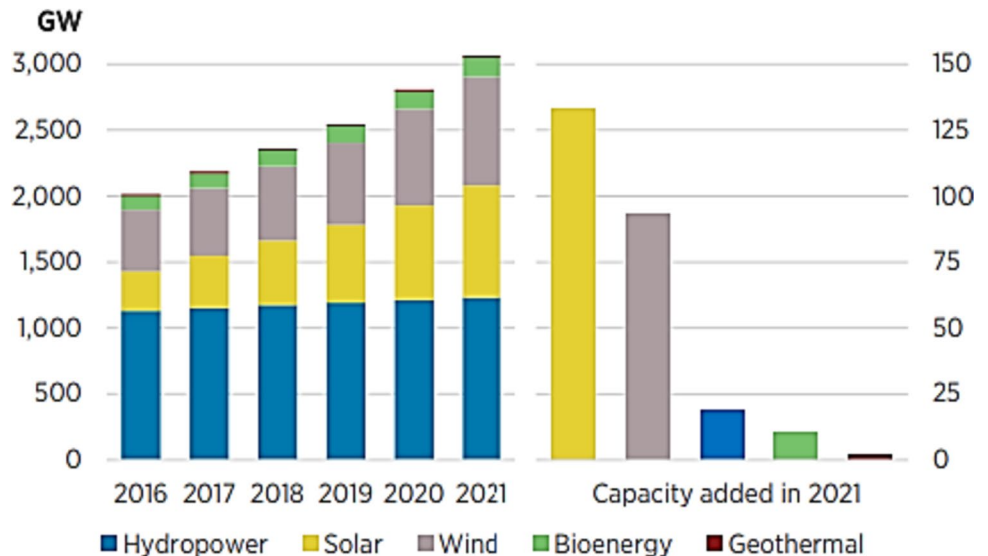
End-of-2021 renewable power capacity was 3064 GW, in accordance with the International Renewable Energy Agency (IRENA) Global, as shown in Fig. 1. With a 1230 GW capacity, hydropower made up the greatest portion of the overall global output, but in the last few years, solar dominated the major portion of the expansion. The remaining energy was split equally between solar and wind power, with capabilities of 849 GW and 825 GW, respectively [2]. In India, according to the Ministry of New and Renewable Energy (MNRE), solar power installed capacity has reached around 61.97 GW as of November 2022. Additionally, 23.14 GW of capacity is in various phases of bidding, while 60.66 GW of capacity is in various stages of implementation [3]. India stands 4th in solar PV deployment across the globe as on the end of 2022. In this work, we have used the meteorological data to estimate the solar power of PV panel technology using machine-learning algorithms. Prior to the machine learning methods, statistical tools are deployed by the researcher for estimation work.

The first linear model presented by Angstrom, presuming a completely clear sky, describes the linear relationship between the worldwide solar irradiation (H) and sunshine hour duration (S). In order to alter it as the Angstrom-PreScott relation, Prescott (1948) substituted the transparent sky given in Eq. 1, for the perfect clear sky assumption [4].

$$\frac{H}{H_{ext}} = a + b \frac{S}{S_{ext}} \tag{1}$$

$H_{ext}$  is the monthly average daily extraterrestrial radiation, while  $S_{ext}$  is the maximum monthly average daily

Fig. 1 Global Renewable Energy Status (Source- International Renewable Energy Agency (IRENA))



sun-shining hour or day length. To anticipate solar radiation, various horizons (the length of the real solar-related meteorological data) have been used as input. Weekly predictions are considered long-term forecasts, and minute-by-minute forecasts of solar power are considered short-term forecasts [5, 6]. Statistical models, including non-linear and linear traditional models like persistence, Autoregressive Integrated Moving Average (ARIMA), Autoregressive Moving Average (ARMA) are employed to capture the stochastic nature of solar energy. Numerous drawbacks exist for different techniques, including their computational cost and inability to modify time-varying time-series systems. The persistence model predicts the future value of the radiation from the present value, but it is inaccurate for more than 1-h estimation and sudden weather changes. ARMA is the combination of autoregressive (AR) and moving average (MA), based on the set of data generated or obtained sequentially, known as time series data, to estimate future values. The main disadvantages of ARMA are its high computational complexity, difficulty in ensuring convergence, and stationary time series. ARIMA was developed to deal with nonstationary time series data. This is the most accurate model, but it is unstable in terms of both fluctuating observations and altered model specifications. ARIMA parameters defined manually, therefore, finding the most accurate fit is a time-consuming process [7].

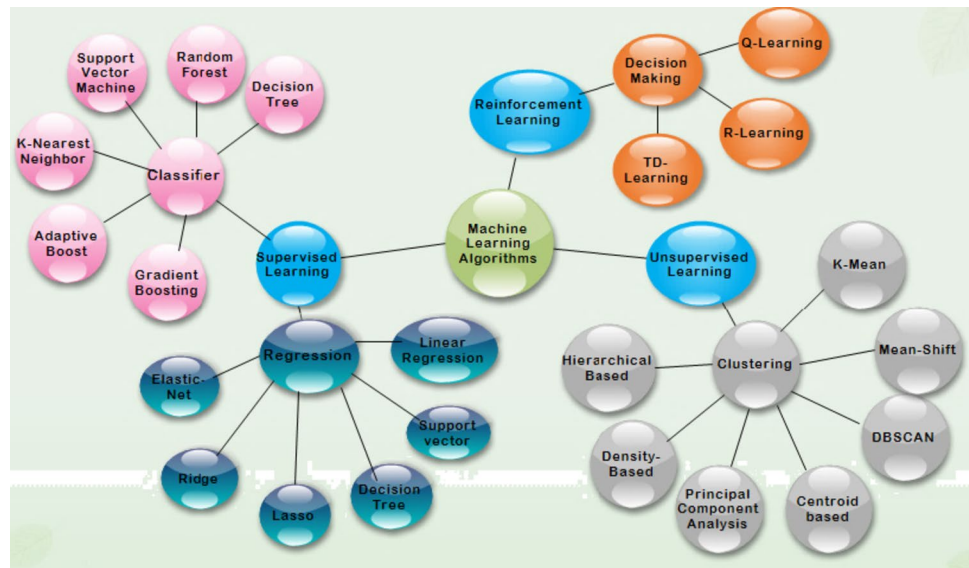
The difficulties and time of computation minimized by adopting Artificial Neural Networks (ANN) techniques. ANN-based techniques have several advantages over traditional ones, which include higher accuracy, simplicity in updating, ease of maintenance, and the ability to handle incomplete inputs [8]. Neural networks are capable of carrying out a wide range of tasks, including nonlinear estimating, grouping, pattern recognition, and optimization. The different ANN techniques for solar energy estimation are utilized [9–11]. Mellit et al. [12] used a multilayer perceptron model to estimate the worldwide sun irradiation 24 h in advance, where daily mean air temperature and daily mean irradiance values are the input parameters. S. K. Aggarwal et al. [13] used a Feedforward Neural Network (FFNN) to predict how much solar energy there is as part of the American Meteorological Society's (AMS 2013–2014) prediction competition. This method works better than Least Squares Regression (LSR) on data from numerical weather forecasts. Silva A. et al. [14] used ANN to predict hourly PV power using the Perceptron type with Multiple Layers (PML) and Radial Base Functions (RBF). They found that the PML method is more accurate than the RBF method. Vakili et al. [15] evaluated the daily surface's global solar radiation at Tehran's using the PLM technique with meteorological data (maximum and minimum daily temperature, relative humidity, and wind speed) including suspended Particulate Matters ( $PM_{10}$  and  $PM_{2.5}$ ) in the atmosphere and found MAPE

1.5% and an absolute fraction of variance of 99%. The main disadvantages of ANN are its large and complete dataset dependency, lack of interpretability, and limited for short-term forecasting.

In an effort to produce an accurate forecast, ML and DL methods have been deployed in recent research work [16–19]. However, it requires large amounts of data for training, validation, and testing purposes. ML algorithms accuracy depends on the amount of training data and the chosen parameters of the model [20]. Researchers suggested various ML approaches to predict solar power, wind power, irradiance, load, and power demand [21–25]. Meng F. et al. [26] suggested a hybrid model that combines the deep Wavelet Transform Package (WTP), Generative Adversarial Networks (GAN), and Dragonfly Algorithm (DA) for solar energy prediction. This approach first break down the irradiance data into its constituent harmonics and then trains a deep GAN-based model. To identify the ideal settings, the generator network adjusted by using an adaptive modified DA technique. The results shown by the proposed method are better as compared to other ML techniques in terms of statistical measures like MAPE and RMSE. Gupta R. et al. [27] used Facebook Prophet and XG Boost to perform Time Series Forecasting (TSF) of solar energy production and determined that the XG Boost model is more effective in terms of precise estimation and more appropriate fitting; the MAPE of XG Boost and Facebook Prophet was 10.9% and 21.8%, respectively. Mutavhatsindi et al. [28] used the convex combination method and Quantile Regression Averaging (QRA) to compare predictions from ML models. They discovered that QRA is better than statistical and traditional ML models. Zhao et al. [29] proposed fault detection and classification of PV modules by using Graph-based Semi-supervised Learning (GBSSL) with the help of only 1% of the total data set and unlabeled data, while Alaraj et al. [30] proposed an ensemble tree approach to ML using meteorological and geographical data from the Kingdom of Saudi Arabia. Other than solar power forecasting, ML and DL techniques, as depicted in Fig. 2, are also used for Maximum Power Point Tracking (MPPT), battery life, load, failure, and tariff prediction [31–34].

