Survey Analysis of Solar Power Generation Forecasting

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

[1]MTech Student, *Computer Science and Engineering, GRIET,* Hyderabad, Telangana, India.
[2]Professor, *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 the conversion of sunlight into electricity using solar photovoltaic cells as a source of energy. There are various applications for solar power; here is information on PV cell generation. We seek to understand the behavior of solar power plants through the data generated by the photovoltaic modules and the power generation in different weather conditions in India. The goal of this survey is to give a thorough assessment and study of machine learning, deep learning and artificial intelligence. Artificial intelligence (AI) models as well as information preprocessing techniques, parameter selection algorithms and predictive performance evaluations are used in machine learning and deep learning models for predicting renewable energies. But in case of time series data we can predict only the errors using a linear regression model, we can also calculate things like root mean square error (RMSE), mean absolute error (MSE), mean bias error (MBE) and mean absolute percentage error (MAPE). By the analysis of weather condition also we can predict the consumption of current by solar for every 15 minutes, 1day, and 1week or even for 1 month and find the accuracy.

1. Introduction

Solar energy has many benefits, including its sensitivity to imitative circumstances such as increasing oil prices, its renewable nature, and its ability to reduce imports and dependence on foreign resources. 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. Since photovoltaic cells generate electricity by converting solar energy into electrical energy, the amount of solar energy produced every day is important to the size of the photovoltaic system, because the amount of solar radiation determines the amount of Major grid [17] integration is difficult because renewable energy is irregular and uncontrollable. Households in India 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 [10] 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.

electricity produced every day. This is influenced by factors such as place, 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 that is in use [13].

Here we seek and analyzed India as a place and the temperature power plants as consideration based on weather condition in India. According to power generation commissioning statistics, solar plants have the largest installed capacity growth dynamics among renewable energy power plants. Making realistic solar generation projections for the day ahead view is becoming an increasingly critical issue in many places throughout the world at the current level of power system development.

Machine-learning techniques are wide applied to several fields where it can separate the weather based power. Machine-learning techniques are wide applied to several fields related to data-driven issues. Machine-learning techniques embrace several knowledge domain areas, like statistics, arithmetic, Ann, data processing, optimization, and computation are all terms that come to mind when thinking about artificial neural networks [18].

Deep learning, a sub-field of machine learning, has been blooming recently due to the rapid expansion of data

technology in hardware and code. Deep Learning calculations have only been used sparingly in the past for determining environmentally friendly power plants. We present these amazing calculations in the field of sustainable power determining by utilizing distinct Deep learning and ANN[16] calculations, such as DBN, Auto Encoder, and LSTM. In our tests, we integrated these calculations to show how effective they are in calculating the energy yield of solar-powered power plants as compared to a traditional MLP and an actual anticipating model. Our outcomes utilizing Deep Learning calculations show a better determining execution thought about than Artificial Neural Networks, like other reference models including such actual models, are artificial neural networks. In the literature, these models are compared with selected machine learning methods and are available predictive models. Using the mean square error, the proposed cloud classification ensemble model reduces the mean square error by 10.49%, 7.78% and 7.95% respectively, with deep belief networks, support vector regression and random forest [15] regression models. The use of weather parameters to forecast solar power generation is suggested as an Artificial Neural Network (ANN). This discussion describes the techniques used to predict renewable energy in machine learning models, including data preprocessing techniques, parameter selection algorithms, and predictive performance measurement.

