

A Review of Machine Learning Models Developed for Solar Radiation Forecasting for Indian Regions

Yogita Ajgar¹ and Dr.N.R.Krishnamoorthy²

¹Department of Electronics, Sathyabhama Institute of science and technology, Chennai, India
Email: yogitaajgar242@gmail.com

²Department of Electronics and communication, Sathyabhama Institute of science and technology, Chennai, India
Email: moorthy26.82@gmail.com

Abstract—These Ever after since the development in technology, many applications are evolved which are based on solar energy utilization and importance of Global solar radiation (GSR) information is consistently increasing. However, due to the changes in solar-radiation and lac of measurement infrastructure, availability of accurate information on GSR data is poor. To deal with this employment of AI based solution is considered. Many machine learning models are developed for forecasting the GSR by virtue of its capacity to handle large and complex nonlinearly correlated data and derive meaningful information from it. This derived information is further utilized for forecasting GSR. Thus, for India, being a country of many climatologically and geographically different zones, approach of employing AI models in GSR prediction becomes very beneficial. This manuscript presents the various machine learning model approaches for GSR forecasting. Recent developments in this view are comprehensively and systematically reviewed considering the aspects like location, selection of input parameters, feature selection and model development. Current situation and future scope of research related to GSR forecasting accuracy improvement across the Indian regions is also discussed.

Index Terms— Solar radiance, Machine learning models, Global Solar radiation.

I. INTRODUCTION

All modern technology and latest developments rely on interrelationship of machineries and nearly impossible to envisage without electricity [1]. Even in today's modern era this electricity is almost entirely generated using fossil fuels and fast growth in technology sectors predicts that the primary electricity demand in the world will continue to rise over the course of time. Therefore, in order to reduce the destructive impact of traditional energy sources (fossil fuel), renewable energy (RE) sources are purposely considered to fulfil this requirement in place of traditional energy sources [2]. In recent decades, significant growth in RE sources development is achieved and by 2050 RE sources are predicted to contribute more than 40 % share in global electricity generation [3]. In particular developing countries lots of research and development is initiate to make RE sources continuously available and accessible. This is due to advancement in technology implementation and globalization in many developing countries, which escalated electricity demand [4]. Therefore, many initiatives are taken imperatively to encourage RE resources development. Amongst many RE sources solar energy is very much abundant, clean and its impact on the nearby environment is very less [5]. Similarly, influence of solar radiation (SR) is also observed on agricultural yield, atmospheric circulation, hydrological processes, public health and ecological

services which highlights the critical need of exploring comprehensive knowledge of SR at that site for environmental sustainability and economic potential [6]. Furthermore, knowledge of GSR is significantly important in many application fields such as meteorology, hydrology, and renewable energy usage [7]. In recent years, due to the need of RE sources as a replacement of fossil fuels, SR becomes one of the most significant variable utilized to predict RE potential at any site. India is rewarded with abundant SR due to its specific geographical location on the globe. India is located between the equatorial and tropic of cancer in the northern hemisphere because of which 2300 to 3200 hours of sunlight is received in most of the Indian regions per year. This is capable of generating nearly 5.5 kWh/m²/day solar radiance [8]. Primarily, solar energy is used to produce electricity.

As on March 2023, India installed many electricity generation plants which are capable of producing 415.4 GW electricity, wherein 236.68 GW is contributed from thermal, 6.78 GW is produced using nuclear, and 171.8 GW is generated from renewable energy sources (<https://powermin.gov.in/en/content/power-sector-glance-all-india>), Central Electricity Authority, Ministry of Power, Government of India.). Astonishingly as on 31 May 2023, RE Sources contribute to produce 126.76904 GW electricity, of which 4.94430 GW is produced using small hydro power stations, 43.19898 GW is generated using wind power, while 10.80401 GW electricity of generated using biomass, and solar energy contributed the highest as 67.82122 GW of electricity production capacity (Installed Capacity Report, Central Electricity Authority, Ministry of Power, Government of India. (<https://cea.nic.in/installed-capacity-report/?lang=en>). Moreover, process is undertaken to install electricity generation power plants for another 82.62 GW of green energy capacity. By 2030, the aim of the Indian Government is to install non-fossil sources-based electricity generation power plants which will be capable of producing 500 GW electricity. With this excellent implementation and vision, India became the fourth globally in RE installed capacity and witnessed highest year on year growth of 9.83% in 2022 in RE source-based electricity production.

Even with this excellent growth, yet India uses thermal energy to produce more than 50 % of its electricity. Thermal energy-based electricity plants are major emitters of CO₂, due to which India ranked as third largest emitter of CO₂ [9], [10]. India is now keen towards changing this situation and therefore target is set to reduce the carbon intensity of the Indian economy by less than 45 % by the end of this decade, to install RE sources based electric power generation plants for contributing more than 50% of cumulative electric power by 2030, and by 2070 attain net-zero CO₂ emission.

