Analysis Of Solar Power Generation Forecasting Using Machine Learning Techniques

K. Anuradha[1],* *Deekshitha* Erlapally[2], *G.* Karuna[3], *V.* Srilakshmi[4], *K.* Adilakshmi[5]

[1]Professor, Computer Science and Engineering, GRIET, Hyderabad, Telangana, India.
[2]MTech Student, Computer Science and Engineering, GRIET, Hyderabad, Telangana, India.
[3]Professor, Computer Science and Engineering, GRIET, Hyderabad, Telangana, India.
[4] Asst.Professor, Computer Science and Engineering, GRIET, Hyderabad, Telangana, India.
[5] Asst.Professor, Computer Science and Engineering, GRIET, Hyderabad, Telangana, India.

Abstract: Solar power is generated using photovoltaic (PV) systems all over the world. Because the output power of PV systems is alternating and highly dependent on environmental circumstances, solar power sources are unpredictable in nature. Irradiance, humidity, PV surface temperature, and wind speed are only a few of these variables. Because of the unpredictability in photovoltaic generating, it's crucial to plan ahead for solar power generation as in solar power forecasting is required for electric grid. Solar power generation is weather-dependent and unpredictable, this forecast is complex and difficult. The impacts of various environmental conditions on the output of a PV system are discussed. Machine Learning (ML) algorithms have shown great results in time series forecasting and so can be used to anticipate power with weather conditions as model inputs. The use of multiple machine learning, Deep learning and artificial neural network techniques to perform solar power forecasting. Here in this regression models from machine learning techniques like support vector machine regressor, random forest regressor and linear regression model from which random forest regressor beaten the other two regression models with vast accuracy.

1 Introduction:

Solar energy has many benefits, but also have their initial investment for installing solar panels is quite high, and not everyone will be able to afford them. Unfortunately, this is a downside of solar panels; nevertheless, as prices continue to decline, the future looks bright. Solar panels are currently relatively costly; but, new government programs and cutting-edge technology are making them cheaper. Despite the fact that photovoltaic cells are recognized as the significant source of potential energy production, their low return on investment and high upfront costs keeps them from becoming widely used.The high initial cost prevents them from becoming widely used. Because photovoltaic cells convert solar energy into electrical energy, the amount of solar energy produced each day influences the size of the photovoltaic system, just as the amount of solar radiation influences the amount of electricity produced each day. This is influenced by factors such as location, time, and weather patterns. Solar irradiance is the power obtained per unit area from the Sun via electromagnetic radiation in the wavelength range of the solar cell in use.

Major grid integration is difficult because renewable energy is irregular and uncontrollable. Households can now use almost any amount of energy due to the recent electric grid at any moment, but it is not equipped for large quantities of uncontrollable generation at this time. As it is converting solar radiance into power we don't get that how much power is emitted for different location, time, and weather. For this type of clarification machine learning techniques are used in order to differentiate it for different conditions. Machine-learning techniques are wide applied to several fields where it can separate the weather based power.

The amount of energy a PV system generates is proportional to meteorological parameters including cloud cover, sun intensity, and site-specific conditions, among other [3]. Solar panel works differently for different weather conditions. In case if its summer seasons then the amount of energy consumed by the panel from sun is very much more. But in case of rainy and windy conditions the energy consumed is pretty much different. Power generation mostly depends on weather conditions so they take weather forecasting into consideration. As a result, the amount of electricity generated is determined by solar irradiance on a given day, which is determined by a number of factors such as location, time, and weather patterns. We concentrate on the problem of automatically generating models that accurately predict renewable generation based on National Weather Service forecasts (NWS). Using historical NWS forecast data and data generated by solar panels, we experiment with a variety of machine learning techniques to develop prediction models.

---

* Corresponding author: author@e-mail.org

Meteorological data, including ambient temperature, humidity, and solar radiation, will be collected by meteorological monitoring stations every three hours.

