A Comprehensive Review of Machine Learning Techniques for Solar Energy Resource Assessment and Prediction

Gopal Bhalani[1], Mukul Mistry[2], Helly KhambhatWala[3], Poojan Mardiya[4], Alok Kumar Singh[5]

[1,2,3,4] Department of Computer Science and Engineering (AI-ML), Adani University, Ahmedabad, India
[5] Electrical engineering department, Adani University, Ahmedabad, India
GOPALBHALANI.cse22@adaniuni.ac.in

Abstract: Solar energy resource assessment plays a critical role in optimizing photovoltaic (PV) deployment, forecasting energy output, and ensuring long-term renewable energy planning. Traditional statistical and physical models often struggle with variability in climate conditions, complex environmental interactions, and heterogeneous datasets. Machine Learning (ML) techniques have emerged as powerful tools capable of addressing these limitations by learning complex nonlinear relationships and integrating multisource data. This review paper presents a comprehensive study of machine learning approaches used for solar irradiance prediction, resource mapping, anomaly detection, and system optimization. It covers traditional ML models, deep-learning architectures such as RNN, GRU, and LSTM, hybrid CNN–LSTM models, ensemble techniques, and satellite-based ML frameworks. Key challenges, recent advancements, comparative model performance, and future research directions are also discussed.

Keywords: Solar Energy, Machine Learning, Irradiance Prediction, Deep Learning, Resource Assessment, CNN-LSTM.

1 Introduction

Maintaining and growing the global economy cannot be separated from supporting electricity, and oil-dependent countries are particularly aware of this problem because their consumption of non-renewable energy sources is high and therefore because their power systems are destined to switch to electricity [1],[2]. Energy resource management is the most important component of modern society's sustainability and progress. As the use of energy is growing exponentially, the efficient use of energy resources has become paramount [3]. This paper delves into the intersection of two major, powerful forces: machine learning and energy resource management [4]. By perfect execution of the capabilities of machine learning, the problem of optimizing energy usage and forecasting can be solved. This paper gave an overview of the significance of energy resource management. This paper tells the role of energy resource management in empowering economies. Moreover, this paper has

R. Mehta et al. (eds.), *Proceedings of the International Conference on Infrastructure Development and Sustainability (ICIDS 2025)*, Atlantis Highlights in Sustainable Development 9,
https://doi.org/10.2991/978-94-6239-685-2_10

demonstrated the challenges faced for solar resource assessment, applications of machine learning to energy resource assessment, and some improvements that can lead to more efficiency of solar resource assessment [5].

The adoption of machine learning (ML) techniques in resource assessment has played a pioneering role in recent years, complementing traditional methods across energy, environmental science, agriculture, and many other fields. By integrating ML algorithms with resource assessment, precision, accuracy, and effectiveness in predictions are improved, and a new set of circumstances arises for taking out valuable insights from huge and diverse datasets [6][7]. The subsequent part of the paper explores the application of energy resource assessment. This paper gives an overview of the latest technologies in machine learning applications for resource assessment, casing theoretical concepts, and real-world implementations [8]. Furthermore, the paper has examined how the predictive model can add to the accuracy and longevity of energy infrastructure, from traditional power plants to renewable energy systems [9][10].

2 Resources for solar energy

Assessing the resources of solar energy involves various factors to discover the potential for solar power. Here are some basic steps and factors involved in the evaluation of solar energy. Solar irradiance refers to the amount of solar radiation received per unit area at a current location. It is measured in kilowatt-hours per square meter per day(kWh/m^2/day). Data on solar irradiance can be collected through ground-level observation, mathematical models, and satellite observations [5]. Solar isolation is defined as the solar radiation consumed on a given ground area over a particular amount of time. Climate data, greenhouse effect, humidity, and cloud cover affect solar energy. Climate change affects the amount of sunlight reaching the Earth's surface.