To work out this plan solar energy is important. Furthermore, solar energy can be utilised for variety of applications. To explore the utility of solar energy its availability information is important which helps to plan out the necessary utilization action. However, ground reality is only 43 metrological stations are available to measure radiation data which is very less as compare to the land available in India. Therefore, to estimate solar energy output in terms of electricity variety of prediction models are designed to forecast global solar radiance.

Researchers are continuously putting their efforts in making these models predictive with high accuracy. Since no single model could be applied to all over the places, according to region and data availability many models are proposed for better and better projected accuracy. Due to this it is challenging to opt for best suitable model for targeted application and location [11].

In this review we intend to analyse GSR prediction models developed for India throughout the various regions. Specific focus is given for AI (machine learning and deep learning) models for GSR modelling or prediction.

A. Motivation

In India, many observatories are established to collect the data of meteorological parameters such as relative humidity, temperature, and sunshine duration. On the other hand, less number of the observatories are available for SR measurement. This is because cost associated with the infrastructure development and measuring equipment itself is very high, along with maintenance cost, the data recording becomes expensive. Therefore, research community is putting efforts towards developing techniques by virtue of which use of available recorded data can be extended to forecast solar radiance for the location where previous measured data is available and also for the location where measured data is not available.

Typically, models developed for predicting SR are categorise according to the forecasting methods such as statistical, empirical, physical, and machine learning models [12], [13]. For exploring and establishing interrelation between SR and other meteorological parameters, physical models are developed along with sky-image-based models [11]. Even after having strong physical basis, complexity in atmospheric conditions enhances structural complexity involved in physical model [14].

Most commonly used empirical models [15] are developed using linear or nonlinear regression-equation [16]. Empirical model is operationally simple and easy but offer limited accuracy of forecast [17]. Further, statistical

correlation strategy is used to build autoregressive moving-average model (ARIMA) [18], [19]. Even though as compared to empirical models, statistical models demonstrated higher accuracy, nonlinear relation of SR and other parameters is not well explained in statistical models.

After the significant development in AI, its application for developing solar radiance prediction models is also explored. The machine learning methods using backpropagation and parameter update techniques can extract meaningful information and therefore shows capability of simulating complex and nonlinear interrelationship mapping task [20], [21]. These capabilities make machine learning models more desirable for forecasting model development. When compared with other models, it is also observed from literature that, prediction or estimation of SR using AI models offer wide range of applications with enhanced accuracy of prediction. Using variety of machine learning models, accurate modelling of SR forecasting for long-term, medium-term, and short-term prediction can be achieved.

In case of SR dataset, most of input parameters show nonlinear and complex relationship and feature extraction from these dataset is critical. However, with the aid of machine learning tools simulating nonlinear and complex relationship mapping in the input dataset is possible, thereafter extracting the meaningful information, models can be designed to learn and forecast. Using these techniques, accurate modelling of SR prediction methods over the time horizon as long-term, medium-term, and short-term are designed [22], [23].

However, in the present study we focused on the AI models typically developed for prediction of SR for short term horizon. Overview of such models is presented for providing a broader scope of further development in this direction.

II. DATA AND ANALYSIS

A. Availability of data

Ministry of Earth Sciences, the Government of India has started India Meteorological Department (IMD), which records meteorological data for different Indian locations. Being the largest organization in India, it collects radiation data at the ground for various locations. However, as compared to 2.9 million km² landmass area of India [24], [25], these active data collection centres set by IMD are critically less. Further details about the information and data made available by IMD is easily accessible from the website [(http://dsp.imdpune.gov.in)]. None the less, IMD apart, several other organizations are recording meteorological data which is easily accessible. Some organizations are listed in Table 1, from where required data can be gathered. The input layers of the dataset tend to vary based on the database from where the dataset is extracted. These datasets are used as different nodes and layers are formed while designing the machine learning models. In most of the machine learning models, three input nodes from datasets are most likely used, namely year, month, and day.

Further depending on the database, other input nodes such as hour or daily sunshine duration, Solar radiation (SR) ambient temperature (T), minimum temperature (T_{min}), maximum temperature (T_{max}), temperature difference (ΔT), wind speed (WS), wind direction (WD), barometric pressure (P), relative humidity (RH), precipitation, rainfall, clear sky index (ky), air pollution index (AQI), sun zenith angle, hour of the day to be forecasted (Hr), month number (Mon), extra-terrestrial radiation (H₀), bright sunshine hours (S), theoretical sunshine hours (S_o) are considered.

While designing a machine learning model, first it may be helpful to identify the correct climatic zone of that location and depending on the climatic zone in which location is situated, other input parameters in the form of nodes can be considered. Using these strategies, parameters can be forecasted as monthly average daily global solar radiation (MADGSR), GSR, Global Horizontal irradiance (GHI)

Similarly, using some software tools like Waikato environmental for knowledge analysis (WEKA), Geospatial Toolkit (NREL), Hybrid Optimization of Multiple Energy Resources (HOMER), RETScreen plus, ESRI ArcGIS 10.x, ERDAS 201x, etc meteorological data can be accessed. Indian Space Research Organisation (ISRO), provides satellite data with high cost for research costly [26]. In general, to develop a SR forecasting model for a particular location, one has to select nearest source of data collection centre to get meteorological data and develop a model.