2. Related Work

Energy is now the primary source of socioeconomic growth. However, because of the increasing rate of environmental concern, renewable energy is attracting a lot of attention. Related to the ongoing depletion of fossil fuels, this alternative energy source is gaining in popularity. That is the energy that comes from the sun, wind, rain, and other natural sources [3].To generate electricity at power plants, electromechanical generators are used, and they are mostly driven by chemical combustion or even nuclear fission heat engines, but also by alternative energy sources such as kinetic energy from flowing water and wind. There are numerous other technologies, such as solar photovoltaic and geothermal power, can and are used to produce electricity. Using thermal power: This method is used in thermocouples, thermopiles, and thermionic converters to convert temperature differences directly to power. Thermal power station is a power plant with a steam-driven primary mover. The water is heated and turned into steam, which powers an electrical generator by spinning a steam turbine. The steam is concentrated in a condenser and returned to where it was heated after passing through the turbine; this is known as the Rankine cycle. Using Wind power: The development of either lift or drag force is one of the two primary physical values by which energy can be extracted from the wind. Hydrostatic pressure, which are the forces felt by a person exposed to the wind, are the most obvious means of propulsion. Buzz forces are the most effective mode of propulsion, but they are less well known than drag forces due to their sensitivity. Using Nuclear power: Charged particles are created and accelerated (examples: beta voltaic or alpha partial emission. In a nuclear power plant, uranium is first formed into pellets, then into long rods. Submerging the uranium rods in water keeps them cold. When they are taken out of the water, a nuclear reaction occurs, resulting in heat. Rising and lowering the thermostat regulates the amount of heat necessary. Using solar power: The photoelectric effect, as seen in solar cells, is responsible for converting light into electrical energy. Since the solar array will stay aligned to the sun, solar monitoring allows for further energy production. To improve performance, a solar system's power production should be maximized. To get the most power out of your solar panels, make sure they're aligned with the sun.

Cluster analysis [8] using nRMSE (Normalized Root Mean Square Error), nMAE (Mean Absolute Error) and nMBE (Mean bias error) solar generation 8.80, 4.06 and -1.01. By using successful feature selection or extraction, you can boost the forecasting efficiency even more. In addition, a [12] new hybrid forecasting method is being developed and tested, which integrates well-known forecasting strategies [7]. The mixed model is created after no discrete method is better than all four seasons. According to RSME, the error of all simulations is 7.210% in summer, 6.921% in early autumn, 8.62% in spring and 9.37% in winter. The results show that a simple combination of many excellent models can provide more reliable predictions than either method. We've got been ready to cut back the error rate from concerning within the persistent ideal, to 15.1% from the Foreign Intelligence Service model, and to 11.8% from the cnn. Similarity features, which are effective in the SVR model, do not improve the deep learning model [9].

For example we take the north china place the winter days based on the climatic characteristics of North China. For each classification, a random forest algorithm-based forecasting model is created [10]. Separately, three other models in addition to the one proposed. Methods for predicting regular power generation and evaluate the performance of the Zhonghu PV station in northern China are described. The model achieves mean absolute percentage errors of less than 2.831 % and 3.1 percent in empirical data due to its capacity to by balancing decision trees, you can reduce the risk of overfitting. The following conclusions are based on empirical findings:
(1) The projected model's MAPE standards for both sunny and cloudy days square measure as low as 2.831 % and 3.890 %.
(2) When it's raining or snowing, the prediction errors of the projected model square measure are greater than they are in other circumstances. In those days, the prediction exactitude may increasing the size coaching samples and

playing subdivision will help. (3) Manual interference with neighbor classifications can increase the accuracy of prediction for sorting transformation days.

(4) In practically every classification, for nearly all error analysis metrics, the predicted model outperforms the other three strategies.

Both the artificial neural network and deep neural network models studied outperform the PPVFM in terms of efficiency [1]. The Auto-LSTM[11] is the best DNN model. With an average RMSE of 0.071, an average MAE of 0.036, and an average absolute deviation of 0:2765, it has the highest score. The greatest BIAS values are seen in the MLP. Using current air temperature and mean daily solar irradiance, they employed Multi-Layer Perceptron (MLP) to estimate the hourly forecast of solar radiation [13]. On unseen test data, the DBN achieves the best correlation, while the Auto-LSTM achieves the best correlation on training data.