3 Challenges Faced During Solar Resource Assessment and Future Directions

3.1 Necessity to Predict Solar Irradiation

The scarcity of hydrocarbons and the rising demand in the price of fossil fuels has led the world to divert towards the use of renewable resources, among which solar-being the most widespread and easiest. But the proper availability of solar energy has been hampered due to Solar irradiation. Solar rays travel a large distance to fall on the PV cells. During the emission of solar radiation, many heavy objects like mountain peaks or huge trees obstruct its way, which in turn deviates the angle of the solar radiation.

3.2 Strengthening Prediction Accuracy

For best results, the availability of accurate and precise data on solar radiation and solar energy-generating potential is crucial to obtain. It has been seen that traditional methods fail to obtain such accurate data, thereby leading to imperfect results. In such cases, Machine Learning algorithms have proved to be supreme in obtaining accurate results through the technology of neural networks, deep learning. Machine learning algorithms can attain huge amounts of data sets, including the previous track of results and data [4].

3.3 Data Fusion and Integration

Data regarding solar energy generation is very difficult to obtain, as to get proper data, it needs to be integrated from various sources such as weather stations, Satellite plants, and Geographical information systems. It is a tedious task to manage such heterogeneous resources and integrate them [2].

3.4 Scalability and Automation

In such cases, Machine learning algorithms provide the facility of Automation, comprising of rapid generation of the solar resource map and providing proper locations for solar installations. It also helps in accessing maximum energy production potential with the aim of least human involvement [3].

3.5 Change of Climate

When the climate changes, it has a vast effect on the solar radiation and solar power plants. Manual methods may not be sufficient for such environmental changes. By implementing machine learning techniques, algorithms are designed that can frequently monitor climate changes and prepare for adaptation. They are also prepared to forecast the changes in solar generation patterns and environment, which can help assist the shareholders in equipping better solar infrastructure and machinery accordingly [11].

4 Applications of Machine Learning in Energy Resources Management

The systems include ML solar plant systems, with specialized ML interactions in solar plant sensor systems and software, deep learning, and ensemble methods. The importance of effective technology integration is emphasized [12]. Moreover, ML algorithms such as TensorFlow and Caffe2 design smart sensor systems for the tools needed to manufacture, integrate ML sensor model functionality with the SCADA system and, even on paper, these build Machine Learning technologies for solar plant systems Emphasize the importance of openness data confirmation and source code, by researchers in the field They also provide a valuable resource for professionals [3]. When it comes to solar energy forecasting, ML methods contributed to more accurate forecasts, models such as RNN-LSTM, ELM, and CNN show better performance in terms of RMSE-R2 Despite this increase, the field is open in digital way The solar energy in smart solar panels document to enable change solar panels In ML is used to solve technology, especially solar panel failure analysis and testing Custom ML [7].

4.1 Solar radiation forecast for the chosen region

Report ML pics include ground air pollution, satellite television for personal computer, TV for pc, television for laptop snapshots (e.g., MODIS, GOES), and weather data. (temperature, humidity, precipitation) [13]. Solar Panel Installation Optimization: Identify locations wherein daylight hours are long and sufficient every day for peak electricity intervals.

4.2 Great Photos of Solar Power

Use ML algorithms (e.g., Support Vector Regression, Random Forests) to create massive dataset strategies inclusive of satellite tv for pc tv for computer pictures, aerial statistics, and virtual elevation models (DEMs) [6]. Create outstanding-resolution sun help maps that cover whole areas or worldwide regions.

Fault detection in post-solar systems: ML algorithms analyse sensor data from post-solar systems to find faults or anomalies in motion systems due to events such as anomalies, mechanical failures, or power outages throughout the ML Day solar energy availability [4].

Dynamic pricing offers: ML channels can analyse real-time market data, weather forecasts, and energy production data to provide better pricing options for solar, the cost to ML customers of supplies based on price conversion mechanisms, and solar energy in energy conversion mechanisms [8].