Machine-learning techniques have been widely used in a range of fields involving data-driven problems in recent decades. Machine-learning approaches encompass a wide range of interdisciplinary topics, including statistics, mathematics, artificial neural networks, data mining, optimization, and artificial optimization. With or without mathematical problem forms, machine learning approaches attempt to find a relationship between input and output data. The process of analysing data is known as data analyzation. ML employs statistical approaches to enable computers to "learn" from data without having to be explicitly programmed. Machine learning has two main application categories: regression and classification. Solar power forecasting necessitates the use of regression methods. Some of the ML regression algorithms that can be used for time series forecasting are Linear Regression (LR), Support Vector Machine Regression (SVMR), and Random Forest (RF).

Weather and physical elements influence the electrical power output of a solar photovoltaic (PV) panel. Solar irradiance, cloud cover, humidity, and ambient temperature are the main meteorological factors that influence solar power generation. Predicted weather parameters can be used as model inputs, while solar power forecasts can be used as the model output. Because of its ongoing training nature, the ML algorithm adjusts to physical parameters.

In machine learning SVM plays a major role in order to classify the data and monitor weather condition according. Combining data from photovoltaic power generation with meteorological conditions, according to the positive position of photovoltaic power generation. For every 3 hours svm gives analyzed data for classification and regression analysis. Using hyperplane we can classify the accurate results from solar panel based on the weather conditions.Random forest, on the other hand, is a classification strategy that uses many decision trees to classify data. In order to generate an uncorrelated forest of trees whose committee forecast is more trustworthy than that of any single tree, bagging and feature randomization are utilised in the development of each individual tree. It gives multiple decisions tress it merges all the decision tress into one form of decision tree it's for different climatic conditions such as for summer, rainy, winter seasons. Error statistics such as mean bias error (MBE), mean absolute error (MAE), root mean square error (RMSE), relative MBE (rMBE), mean percentage error (MPE), and relative RMSE are used to assess the model's validity (rRMSE). Linear regression is a supervised learning-based machine learning approach. It does a regression analysis. Based on independent variables, regression models a goal prediction value. It's generally used in forecasting to figure out how variables are related. Regression models differ in terms of the sort of link examined between dependent and independent variables, as well as the number of independent variables used.

2. Related Work:

Solar energy forecasts can be categorised in a variety of ways. The persistence or smart persistence model, which uses historical data to forecast future power generation over a short period of time, is the most basic method (2-3 hours). This method can be used to set a standard against which other forecasting methods can be measured. In most cases, a prediction is completed in two stages. A NWP is designed for a specified time period and location to begin with. The generated NWP is then utilised to forecast power generation using forecasting algorithms. It is possible to employ a physical model, a statistical method, or a machine learning methodology [1].For prediction, ML algorithms are compared to the Smart Persistence (SP) approach, with ML models outperforming the SP model. The unpredictability of solar resources has hampered grid management as solar diffusion rates have increased. Unpredictability and intermittent electricity delivery are two of the most difficult aspects of integrating renewables into the system. As a result, solar power forecasting is becoming increasingly important for grid stability, optimal unit commitment, and cost-effective dispatch. To overcome the problem, we employ machine learning techniques to sift through extraordinary solar radiation predicting models. For developing prediction models, a variety of regression algorithms are tested, including linear least squares and support vector machines with various kernel functions. We use day-ahead sun radiation data forecasts in these tests to show that a machine learning approach can correctly anticipate short-term solar power [2]. A hybrid or mixed forecasting method was developed by combining clustering, classification, and regression approaches to produce a forecasting model. Based on the weather forecast for the next day, the model (with the closest weather condition) is chosen to forecast the power output using cluster-wise regression [3].

Renewable energy sources are progressively being integrated into electric networks alongside nonrenewable energy sources, posing significant issues due totheir sporadic and erratic nature In order to address these issues, soft-computing solutions for energy prediction are essential.We apply a number of data mining methodologies, including preparing historical load data and analysing the features of the load time series, because electricity consumption is entangled with the usage of other energy sources like natural gas and oil. The trends in power consumption from renewable and nonrenewable energy sources were examined and contrasted. A novel machine learning-based hybrid technique (SVR) uses multilayer perceptron (MLP) and support vector regression [5].Using SVM regression, solar power generation produces acceptable results [6]. However, it lacks a detailed examination of solar power generation and meteorological data, and hence is restricted in its capacity to accurately predict other data sets by merely using different SVM kernels after some basic statistical data processing [8].