Whereas the cloudy version suits the overall sample of the climate, it often predicts incorrectly. When the climate shifts, which occurs nearly each day, PPF's effects are inaccurate [2]. SVM-RBF with decreased features, on the alternative hand, produces a miles extra correct version. The RMSE mistakes for every version imply this: SVM-RBF with 4 dimensions has an RMS-Error of 128 amps, at the same time as hazy and PPF have RMS-Errors of a hundred seventy five and 261, respectively. As a consequence, the 4-dimensional SVM-RBF version is 27% extra correct than the easy cloudy version and 50.9% extra correct than the PPF version [3]. Using the fuzzy logic method, the average error was calculated is 1.92 %, while the average error obtained using the ANN method is 2.62 %.

Random forest is part of machine learning algorithm a random forest is an estimator of some sort, that can fit several decision trees for different sub-samples of a dataset [4]. To get the best algorithm parameters, the hyper parameters tuning in Random Forest Classification approach was utilized [15]. In our case, both the number of decision trees in a forest and the number of landscapes fitted by a single decision tree are hyper parameters. This algorithm is sufficiently accurate, with a state recognition accuracy of 73%.

The Kernel function is used in Support Vector Regressor (SVR) to map input forms into a higher-dimensional space where output patterns can be found and become linearly divisible, allowing pattern extraction by fitting linearly. The aim is to find the best possible match, which can be challenging [9]. Assumed the shape of an optimization problem. Similar weather conditions, according to KNN, may result in alike PV power generation. As a result, the historical dataset can be viewed as a series of cause-and-effect relationships. Predicting future Photovoltaic power generation can be reduced to finding K feature vectors in the past database that are the closest neighbors to those who were present at the time

interest. Knn has a lower percentage of error than nrsme, which is around 6.2 % and provides an approximate accuracy of 70% to 80%. A fixed input network with a single hidden layer is used as a back propagation neural network [14]. The input and output vectors' dimensions can be used to measure the neurons in the input and output layers automatically [1]. The initialization of weights and thresholds is performed at random, with an error rate of 8%.

To combine the outcomes of approximately the well-performing uncorrelated techniques into a weighted average as a single forecast the first step is to implement and evaluate all ten forecasting models. The hybrid model has the highest average performance of all the methods, with a yearly nRMSE of just 6.74 percent. CNNs have also been shown to improve efficiency while requiring less memory, all neurons in adjacent layers are connected in completely connected neural networks [8]. As a result, completely related neural networks have an enormous number of model parameters.

Long short term memory (LSTM) is a recurrent neural network (RNN) type that has successfully addressed the conventional RNN's vanishing gradient problem, which makes it difficult to detect and use useful features occurring early in input sequences. Here the combination of cnn+lstm gives MAPE 19.2%.The most basic methodology for estimating the next day's power production is calculated using linear or non-linear regression methods such as logistic regression or SVR and is dependent on inputs such as irradiation, temperature, and power output obtained the day before. Whereas svr is having a MAPE of 49.1%.
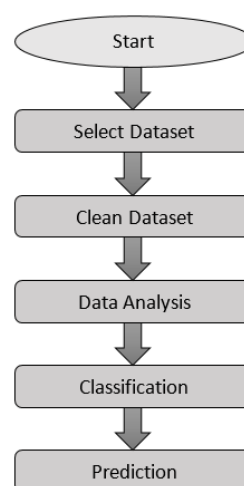
3. Flow chart for solar power generation

Fig.1 Flow chart

4. Methodology

4.1 Deep Learning Techniques

Consumption of power has become as additional and there are more regenerative generators included into the power grid, there has been an increase of interest in methods for regenerative power forecasting over the last decade [1].

