**Comparative analysis**

**Introductory Paragraphs**

Quej et al. [1], worked on predicting daily global solar radiation at Yucatán Peninsula, Mexico. The data required for training ML models was collected from Mexican National Meteorological Service and Instituto Nacional de Investigaciones Forestales Agrícolas y Pecuarias (INIFAP).

Yadav & Sethy [2], presented methods for predicting solar irradiation for the USA. The dataset used was collected by the National Renewable Energy Laboratory, USA

The research by Suyono et al. [3], compared the Solar Radiation Intensity predicted by ANFIS and MLR for Basel City, Switzerland. The dataset used for the study was taken from NASA Meteoblue Climatology website

Alqudah et al. [4], worked on Prediction of Solar Radiation Based on Spatial and Temporal Embeddings for San Antonio, Texas. The National Solar Radiation Database was used for training ML models.

Ivanova et al. [5], proposed the usage of NARX Model for Solar Radiation Prediction at Barcelona. The data was provided by an industrial company located in Northern Barcelona.

Gupta & Singhal [6], worked on Prediction of Global Solar Radiation of Pune, India. The data for the research was collected from the Indian Meteorological Department, Pune (IMD,Pune).

Noriega-Angarita et al. [7], proposed methods for Solar Radiation Prediction for Dimensioning Photovoltaic Systems. The data collection was carried out at Columbia weather stations in the Atlantic Coast of Columbia.

Works by Madugu et al. [8], focused on Solar radiation prediction for household purposes. The study was completed in Kano city of Nigeria. The dataset provided by the Nigerian Meteorological Agency (NiMet) was used in ML modelling.

**Other Literature Review and Findings**

There are multiple researches proposing methodologies for predicting direct or beam solar radiation. But predicting diffused solar radiation is quite difficult. In the research by Lou et al. [19], a logistic regression algorithm was employed to predict the horizontal sky-diffuse irradiance and conduct sensitivity analysis for the meteorological variables. The study focused on some important parameters for predicting diffused solar radiation such as clearness index, solar altitude, air temperature, cloud cover and visibility. For Hong Kong and Denver, USA, the mean absolute error (MAE) was less than 21.5 W/m2 and 30 W/m2, respectively. The suggested model is suited to estimate long-term diffuse solar radiation and research climate change. Prediction of Global solar radiation (GSR) is a critical variable for designing photovoltaic systems. Dhakal et al. [20], proposed a cost effective approach to this problem. Readily available meteorological data at Biratnagar Airport, Nepal was utilized to predict GSR. Extraterrestrial solar radiation, sunshine duration, maximum and minimum ambient temperatures, precipitation, and relative humidity were all factors considered in the study. An Artificial Neural Network along with five other statistical ML models was used for prediction of GSR. The R-Squared calculated was 0.8870 which is very good for cost effective data collection and computationally efficient statistical ML model training.

**Abbreviations**

| Abbreviation | Full - form |
| --- | --- |
|  |  |
| AI | Artificial Intelligence |
| ANFIS | adaptive-network-based fuzzy inference system |
| ANN | Artificial Neural Network |
| ARIMA | autoregressive integrated moving average |
| ARMA | Auto Regressive Moving Average |
| ARMAP | autoregressive moving average process |
| DBN | deep belief network |
| DNN | Deen Neural Network |
| DT | Decision Tree |
| HELmanNN | Elman Neural Network |
| ELMNN | Extreme Learning Machine Neural Network |
| GBM | Gradient Boosting Machine |
| GHI | Global Horizontal Irradiation/Irradiance |
| GRNN | general regression neural network |
| GTI | Global Tilted Irradiation/Irradiance |
| LSTM | Long Short-Term Memory |
| MABE | Mean Absolute Bias Error |
| MAE | Mean Absolute Error |
| MAPE | Mean Absolute Percentage Error |
| ML | Machine Learning |
| MLP | Multi-layer Perceptron |
| MSE | Mean Squared Error |
| NARX | nonlinear autoregressive model which has exogenous inputs |
| PV | Photovoltaics |
| RF | Random Forests |
| RMSE | Root Mean Squared Error |
| SSE | Sum of Squared Errors |
| TDNN | Time delay neural network |
| WVC FM | Weight-Varying Combination Forecast Mode |
| XGBR | XGBoost Regressor |

**Conclusion**

For this comparative analysis of different methodologies in predicting beam, diffused, and reflected solar radiation, three types of Machine Learning algorithms were taken into consideration - Neural Network based models, Statistical or Probabilistic models, and Time-Series models. As a gesture of inclusiveness, this review includes different approaches proposed by the latest (2016 to 2021) research works in diverse meteorological and geographical conditions. The findings of this study indicate that the most Important parameters in solar radiation prediction are temperature(C), humidity(mm), clear sky index, wind speed(m/s), the angle of incidence(ω), azimuth angle(φ), zenith angle(Z), atmospheric pressure(atm), Albedo, solar insolence(kWh/m^2), tilt angle(α), terrain elevation(m), global horizontal irradiance(W/m^2), and the size of PV installation(kW). Solar radiation prediction can be carried out for short-term (hourly), diurnal (daily) and long-term (monthly or seasonally) purposes. For short-term solar radiation forecasting, NARX models having exogenous inputs and GRNN trained with Genetic algorithm optimizer outperformed other approaches. For daily solar radiation prediction purposes, ARMA models, ANN with hidden neurons having purelin activation function, and Gaussian or Sinusoidal wave functions produced the most precise aggregates. Random Forests and ARIMA were found to be the most reliable models for long-term solar radiation forecasting. The variable or unprecedented nature of most of the important features in forecasting solar radiation makes this domain challenging yet promising.