5 Machine learning techniques for energy efficiency

5.1 Predict solar irradiance:

Solar irradiance is a power in the form of electromagnetic radiation that is received from the sun. It is normally expressed in watts per square meter(W/m^2). There are three types of irradiations shown in Figure 1. There are three major types of solar irradiance, which are shown in the figure below [1]:

1) Horizontal irradiance

2) Global Tilt irradiance

3) Direct normal irradiance

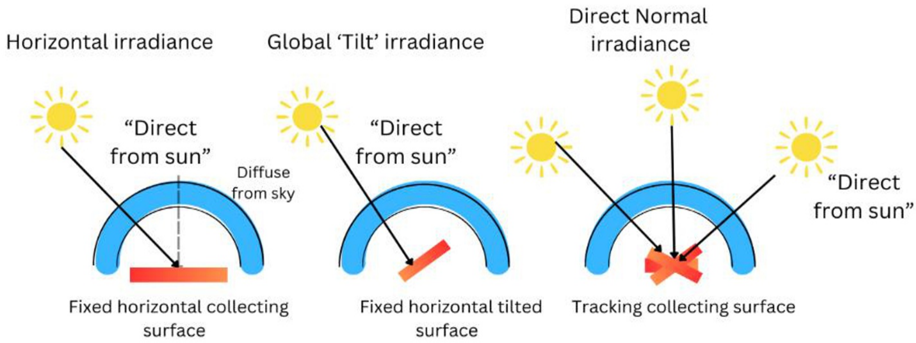

Fig. 1. Types of Solar Irradiance

We can predict solar irradiance using previous data of environmental parameters such as wind speed, solar irradiance, temperature, atmospheric pressure, etc. We used the R^2 method to predict solar irradiance. R^2 coefficient of determination, which is calculated as a ratio of explained variance and total variance. R^2 is also equivalent to 1- the ratio of the error sum of squares and total sum of squares. If R^2 is equal to 0, then the regression line is horizontal, which indicates no variation in output. If r^2 is equal to one, that means predication is the same as variation. The value of r^2 is unique for various machine learning algorithms that as shown in the table below [1].

5.2 Machine learning using satellite:

Machine learning in sun useful resource evaluation using satellite television for personal computer tv for laptop pictures entails leveraging computational algorithms to investigate satellite television for laptop tv for computer imagery statistics for the purpose of assessing the sun strength functionality of locations. As shown in Figure 2, this approach generally consists of collecting satellite television for laptop images that capture numerous environmental elements applicable to solar energy technology, which include cloud cover, terrain capabilities, and flora density. These predictions can help find out suitable locations for solar electricity equipment, optimize the location and orientation of solar panels, and forecast power generation [14].