Convolutional Neural Network (CNN) is a type of Deep Neural Network (DNN) that is mainly used for computer vision and image identification; however, it can also solve problems with sequential data, like time-series data. There are a few different types of neural networks that Alam et al. [35] suggested for short-term PV power forecasting. These are CNN, CNN- Long Short-Term Memory (LSTM), multi-headed CNN, and traditional methods like ARMA and MLR. Heo J. et al. implemented a multi-channel CNN model on meteorological data and geographical datasets to forecast monthly PV solar power. This method utilizes geographical and meteorological features of PV sites from raster image

**Fig. 2** Some popular Machine Learning Algorithms



datasets. A MAPE of 8.639% was observed for the applied technique, which is better than other conventional methods such as multiple linear regression (16.187%) and ANN (15.991%) [36]. Cheng et al. [37] suggest power prediction based on intra-hour satellite measurement using a spatial temporal Graph Neural Network (GNN) and conclude that it is more accurate than statistical and CNN-based models. A CNN model uses infrared (IR) thermographic images and the solar panel's temperature to assess the malfunctioning of solar PV modules [38, 39].

LSTM (Long Short-Term Memory) networks are another type of DNN that is extensively used by researchers to estimate solar power due to their extraordinary capability to manage sequential data [40]. Obiora C. et al. [41] proposed a ConvLSTM model for irradiance forecasting to mitigate the effects of solar PV power fluctuations in Johannesburg. The statistical measure in terms of nRMSE was reported 1.51% (for a ten-year dataset). A similar study reported in China, where a hybrid model, LSTM-CNN, is used for the estimation of solar power and extracts the temporal-spatio features of PV data [42]. Gao M. et al. [43] suggested an LSTM base model for day-ahead power forecast using Numeric Weather Prediction (NWP) data, while a frequency domain decomposition and LSTM model were proposed by Wang L. et al. [44]. Sarmas E. et al. [45] developed a stacked Long Short-Term Memory (LSTM) model with three Transfer Learning (TL) strategies to forecast accurate solar PV plant production using a limited data set. Bui L. et al. [46] demonstrated a similar technique to predict the output of a large solar PV power plant in Vietnam in conditions of curtailment. Djaafari A. et al. [47] proposed a unique hybrid model where the Balance-Dynamic Sine-Cosine Algorithm is combined with an LSTM predictor for accurate estimation of Direct

Normal Irradiation (DNI) with a relative RMSE value of less than 2.07% and all correlation coefficients greater than 0.99.

The objective of this paper is to estimate solar power using a ML algorithm using meteorological data as input. After going through the literature, we have observed that solar irradiance is one of the most important meteorological parameters; however, it is not easily available at every solar site. To overcome this problem, we considered two approaches. First, consider solar irradiance as one of the input parameters, and the second approach is without considering solar irradiance as the input. In addition to the hourly forecast, we have also estimated the solar power generation for different seasons (summer, winter, monsoon, autumn, and spring). The results are compared with the standard techniques, which were found to be in good agreement in terms of various statistical measures like RMSE, MAPE, MAE, and coefficient of determination  $R^2$ . The paper is organized as follows:

Section "[Data Description and Methodology](#)" describes the data and the solar PV site used for the proposed research work. This section also explains the methodology used in the research work. The tools used for the estimation and the various pre-processing techniques used are elaborated on in this section.

Section "[Results and Discussion](#)": This section presents the results obtained from the various machine-learning algorithms used in the proposed research work. The results compared with the existing similar work done by the research community in the same field.

Section "[Conclusion](#)": This section concludes the research perspective along with the future direction.



## Data Description and Methodology

### Data Description

Data obtained from a solar power plant located in Dhar, Madhya Pradesh, India, for the amorphous silicon technology shown in Fig. 3(a). The total power generation capacity of this plant is 79.95 kW, as shown in Fig. 3(b). Three-year data collected from this site, covering 1096 days from January 1, 2020, to December 31, 2022. We collected 14,248 observations from the above site from 6:00 to 18:00. The above data is available on the cloud portal of AVAADA Energy Company (service provider: Intello Tech AMC Pvt. Ltd., website: <https://portal.inteltechsolutions.co.in>), as shown in Fig. 4. The meteorological parameters recorded at the plant location are year, month, time (hours), temperature (temp), and irradiance. However, the parameters wind speed, surface pressure, and humidity, which are not available at the planned location, are collected from the NASA website (<https://power.larc.nasa.gov>) and modified according to the requirements of the model. The data obtained from the site is in rough form, so before using it for estimation, data pre-processing is required. The major task of pre-processing includes the removal of ambiguous values, removing duplicate values, considering sunshine hour data, dividing the dataset into five seasons, and obtaining the missing values. The other important part is to convert the hourly data set into a daily data set using averaging techniques. The daily data set converted into a monthly data set by a similar approach. Estimation of solar power in various season done by considering appropriate months in that particular season.

- **Software & Equipment:** The software used for the analysis is Python 3.8.8, and the version of the notebook

server is 6.3.0. Processing is done via a Nvidia GeForce RTX 3090 24 GB GPU processor with 128 GB of RAM (Corsair Vengeance RGB PRO DDR4 3200 MHz) and a 3.2 GHz Intel i9 processor (12 cores, 24 threads). Python is the latest open-source tool used by researchers. A different machine learning model algorithm library is available, which is utilised for the estimation of solar power.

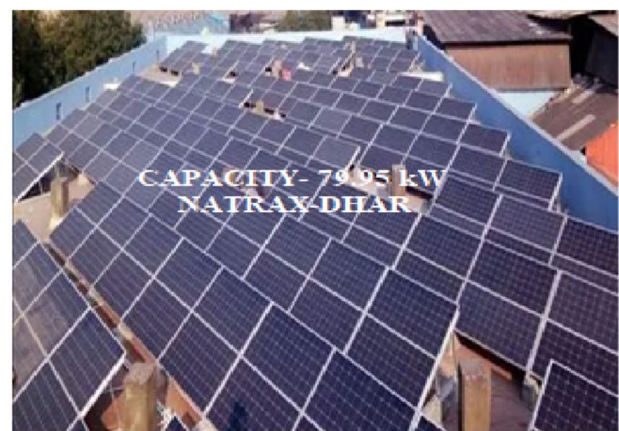
### Methodology

The methodology to estimate solar power uses different machine-learning and deep-learning techniques. Figure 5 explained the complete process flow for solar power estimation. The estimation work is divided into the following steps:

#### Step-1: Data collection and pre-processing

To avoid simulating periods of darkness at night and diminished brightness, the original dataset was segmented to only cover the timeframe of 6:00–18:00. This restriction also reduced computational time in data processing. The data obtained from the site is in rough form, so before using it for estimation, data pre-processing is required. The major tasks of pre-processing include the removal of duplicate values, interpolation for missing values, normalisation of data, conversion of hourly data to day-wise and month-wise data using the averaging technique, and removal of ambiguous values.

- **Removal of duplicate and Interpolation for the missing values:** The meteorological data collected from the plant exhibited some missing and duplicated values. Initially, the duplicate values eliminated, and the missing values were estimated using linear interpolation as part of the pre-processing procedure.