To study the association between expected weather conditions and power output created as a historical time series, artificial intelligence (AI) approaches are applied.

AI approaches use algorithms that can implicitly characterise the nonlinear and highly intricate relationship between input data (NWP predictions) and output power instead of formal statistical analysis. The ANN is a brain model that is based on biology. They're employed in a range of applications that use AI approaches including supervised, unsupervised, and reinforcement learning. The ANN learns from data in the supervised learning approach by being trained to approximate and estimate the function or relationship. [6].

Their models have been improved to predict PV plant power generation [4–7]. Even with the cloud graph from synchronous meteorological satellites, the significant unpredictability in critical components, particularly the diffuse component from the sky hemisphere, makes solar irradiance far less predictable than temperature. PV systems including a large number of different tiles deployed over a large area have additional challenges [12]. Because it is impossible to examine all connected meteorological forecasts in a practical context, many alternative alternatives have been devised. Weather forecasts from meteorological websites [8] were considered by some. Others used nonlinear modelling approaches like artificial neural networks to try to simplify the solar forecast model (ANN). Two types of networks are commonly used to forecast global solar radiation, solar radiation on titled surfaces, daily solar radiation, and short-term solar radiation: radial basis function (RBF) and multilayer perception (MLP).

In a three-layer feed forward model, back-propagation is the neural network training technique. To reduce forecast error, the input layer provides an error correction factor depending on the projected output for the previous 5 minutes.

An LSTM network will learn a function that accepts a sequence of previous solar irradiance values as input and returns a solar irradiance value as output. Deep neural networks, such as the Deep Belief Network (DBN), will learn a function that takes a sequence of historical sun irradiance values as input and outputs a solar irradiance value. If a series of observations are converted into a variety of occurrences, an LSTM network can learn from them. The sequence is partitioned using LSTM for prediction purposes.

3 PROPOSED WORK:

For knowing how much power is generated from solar we have the dataset showing daily average temperature in Celsius, distance from solar noon, wind speed, wind direction, sky cover, and humidity and then the power generated. Here we are calculating how much power is generated in different weather condition for India dataset . We have taken Indian dataset with different temperature readings. The available dataset is based on hourly weather parameter values. To convert the data to mean values per day, the average of the 24-hour data was used. From 2019 to 2020, several weather factors were collected to investigate the relationship between mean solar irradiance and meteorological data in order to accurately estimate power generated.

The proposed work's System Architecture is to first consider the dataset and preprocess the data, then divide it into train and test data, apply classification techniques, and predict the results. Solar power weather dataset is used for forecasting purposes in this case. Data preprocessing methods include cleaning, integration, reduction, and transmission. We must purge any data that is no longer absolutely necessary. Data cleaning is the process of identifying and removing inaccurate or incorrect records from a dataset. Data from the real world frequently contains noise and missing values, and it may be in an unusable format that cannot be directly used for DL models. Data preprocessing is required to clean data and prepare it for various Deep Learning models, increasing accuracy and efficiency. Training and testing data are separated from the preprocessed data. The model is trained using training data, and its predictions are validated using testing data. Data splitting is the process of dividing available data into two halves, usually for cross-validator purposes. The first set of data is used to build a predictive model, while the second set is used to evaluate the model's performance. In analyzing data mining algorithms, separating data into training and testing sets is crucial. The training percentage is set at 80% and the test percentage is set at 20%. When a data set is divided into a training set and a testing set, the majority of the data is used for training and only a small portion is used for testing. To train any model, no matter what type of dataset is used, the dataset must be divided into training and testing data. The dataset will be examined for null values and outliers during the data preprocessing step, and the model will be trained using three hours of data before being used to forecast solar power generation value. The power generated radiance phase will be estimated using machine learning (ML) methods (e.g., support vector regression, linear regression, elastic net regression, and random forest) as shown in below Fig 1.
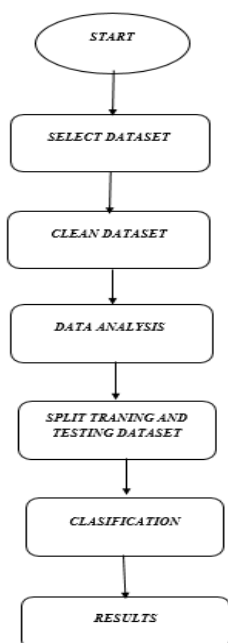
Fig.1 Flow chart for solar power generation