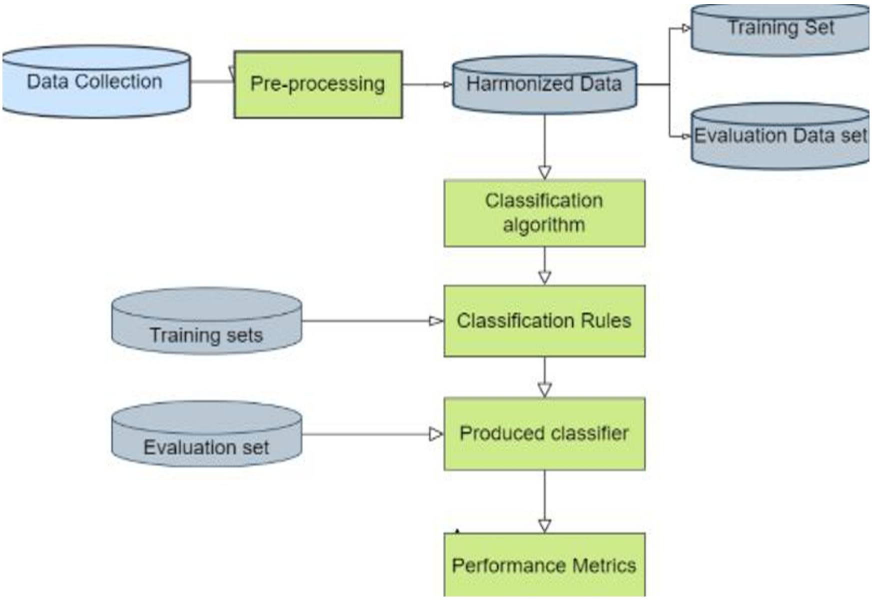

Fig.2. Data Collection flow chart

5.3 Methodological Frameworks for Solar Power Prediction

Recent developments in forecasting solar power have involved different machine learning and deep learning techniques that are applied to time series data. Recurrent Neural Networks (RNN), Gated Recurrent Units (GRU), and Long Short-Term Memory (LSTM) are popular model architectures that are used to learn time-dependent dependencies and complex sequential behavior of solar irradiance and power generation. Hybrid networks, such as CNN-LSTM, are structures that integrate both spatial and temporal feature representation and provide additional predictive power of the photovoltaic generation and irradiance forecasting tasks. The approaches discussed here are similar to those that are prevalent in scientific literature.

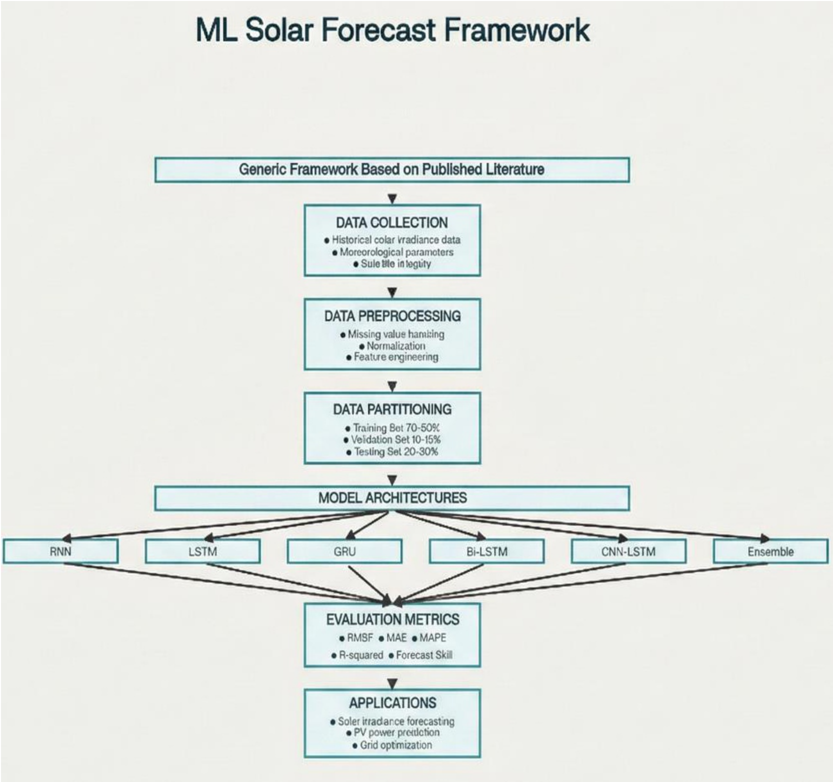

Fig.3. Flow chart

Recurrent Neural Network (RNN):

Figure 4 describes the basic form of an RNN cell. The RNN cells face problems like declining gradient over a very long sequence [10], as shown in equation (1). The RNN cells are incapable of preserving long-term dependence.

The equations for RNN are given below:

$$H_t = \sigma\,(P_h\, H_t - 1 + P_x\, X_t + B_a) \qquad (1)$$

$$Y_t = \tanh\,(P_0 * H_t + B_0) \qquad (2)$$

H_t-1 denotes the pre-hidden state, and H_t denotes the hidden state of the RNN. At the time ‘t’, input and output are represented as X_t and Y_t, individually. P_h, P_x, and P_0 denote weight matrices, and B_a and B_0 represent, respectively, the bias vector for the concealed state as shown in equations (1) and (2) [7].