B. Choosing the correct forecasting Methods

After selection of location of interest, database required to develop AI model is carefully selected. Here after, the most important task is to select forecasting method. Using on input parameters of database choice of dataset is made and a method is developed for AI models. In general, for GSR forecasting the models are developed in three different ways [27].

- using meteorological and geographical parameters as input datasets; structural models are developed (exogenous forecasting)
- considering historically recorded data of solar irradiance as input features; time-series models are developed (endogenous forecasting);
- considering both, solar irradiance and meteorological parameters as input datasets; hybrid models are developed.

It is important to mention here that; different AI methods show different accuracy according the forecasting horizon [27]. Therefore, one of the important aspects of choosing a correct forecasting method is to consider the forecasting horizon.

TABLE I. LIST OF ORGANIZATIONS AND CORRESPONDING RECORDED DATA PARAMETERS

Sr. No	Data Collection Centers	Parameters
1	National Air Quality Index-IIT Kanpur	AQI
2	Electrical Engineering Department, Madan Mohan Malaviya University of Technology, Gorakhpur	T, WS, Rainfall, Dew point, SR, P, Solar azimuth
3	Solar Radiation Resource Assessment (SRRA)	SR, T, RH, P, and WS
4	NIT, Tiruchirappalli	T, Sunshine duration, and SR
5	National renewable energy laboratory (NREL)	T, P, RH, Solar zenith angle, Precipitation, WD, WS, and SR
6	National Institute of Solar Energy (NISE) Gurugram, India.	T, RH, P, WS, and SR
7	Energy and Environment Center, National Institute of Technology, Hamirpur	T, RH, P, Rainfall, Sunshine duration, and SR
8	Central Research Institute for Dryland Agriculture (CRIDA), Hyderabad	Rainfall, RH, T, Sunshine Duration, and SR
9	International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Hyderabad	Rainfall, RH, T, Sunshine Duration, and SR
10	Automatic Weather Station, Tamilnadu Agriculture University	RH, P, WS, T, and SR
11	Langley Research Center of National Aeronautics and Space Administration (NASA)	Tmax and Tmin (Average)

C. Statistical Indicators and Benchmarking Techniques

Solar irradiance forecast accuracy is assessed using benchmarking studies and various forecasting approaches are compared. Some parameters such as statistical indicators are often used for this, however, time period over which the accuracy is evaluated varies significantly. Thus, a potential seasonal bias is introduced when some models are assessed over a period of one or several years, while some are analysed over a period of some weeks. In such situation selection of right method for fare comparison and accuracy analysis is challenging task.

Few statistical measures are most widely used to evaluate predictability of the developed machine learning model. Those statistical indicators are Uncertainty at 95% (U95), Mean Absolute Error (MAE), Mean Percentage Error (MPE), Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), Relative Root Mean Square Error (RRMSE), Root Mean Square Relative Error (RMSRE), Mean Bias Error (MBE), Maximum Absolute Relative Error (MARE), Correlation Coefficient (R2) and t -static. Table 2 is provided with details of statistical indicators.

III. A REVIEW OF MACHINE LEARNING MODELS FOR SOLAR RADIANCE FORECAST FOR INDIAN REGIONS

With the progression in technology, new methods such as Machine learning, deep learning and ensemble modelling are explored for GSR forecasting. These methods are capable of handling the correlation of dependent and independent input parameters which are very complex and nonlinear and thereafter extracting features out of it. Fig. 1 shows the approach to develop machine learning model for better solar radiance forecasting accuracy. First, the input parameters are listed which are generally used.

Further, one or more algorithms are chosen to develop a machine learning based forecasting model. Strategical development of the machine learning model is necessary to achieve trade-off between better forecasting accuracy and model complexity. Various statistical analysis tools are used to analyse forecasting accuracy. From this perspective, in this review, further discussion is presented on the research work carried out for Indian regions.

A group of researchers [28] formulated the hourly GSR forecasting models 1-to-6-day-ahead forecasting. The machine learning models such as M5, Multivariate Adaptive Regression Splines (MARS), random forest,