4. Methodology:

The current dataset is based on hourly weather parameter values. To convert the data to mean values per day, the average of the 3-hour data was used. Various weather characteristics were gathered in order to investigate the relationship between mean solar irradiance and meteorological data in order to accurately estimate mean solar irradiance. The average daily values of air temperature, humidity, wind speed, wind direction, visibility, average pressure, average wind speed, and electricity generated are among the data collected. The direction of the wind, on the other hand, indicates how high the sun is. It's also expressed in degrees.

Machine Learning (ML) Models are used for forecasting the solar power generation weather analysis. The Regression techniques here proposed are Support Vector Machine, Random Forest, Linear Regression are various ML Models used in this paper.

4.1 Forecasting models:

In this study, we used the chosen dataset to evaluate individual performance using a number of meteorological attributes utilising three commonly used machine learning algorithms. The output of the unseen test sample is predicted to be the mean of these K closest matches because our prediction variable is continuous valued. We investigated a variety of K values, however only the results for K=3 and K=5 are presented. When K is more than 3, the RMS error increases.

Support vector regression (SVR) using a radial basis function as the kernel and random forest (RF) approaches are used to create the models. Because of the non-linearity of the dataset, we used the models indicated above instead of linear models. The most basic and widely used regression method is linear regression

(LR) [10].It uses linear predictor functions to represent the relationship between the input and output variables, and a least squares approach is used to estimate the unknown model parameters from the data. A set of linear equations or an iterative method like gradient descent can be used to estimate parameter values. We employed the characteristics provided in the, followed by feature scaling, to standardise the input data. SVR's precision varies depending on the kernel function and other variables. To discover the optimal settings, we employed the Grid search approach. To evaluate the models' performance on the test set, we computed the Root-Mean-Square Error (RMSE) and R squared values. Before choosing the models with the lowest root mean squared error and highest R squared values, we fine-tuned the model hyperparameters. Nonlinear relationships can be mapped using these methods. In data science challenges of various kinds, methods including decision trees, RF, and gradient boosting are commonly utilised.The RF method is a tree-based machine learning approach that can be used for regression and classification. It also performs dimensional reduction, controls missing and outlier values, and performs a variety of additional data exploration activities. The bagging approach is used to train RFs. This method allows for the usage of numerous instances for the training stage because the dataset is sampled with a replacement. Linear regression is a method for demonstrating the link between a dependent variable and one or more independent variables by using the best-fit linear curve. It is concerned with determining the best-fit line with the data by attaining a perfect slope and intercept value.The best model for forecasting solar power system output based on numerous weather parameters was then created. The models that gave the greatest results on the dataset were support vector regression, random forests, and linear regression, and these models were then utilized to anticipate PV system performance for 2019. Thanks to the predictive analysis, the estimated production in this situation ranges from 0 to 1000 Watt hours. These models were then evaluated using the test data. The SVR model has an RMSE of 135.7, while the random forest model has an RMSE of 28.62 and the SVR model has an RMSE of 58.24. The random forest model's points are close to the regressed diagonal line, however the SVR model's points are not.