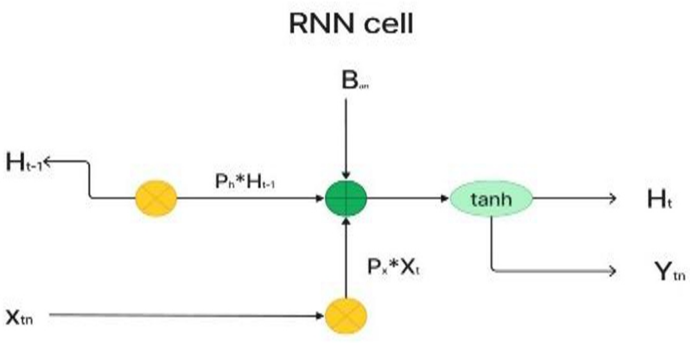

Fig.4. RNN Cell

Gated Recurrent Unit (GRU):

As mentioned in Figure 5, the important fact about GRU is that it comes under the category of RNN cells. Input X_t and the last hidden State H_{t-1} are required for GRU.

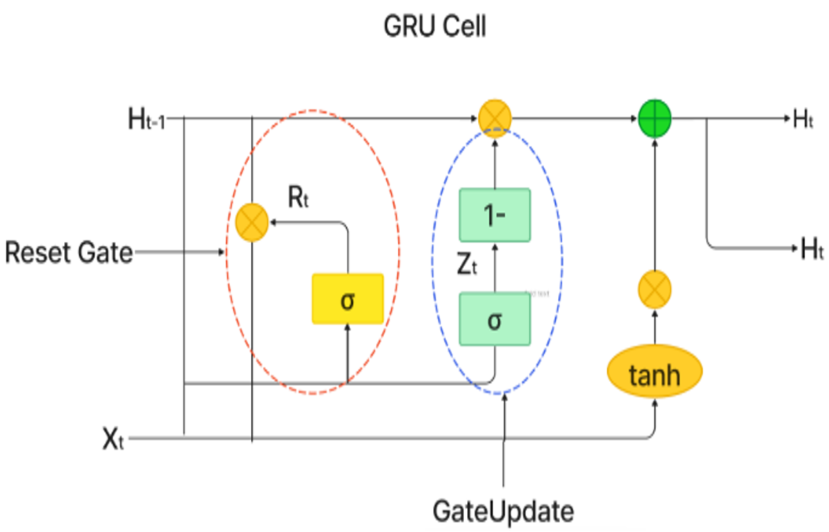

Fig.5. GRU CELL

The initial step involves calculating the value of the sigmoid function(σ), which ranges from 0 to 1, determined in equation (3).

$$R_t = [\sigma\ (W_{rh} * H_{t-1}) + (W_{rx} * X_t)] \qquad (3)$$

The second step is known as the update gate,

$$Z_t{=}\sigma\ (W_{zh}{*}H_{t-1}{+}W_{zx}{*}X_t) \qquad (4)$$

In equations (3) and (4), W_{zx} and W_{zh} are the loads of the update gate of the hidden state H_{t-1}, and X_t is the related input. The reset gate is represented by R_t, and W_{rh} and W_{rx} are the weights associated with the reset gate. [7].

To calculate H't, firstly, calculate the product of X_t and Wh'x. after that product of R_t and (H_{t-1}*Wh'h) is evaluated, and then apply function "tanh". This is shown in the equation.

$$H_t'{=}\ \tanh\ [W_{h'h}(R_t{*}H_{t-1}) + (W_{h'x}{*}X_t)] \qquad (5)$$

Two required entries are used for the converted gates: the product of Z_t and H_t`, and the product of Ht-1 and (1-Z_t) are generated, and with derivatives. The inclusion of these factors will decide the values of H_t' and H_t (5) and (6).

$$H_t = [H_{t-1}(1-Z_t) + (Z_t * H_t')] \qquad (6)$$

Long Short-Term Memory (LSTM):

