10.48047/jocaaa.2020.28.04.2

Evaluation of Solar Power Generation, and Prediction Using Machine Learning

[1*]Kamlesh Gupta, [2]Firoz Mohammed, [3]Namrata Nebhnani, [4]Priya Pal

[1*]Assistant Professor, Department of Electrical & Electronics Engineering, IPS Academy, Institute of Engineering and Science, Indore
[2]Assistant Professor, Department of Electrical & Electronics Engineering, IPS Academy, Institute of Engineering and Science, Indore
[3]Assistant Professor, Department of Electrical & Electronics Engineering, IPS Academy, Institute of Engineering and Science, Indore
[4]Assistant Professor, Department of Electrical & Electronics Engineering, IPS Academy, Institute of Engineering and Science, Indore

Abstract: -Solar photovoltaic (PV)Power system backbone of the renewable energy system. Energy system is depended on weather conditions such as temperature and radiation intensity. The role of machine learning (ML) for solar energy generation and radiation forecasting. This paper presents ML algorithm or methods review for prediction of solar energy generation and radiation. This paper also presents the state of art on different ML methods and parameters for forecasting solar energy production and radiation.

Keywords: machine learning, photovoltaic generation, forecasting, neural networks, regression

I. INTRODUCTION

In India energy demand is increasing day to day. The present energy generation is depended on the fossil fuels. Deposit fuels generation about 70% in India. Fossil fuse are finite resources and produce harmful gases. these harmful gases have increased the health risk as well as climate change. So alternate energy source is required like renewable energy sources (RES). The main advantage of RES are infinite source of energy, pollution free, easily available, free available source of energy [1-2]. Energy is the future of the energy in terms of electricity generation. RES based energy system fulfil the electricity demand and reduces the dependency of the fossil fuels. The main problem of RES, generation and weather condition is unpredictable. The forecasting technologies are adapted for efficient functioning of the RES system. Solar energy is the most prominent energy source of energy compared to wind, biomass, ocean etc.ML methods are help for prediction of the PV energy system generation and weather forecasting [3-4].

In the field of computer science known as machine learning, it is possible to train computers to act naturally by giving the m experience. This indicates that information is being run int o data by the computer system. Without using a model, extra ct information from the data. Machine learning use algorith ms to identify patterns in data, then applies the model that do es so to fresh data [5–6].
The forecasting of energy production, sun radiation, wind sp eed, and other things is made possible by machine learning a nd its prediction algorithm. Numerous papers about the use o f ml in the field of PV energy systems have been published i n the preceding year [7-8].
This paper presents review of recent advances in ML for PV energy system, provide state of art of current algorithm to academic and researches.
The main contributions of this paper are:

- Identify the ML technique for forecasting the PV energy system generation.
- Identify the ML technique for forecasting for solar radiations.
- Review different types of ML methods.

II. MACHINE LEARNING

ML lies at the intersection of computer science, engineering and statistics and often appears in other discipline as shown in Fig. 1. It can be applied to many field e.g. Marketing, transportation, politics, banking sector, RES etc. Machine learning uses algorithms to find patterns in data and then uses in order that recognizes those patterns to make prediction on new data. ML can improve the accuracy and reliability of the different field [9-10].

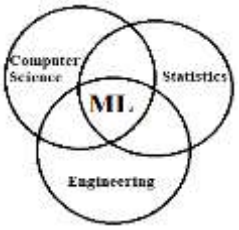
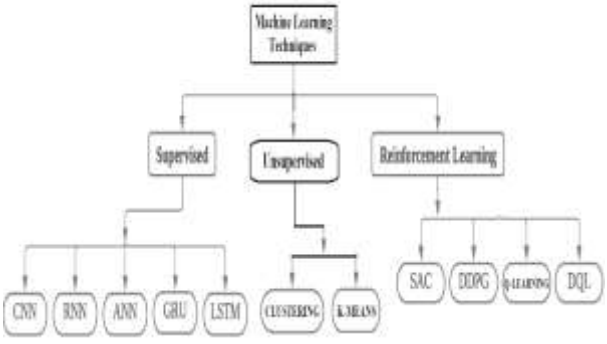


Fig. 1. Intersection of ML [9]

III. TYPES OF MACHINE LEARNING

The three machine learning techniques that are most frequently used are reinforcement learning, unsupervised learning, and supervised learning. Give a learning buddies algorithm based on labelled example input and output data. Unsupervised learning gives Gowtham no level data so that it may discover structure and patterns in the input data [10–12].