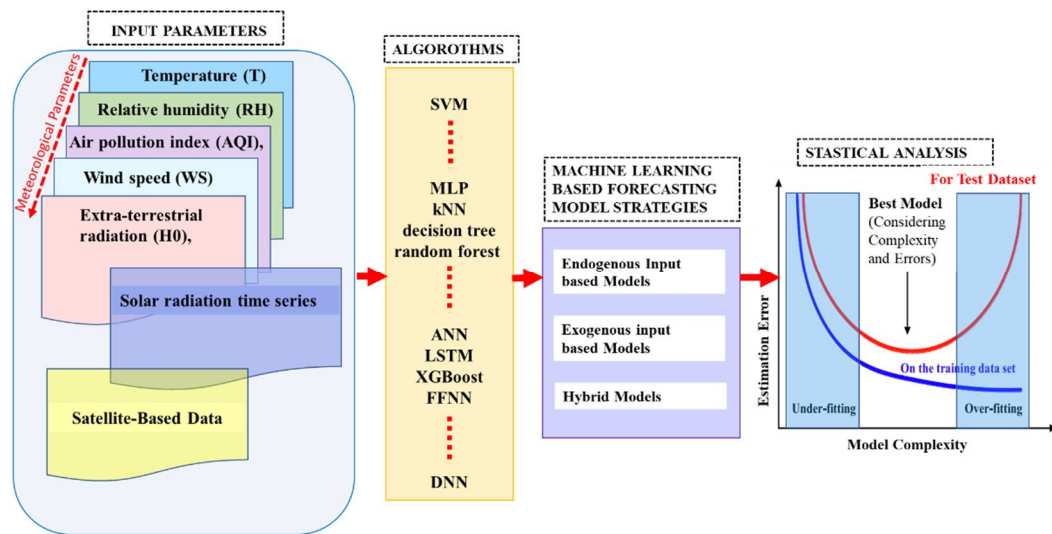


Figure 1. Machine learning model approach for better solar radiance forecasting accuracy

Classification And Regression Tree (CART) models are built for GSR forecasting for Gorakhpur region in India. Nine meteorological variables are used as input dataset wherein random forest shows best performance CART performed worst. However, it is observed that with $\sim 95\%$ confidence interval each of the four algorithms may be used.

Using different combinations of four meteorological parameters, a group of researchers developed Support vector regression (SVR) model for Gurgaon, India region [29]. The meteorological input datasets are collected from the NISE, Gurgaon, India. SVR is observed to produce better forecasting accuracy than other state of art techniques such as Adaptive Neuro-Fuzzy Inference System (ANFIS), Hidden Markov model (HMM), and artificial neural network (ANN).

General fuzzy based model and ANFIS model are developed using hourly solar irradiance data collected at IMD to forecast short term solar energy [30]. Results obtained revealed the better performance of ANFIS algorithm.

A group of researchers [31] analysed IMD meteorological data at various Indian location using the WEKA software tool, to estimate effect of various parameters on GSR and for deriving most influential parameter which affects the GSR predictability of the model. The study revealed that parameters that affect GSR are month, S0, latitude, and Tmax, while the least affecting parameter is RH. Temperature (being most influencing parameter) is used to develop empirical model, ANN, and Support Vector Machine (SVM) based models to forecast GSR at various Indian cities, of which SVM shows better performance.

Rao et al., [20] developed six different ANN models by varying the combinations input variables. Input dataset contains seven different parameters. Among all, the two ANN models trained with input dataset as [Tmin and H0] and [Tmin, Ho and S0] produced better GSR forecasting accuracy.

Based on an ensemble learning technique, multistage intelligent method is developed for four different locations in India [32]. By utilizing the two-way interaction model, correlation between features is established. For developing the forecasting models, methods like SVM, Least Absolute Shrinkage and Selection Operator (LASSO), Group Regularised Estimation under Structural Hierarchy (GRESH) are used. Similarly, hybrid model is formed by combining Glowworm Swarm Optimization (GSO) and LASSO methods. Comparative analysis of these methods revealed the outperformance of the hybrid model of GSO and LASSO. Makade et al. [33] used theory of experimentation for prediction of GSR, by. Analysis of the results suggested that increase in number of input parameters does not cause significant increase in prediction accuracy. However, location of interest influences the accuracy. Further, the relative humidity, altitude, and sunshine duration are identified as most influencing input parameters.

Long Short-Term Memory (LSTM), model is developed by P Kumari et al. [34], to forecast hourly GHI at seven locations in India. Five different input parameters are obtained from SUNY dataset. Solar irradiance satellite data is also used in this study which is available for many geo-locations. Using National Solar Radiation Database (NSRDB) data, efficiency of this model is evaluated at seven cities across India having different climatic conditions. Performance assessment of the developed model is carried out against Extreme Gradient Boost

(XGBoost), Feed Forward Neural Network (FFNN) and Persistence model at broader coverage of geographical regions. The assessment revealed outperformance of LSTM model.

A new ensemble model is developed on the basis of extreme gradient boosting forest and deep neural networks (XGBF-DNN) to predict hourly global horizontal irradiance forecast [35]. From SUNY dataset, four different input parameters used. The forecasting accuracy of the developed model tested for three climatologically different locations. Test results shows that irrespective of seasonal variation in meteorological parameters, proposed model produced better stability and prediction accuracy.