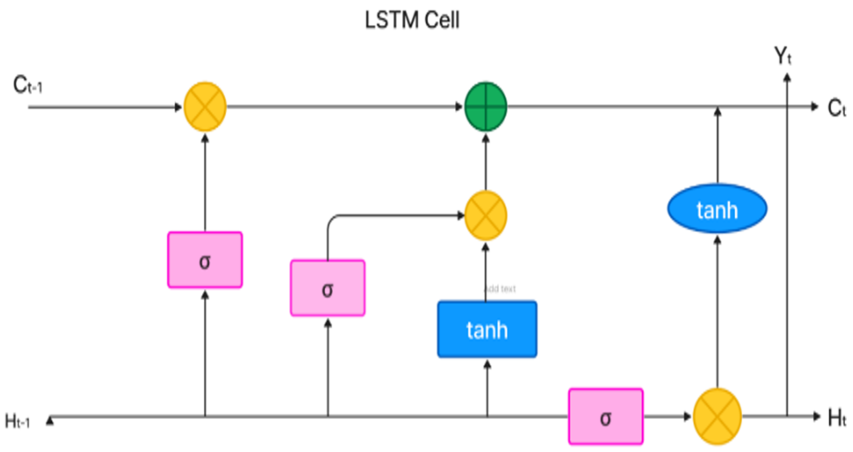

Fig.6. LSTM cell

The LSTM method was invented to deal with problems of explosive and declining gradients; the sample picture of an LSTM cell is shown in Fig. 6. LSTM has three gates, such as forget gate, input gate, and output gate. The sigmoid function takes values within 0 and 1, and it decides whether values should be memorized by product of them by one. The emission of a forget gate in the mathematical form is shown in equations (7) and (8) [15].

$$F_t = \sigma \{W_f (H_{t-1}, X_t)\} \quad (7)$$

$$It = \sigma W_i (H_{t-1}, X_t) \quad (8)$$

The input gates decide whether parts of the existing cell state (Ct) are surely associated there. The system decides the values to be changed. The first system with "tanh" layer generates additional potential state values (C't) as shown in equation (9).

$$C_t' = \tanh(W_c[H_{t-1}, X_t]) \quad (9)$$

The update gate C_t, multiplication of last cell $C_{t-1,}$ and forget gate $F_{t,}$ and multiplication of cell state C_t' with the value of the input gate that is derived in equation (10).

$$C_t = (F_t * C_{t-1}) + C_t * C_t' \quad (10)$$

Hybrid and Ensemble Architectures

New methods of solar forecasting incorporate a combination of more than one type of neural network to increase the accuracy of prediction. CNN-LSTM models combine the use of convolutional layers to extract spatial features and LSTM layers to model time sequences, as shown in the figure. 7 [16][17][18]. The logic behind hybrid designs is that convolutional layers are good at detecting spatial patterns in satellite or gridded meteorological data, whereas LSTM layers are good at detecting time series [17][18]. Hybrid designs with much better performance indicators, such as CNN-LSTM, can be applied to solar data; direct normal irradiance prediction with these models has shown an R^2 of 0.9992, an RMSE of 0.0833, and an MAE of 0.0679 [16]. These are significant gains over single-architecture models and show the interaction between spatial and temporal feature extraction.

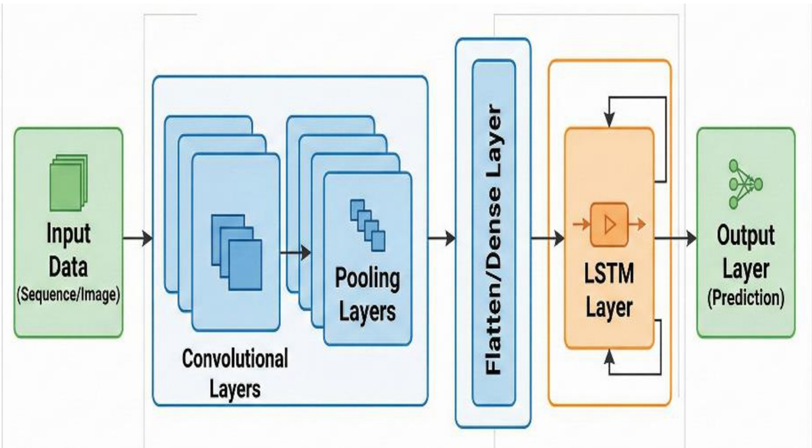

Fig.7. CNN-LSTM models

Ensemble techniques, which combine the forecasts of many different models, also find their way to the solar forecasting literature [17][18][19]. Ensemble methods make use of the synergies of the various architectures to come up with stronger and better-performing predictions. The research on the use of ensemble methods on solar irradiance prediction has found that better generalization and less prediction variance are achieved utilizing ensemble approaches than an individual model [19][20].