4.1.1 Auto LSTM:

The Auto-LSTM algorithm blends an Auto Encoder's function learning with an LSTM's temporal context utilization. The Auto-error LSTM's increases as more time measures are considered during the forecast. It's worth noting that when two previous time phases are factored into the prediction, the RMSE jumps dramatically [1].CNN is a deep learning neural network used to process organized data matrices, such as views. Lines, gradients, circles, and even eyes and faces are very effective in recognizing the design in the input image [5].
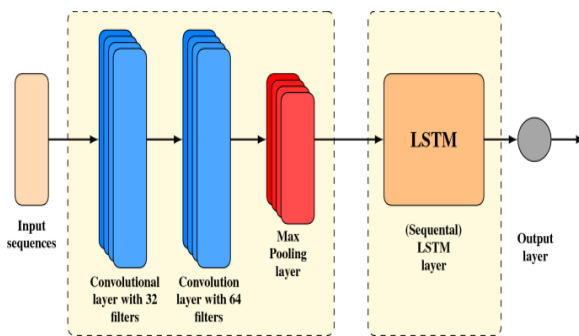
Fig.2 CNN and LSTM architecture

4.1.2 Deep Neural Networks:

 DNN has proven its worth in a variety of fields, including representation learning and time series forecasting. To name a few, DNN claims Speech and picture recognition, machine translation, and financial time series forecasting are just a few of the applications. It is used to forecast solar power output [11].
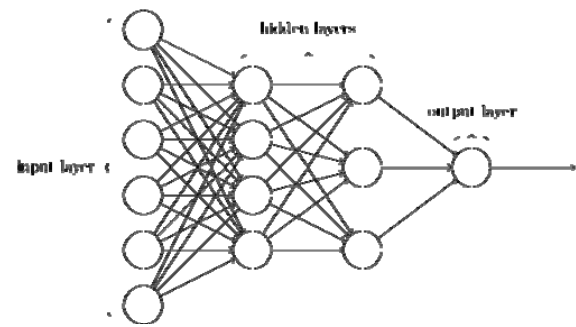
Fig.3 Layers differentiation of DNN

4.1.3 Deep Belief Network (DBN):

Deep belief network (DBN) models began to gain popularity. To forecast monthly solar power output data, a seasonal deep belief network (SDBN) was created in this study. By merging the seasonal decomposition approach and the DBN, the SDBN was created. Furthermore, data on monthly solar power generation was employed in this study.

Foretelling is a two-step approach [1]:

1) To reduce the dimensionality of the input data set, the DBN uses feature learning.

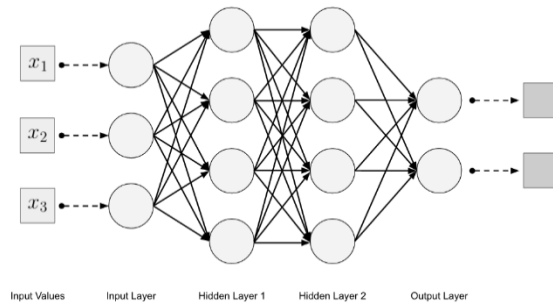2) To perform forecasting, a new layer is added, such as the linear layer.

Fig.4 Deep belief network model

4.2 Machine Learning Techniques

4.2.1 Support Vectors Machine:

Then, we investigate a variety of supervised learning methods based on SVM. These methods, which create hyperplanes in a multidimensional space, have currently been approved to be used in indexing and regression analysis. The kernel function and parameters used determine the accuracy of svm regression. In our work, we investigated three distinct SVM kernel functions are Linear Kernel, a Polynomial Kernel and a Radial Basis Function (RBF) kernel [2].
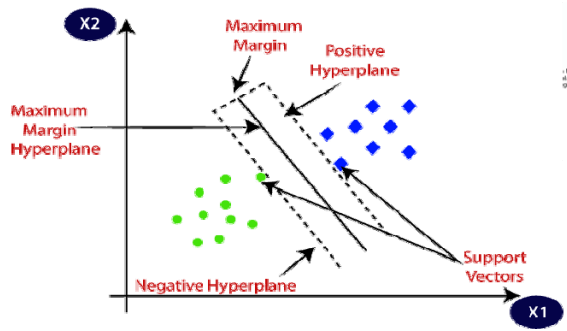
Fig.5 Hyperplane separating data points in SVM