A group of researchers [36] developed five different ANN models by varying combination of input dataset. Five different input parameters gathered from IMD are utilized as input data parameters to build ANN models. These models are specially tested to forecast GHI at mountainous state Uttarakhand, India. Result analysis showed that the ANN models formed considering input parameters as [temperature difference and bright sunshine hours] and [minimum temperature, maximum temperature, temperature difference] perform to yield better accuracy.

In a study [37], eight different machine learning models are developed to predict hourly global horizontal irradiance. Models are multivariate linear regression, ridge, LASSO, elastic net, multilayer perceptron (MLP), k-nearest neighbors, decision tree, and random forest. Nine different input parameters are gathered from National Solar Radiation Database (NSRDB). Forecasting models are tested for accuracy at 21 climatologically and geographically different locations. Study on the similar line is conducted by the same author for different location.

NASA's POWER database contains satellite data which is utilized to develop a daily solar irradiance data forecasting model [38]. To extract and learn the symmetry latent in data patterns, deep neural network (DNN) models such as bidirectional LSTM and attention-based LSTM models are developed and trained for predicting solar irradiance in two locations of India.

Further, few researchers [39] developed a prediction model which can be taken in to account to forecast solar radiance for all parts of India. To forecast monthly average MADGSR, a higher-order regression model is developed using relative humidity sunshine duration, altitude and latitude as input meteorological parameters. In order to validate this model, forecasting accuracy performance of this model is compared with the model developed by Samanta et al., [40], and study revealed the comparable performance with model [40] for forecasting GSR across Indian regions. In general, GSR prediction models to perform well in forecasting across all climatologically and geographically varying Indian regions, uses sunshine duration, relative humidity, latitude, and altitude as input parameter. However, a research group [41] developed a GSR forecasting model using temperature as only input parameter, owing to its readily data availability across all regions of India compared to other parameters like sunshine duration and solar radiance.

Husain S et al., [42] developed six different AI based GSR forecasting models using single input meteorological parameter as air temperature. These models are designed for GSR forecasting across geographically and climatologically diverse zones of India. Using six different machine learning algorithms such as SVM, MLP, random forest, k-nearest neighbors, Gaussian process regression, and XGBoost forecasting models are developed and compared for prediction accuracy. Amongst all k-nearest neighbors (KNN) and XGBoost outperforms all the models. It is suggested by the authors that combining KNN and XGBoost with air temperature forecasts can be used to develop accurate future GSR forecasting.

Table 3. summarises different algorithms used to estimate the SR forecasting across the Indian regions. It covers detail information about the input parameters used by various authors. Using same sets of input parameters, different machine learning algorithms are developed and compared by every group of researcher. Table presents a bird eye view to compare all this information thoroughly.

TABLE II. DIFFERENT MACHINE LEARNING MODELS DEVELOPED FOR SOLAR RADIANCE FORECASTING FOR INDIAN REGIONS

No	Time horizon	Location	Dataset	Input Parameters	Algorithm	Error	Ref
1	1-day-ahead to 6-day-ahead hourly solar radiation	Government of India in Gorakhpur, India.	IMD	Tmax, Tmin, Tavg, WS, rainfall, dew point, GSR, P, solar azimuth.	MARS, CART, M5 and random forest model	Standard t-static error for all models between 1.9 to 2	[28]
2	MADGSR	NISE, Gurugram	National Institute of Solar Energy (NISE) Gurugram,	T, RH, P, WS	SVR	RMSE = 14.3	[29]