Data Collection and Preprocessing

Solar power prediction datasets normally combine a variety of data sources [20]. The historical solar irradiance data are obtained in various forms, which include: Global Horizontal Irradiance (GHI), Direct Normal Irradiance (DNI), and Diffuse Horizontal Irradiance (DHI) [16][20]. The measurements on the ground are complemented with meteorological measurements such as ambient temperature, relative humidity, wind speed, and atmospheric pressure [16]. Additional forecasting data is provided by satellite-derived cloud cover and aerosol optical depth measurements [20]. Ground-truth data is typically complemented by reanalysis data or satellite data like MODIS or GOES to expand the area coverage [20]. Preprocessing tasks are aimed at dealing with data quality problems and preparing the data to be used in the training of machine learning models. Other frequent preprocessing steps involve dealing with missing data using interpolation or imputation techniques that are suitable for time series data [16][19]. To bring the input values of the different variables into the same range, Min-max scaling or Z-score standardization is used to normalize the features [16][21]. The outliers are detected and eliminated or tagged based on the domain knowledge of their legitimacy. The additional information to the model can be the engineering of derived features like lagged predictors (e.g., irradiance values of the last few hours), temporal aggregations (e.g., moving averages), and rolling statistics (e.g., standard deviation) [16][19][22]. These data pre-processing procedures lead to unvarying data traits in regards to geographic locations and time spans, thereby enhancing the performance of model generalization.

Table 1.: Study results

Study	Model	RMSE	MAE	R^2
Short-term solar energy forecasting: Integrated computational intelligence of LSTMs and GRU [23]	Bi-LSTM	0.0315	0.0135	—
An ensemble method to forecast 24-h solar irradiance using wavelet decomposition and BiLSTM deep learning network [18]	WT-BiLSTM	45.61 W/m^2	—	0.94
Hybrid deep learning CNN-LSTM model for forecasting direct normal irradiance [16]	CNN-LSTM	0.0833	0.0679	0.9992

Machine learning forecasting of solar PV production using satellite imagery [24]	LSTM	—	—	0.97
The comparison of GRU and LSTM in solar power forecasting [25]	GRU	—	1.62	0.9836
The comparison of GRU and LSTM in solar power forecasting [25]	LSTM	—	1.9642	0.9733
Deep learning application in power system with a case study on solar irradiation forecasting [7]	LSTM	—	—	0.97

Data Partitioning and Model Training

Predictive models are trained and evaluated by dividing datasets into two subsets: one used in learning, and another used in validation. It is a common practice in machine learning to use 70-80 percent of data as part of the training set, and the rest 20-30 percent as part of the test set [16][21][24][26]. The model parameters are optimized with the help of the training set and the help of iterative learning algorithms. The test set, which is stored entirely during training, gives an objective evaluation of the model's performance on the unseen data. Other works use an external validation set (10-15%) to tune hyperparameters and early-stop, especially when using deep learning, where many hyperparameters must be optimized [16][21]. The validation set helps in decision-making concerning learning rate, the strength of regularization, and when to cease the training process to avoid overfitting.

6 Conclusion

Machine learning has become a transformative enabler for improving the accuracy, reliability, and efficiency of solar energy resource assessment. From classical ML techniques to advanced hybrid deep learning architectures, ML frameworks offer significant improvements in irradiance prediction, site evaluation, anomaly detection, and operational optimization. Integrating satellite imagery, climate data, and sensor information further enhances prediction capabilities. Future advancements in deep learning, explainable AI, and digital twins are expected to drive next-generation solar forecasting and smart energy systems.

References

1. Javed, A., Kasi, B.K., Khan, F.A.: Predicting solar irradiance using machine learning techniques. In: 15th International Wireless Communications & Mobile Computing Conference (IWCMC), pp. 1645–1650. IEEE, (2019)
2. Myers, D.R.: Solar radiation resource assessment for renewable energy conversion. E3S Web of Conferences.