10.48047/jocaaa.2020.28.04.2

Fig. 2. Some of ML techniques [11-12]

3.1Supervised learning

In supervised learning, sample input data are given to the computer and are labelled with the intended outputs. The goal of this technique is to enable the algorithm to learn by comparing its real output with the taught outputs in order to identify faults and adjust the motor as necessary. There are four ways to secure the labour values on extra unlabeled data using supervised learning [13].

Classification or regression is a fundamental component of all supervised learning approaches. Regulation approaches forecast continuous responses and classification techniques modify discrete answers.

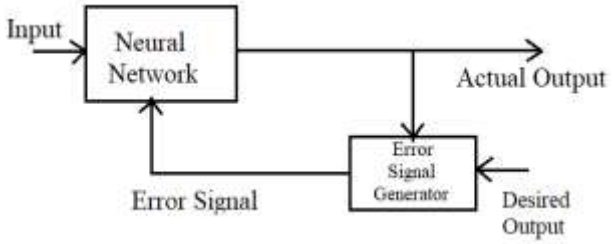


Fig.3 supervised learning [14]

The important supervised learning algorithms are artificial neural network (ANN), linear regression (LR), logistic regression, decision tree, support vector machine (SVM), bayesian classification, hidden Markov model etc.

A) Artificial neural network

ANN ability to perform computations at home that can replicate some of the flexibility and power of the human brain by artificial means. It attempts to mimic the structure and function of our nervous system. ANN is used as an information processing methodology and its methods are derived from the biological nervous system. The system consists of in numerable highly interconnected neurons working together to solve different kind of problems. Learning is biological systems takes place why adjusting is synaptic condition that exist between the neurons [15-16].

B) Linear regression

Supervised learning is learning in which a supervisor teaches a machine using level data with the correct answer. The machine will then receive a new data set so that the supervised learning algorithm can generate the correct results from the labeled data using the roll example. Regression analysis for unsupervised machine learning. The most important features of regression are: The answers obtained from the model are always quantitative in nature and the model can only be built if previous data are available [16-17]. The rule has an equation of the form [17-18]

$x = c + (d \times y)$ (1)

Where, yand x aredescriptiveand dependent variable respectively, d is slope.

C) Logistic regression

Logistic regression is a binary based method (0,1). Therefore, the output $Z_j$is the implemented random variable, Z_j can take the values 1 and 0 with probabilities$p_j$and $1 - p_j$, respectively [19]. Let p be the probability that $Z = 1$when$X = x$. if we use the standard linear model to describe p, as shown in equation (2),

the probability model will be [19-20],

$P = P(Z = 1|X = x) = \beta_o + \beta_1 x$(2)

D) Decision tree

Decision tree learning is a method of approximating a discrete-valued objective function, where the dealer's function is represented by a decision tree. The learning tree can also be represented as a set of if-then rules to improve human readability. These learning methods are among the most popular inductive inference algorithms and have been successfully applied to a wide range of tasks from learning how to diagnose medical cases to learning how to accept applications. renewable energy [21]. The classifier decides on the cases by arranging them in the rule tree into a note that provides the instance classification. Each node of the assignment tree checks one property of the instance, and each descending branch from this node corresponds to one of the possible values for the attribute t. File classification instances start with the root node of the tree, the test and attribute specified by this node then go down the branch of the tree corresponding to the value of the attitude [22].

E) Bayesian learning

Bayesian learning provides a probabilistic approach to reasoning. It is based on the assumptions that the quantity of interest is governed by a probability distribution and that optimal decisions can be made by the area around these probabilities with observed data. It is important for machine learning because it provides a quantitative approach to evaluating evidence that supports an alternative hypothesis. Bayesian inference provides a basis for learning algorithms that directly manipulate probability as well as a framework for analyzing the behavior of other algorithms that do not explicitly manipulate probability [23].

3.2 Unsupervised learning

In supervised learning, the data is not labeled, so the learning algorithm must find common ground among its inputs. Since unlabeled data is more abundant than labeled data, machine learning methods that support and monitor learning are especially valuable[24]. Most unsupervised learning techniques are a form of cluster analysis. Clusters are formed so that objects within the same cluster are very similar and objects from different Valley restaurants are distinct. Clustering algorithms are divided into two broad groups a) hard clustering where each data point belongs to a single

10.48047/jocaaa.2020.28.04.2

cluster and b) soft clustering where a data point can belong to more than one cluster. Hard and soft clustering techniques can be used if you already know how to cluster possible data.