**Fig. 3** a Graphical representation of the Dhar district, Madhya Pradesh- India (b) Solar Site View

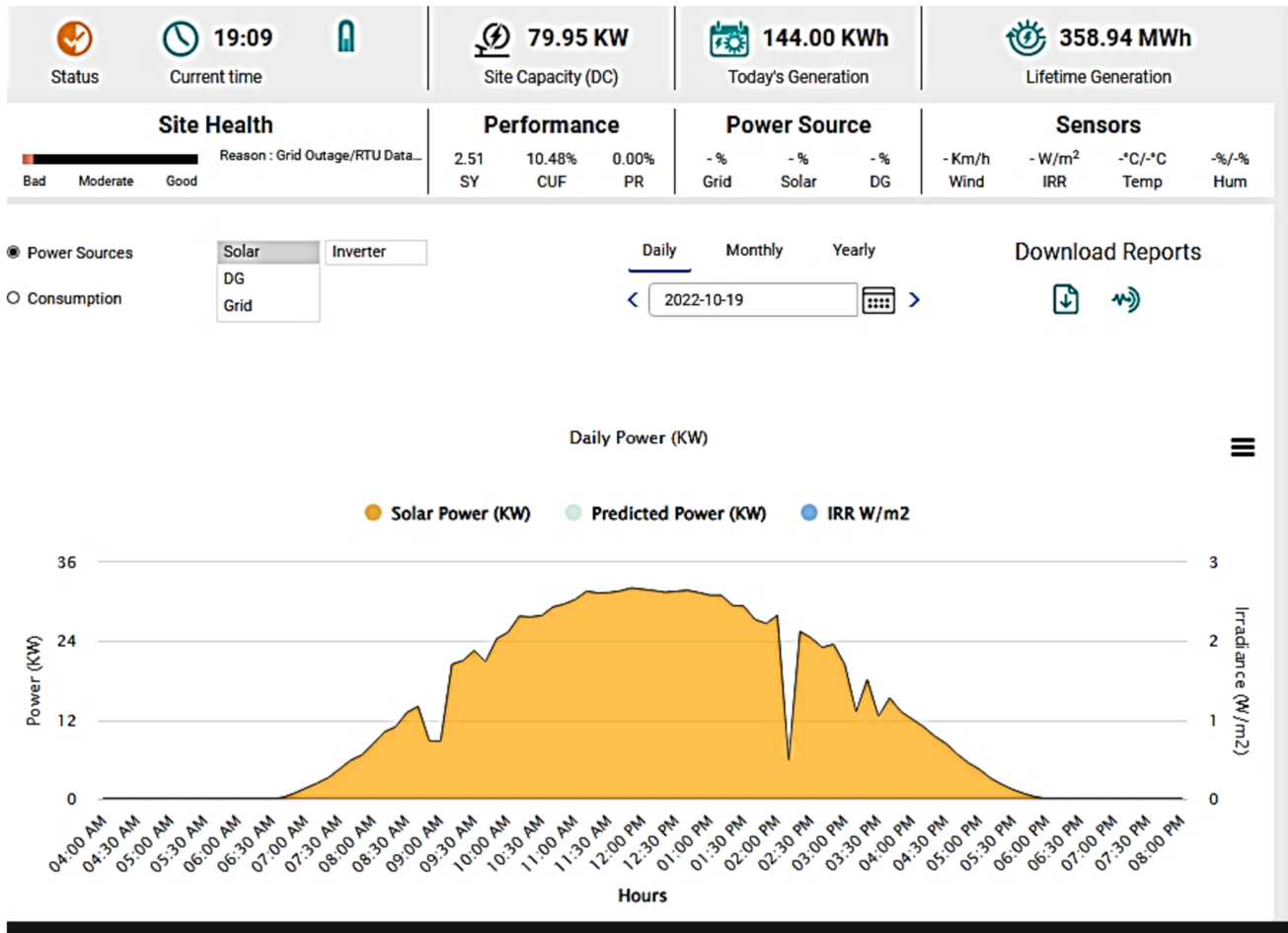


Fig. 4 Online cloud portal view of AVAADA energy company

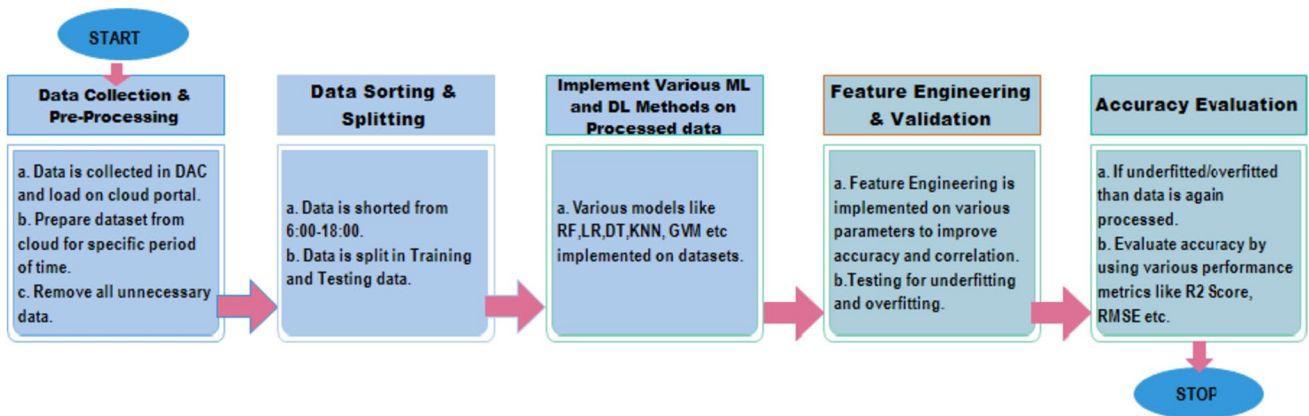


Fig. 5 Flowchart showing the methodology in forecasting solar power using AI techniques

- **Normalization of data:** The input data exhibits significant heterogeneity among the various parameter values, making it challenging to model. The significant range of data, as indicated by the Min (minimum) and Max (maximum) values, is evident from Table 3. In order

to standardise the data within a consistent range of values between 0 and 1, it was subjected to normalisation using the max–min normalisation technique, as defined by the following Eq. (2):

$$X' = \frac{X - X_{Min}}{X_{Max} - X_{Min}} \quad (2)$$

Where  $X'$  represents the normalized value of the given data,  $X$  represents the input meteorological value, and  $X_{Min}$  and  $X_{Max}$  represent the minimum and maximum values of the data. Conversion of data from hour to daily and monthly format: The Data Acquisition Card (DAC) consistently captures meteorological variables and power generation data on an hourly frequency. The proposed study requires data in both daily and monthly formats. In order to accomplish this, the available data processed by utilizing an averaging methodology and subsequently converted into daily, monthly, and eventually seasonal formats.

- Step-2: Data sorting and splitting

Our dataset contains 24 h of data for solar power generation and meteorological parameters. Data sorting is required to remove the unnecessary values of data that are present in our dataset. In the process of data sorting, we have removed the values from 18:00–6:00 for which solar radiation is either very low or not available. This will also reduce computational time in data processing. After sorting the data, the next step is to split the data into different parts, i.e., training, testing, and validation. The dataset contains 14,248 observations, out of which about 80% are used for training purposes. The training data has 11,398 observations, while 10% of the data is used for testing, and the remaining 10% of the data is used for validation purpose, i.e., 1425 observations in each dataset.

- Step-3:- Machine Learning (ML) & Deep-Learning Algorithm Implementation

Modern methods for very short-term (VST-1 min-1 h) and short-term (ST-1 h to 1 day) forecasting include ML algorithms and meta-heuristic optimisation methods inspired by the DNN model. In recent years, ML methods have outperformed conventional empirical methods for solar power forecasting in terms of results [37]. In the literature, there are numerous examples of ML methods used for estimating solar power by researchers [48–51]. So we have implemented Deep-Learning-Feedforward Neural Network (DL-FFNN) and Machine Learning techniques in the present study: Linear Regression (LR), Ridge, Random Forest (RF), Decision Tree (DT), Gradient Boosting Classifier (GBC), Least Absolute Shrinkage and Selection Operator (Lasso), Adaptive Boost Classifier (ADC), Support Vector Regression (SVR), K-Nearest Neighbour (KNN) and Elastic Net (EN). In DL-FFNN method has 10 layers with 512 neurons in each, and the Rectified Linear Unit (ReLU) activation function is used for solar power prediction. In this technique, Adam Optimizer has been used. The algorithm for ML and DL models utilised in

the present study is available in the Python library. They are implemented using Python software on the notebook server 6.3.0. Processing is done via a Nvidia GeForce RTX 3090 24 GB GPU processor with 128 GB of RAM (Corsair Vengeance RGB PRO DDR4 3200 MHz) and a 3.2 GHz Intel i9 processor (12 cores, 24 threads).

- Step-4 Feature Engineering: Estimating power for seasonal variation

Feature engineering involves extracting relevant features from meteorological data to estimate solar power production in various seasons. As one of the objectives of the proposed work is to find out the power variation in the different seasons, feature engineering provides valuable input parameters that are required for seasonal power generation. We use the one-hot encoding approach to encode the categorical variables season. Another feature of engineering includes extracting data only between 06:00 and 18:00 and removing dark hours. The next feature implemented is the extraction of cyclic patterns in hours and months, which used for estimating daily and monthly power generation. In python language pandas, numpy, sklearn, seaborn and matplotlib.pyplot library used in coding.

In order to achieve precise predictions in deep learning, researchers have experimented with numerous layers containing different numbers of neurons, using various optimisation techniques. It has been determined that a DL-FFNN approach with 10 layers, each containing 512 neurons, utilizing the rectified linear unit (ReLU) activation function and the Adam optimizer, yields favourable results.

- Step-5 Accuracy Evaluation

- To evaluate the accuracy of the machine learning models used in this study, we have applied the statistical metrics listed in Table 1. Table 1 includes three error metrics: MAE, MAPE, MSE, and RMSE. In order to achieve an accurate power forecast, it is crucial that the values of the error measure are minimised.

Where,  $N$  is the number of observation and  $i$  is the  $i$ th observation.  $SS_{Regression}$  and  $SS_{Total}$  is the sum squared regression error and sum squared total error respectively, which can be evaluated as

$SS_{Regression} = \sum (y_i - y_{Regression})^2$	Squared difference between each data points values and the regression values
$SS_{Total} = \sum (y_i - \bar{y})^2$	Squared difference between each data points values and the mean values

Table 2 elaborate the importance of coefficient of determination i.e.  $R^2$  value. The value equal to 1 indicates

**Table 1** Validation Metrics with formulae and description

Metric	Formulae	Description
R <sup>2</sup> Score	$R^2 = 1 - \frac{SS_{Regression}}{SS_{Total}}$	The amount of fluctuation in the output parameter that can be predicted using the input parameter (s)0.6{
Mean Absolute Error	$MAE = \frac{1}{N} \sum_{i=1}^N  (y_i - x_i) $	Calculates the discrepancies between the model's predicted values (x <sub>i</sub> ) and the actual values (y <sub>i</sub> )
Mean Absolute Percent Error	$MAPE = \frac{1}{N} \sum_{i=1}^N \left  \left( \frac{y_i - x_i}{y_i} \right) \right  * 100$	The percent discrepancies between the model's predicted values (x <sub>i</sub> ) and the actual values (y <sub>i</sub> )
Mean Square Error	$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2$	Mean of the square of the discrepancy between the estimated and actual values
Root Mean Square Error	$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}$	Square root of the mean of the discrepancy between the estimated and actual values
Variance	$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2}$	Low variance indicates that even the slightest change in the data set will affect the outcomes of the target function

that the model estimation is accurate i.e. no error. If the value is zero or less than zero, model performance is worst. Practically the value of R<sup>2</sup> lies in between 0 and 1.

- Overfitting and Under fitting:** Overfitting occurs when our ML model attempts to include all (or more) of the data points in the dataset. As an outcome, the model starts to cache faulty values and noise from the dataset, reducing its efficiency and accuracy. Cross-validation, training with extra data, deleting features etc. used to reduce overfitting. Under fitting occurs when our ML model is not able to identify the underlying trend in the data. It happens when a model is unable to learn properly from the training dataset, resulting in reduced accuracy and inaccurate predictions.