3. Datta, A., Karakoti, I.: Solar resource assessment using GIS and remote sensing techniques. Semantic Scholar, (2010)
4. Li, J.: Assessment and rational utilization analysis of regional solar power resources. Academic Journal of Science and Technology 3(2), (2022)
5. Chaudhary, P.: Forecasting solar radiation using machine learning algorithms. North Carolina A&T State University.
6. Jäger-Waldau, A.: Solar energy and photovoltaics. In: Encyclopedia of Inorganic and Bioinorganic Chemistry, pp. 1–10. Wiley, (2015)
7. Abubakar, M., Che, Y., Faheem, M.: Prediction of solar power generation based on machine learning algorithms. In: IEEE Conference Publication. IEEE, (2024)
8. Simankov, V., Buchatskiy, P.: A solar and wind energy evaluation methodology using artificial intelligence technologies.
9. Google: Project Sunroof. http://sunroof.withgoogle.com
10. Tiwari, S., Jain, A., Sid Ahmed, N.M.O., Alkwai, L.M., Dafhalla, A.K.Y., Hamad, S.A.S.: A comprehensive review of machine learning techniques for solar power forecasting. Expert Systems with Applications 177, 113832 (2021)
11. Liang, Y., Shen, Y., Bai, L.: Brief introduction of assessment and development potential of solar power resources in South China. Journal of Applied Meteorological Science 28(4), 12 (2017)
12. Engel, E., Engel, N.: A review on machine learning applications for solar plants. MDPI 22, (2022)
13. Wald, A., et al.: Machine learning methods for solar radiation prediction—A review. Renewable and Sustainable Energy Reviews 47, 498–509 (2015)
14. Escuela de Ingeniería de Sistemas y Computación (EISC): Machine learning for solar resource assessment using satellite images. Semantic Scholar
15. Gao, M., Shi, G., Li, S.: Online prediction of ship behavior with AIS sensor data using bidirectional LSTM recurrent neural networks. Sensors 18(12), (2018)
16. Ladjal, B., Mohamed, N., Bechouat, M., Hadroug, N., Sedraoui, M., Rabehi, A., Guermoui, M., Agajie, T.: Hybrid deep learning CNN–LSTM model for forecasting direct normal irradiance: A study on solar potential in Ghardaia, Algeria. Scientific Reports 15, (2025)
17. Journal of Information Systems Engineering and Management (JISEM): Combination approach of LSTM and CNN in solar energy production prediction. JISEM Journal 10(1), (2025)
18. Singla, P., Duhan, M., Saroha, S.: An ensemble method to forecast 24-h ahead solar irradiance using wavelet decomposition and BiLSTM deep learning network. Earth Science Informatics 15(1), 291–306 (2021)
19. TechScience Journal: Deep-ensemble learning method for solar resource assessment. TechScience 38(4), 1–15 (2024)
20. MDPI Atmosphere: Machine learning dynamic ensemble methods for solar irradiance and wind speed predictions. Atmosphere 14(11), 1635 (2023)

21. Energies: A comprehensive review of smart grid technologies for renewable energy integration and future trends: The role of machine learning and energy storage systems. Energies 17(16), 4128 (2024)
22. Renewable and Sustainable Energy Reviews: A systematic review on predicting PV system parameters using machine learning. Renewable and Sustainable Energy Reviews 177, 113241 (2023)
23. Zameer, A., Jaffar, F., Shahid, F., Muneeb, M., Khan, R., Nasir, R.: Short-term solar energy forecasting using integrated computational intelligence of LSTMs and GRU. PLoS ONE 18(10), e0285410 (2023)
24. Heliyon: Machine learning forecasting of solar PV production using satellite imagery. Heliyon 10(4), e28945 (2024)
25. International Journal of Science and Research Archive (IJSRA): The comparison of GRU and LSTM in solar power forecasting. IJSRA 13(1), 1360–1370 (2024)
26. Open Access Journal of Artificial Intelligence and Machine Learning (OAJAI ML): Data-driven forecasting of solar PV output using machine learning approaches. OAJAI ML 8(2), 45–68 (2024)