			India.				
3	hourly GSR	across India	IMD	hourly data of solar irradiance	fuzzy logic and ANFIS	ARE= 0.001 (ANFIS) ARE = 0.01 (FUZZY)	[30]
4	MADGSR	Bhubaneswar, Chennai, Hyderabad, Patna, Nagpur, New Delhi, Mangalore and Trivandrum and ten cities of Tamil Nadu, India	India Meteorological Department (IMD), Pune	RH, Tmax, Tmin, H0, attitude, longitude.	Support Vector Machine (SVM), Artificial Neural Network (ANN) and empirical solar radiation models	Range RMSE = Range 2.05 to 2.68 (ANN), RMSE = Range 1.77 and 2.02 (SVM)	[31]
5	MADGSR	Tiruchirappalli	Self measured data at NIT, Tiruchirappalli.	daily GSR, daily Tmin, daily Tmax, difference between temperature (DT), S, S0 and H0	six different ANN models	RRMSE= 3.96% for ANN [Tmin, H0]	[20]
6	5 days global horizontal radiation	at four given locations of India	IMD	T, P, RH, WD, WS, solar zenith angle, precipitation, and GHR	DCGSO-LASSO, LASSO, GRESH, and SVM	MAPEs of DCGSO-LASSO = Range 7% and 13%	[32]
7	Hourly (GHI)	7 different cities of different climate zone as Dehradun, Dharamshala, Gandhinagar, Guwahati, Jaipur, Pune, Thiruvananthapuram	SUNY data	T, RH, P, WS and dew point	FFNN, LSTM and XGBoost	nRMSE for FFNN = 25.4, LSTM = 22.8, XGBoost = 24.6	[34]
8	Hourly (GHI)	New Delhi, Jaipur, Gangtok	SUNY data	T, RH, P, WS, WD, ky measured at hour t, Hr and Mon	XGBoost models and a DNN model XGBF-DNN comprising ensemble models	forecast skill score = 33% to 40% (reduction in error)	[35]
9	daily GHI	Almora, Bageshwar, Chamoli, Champawat, Dehradun, Haridwar, Nainital, Pauri Garhwal, Pithoragarh, Rudrapur, Tehri Garhwal, Udham Singh Nagar, Uttarkashi	IMD, Pune.	Tmin, Tmax, ΔT , GHI, H0, and bright sunshine hours (S)	5 different ANN models	For ANN-3, RMSE = 2.03	[36]
10	hourly GHI	21 locations of different climatic and geographic regions Agartala Ahmedabad Bhopal Kochi Chennai Coimbatore Davangere Diu Guwahati Bhubaneswar Indore Jaipur Kakinada Vishakhapatnam Ludhiana Solapur Pune New Delhi Surat Udaipur Imphal	National Solar Radiation Database (NSRDB)	GHI, T, RH, WS, WD, Dew point, P, Precipitation	MLR, ridge, LASSO, elastic net, multilayer perceptron, k-nearest neighbors, decision tree, and random forest	Forecasting performance random forest > kNN > MLP > decision tree > elastic net > ridge > lasso > MLR	[37]
11	daily solar irradiance data	India	NASA's POWER database	Multisite daily solar irradiance data	DNN such as bidirectional LSTM and attention-based LSTM	$R^2 = 70\%$ and 73%	[38]
12	MADGSR	all parts of India	IMD	sunshine duration, latitude, altitude, and RH	higher-order regression model	MPE = 11.8935%, $R^2 = 0.8454e0.9853$, MBE = -2.0736, RMSE = 2.3227	[39]
13	MADGSR	all parts of India	IMD	T	SVM, MLP, random forest, k-nearest neighbors, Gaussian process regression, and XGBoost.	RMSE = 2.04	[42]

IV. CONCLUSION AND FUTURE SCOPE FOR WORK

As India consists of numerous climatologically and geographically distinct zones, the approach of employing AI models for GSR forecasting becomes extremely advantageous. Present manuscript discusses the numerous machine learning model approaches for GSR forecasting. Recent developments in this perspective are reviewed thoroughly and methodically, considering the factors such as location, selection of input parameters, feature selection, and model development. Review of the literature highlighted that in most of the cases sunshine duration is most prominently used input parameter after which temperature is preferred. Thereafter, input parameters such as pressure, relative humidity, solar zenith angle, precipitation, wind direction and wind speed are used in general to train machine learning models. More often the choice of input parameters is driven by the climatological and geographical zone of location and data availability. Basic machine learning model is chosen on the basis of its nonlinear correlation capability of input parameters. This is also emphasized when different accuracies are offered by different machine learning models over same input datasets from same database. Few researchers also presented the models developed for GSR forecasting across all regions of India.

In the view of this observation, future scope of the research work can be presented. In most of the cases, using the same input datasets for training the models, various machine learning models are developed and tested for forecasting accuracy, which tend to vary. In such cases, depending on the zone and data type, correction factors might be derived from other more accurate models to enhance accuracy of less accurate models. However, further use of various combinations of different machine learning models to formulate hybrid models might improve accuracy of prediction over larger zones across India.