## Results and Discussion

This part provides a statistical description of numeric variables as well as a correlation analysis between different factors in the dataset. It also explains the performance metrics for various machine-learning approaches. Subsequently, a thorough evaluation conducted by comparing the current study with previous research, utilising diverse performance indicators such as the R<sup>2</sup> score, RMSE, and others.

### Statistical Information

Table 3 describes the range of different input and output parameters used for the present work. Meteorological parameters considered as input are hours, humidity, temperature, irradiance, wind speed, and pressure, while solar power used as output. Table 3 highlights the minimum and maximum values of each parameter, along with the mean value of each

parameter. Prior to being used as input for machine learning models, the data-scaling method performed to prepare for the varying ranges of each parameter. The mean irradiance is approximately 382.97 Wh/m<sup>2</sup>, and the temperature is 27.68 °C, which shows that the site is appropriate for solar power generation.

### Correlation Coefficient

The correlation method used to determine the relationship between different input variables and the output. The outcome obtained from the Pearson correlation coefficient shown in Fig. 6. Correlation coefficients indicate the degree of correlation between two variables. Equation 3 represents the mathematical expression for Pearson's correlation coefficient.

$$\text{Pearson correlation Coefficient} = r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n \sum x^2 - (\sum x)^2][n \sum y^2 - (\sum y)^2]}} \tag{3}$$

Where x is the independent variable, y is the dependent variable, n indicates the sample size, and Σ denotes the sum of all values. The relationship of solar power with ambient temperature and irradiance is strong, with correlation coefficients of 0.36 and 0.59, respectively. Nevertheless, the

**Table 2** Performance relation to R<sup>2</sup> value

Value of R <sup>2</sup>	Performance
Equal to 1	no error in regression
1 > R <sup>2</sup> > 0	Moderate error in regression
R <sup>2</sup> = 0	regression line is not better than the mean
R <sup>2</sup> < 0	regression line is even worse than the mean value



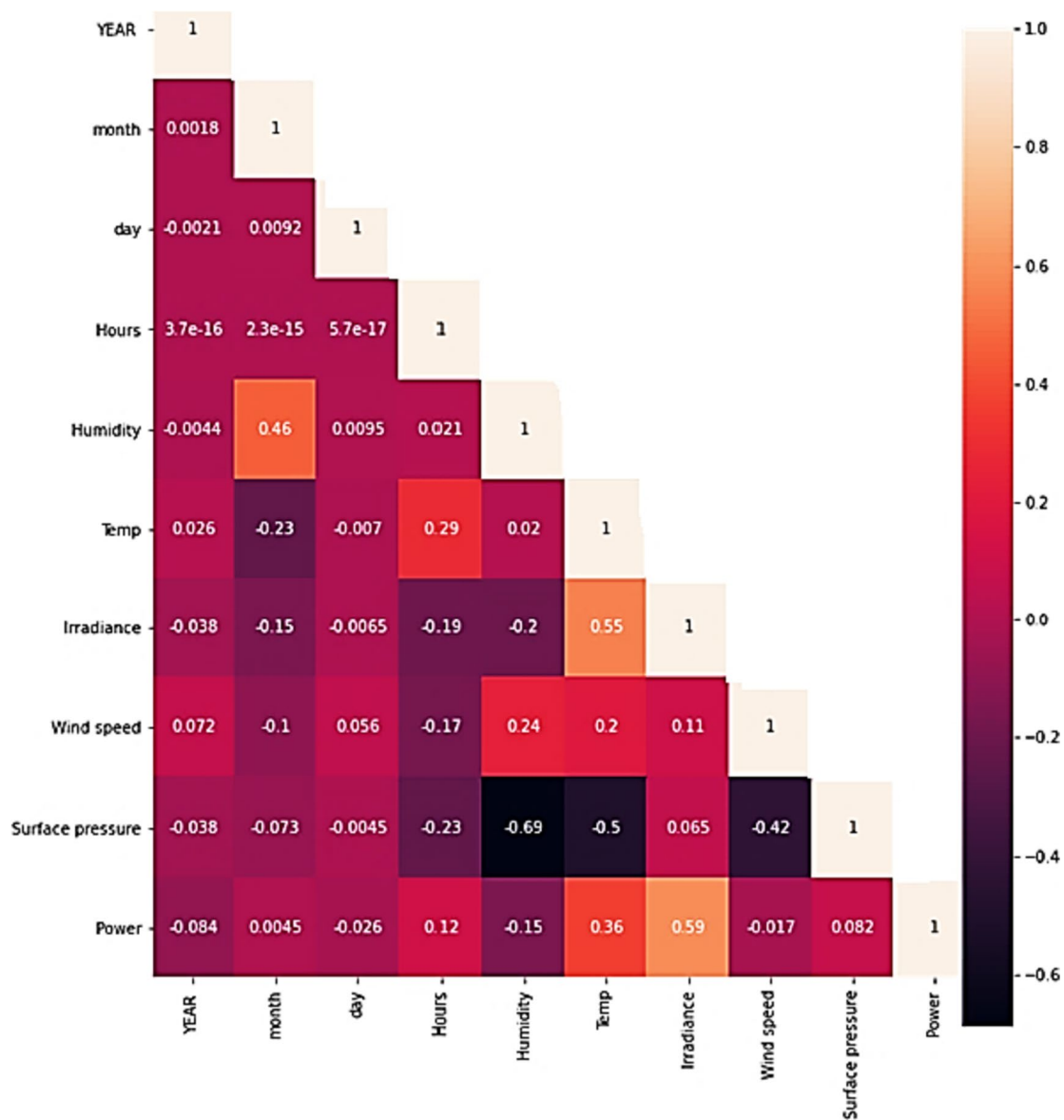
**Table 3** Range for meteorological variable

Unit	Hours	Humidity	Temp	Irradiance	Wind speed	Surface pressure	Power
	Hr	Percentage	Celsius	Wh/m <sup>2</sup>	m/s	kPa	kW
Mean	12	46.25	27.68	382.97	3.82	95.08	15.48
Minimum Value	6	5.12	3.52	0	0.02	93.52	0
Maximum Value	18	89.6	43.44	1015.7	14.76	96.31	69.87

variables of wind speed, humidity, year, and hours exhibit a rather weak correlation with output power.

The next step is to extract features from available input parameters, which are required for seasonal power estimation. Therefore, four new variables, sine\_hr(hour),

sine\_mon(month), season monsoon and season summer, were created as new features that enhanced the performance of machine learning models used for seasonal power estimation. The correlation of this new parameter sine\_hr has a high correlation with power, i.e., 0.83, as shown in Fig. 7.



**Fig. 6** Correlation analysis including encoded features for District Dhar, MP India

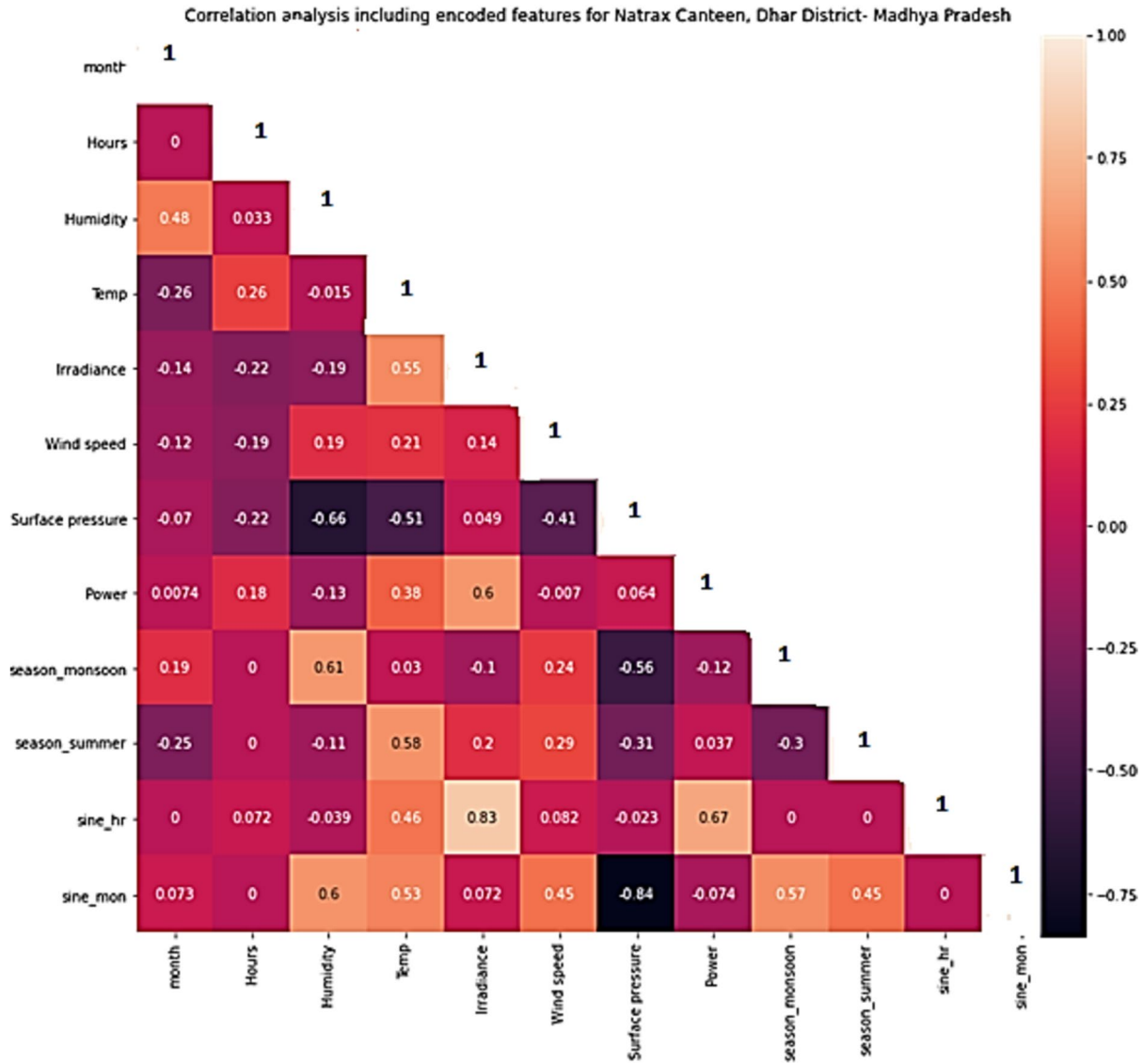


Fig. 7 Correlation analysis for the Dhar District (MP-India) site, including encoded features