4.2.2 K Nearest Neighbors (Knn):

KNN assumes that similar temperature variations will occur in the future. Would result in PV energy clusters that are comparable. Forecasting future PV power generation with an accuracy of about 81 % can be reduced to searching the past database for K feature vectors that are closest neighbors to those of the time of interest.
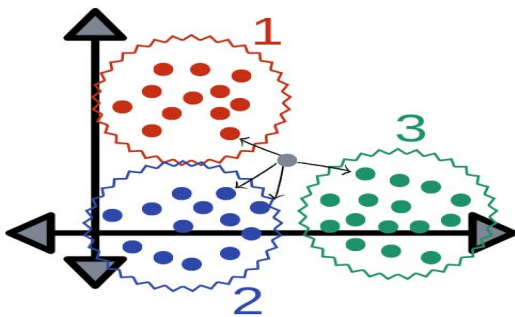

Fig.6 Partition and finding the nearest value in KNN

4.2.3 Random Forest:

Random Forest is an ensemble technique that can be used to classify or predict data productivities of a variety of unrelated decision tree subsystems. The random forest estimates that the precision is about 79%. These techniques can be used to map nonlinear interactions again [19].
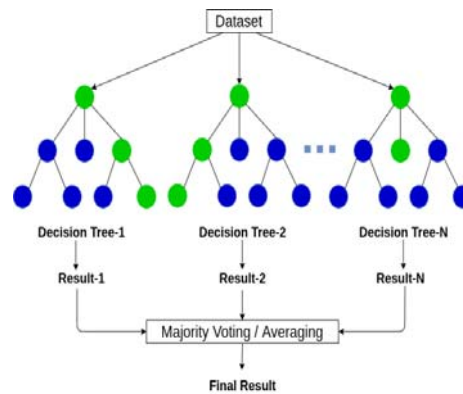

Fig.7 Random forest final results based on decision trees

4.2.4 Linear Regression Model:

Linear regression is a machine learning approach for supervised learning. It performs a regressive mission. Using the independent variables as a guide, regression models attempt to predict the value. Its main use is for forecasting and determining the relationship between variables.
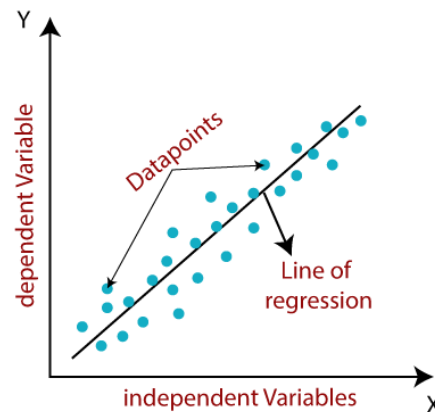

Fig.8 Linear regression model

4.3 Artificial Neural Networks

MATLAB was used to carry out the research. Fuzzy logic and Ann techniques[18] were both found to be effective in that their findings matched the characteristic of a real solar PV array [3]. These will be put to use to forecast and test the overall reliability of the solar power plant. They perform on three sets of data: training, testing, and validation. The MLR model's performance is substantially improved when classification variables and interactions between variables are used, but this is not the case with the ANN model [7].

4.3.1 Fuzzy Logic:

When combined with high-performance computer processors, allows for reasonable accuracy estimates of solar plant outputs while also allowing for system flexibility to adjust for natural events. The MATLAB software was used to construct the load forecasting system, which is based on fuzzy logic. The model has two inputs and one output and represents solar irradiance, ambient temperature, humidity and output power [3].
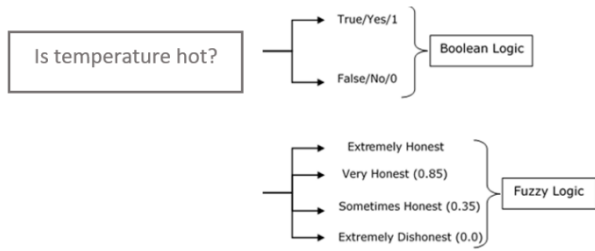