REFERENCES

- [1] D. Guijo-Rubio et al., "Evolutionary artificial neural networks for accurate solar radiation prediction," *Energy*, vol. 210, Nov. 2020, doi: 10.1016/j.energy.2020.118374.
- [2] S. A. Sarkodie, S. Adams, and T. Leirvik, "Foreign direct investment and renewable energy in climate change mitigation: Does governance matter?," *J. Clean. Prod.*, vol. 263, Aug. 2020, doi: 10.1016/j.jclepro.2020.121262.
- [3] "Key World Energy Statistics 2018 - Event - IEA." <https://www.iea.org/events/key-world-energy-statistics-2018> (accessed Jan. 27, 2023).
- [4] T. R. Govindasamy and N. Chetty, "Machine learning models to quantify the influence of PM10 aerosol concentration on global solar radiation prediction in South Africa," *Clean. Eng. Technol.*, vol. 2, Jun. 2021, doi: 10.1016/j.clet.2021.100042.
- [5] S. Ghimire, R. C. Deo, N. Raj, and J. Mi, "Wavelet-based 3-phase hybrid SVR model trained with satellite-derived predictors, particle swarm optimization and maximum overlap discrete wavelet transform for solar radiation prediction," *Renew. Sustain. Energy Rev.*, vol. 113, p. 2019, Oct. 2019, doi: 10.1016/j.rser.2019.109247.
- [6] O. Abedinia, M. Zareinejad, M. H. Doranehgard, G. Fathi, and N. Ghadimi, "Optimal offering and bidding strategies of renewable energy based large consumer using a novel hybrid robust-stochastic approach," *J. Clean. Prod.*, vol. 215, pp. 878–889, Apr. 2019.
- [7] J. Dong et al., "Novel stochastic methods to predict short-term solar radiation and photovoltaic power," *Renew. Energy*, vol. 145, pp. 333–346, Jan. 2020.
- [8] S. Saud, B. Jamil, Y. Upadhyay, and K. Irshad, "Performance improvement of empirical models for estimation of global solar radiation in India: A k-fold cross-validation approach," *Sustainable Energy Technologies and Assessments*, vol. 40, Aug. 2020.
- [9] S. Solaymani, "CO2 emissions patterns in 7 top carbon emitter economies: The case of transport sector," *Energy*, vol. 168, pp. 989–1001, Feb. 2019.
- [10] S. C. Gowd et al., "Economic perspectives and policy insights on carbon capture, storage, and utilization for sustainable development," *Science of The Total Environment*, vol. 883, p. 163656, Jul. 2023, doi: 10.1016/J.SCIOTENV.2023.163656.
- [11] F. Besharat, A. A. Dehghan, and A. R. Faghih, "Empirical models for estimating global solar radiation: A review and case study," *Renewable and Sustainable Energy Reviews*, vol. 21, pp. 798–821, 2013.
- [12] Y. Zhou, Y. Liu, D. Wang, X. Liu, and Y. Wang, "A review on global solar radiation prediction with machine learning models in a comprehensive perspective," *Energy Convers. Manag.*, vol. 235, May 2021.
- [13] R. G. Makade, S. Chakrabarti, and B. Jamil, "Development of global solar radiation models: a comprehensive review and statistical analysis for Indian regions," *J. Clean. Prod.*, vol. 293, Apr. 2021, doi: 10.1016/j.jclepro.2021.126208.
- [14] C. Rigollier, M. Lefèvre, and L. Wald, "The method Heliosat-2 for deriving shortwave solar radiation from satellite images," *Solar Energy*, vol. 77, no. 2, pp. 159–169, 2004.
- [15] H. B. Tolabi, S. B. M. Ayob, M. H. Moradi, and M. Shakarmi, "New technique for estimating the monthly average daily global solar radiation using bees algorithm and empirical equations," *Environ Prog Sustain Energy*, vol. 33, no. 3, pp. 1042–1050, Oct. 2014.