Table 4 Performance Parameters of various Machine Learning with and without Irradiance

Rank No	Model	Metrics						
		R <sup>2</sup> Score (without Irradiance)	R <sup>2</sup> Score (with Irradiance)	MAE	MAPE	MSE	RMSE	Variance
1	LR	<b>0.99994</b>	0.99995	<b>0.0091</b>	<b>0.91</b>	<b>0.0146</b>	0.121	<b>0.99996</b>
2	Ridge	<b>0.99991</b>	0.99994	0.0097	0.97	0.0148	0.122	0.99994
3	RF	<b>0.99990</b>	0.99993	0.0146	1.46	0.0256	0.166	0.99993
4	DT	<b>0.99984</b>	0.99990	0.0194	1.94	0.0327	0.180	0.99991
5	GBC	<b>0.99986</b>	0.99987	0.0943	9.43	0.0378	0.194	0.99988
6	DL-FFNN	<b>0.99710</b>	0.99874	0.0462	4.62	0.0029	0.0542	0.99882
7	Lasso	<b>0.99626</b>	0.99626	0.8226	82.26	1.0583	1.028	0.99647
8	ADC	<b>0.99332</b>	0.99467	0.9827	98.27	1.5090	1.228	0.99394
9	SVR	<b>0.98735</b>	0.99884	0.7653	76.53	5.6825	2.383	0.98567
10	KNN	<b>0.93932</b>	0.95964	2.2523	225.23	10.7234	3.274	0.96084
11	EN	<b>0.86992</b>	0.87024	4.6947	469.47	35.3035	5.941	0.88106
AVERAGE (Best 5)		<b>0.99989</b>	<b>0.99992</b>	<b>0.02942</b>	<b>2.942</b>	<b>0.0251</b>	<b>0.155</b>	<b>0.99993</b>
AVERAGE (All)		0.98024	0.98345	0.88280	88.280	4.9459	1.335	0.98331

**Table 5** Comparison of the proposed work with existing work

Authors/ Parameters	R <sup>2</sup> Score	RMSE	MAE	Location	Technique [Reference]
Proposed Work (Best 5)	<b>0.9998</b>	<b>0.155</b>	<b>0.0294</b>	Dhar- India	ML
AVERAGE (All)	0.98345	1.335	0.88280	Dhar- India	ML & DL
Proposed DL-FFNN	<b>0.9987</b>	<b>0.0542</b>	<b>0.0462</b>	Dhar- India	DL-FFNN
Mutavhatsindi et al. (average)	0.9971	0.0320	0.0313	South Africa	ML [28]
Silva et al. (average)	0.9933	0.3795	NA	Fortaleza—CE	ANN [14]
Kim et al. (average)	0.9862	0.1408	0.0213	Texas-USA	DL [52]
Pasion et al. (average)	0.9518	0.1223	0.0363	12 sites-USA	ML [53]
Zhang et al. (average)	0.9400	3.1748	2.2017	Eugene-USA	PLM [54]
Asensio et al. (average)	0.8880	NA	NA	Aragon-Spain	ML [34]
Vakili et al. (average)	0.8334	NA	NA	Tehran	ANN [15]
Karimi et al. (average)	0.8150	0.1107	0.0365	Cleveland, Ohio	CNN [55]
Alaraj et al. (average)	NA	0.8275	0.5242	Saudi Arabia	ML [30]
Queen et al. (average)	NA	0.4264	0.0303	Coimbatore- India	ML [21]

\*NA-Not Available

### Metrics Results for Various ML Models

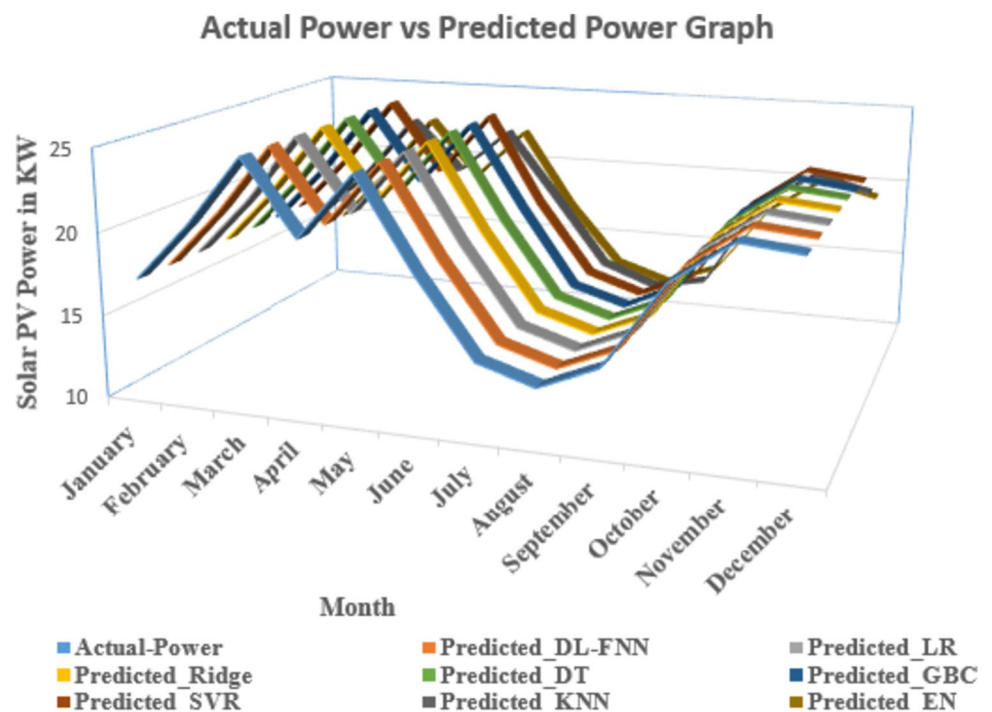
#### Prediction with Solar Irradiance as Input

Table 4 elaborates on the results of the various algorithms deployed for solar power estimation. Among all ML techniques, the LR approach has the highest accuracy. The performance of the ML model evaluated with the R<sup>2</sup> score, MAE, MAPE, RMSE, and variance. The R<sup>2</sup> scores of LR, Ridge, RF, and DT equal to 0.9999, which means that the accuracy of the estimation is approaching the actual

value. However, the EN model had the lowest R<sup>2</sup> score of 0.870249. We have used MAE, MSE, and RMSE metrics, which give the error in estimation. The LR model exhibited the lowest error values for MAE, MSE, and RMSE, i.e., 0.0091, 0.0146, and 0.0513, respectively. Whereas EN models had the highest error values of 4.6947, 35.3035, and 5.8085, respectively.

We have compared the results of the proposed technique with similar work done by researchers in the same field. Table 5 compares the results of the current research approach with similar studies performed at different locations using