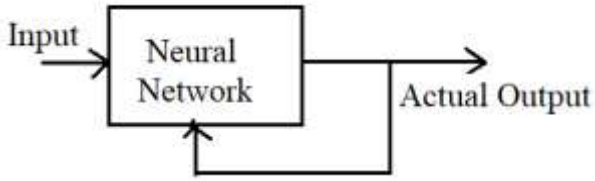


Fig. 4.unsupervised learning [25]

A) Clustering

Data clustering is a process of grouping the objects into a cluster so that objects within the cluster are closer in meaning than objects outside the cluster. Clustering is sometimes referred to as: data, information, or knowledge combination; "outlier" elimination; and "data cleaning. The similarities are excess please on the attribute that describe the objects. Similarities is measured by distance matrix. Partitioning of clusters is not done by humans. It is done with the help of algorithms. Flowers to drive some useful information from the data which was previously unknown. Clustering is also called data segmentation because it partitions suggestions large data set into groups according to their similarities [26].

Clustering is known as unsupervised learning because the class level information is not present. You have already seen in supervised learning algorithms that every input has a corresponding output which helps in designing in model[27].

3.3 Reinforcement learning

Learning from examples is like supervised learning in that desirable information is available. The important difference is that in the case of reinforcement learning, only critical information is available. The process of extracting real information from critical information is term as reinforcement learning as shown in figure 5.

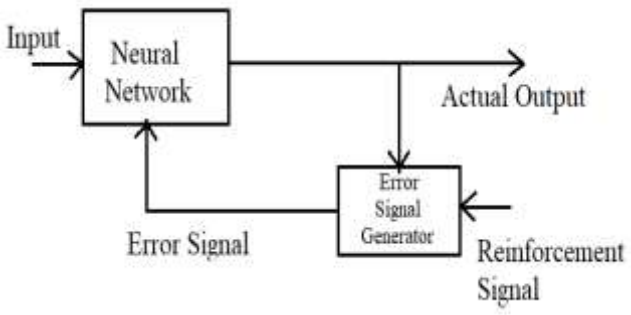


Fig. 5.Reinforcement learning [28]

IV. ML FOR WEATHER FORECASTING OF PV SYSTEM

A major prerequisite for ensuring the stability and dependability of the grid is precise forecasting of PV production. With a primary focus on the metaheuristic and machine learning methodologies, Akhter [29] offers a thorough and critical analysis of the techniques used to anticipate PV power output.

Tina et al. [30] reviewed ML models for solar PV energy demand forecast for predicting solar irradiance, power generation, electricity prices, and energy demand. PV modules, extraction of PV design parameters, monitoring the maximum power, and optimization of PV systems efficiency are further ML applications in PV industry that are taken into consideration. Mainly neural networks techniques are used for forecasting the solar irradiation. The long short-term memory (LSTM) and gated recurrent units (GRU) networks model designed for forecasting. Results shows that ML is better than conventional methods.

ML is more reliable and accurate method for forecasting renewable energy selection. ML method was calculated accurate prediction based on mean square error(MSE), root mean square error (RMSE) and mean absolute error (MAE) parameters [31]. Villegas-Mier et al. [32] presented Queretaro, Mexico areas solar insolation level prediction using Random Forest (RF) algorithm. The results shows that RF based prediction accuracy is high as compared convention ML methods. Solano et al. [33] compared ML methods e.g.support vector regression (SVR), extreme gradient boosting (XGBT) and voting-average (VOA) for prediction of solar radiation in Brazil. Results shows VOA gives better prediction compared to others.