- [16] Y. Jiang, "Computation of monthly mean daily global solar radiation in China using artificial neural networks and comparison with other empirical models," *Energy*, vol. 34, no. 9, pp. 1276–1283, Sep. 2009, doi: 10.1016/j.energy.2009.05.009.
- [17] J. Huertas-Tato, R. Aler, I. M. Galván, F. J. Rodríguez-Benítez, C. Arbizu-Barrena, and D. Pozo-Vázquez, "A short-term solar radiation forecasting system for the Iberian Peninsula. Part 2: Model blending approaches based on machine learning," *Solar Energy*, vol. 195, pp. 685–696, Jan. 2020.
- [18] A. Shadab, S. Ahmad, and S. Said, "Spatial forecasting of solar radiation using ARIMA model," *Remote Sens Appl*, vol. 20, p. 100427, Nov. 2020, doi: 10.1016/J.RSASE.2020.100427.
- [19] A. Shadab, S. Said, and S. Ahmad, "Box–Jenkins multiplicative ARIMA modeling for prediction of solar radiation: a case study," *International Journal of Energy and Water Resources*, vol. 3, no. 4, pp. 305–318, Dec. 2019, doi: 10.1007/S42108-019-00037-5.
- [20] D. V. S. K. Rao K, M. Premalatha, and C. Naveen, "Analysis of different combinations of meteorological parameters in predicting the horizontal global solar radiation with ANN approach: A case study," *Renewable and Sustainable Energy Reviews*, vol. 91, pp. 248–258, Aug. 2018.
- [21] K. D. V. Siva Krishna Rao, M. Premalatha, and C. Naveen, "Method and strategy for predicting daily global solar radiation using one and two input variables for Indian stations," *Journal of Renewable and Sustainable Energy*, vol. 10, no. 1, Jan. 2018, doi: 10.1063/1.4995035.
- [22] S. Y. Heng et al., "Artificial neural network model with different backpropagation algorithms and meteorological data for solar radiation prediction," *Sci Rep*, vol. 12, no. 1, Dec. 2022, doi: 10.1038/S41598-022-13532-3.
- [23] N. Premalatha and A. Valan Arasu, "Prediction of solar radiation for solar systems by using ANN models with different back propagation algorithms," *Journal of Applied Research and Technology*, vol. 14, no. 3, pp. 206–214, Jun. 2016, doi: 10.1016/j.jart.2016.05.001.
- [24] N. Rathore, N. L. Panwar, A. Gama, and F. Yettou, "Solar map of India under clear sky conditions," *International Journal of Sustainable Energy*, vol. 38, no. 5, pp. 415–446, May 2019, doi: 10.1080/14786451.2018.1527334.
- [25] P. K. S. Rathore, S. Rathore, R. Pratap Singh, and S. Agnihotri, "Solar power utility sector in india: Challenges and opportunities," *Renewable and Sustainable Energy Reviews*, vol. 81, pp. 2703–2713, Jan. 2018.
- [26] S. Samanta et al., "Generation of common coefficients to estimate global solar radiation over different locations of India," *Theor Appl Climatol*, vol. 136, no. 3–4, pp. 943–953, May 2019, doi: 10.1007/S00704-018-2531-4.
- [27] S. K. Aggarwal and L. M. Saini, "Solar energy prediction using linear and non-linear regularization models: A study on AMS (American Meteorological Society) 2013–14 Solar Energy Prediction Contest," *Energy*, vol. 78, pp. 247–256, Dec. 2014, doi: 10.1016/j.energy.2014.10.012.
- [28] R. Srivastava, A. N. Tiwari, and V. K. Giri, "Solar radiation forecasting using MARS, CART, M5, and random forest model: A case study for India," *Heliyon*, vol. 5, no. 10, p. e02692, Oct. 2019.
- [29] P. Bhola and S. Bhardwaj, "Estimation of solar radiation using support vector regression," *Journal of Information and Optimization Sciences*, vol. 40, no. 2, pp. 339–350, Feb. 2019, doi: 10.1080/02522667.2019.1578093.
- [30] Viswavandya Meera, Sarangi Bakdevi, and Mohanty Asit, "Fuzzy Logic and ANFIS based Short Term Solar Energy Forecasting," *Int. J. Futur. Revolut. Comput. Sci. Commun. Eng.*, vol. 4, pp. 631e636, vol. 4, no. 4, pp. 631–636, Apr. 2018.
- [31] R. Meenal and A. I. Selvakumar, "Assessment of SVM, empirical and ANN based solar radiation prediction models with most influencing input parameters," *Renew Energy*, vol. 121, pp. 324–343, Jun. 2018.
- [32] H. Jiang, Y. Dong, and L. Xiao, "A multi-stage intelligent approach based on an ensemble of two-way interaction model for forecasting the global horizontal radiation of India," *Energy Convers Manag*, vol. 137, pp. 142–154, Apr. 2017, doi: 10.1016/J.ENCONMAN.2017.01.040.
- [33] R. G. Makade, S. Chakrabarti, B. Jamil, and C. N. Sakhale, "Estimation of global solar radiation for the tropical wet climatic region of India: A theory of experimentation approach," *Renew Energy*, vol. 146, pp. 2044–2059, Feb. 2020.
- [34] Pratima Kumari and Durga Toshniwal, "Hourly Solar Irradiance Prediction From Satellite Data Using Lstm," in *Innovative Solutions for Energy Transitions*, 2019.
- [35] P. Kumari and D. Toshniwal, "Extreme gradient boosting and deep neural network based ensemble learning approach to forecast hourly solar irradiance," *J Clean Prod*, vol. 279, p. 123285, Jan. 2021, doi: 10.1016/j.jclepro.2020.123285.
- [36] P. Kumari and D. Toshniwal, "Analysis of ANN-based daily global horizontal irradiance prediction models with different meteorological parameters: a case study of mountainous region of India," *Int J Green Energy*, vol. 18, no. 10, pp. 1007–1026, Aug. 2021, doi: 10.1080/15435075.2021.1890085.
- [37] Kumari P, "Machine learning techniques for hourly global horizontal irradiance prediction: A case study for smart cities of India," in *Sustainable Energy Solutions for a Post-COVID Recovery towards a Better Future*, 2021.
- [38] B. Brahma and R. Wadhvani, "Solar Irradiance Forecasting Based on Deep Learning Methodologies and Multi-Site Data," *Symmetry (Basel)*, vol. 12, no. 11, p. 1830, Nov. 2020, doi: 10.3390/sym12111830.
- [39] R. G. Makade, S. Chakrabarti, and B. Jamil, "Prediction of global solar radiation using a single empirical model for diversified locations across India," *Urban Clim*, vol. 29, p. 100492, Sep. 2019, doi: 10.1016/j.uclim.2019.100492.
- [40] S. Samanta et al., "Generation of common coefficients to estimate global solar radiation over different locations of India," *Theor Appl Climatol*, vol. 136, no. 3–4, pp. 943–953, May 2019, doi: 10.1007/s00704-018-2531-4.
- [41] B. Jamil and E. Bellos, "Development of empirical models for estimation of global solar radiation exergy in India," *J Clean Prod*, vol. 207, pp. 1–16, Jan. 2019.

- [42] S. Husain and U. A. Khan, "Development of machine learning models based on air temperature for estimation of global solar radiation in India," *Environ Prog Sustain Energy*, vol. 41, no. 4, Jul. 2022, doi: 10.1002/ep.13782.