Fig.9 Fuzzy logic model

4.3.2 Multiple Linear Regression:

An ANN[16], as opposed to a multiple linear regression (MLR) model, a statistical method for assessing a non-linear connection between one or more inputs and outcomes. The artificial neural network[6] has been used to model, identify, and forecast complex systems.
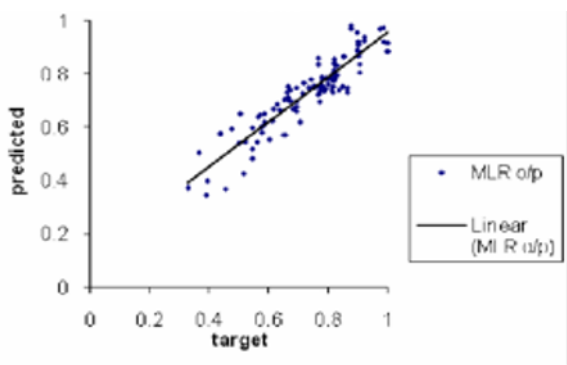

Fig.10 Multi linear regression

5. Comparison

5.1 Comparison of methods used:

Table 1. Model Prediction of RMSE, MAE, MAPE, and MBE

Model	RMSE	MAE	MAPE	MBE	Reference
MLP(mul tilayer perceptron)	0.0889	-	-	-	1
LSTM	0.0816	-	-	-	1
Clustering based on solar generation	8.80	4.60		-1.01	5
Clustering based on weather variables	8.61	3.77	-	-0.53	5
ANN	0.067	-	-	--	6
MLR(multi linear regression)	0.738	-	-	-	6
SVR(support vector regressor)	0.122	0.099	40.03	-	8
RFR(random forest regressor)	0.178	0.1378	67.58	-	8
CNN+LSTM	0.098	0.0568	13.42	-	8
SP(SMART PERSISTNACE)	0.135	0.071		0.023	11
DBN(deep belief network)	0.0390	0.0138		0.00244	11
SVM(support vector machine)	0.034	0.012		0.00777	11
RF(Random Forest)	0.033	0.0099		0.000351	11

5.2 Graphs comparison based on weather conditions with temperature values:
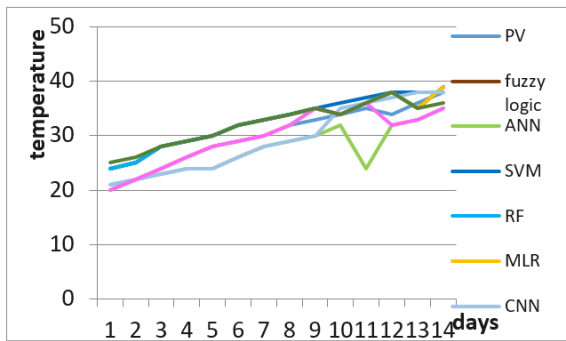

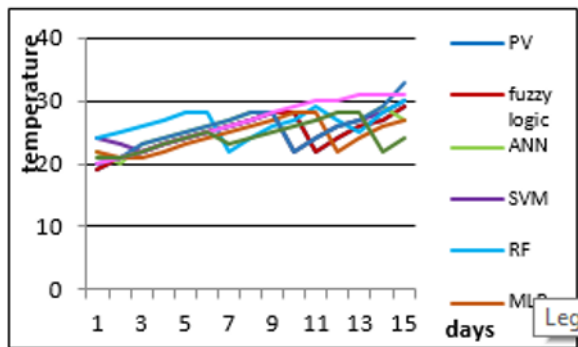**Fig.11** Mean of solar radiance in summer season.