5 RESULTS:

A matrix of pair correlation coefficients is generated for a set of features under investigation in order to find collinear factors as shown in below Fig 1, Fig 2,Fig 3 and Fig 4.
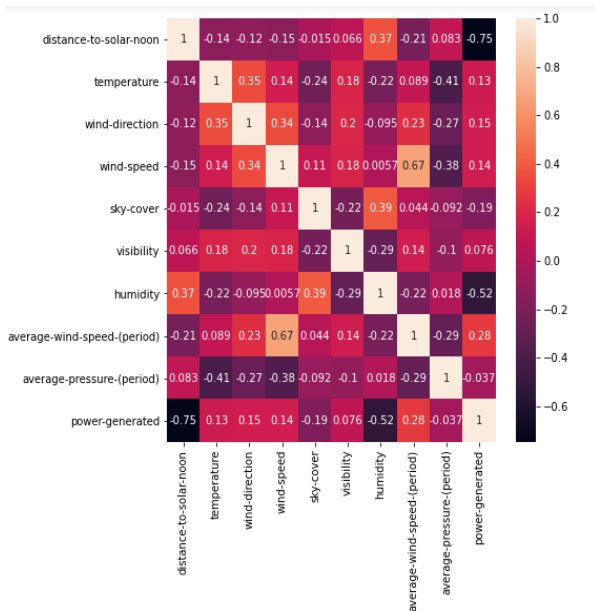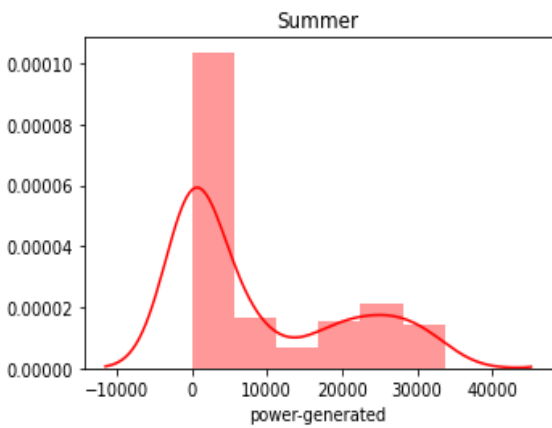
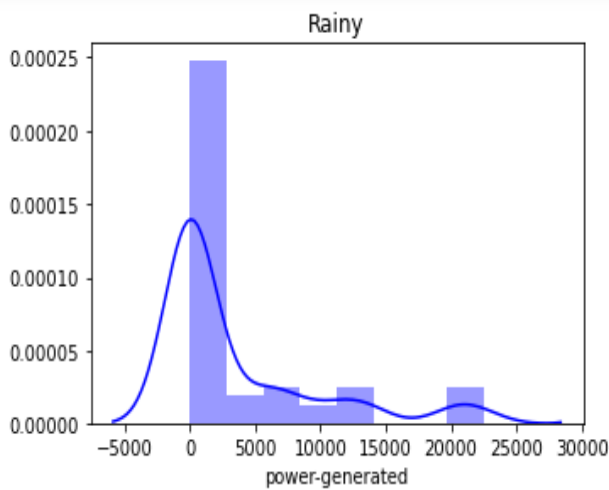**Fig 2** correlation matrix

Fig 3:summer season analysis

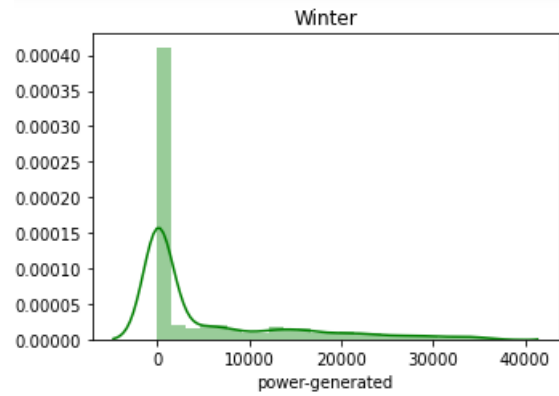**Fig 4**: rainy season analysis
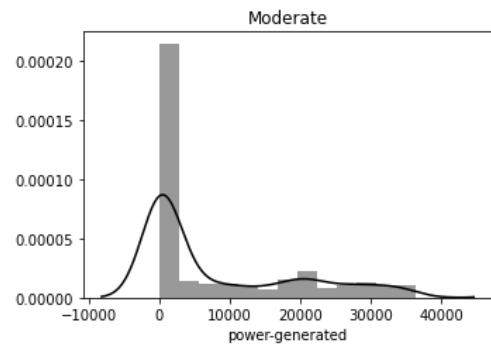