**Fig. 8** Month-wise actual and predicted power by various ML models



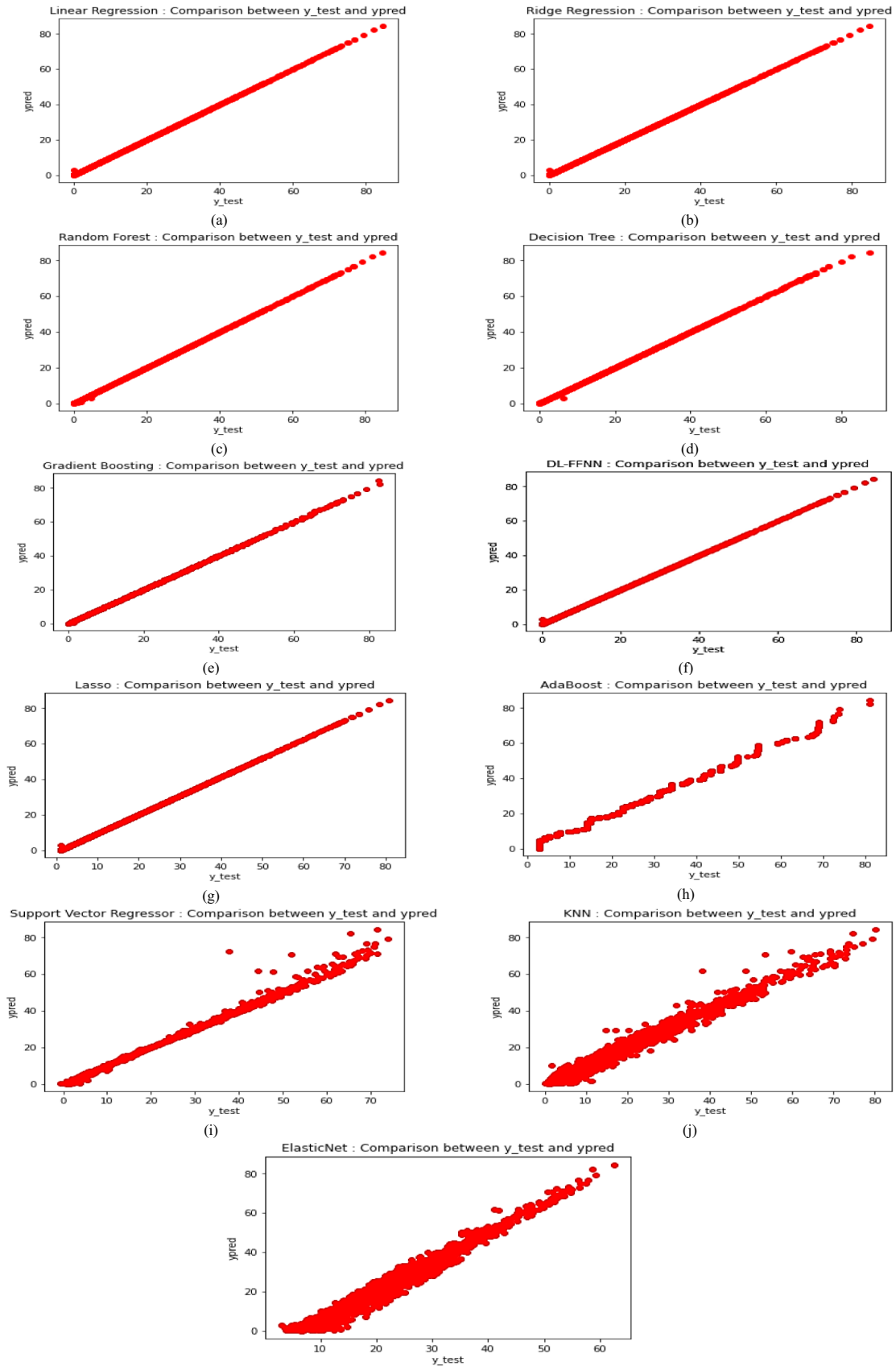


Fig. 9 Comparison Chart of testing data with predicted data



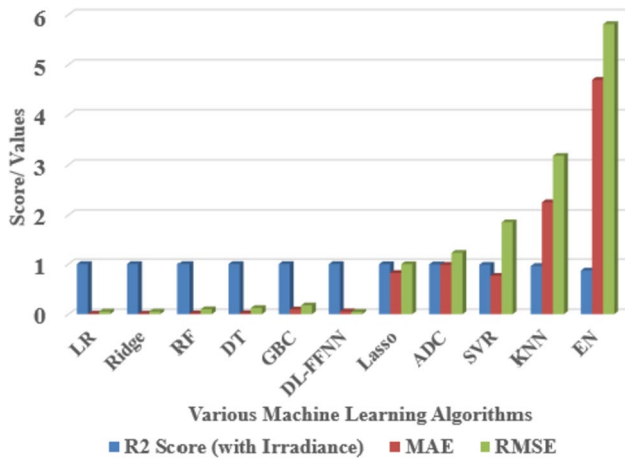


Fig. 10 Relationship of performance metrics parameters with different ML Algorithms

machine-learning models and deep learning approaches. The results are in good agreement with the results achieved by different people.

Figure 8 illustrates the actual and forecasted PV power produced by different ML algorithms used in the present work. The estimations obtained using the DL-FFNN, LR, Ridge, DT, GBC, and SVR methods exhibit a higher degree of accuracy when compared to the predictions made by the KNN and EN models, as shown in Fig. 8.

Figure 9 shows the predicted versus actual power for testing data for all the ML algorithms. The LR, Ridge, RF, DT, GBC, and Lasso algorithms (a–f) make a straight-line graph. The ADC, SVR, KNN, and EN algorithms (g–j), on the other hand, make a graph that spread out between predicted and test values across the whole range. The LR algorithm

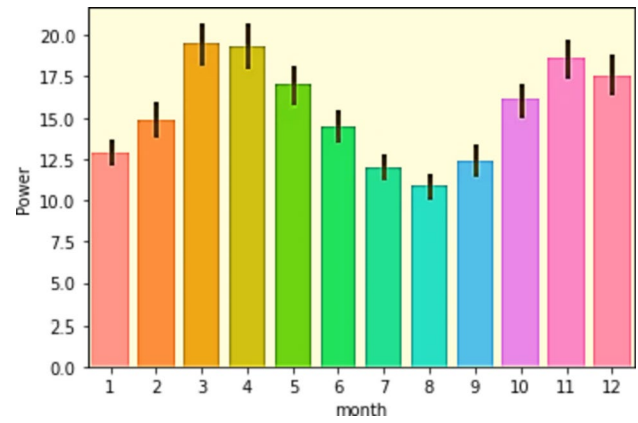


Fig. 12 Month-wise solar power forecasting

graph is extremely linear due to its highest accuracy (highest  $R^2$  score), whereas the EN algorithm graph is the least linear due to its lowest accuracy (lowest  $R^2$  score).

Figure 10 shows the performance metrics parameters for all proposed ML algorithms. Graphs show that the  $R^2$  score and error parameters like RMSE and MAE are inversely proportional to each other. LR and Ridge algorithms show (almost the same) lowest error and highest  $R^2$  score, which means the most accurate model. The arrow indicates that the error metrics increase as we move from the LR to the EN ML approach.

### Prediction without Solar Irradiance

One of the key objectives of this research is to determine solar power without using solar irradiance as an input parameter. The high cost of measuring equipment makes

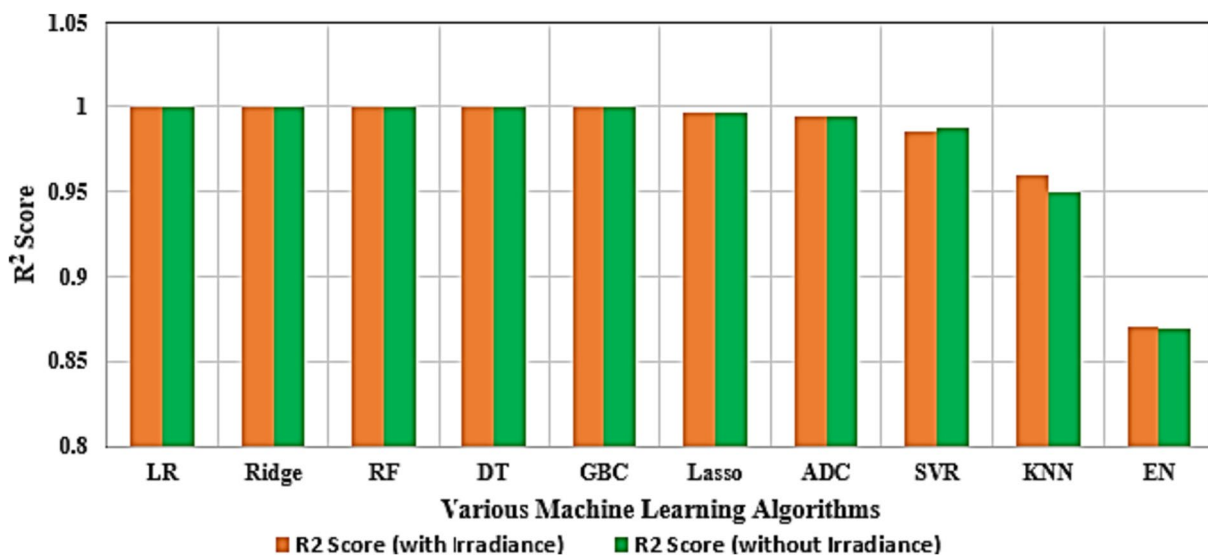
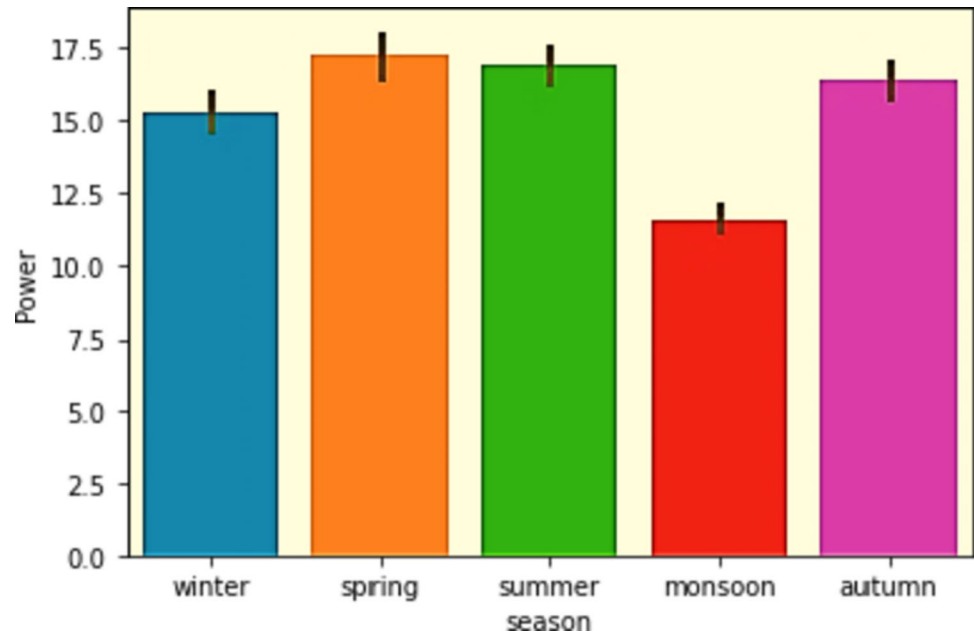


Fig. 11 Relationship between  $R^2$  score (with Irradiance) and  $R^2$  score (without irradiance)

**Fig. 13** Seasonal power forecast

it difficult to obtain solar irradiance data at each plant location. Therefore, we have estimated the solar power for all machine-learning algorithms without including solar irradiance as an input parameter. Table 4 clearly shows a decline in the performance of the machine-learning model based on statistical measurements. However, the results indicate minimal deviation in terms of the  $R^2$  score. The  $R^2$  score, which represents the coefficient of determination, is 0.99995 when solar irradiance is used as input, and without irradiance, the  $R^2$  score is 0.99994 for the LR model. The root mean square error (RMSE) value is 0.121 while considering sun irradiance and 0.125 in the absence of solar irradiance. Figure 11 illustrates the correlation between the  $R^2$  score (with irradiance) and the  $R^2$  score (without irradiance) for different machine learning models. It is evident from the graph that the performance of the model decreases significantly when compared to other machine learning models, achieving a score of 0.95964 when considering solar irradiance and 0.93932 in the absence of solar radiation. Therefore, it can be concluded that we can estimate solar power without solar irradiance if the data for solar is not available at the plant location using the proposed techniques.