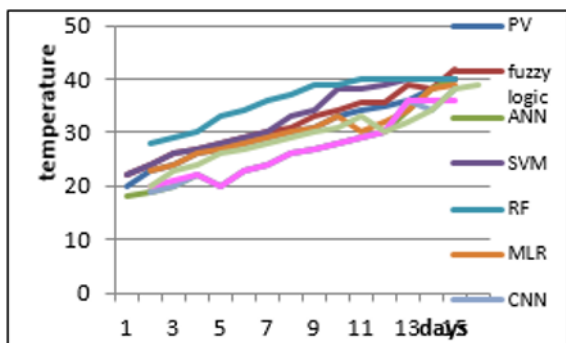**Fig.12** Mean of solar radiance in rainy season

Fig.13 Mean of solar radiance in winter season

Here fig 11, 12, 13 gives the comparative analysis of different models that are used like fuzzy logic, artificial neural network, support vector machine, random forest, multiple linear regressor and Convolutional Neural Network results for different temperature based on seasonal effects with power generated for each day.

7. Conclusion

In this we conclude that it gives the partial analysis of solar power generation or forecasting based on different weather conditions. Different types of solar and wind dataset are considered for different seasons like summer (April to June), rainy (July to November) and winter (December to January) these are all about Indian weather condition. In order to get best results for diverse climatic condition for every hour for future betterment of analyzing power generation in different climatic condition machine learning give better results according to deep learning and artificial intelligence.

Reference

1. A. Gensler, J. Henze, and B. Sick, 2016 IEEE International (2016)
2. N. Sharma, P. Sharma, D. Irwin, and P. Shenoy, in 2011 IEEE International Conference on Smart Grid Communications (SmartGridComm), pp. 528–533, (2011)
3. Z. P. Ncane and A. K. Saha, in 2019 Southern African Universities Power Engineering Conference/Robotics and Mechatronics/Pattern Recognition Association of South Africa (SAUPEC/RobMech/PRASA), pp. 518–523,(2019)
4. A. Khalyasmaa, S. A. Eroshenko, T. P. Chakravarthy, V. G. Gasi, S. K. Y. Bollu, R. Caire, S. K. R. Atluri, and S. Karrolla, in 2019 International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON), pp. 0780–0785, (2019)
5. C. Pan and J. Tan, IEEE Access 7, 112921 (2019).
6. P. Kora and S. R. Kalva, Springer plus 4, 481 (2015)
7. M. Abuella and B. Chowdhury, in 2015 North American Power Symposium (NAPS), pp. 1–5, (2015)
8. K. Prasanna Lakshmi and C. R. K. Reddy, in 2010 International Conference on Networking and Information Technology, pp. 451–455(2010)
9. D. Su, E. Batzelis, and B. Pal, in 2019 International Conference on Smart Energy Systems and Technologies (SEST), pp. 1–6, (2019)
10. Swaraja K, Multimed. Tools Appl. 77, 28249 (2018)
11. W. Lee, K. Kim, J. Park, J. Kim, and Y. Kim, IEEE Access 6, 73068 (2018)
12. S. Kumar, P. Reddy, G. Ramesh, and V. Maddumala, Trait. Du Signal 36, 233 (2019)
13. F. Jawaid and K. NazirJunejo, in 2016 Sixth International Conference on Innovative Computing Technology (INTECH), pp. 355–360, (2016)

14. C. U. Kumari, S. Jeevan Prasad, and G. Mounika, 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC) (2019)

15. X. Wang, D. Luo, and C. Li, in 2019 2nd International Conference on Artificial Intelligence and Big Data (ICAIBD) , pp. 97–101, (2019)

16. B. Dhanalaxmi, G. A. Naidu, and K. Anuradha, Procedia Comput. Sci. 46, 432 (2015)

17. Application of Machine Learning Algorithms for Solar Power Forecasting in Sri Lanka

18. P. Nayak, G. K. Swetha, S. Gupta, and K. Madhavi, Measurement 178, 108974 (2021)

19. M. Z. Hassan, M. E. K. Ali, A. B. M. S. Ali, and J. Kumar, in 2017 4th Asia-Pacific World Congress on Computer Science and Engineering (APWC on CSE) , pp. 252–258, (2017)