Fig 5: winter season analysis
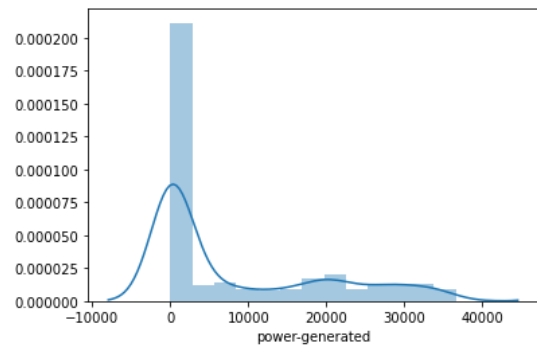

Fig 6: moderate analysis

Fig7: overall season's comparative analysis

Here fig6 shows that how power is generated comparatively in different weather conditions where the x axis shows the power generated and y axis shows the distance from solar noon. It is nothing more than Solar noon occurs when the Sun passes through a location's meridian (a meridian is an imaginary line that runs from the North Pole to the South Pole along the Earth's surface.) and ascends to its highest point in the sky. In most cases, it does not occur at 12 p.m and as shown in above Fig 5 and Fig 6 and Below Fig 7.
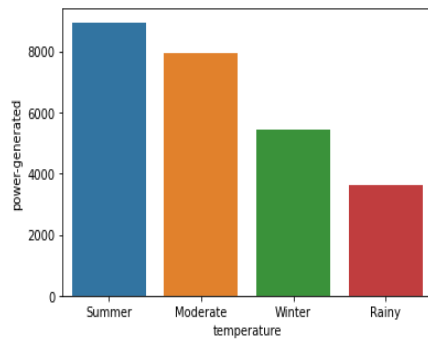
Fig 8: Different temperature analysis results of how much power is generated.

Fig6 says the power generated is read as Jules here where as temperature as in summer, rainy, winter and the moderate is said as the different weather condition. Dataset considered is the numerical dataset so we replaced them with string type where in dataset it is shown as 0 for rainy 1 for winter, 2 for moderate,3 for summer as shown in below Fig 9 and Fig 10.
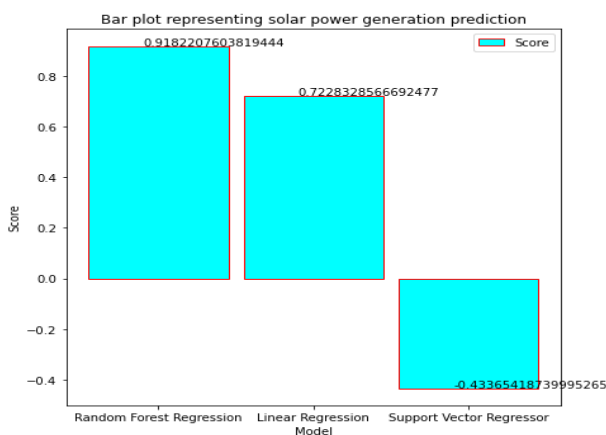

Fig 9: score of all the three model used in solar power analysis.

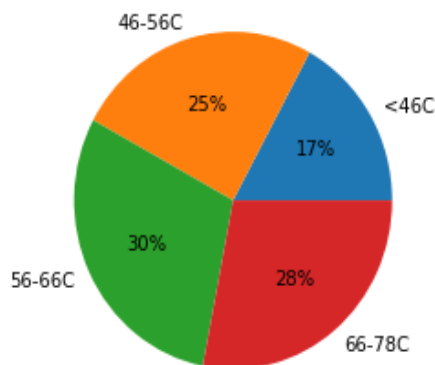This fig8 shows that score of three different models used in this paper for solar power generation

Fig 10: temperature difference and their power generated percentage accordingly.