Table-1: Summary of solar radiation prediction using ML

Input variables	methods	Forecast duration	Parameters	outcomes	Ref
Solar irradiation, wind flow, humidity, temperature	ANN	90 Days	MSE and MAE	Accuracy improved = 96%	*Akhter et al. [29]*
GHI,DNI, wind seed	Deep learning	175 hours	MSE, RMSE, MAE	ML result is 38% better than traditional methods	Narvaez [34]
irradiation, wind speed, temperature,	SVM, LR, ANN	24 hours	MSE, RMSE, MAE	forecasting solar irradiance ML is reliable	Toubeau et al.[31]
GHI, Humidity, Temperature, wind speed	Feature selection, regression, numerical models	80 Days	RMSE, MAE, R-squared	Improves prediction of GHI values	Hedar et [35
Solar irradiation	RF, recurrent neural networks (RNN)	90 days	MSE, RMSE, MAE	Accuracy improved = 95.98%	Villegas-Mier et al. [32]

10.48047/jocaaa.2020.28.04.2

Solar radiation	SVR, VOA,XGBT	31 Days	MSE, RMSE, MAE, R^2	VOA accuracy is high	Solano et al. [33]

Table-2 shows summary of solar radiation forecasting using ML techniques. The numbers of different parameters have been taken for review of ML application in solar radiation prediction. Table-2 shows that ML techniques gives better results in terms of solar irradiation prediction.

V. ML FOR PV OUTPUT POWER ASSESSMENT

There are many more forecasting methods, including hybrid models and artificial intelligence systems. Pre-processing techniques are used by machine learning algorithms to precisely forecast the yields of renewable energy sources [36]. Mittal et al. [37] studies forecasting of PV output power using ML methods e.g. bagging model, deep learning, genetic algorithm (GA), RF, GBT and ANN. The presentation of numerous ML methods for estimating PV systems output energy is being evaluated. Range of techniques were investigated, including regression trees (RT), SVR, and ANN, using a variety of hyper-parameters and features. On real PV production data sets collected over the course of a year, the output power forecast performance of each model was evaluated and contrasted with an existing persistence model (PM) [38].Ramadhan et al. [39] compared physical model with ML model in terms of solar irradiation, modeling, and PV generator power. ML method reduce the mean bias difference (MBD) than the root mean square difference (RMSD), to determine if a 1.50kW PV system installed at the Nairobi in Kenya can accurately forecast the PV power of the system in real time, SVR and RF models were developed. With a MAE (32.57) and RMSE (43.16) adjusted R^2, the SVR model surpasses the random forest regression (RFR) model in the validation. The dataset for the RFR model has an RMSE of 85, an adjusted R^2 of 0.901, and an MAE of 69.0 [40].

Table-2:Summary of PV generator power prediction using ML

Input variables	methods	Parameters	Forecast duration	outcomes	Location	Samples	Ref
PV power, loss	deep RNN	RMSE, MAE, R^2	few days	measured forecasted power and loss	Puglia (Italy)	7320 samples	Mellit and Kalogirou [23]
PV power	ANNs, SVR and RT		50 days	Lowest RMSE			Theocharides et al. [38]
PVpower and irradiation	SVR model and RNN, k model			SAM =3.06 kWh, SVR =3.75k Wh, RNN =3.52k Wh	Kookmin University	887, 336 data sets	Ramadhan et al. [39]
PV power	RFR, SVR	RMSE, MAE, R^2	21 Days	RMSE=86, MAE=69, R^2=0.9	Kenya		Mwende et al. [40]
Photovoltaic Power	SVR	hyper-parameter	90 Days		Italy	1376	Polo [17]

Table-2 shows summary of PV generation forecasting using ML techniques. The numbers of different parameters have been taken for review of ML application in solar power generation prediction. Table-2 shows that the ML techniques give better results in terms of solar power generation prediction.

VI. CASE STUDY

This work is used location latitude (24.47755º) and longitude (77.75465º) for prediction of irradiation and PV power generation. The prediction of irradiation result is tabulated in Table-3 using RFR, SVM, LR, R^2. The prediction of PV power result is tabulated in Table-4 using RFR, SVM, LR, R^2.

Table-3: prediction of irradiation using ML algorithm

Algorithm	MAE (w/m^2)	MSE (w/m^2)	R^2
RFR	40.52	135.87	0.95
SVM	50.12	136.84	0.75
LR	45.21	133.54	0.70
ANN	48.36	131.58	0.74

Table-4: Prediction of solar PV generation using ML algorithm

Algorithm	MAE (kWh)	MSE (kWh)	R^2
RFR	0.57	2.25	0.65
SVM	0.56	2.89	0.61
LR	0.58	2.46	0.56
ANN	0.60	2.68	0.59

Results shows that ML algorithm results are more accurate. The present work is compared with reference [41] shown in Table-5. The present work explains ANN, LR, SVM and RFR method for forecasting the PV performance. The performance parameters are MSE, MAE and R^2 taken in this work.