### Prediction for Seasonal Variation

Another goal of the present research is to forecast the power output for various seasons. The plant location in Thar district, Madhya Pradesh, experiences five distinct seasons: winter (December and January), spring (February and March), summer (April, May, and June), monsoon (July and August), and autumn (September, October, and November). Since each season occurs in different months throughout

the year, it is necessary to have data organised month-wise in order to estimate seasonal power. Statistical averaging techniques used for monthly data compilation. Figure 12 displays the monthly forecast for different months throughout the year, and Fig. 13 represents the seasonal power variation. After completing the monthly forecast, the subsequent stage is to evaluate the seasonal power by employing feature engineering and an averaging technique. Our investigation demonstrates that the spring season, with an average power generation of 17.38 kW per day, is the most favourable period. Conversely, the monsoon season, with an average power generation of 11.26 kW per day, is the least appropriate. This monthly and seasonal power generation estimate is helpful to organise the solar power demand to schedule resources when power generation is poor in advance.

### Conclusion

This paper focuses on the prediction of solar power generation in Dhar District, Madhya Pradesh, India. Various meteorological data characteristics are utilised to estimate output power, both with and without considering irradiance. Multiple machine learning techniques were employed to model data covering a period of 36 months. The cross-correlation analysis revealed that temperature and irradiance are the two main meteorological variables that significantly influence the prediction of solar power. Based on the implementation of multiple machine-learning models, it can be concluded that the LR model has an  $R^2$  score of approximately 0.99995, with RMSE and MAE values of 0.121 and 0.0091, respectively. Similarly, the Ridge model

has  $R^2$  score, RMSE, and MAE values of 0.99994, 0.122, and 0.0097, respectively, with solar radiation as the input variable. The EN model has the lowest performance among all models, with an  $R^2$  score of 0.87024, an RMSE of 5.941, and an MAE of 4.6947. In DL, the FFNN method has 10 layers with 512 neurons in each, and the Rectified Linear Unit (ReLU) activation function with Adam Optimizer has an  $R^2$  score of 0.9987 with RMSE and MAE values of 0.0542 and 0.0462, respectively. Solar PV power generation is now a measurable part of overall power generation from multiple sources. This forecast helps in effectively managing surplus or insufficient power supply throughout different months or seasons. By employing precise forecasting methods and effectively regulating the equilibrium between supply and demand, we can prevent issues such as power outages or excessive power generation. Additional meteorological factors, such as cloud ceiling, precipitation, altitude, sunshine length, visibility, air quality index (AQI), and others, can be considered to determine the relationship between expected power and to improve the model's accuracy, we can also combine these algorithms using the ensemble and stacking methods.

**Authors' Contributions** All authors contributed equally.

**Funding** None.

**Data Availability** Will be made available on suitable request.

**Code Availability** NA.

## Declarations

**Ethics Approval** NA.

**Consent to Participate** NA.

**Consent for Publication** Yes.

**Conflicts of Interest/Competing Interests** NA

## References

- IEA (2022) World Energy Outlook 2022, IEA, Paris, License: CC BY 4.0 (report); CC BY NC SA 4.0 (Annex A). <https://www.iea.org/reports/world-energy-outlook-2022>. Accessed 6/6/2023
- Renewable Power Generation Costs (2021) International Renewable Energy Agency, Abu Dhabi. ISBN 978–92–9260–452–3
- MNRE Annual Report (2021) [https://mnre.gov.in/img/documents/uploads/file\\_f-1618564141288.pdf](https://mnre.gov.in/img/documents/uploads/file_f-1618564141288.pdf). Accessed 6/6/2023
- Paulescu M, Stefu N, Calinoiu D, Paulescu E, Pop N, Boata R, Mares O (2016) Ångström-Prescott equation: Physical basis, empirical models, and sensitivity analysis. *Renew Sust Energy Rev* 62:495–506. <https://doi.org/10.1016/j.rser.2016.04.012>. (Elsevier)
- Nikitha MS, Nisha KCR., Gowda MS, Aithal P, Mudakkayil NM (2022) Solar PV Forecasting Using Machine Learning Models. In: Proc. Second ICAIS-2022. pp 109–114. <https://doi.org/10.1109/ICAIS53314.2022.9742889>
- Kostylev V, Pavlovski A (2011) Solar power forecasting performance—towards industry standards. In 1st international workshop on the integration of solar power into power systems, Aarhus, Denmark. Energynautics GmbH Mühlstraße, Langen, Germany, pp 1–8
- Liu LM, Hudak GB, Box GE, Muller ME, Tiao GC (1992) Forecasting and time series analysis using the SCA statistical system. Scientific Computing Associates, DeKalb
- Garg U, Chohan DK, Dohal DC (2021) The Prediction of Power in Solar Panel using Machine Learning, In Int Conf on Computat Perfo Eval ComPE. pp 354–358. <https://doi.org/10.1109/ComPE53109.2021.9751901>
- Yadav AK, Chandel SS (2014) Solar radiation prediction using Artificial Neural Network techniques: A review. *Renew Sustain Energy Rev* 33:772–781. <https://doi.org/10.1016/j.rser.2013.08.055>
- Khatib T, Mohamed A, Sopian K (2012) A review of solar energy modeling techniques. *Renew Sustain Energy Rev* 16(5):2864–2869. <https://doi.org/10.1016/j.rser.2012.01.064>
- Kalogirou SA (2001) Artificial neural networks in renewable energy systems applications: a review *Renew. Sustain Energy Rev* 5(4):373–401. [https://doi.org/10.1016/S1364-0321\(01\)00006-5](https://doi.org/10.1016/S1364-0321(01)00006-5)
- Mellit A, Pavan AM (2010) A 24-h forecast of solar irradiance using artificial neural network: Application for performance prediction of a grid-connected PV plant at Trieste. *Sol Energy* 84(5):807–821. <https://doi.org/10.1016/j.solener.2010.02.006>
- Aggarwal SK, Saini LM (2014) Solar energy prediction using linear and non-linear regularization models: A study on AMS (American Meteorological Society). *Solar Energy Pred Cont* 78:247–256. <https://doi.org/10.1016/j.energy.2014.10.012>
- Silva AWDB et al (2022) Methodology Based on Artificial Neural Networks for Hourly Forecasting of PV Plants Generation. *IEEE Lat Am Trans* 20(4):659–668. <https://doi.org/10.1109/TLA.2022.9675472>
- Vakili M, Yazdi SRS, Kalhor K, Khosrojerdi S (2015) Using Artificial Neural Networks For Prediction Of Global Solar Radiation In Tehran Considering Particulate Matter Air Pollution. *Energy Procedia* 74:1205–1212. <https://doi.org/10.1016/j.egypro.2015.07.764>
- Elsaraiti M, Merabet A (2022) Solar Power Forecasting Using Deep Learning Techniques. *IEEE Access* 10:31692–31698. <https://doi.org/10.1109/ACCESS.2022.3160484>
- Panda S, Dhaka R, Panda B, Pradhan A, Jena CV, Nanda L (2022) A review on application of Machine Learning in Solar Energy & Photovoltaic Generation Prediction, In:Proc. of the Int. Conf. on Electr. and Renew. Sys . ICEARS 2022. 1180–1184. <https://doi.org/10.1109/ICEARS53579.2022.9752404>
- Khelifi, B, Zdiri MA, Salem FB (2021) Machine Learning for Solar Power Systems-A short tour, In: 12th Int. Renew. Energy Congress IREC 2021.1–6. <https://doi.org/10.1109/IREC52758.2021.9624896>
- Chang R, Bai L, Hsu CH (2021) Solar power generation prediction based on deep Learning. *Sust Energy Tech Ass* 47:101354. <https://doi.org/10.1016/j.seta.2021.101354>
- Natarajan V, Karatampati P (2019) Survey on renewable energy forecasting using different techniques, In:Proc. of 2nd Int. Conf. on Power and Embedded Drive Control (ICPEDC). pp 349–354. <https://doi.org/10.1109/ICPEDC47771.2019.9036569>
- Queen HJ, Jayakumar J, Deepuika TJ, Babu KVSM, Thota SP (2021) Machine learning-based predictive techno-economic analysis of power system. *IEEE Access* 9:123504–123516. <https://doi.org/10.1109/ACCESS.2021.3110774>
- Tan H, Li Z, Wang Q, Mohamed MA (2023) A novel forecast scenario-based robust energy management method for integrated