This shows that how much percentage of power is generated ass the temperature rises less than 46 Celsius is 0 that's the rainy season temperature, 46 o 56 Celsius

is 1 that's winter season temperature,56 to 66C is 2 that's moderate temperature and 66 to 78C is 3 that's summer season and as shown in below Table 1.

Table 1: Model Prediction of RMSE, MAE, MAPE, and MBE and Accuracy

Model	RMSE	MAE	MSE	Accuracy
Support vector machine regressor	131.44	77.16	172.76	-44.92
Linear regression	58.57	48.39	343.08	72.4
Random forest regressor	27.32	12.45	746.48	94.01

6. Conclusion:

We presented a machine learning-based approach for solar power generation analysis in this paper, which accurately forecasts power generated across India's states based on environmental data. Most importantly, our methodology went beyond prediction by delivering key results that aided in the understanding of solar power analysis (variable importance by time period). By a wide margin, the proposed method outperformed other popular methods, such as Random forest. The proposed models are SVR, LR, and RF. Compared to the temperature with the given data. 56-55F --30% power generation is increasing compared to other temperatures. Temperature <46F is 17% temperature average of it. As the above results, we can see that Random Forest Regressor model is performing better with 94.01% accuracy and hence that model is preferred for deployment.

Reference:

1. P. A. G. M. Amarasinghe and S. K. Abeygunawardane, "*Application of Machine Learning Algorithms for Solar Power Forecasting in Sri Lanka*" (2nd International Conference On Electrical Engineering (EECon), Colombo, Sri Lanka, 87 2018).

2. M. Z. Hassan, M. E. K. Ali, A. B. M. S. Ali and J. Kumar, "*Forecasting Day-Ahead Solar Radiation Using Machine Learning Approach*" (4th Asia-Pacific World Congress on Computer Science and Engineering (APWC on CSE), Mana Island, Fiji, 252 2017).

3. A. Bajpai and M. Duchon, "*A Hybrid Approach of Solar Power Forecasting Using Machine Learning*" (3rd International Conference on Smart Grid and Smart Cities (ICSGSC), 108 2019).

4. A. Khan, R. Bhatnagar, V. Masrani and V. B. Lobo, "*A Comparative Study on Solar Power Forecasting using Ensemble Learning*," (4th International

Conference on Trends in Electronics and Informatics (ICOEI), 224 2020).

5. Khan, P.W.; Byun, Y.-C.; Lee, S.-J.; Kang, D.-H.; Kang, J.-Y.; Park, H.-S. Energies, **13**, 4870 (2020).

6. Faquir, Sanaa & Yahyaouy, Ali & Tairi, H. & Sabor, Jalal. International Journal of Fuzzy System Applications. **4**, 10 (2015).

7. Aler R., Martín R., Valls J.M., Galván I.M. Intelligent Distributed Computing VIII. Studies in Computational Intelligence, vol **570** (2015).

8. Y. Wang, G. Cao, S. Mao and R. M. Nelms, "*Analysis of solar generation and weather data in smart grid with simultaneous inference of nonlinear time series*," (IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), 600 2015).

9. Carrera B, Kim K. Sensors (Basel). **20**, 3129 (2020).

10. Jawaid F, NazirJunejo K. *Predicting daily mean solar power using machine learning regression techniques*. (Sixth International Conference on Innovative Computing Technology (INTECH) 355 2016).

11. Batcha RR, Geetha MK. *A survey on IOT based on renewable energy for efficient energy conservation using machine learning approaches*. (3rd International Conference on Emerging Technologies in Computer Engineering: Machine Learning and Internet of Things (ICETCE) 123 2020).

12. Li, Zhaoxuan & Rahman, Sm Mahbobur & Vega, Rolando & Dong, Bing. Energies. **9**, 55 (2016).

13. Lai JP, Chang YM, Chen CH, Pai PF. Applied Sciences; **10**, 5975 (2020).

14. Brahma, B.; Wadhvani, R. Symmetry, **12**, 1830 (2020).