Table-5: Compare present work to Widodo et al [41].

Parameters	Widodo et al [41].	Present work
Aim	Renewable energy power generation using Deep Neural Network (DNN).	Solar PV energy generation and weather prediction using ML
Location	Florida	India
Algorithm	DNN	ANN, LR, SVM, RFR
Measurement	power generation	Power generation and irradiation
Performance parameters	MAE, MSE	MSE, MAE, R^2

10.48047/jocaaa.2020.28.04.2

Future aspects	No	yes
Outcomes	DNN gives better results	ML algorithm gives the reliable and precise results.

VII. CONCLUSION AND FUTURE WORK

This paper presents literature review of advance a method on prediction of solar radiation and PV energy system generation have been carried out. The area of predicting PV generation has, by far, received the most research attention, and numerous ML-based models have been put forth in this area. Historical and real time weather data are used for prediction the weather condition using ML algorithm.ML methods (ANN, SVM, regression) were used for forecasting PV energy production. The main parameters are RMSE, MAE and R^2. ML methods (ANN, SVM, RN, feature selection) were used for forecasting solar radiation. The main prediction parameters are RMSE, MAE and MSE. The prediction reliability and accuracy are dependent on having historical data of output PV energy generation. The main advantage of forecasting the energy generation information, system operator can control the energy flow and cut the operational cost also. Prediction of solar irradiance and PV generation is reliable and precise using ML algorithm on given location.

In future work, ML helps to improve the efficiency of the solar energy-based energy system. To enhance performance, accuracy and reliability using deep learning methods. ML also help to controlling and smart infrastructure of the solar PV energy system. In future, ML also help in generation, transmission, distribution and reducing transmission losses.

REFERENCES

[1.] S. S. Raghuwanshi and Vikas Khare, "Sizing and modelling of stand-alone photovoltaic water pumping system for irrigation", Energy & Environment (Sage Pub), Vol. 29(4) 473–491, 2018. DOI: 10.1177/0958305X17752739.

[2.] L. F. Jim, D. A. Hern, and R. Espinosa, "Renewable Energy Prediction through Machine Learning Algorithms," 2020.

[3.] S. S. Raghuwanshi, R. Arya, "Economic and reliability evaluation of hybrid photovoltaic energy systems for rural electrification", International Journal of Renewable Energy Research, Vol. 9, No. 1, PP.-515-524, March, 2019.

[4.] I. Antonopoulos et al., "Artificial intelligence and machine learning approaches to energy demand-side response : A systematic review," Renew. Sustain. Energy Rev., vol. 130, no. February, p. 109899, 2020, doi: 10.1016/j.rser.2020.109899.

[5.] H. Demolli, A. Sakir, A. Ecemis, and M. Gokcek, "Wind power forecasting based on daily wind speed data using machine learning algorithms," Energy Convers. Manag., vol. 198, no. July, p. 111823, 2019, doi: 10.1016/j.enconman.2019.111823.

[6.] K. Shivam, J. Tzou, and S. Wu, "A multi-objective predictive energy management strategy for residential grid-connected PV-battery hybrid systems based on machine learning technique," Energy Convers. Manag., vol. 237, no. 195, p. 114103, 2021, doi: 10.1016/j.enconman.2021.114103.

[7.] S. S. Raghuwanshi, R. Arya, Reliability evaluation of stand-alone hybrid photovoltaic energy system for rural healthcare center, Sustainable Energy Technologies and Assessments, Volume 37, February 2020, 100624.

[8.] S. Phommixay, M. L. Doumbia, and D. Lupien St-Pierre, "Review on the cost optimization of microgrids via particle swarm optimization," Int. J. Energy Environ. Eng., vol. 11, no. 1, pp. 73–89, 2020, doi: 10.1007/s40095-019-00332-1.

[9.] M. N. Akhter, S. Mekhilef, H. Mokhlis, and N. M. Shah, "Review on forecasting of photovoltaic power generation based on machine learning and metaheuristic techniques," IET Renew. Power Gener., vol. 13, no. 7, pp. 1009–1023, 2019, doi: 10.1049/iet-rpg.2018.5649.

[10.] D A Widodo et al, "Renewable energy power generation forecasting using deep learning method", Conference IOP Conf. Series: Earth and Environmental Science 700, 012026IOP, 20121, doi:10.1088/1755-1315/700/1/0120262.