- rural energy systems with greenhouses. *Appl Energy* 330:120343. <https://doi.org/10.1016/j.apenergy.2022.120343>
23. Mahmud K, Azam S, Karim A, Zobaed S, Shanmugam B, Mathur D (2021) Machine Learning Based PV Power Generation Forecasting in Alice Springs. *IEEE Access* 9:46117–46128. <https://doi.org/10.1109/ACCESS.2021.3066494>
  24. Chen Z, Jin T, Zheng X et al (2022) An innovative method-based CEEMDAN–IGWO–GRU hybrid algorithm for short-term load forecasting. *Electr Eng* 104:3137–3156. <https://doi.org/10.1007/s00202-022-01533-4>
  25. Dayalan S, Gul S, Rathinam R, Savari GF, Aleem SHEA, Mohamed MA, Ali ZM (2022) Multi-Stage Incentive-Based Demand Response Using a Novel Stackelberg–Particle Swarm Optimization. *Sustainability* 2022(14):10985. <https://doi.org/10.3390/su141710985>
  26. Meng F, Zou Q, Zhang Z, Wang B, Ma H, Abdullah HM, Almalaq A, Mohamed MA (2021) An intelligent hybrid wavelet-adversarial deep model for accurate prediction of solar power generation. *Energy Reports* 7:2155–2164. <https://doi.org/10.1016/j.egyr.2021.04.019>
  27. Gupta R, Yadav AK, Jha SK, Pathak PK (2022) Time Series Forecasting of Solar Power Generation Using Facebook Prophet and XG Boost, In: Proc. of Int. Conf. on Delhi Section Conf. (DELCON). 1–5. <https://doi.org/10.1109/DELCON54057.2022.9752916>
  28. Mutavhatsindi T, Sigauke C, Mbuva M (2020) Forecasting hourly global horizontal solar irradiance in South Africa using machine learning models. *IEEE Access* 8:198872–198885. <https://doi.org/10.1109/ACCESS.2020.3034690>
  29. Zhao Y, Ball R, Mosesian J, Palma JD, Lehman B (2015) Graph-based semi-supervised learning for fault detection and classification in solar photovoltaic arrays. *IEEE Trans Power Electron* 30(5):2848–2858. <https://doi.org/10.1109/TPEL.2014.2364203>
  30. Alaraj M, Kumar A, Ilsaidan I, Jamil M (2021) Energy production forecasting from solar photovoltaic plants based on meteorological parameters for Qassim Region, Saudi Arabia. *IEEE Access* 9:83241–83253. <https://doi.org/10.1109/ACCESS.2021.3087345>
  31. Avila L, Paula MD, Carlucho I, Sanchez C (2019) MPPT for PV systems using deep reinforcement learning algorithms. *IEEE Lat Am Trans* 17(12):2020–2027. <https://doi.org/10.1109/TLA.2019.9011547>
  32. Bansal RC, Pandey JC (2005) Load forecasting using artificial intelligence techniques: a literature survey. *Int J Comput Appl Technol* 22(3):109–119. <https://doi.org/10.1504/IJCAT.2005.006942>
  33. Singh R, Bansal RC, Singh A, Naidoo A (2018) Multi-objective optimization of hybrid renewable energy system using reformed electric system cascade analysis for islanding and grid connected modes of operation. *IEEE Access* 6:47332–47354. <https://doi.org/10.1109/ACCESS.2018.2867276>
  34. Asensio AG, Gorrachategui IS, Nuez AB, Bernal C, Alcaine JMS, Cebolla FJP (2021) Energy shortage failure prediction in photovoltaic standalone installations by using machine learning techniques. *IEEE Access* 9:158660–158671. <https://doi.org/10.1109/ACCESS.2021.3129930>
  35. Alam AM, Masood NA, Raze IA, Zumaed M (2021) Solar PV power forecasting using traditional methods and machine learning techniques, In: Proc. of 2021 IEEE KPEC. pp 1–5. <https://doi.org/10.1109/KPEC51835.2021.9446199>
  36. Heo J, Song K, Han S, Lee DE (2021) Multi-channel convolutional neural network for integration of meteorological and geographical features in solar power forecasting. *Appl Energy* 295:117083. <https://doi.org/10.1016/j.apenergy.2021.117083>
  37. Cheng L, Zang H, Wei Z, Ding T, Sun G (2022) Solar power prediction based on satellite measurements – a graphical learning method for tracking cloud motion. *IEEE Trans Power Syst* 37(3):2335–2345. <https://doi.org/10.1109/TPWRS.2021.3119338>
  38. Hwang HP, Ku C, Chan JC (2021) Detection of Malfunctioning Photovoltaic Modules Based on Machine Learning Algorithms. *IEEE Access* 9:37210–37219. <https://doi.org/10.1109/ACCESS.2021.3063461>
  39. Bhardwaj G, Bhardwaj S, Agarwal R (2022) An efficient speaker identification framework based on Mask R-CNN classifier parameter optimized using hosted cuckoo optimization (HCO). *J Ambient Intell Human Comput* 2022. <https://doi.org/10.1007/s12652-022-03828-7>
  40. Kumar D, Mathur HD, Bhanot S, Bansal RC (2019) Forecasting of solar and wind power using LSTM RNN for load frequency control in isolated microgrid. *Int J Modell Simul* 41(4):311–323. <https://doi.org/10.1080/02286203.2020.1767840>
  41. Obiora CN, Hasan AN, Ali A, Alajarmeh N (2021) Forecasting Hourly Solar Radiation Using Artificial Intelligence Techniques. *IEEE Can J Electr Comput Eng.* 44(4):497–508. <https://doi.org/10.1109/ICJECE.2021.3093369>
  42. Wang K, Qi X, Liu H (2019) Photovoltaic power forecasting based LSTM-Convolutional Network. *Energy* 9:116225. <https://doi.org/10.1016/j.energy.2019.116225>
  43. Gao M, Li J, Hong F, Long D (2019) Day-ahead power forecasting in a large-scale photovoltaic plant based on weather classification using LSTM. *Energy* 187:115838. <https://doi.org/10.1016/j.energy.2019.07.168>
  44. Wang L, Mao M, Xie J, Liao Z, Zhang H, Li H (2023) Accurate solar PV power prediction interval method based on frequency-domain decomposition and LSTM model. *Energy* 262:125592. <https://doi.org/10.1016/j.energy.2022.125592>
  45. Sarmas E, Dimitropoulos N, Marinakis V et al (2022) Transfer learning strategies for solar power forecasting under data scarcity. *Sci Rep* 12:14643. <https://doi.org/10.1038/s41598-022-18516-x>
  46. Bui LD, Nguyen NQ, Doan BV, Sanseverino ER (2022) Forecasting energy output of a solar power plant in curtailment condition based on LSTM using P/GHI coefficient and validation in training process: a case study in Vietnam. *Electric Power Sys Res* 213:108706. <https://doi.org/10.1016/j.epsr.2022.108706>
  47. Djaafari A, Ibrahim A, Bailek N, Bouchouicha K, Hassan MA, Kuriqi A, Al-Ansari N, El-Kenway MES (2022) Hourly predictions of direct normal irradiation using an innovative hybrid LSTM model for concentrating solar power projects in hyper-arid regions. *Energy Rep* 8:15548–15562. <https://doi.org/10.1016/j.egyr.2022.10.402>
  48. Obando E, Carvajal S, Pineda J (2019) Solar Radiation Prediction Using Machine Learning Techniques: A Review. *IEEE Lat Am Trans* 17(4):684–697. <https://doi.org/10.1109/TLA.2019.8891934>
  49. Liu X, Bansal RC (2014) Integrating Multi-objective optimization with computational fluid dynamics to optimize boiler combustion process of a coal fired power plant. *Appl Energy* 114:658–669. <https://doi.org/10.1016/j.apenergy.2014.02.069>
  50. Gaurav, Bhardwaj S, Agarwal R (2023) Two-Tier Feature Extraction with Metaheuristics-Based Automated Forensic Speaker Verification Model. *Electronics* 12(10):2342. <https://doi.org/10.3390/electronics12102342>
  51. Bholra P, Bhardwaj S (2019) Estimation of solar radiation using support vector regression. *J Inf Optim Sci* 40(2):339–350. <https://doi.org/10.1080/02522667.2019.1578093>
  52. Kim N, Lee H, Lee J, Lee B (2021) Transformer based prediction method for solar power generation data. In: Proc Int Conf Inf Comm Tech Convergence (ICTC). 7–9. <https://doi.org/10.1109/ICTC52510.2021.9620897>
  53. Pasion C, Wagner T, Koschnick C, Schuldt S, Williams J (2020) Hallinan K (2020) Machine Learning Modeling of Horizontal Photovoltaics Using Weather and Location Data. *Energies* 13(10):2570. <https://doi.org/10.3390/en13102570>



54. Zhang C, Zhang Y, Su J, Gu T, Yang M (2020) Modeling and prediction of PV module performance under different operating conditions based on power-law I-V model. *IEEE J Photovolt* 10(6):1816–1827. <https://doi.org/10.1109/JPHOTOV.2020.3016607>
55. Karimi AM, Fada JS, Parrilla NA, Pierce BG, Koyuturk M, French RH, Braid JL (2020) Generalized and mechanistic PV module performance prediction from computer vision and machine learning on electroluminescence images. *IEEE J Photovoltaics* 10(3):878–887. <https://doi.org/10.1109/JPHOTOV.2020.2973448>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.