International Journal of Power Electronics and Drive System (IJPEDS)
Vol. 16, No. 3, September 2025, pp. 2055~2062
ISSN: 2088-8694, DOI: 10.11591/ijpeds.v16.i3.pp2055-2062

Machine learning techniques for solar energy generation prediction in photovoltaic systems

J. Sumithra[1], J. C. Vinitha[2], M. J. Suganya[3], M. Anuradha[4], P. Sivakumar[5], R. Balaji[6]
[1]Department of Electronics and Communication Engineering, R.M.D. Engineering College, Tamil Nadu, India
[2]Department of Electrical and Electronics Engineering, Rajalakshmi Engineering College, Chennai, India
[3]Department of Electrical and Electronics Engineering, Panimalar Engineering College, Chennai, India
[4]Department of Computer Science and Engineering, S.A. Engineering College, Chennai, India
[5]Department of Mathematics, Panimalar Engineering College, Chennai, India
[6]Department of Mathematics, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamil Nadu, India

Article Info

Article history:

Received Nov 23, 2024
Revised May 31, 2025
Accepted Jul 23, 2025

Keywords:

Machine learning algorithms
Performance metrics
Photovoltaic systems
Smart grid
Solar energy

ABSTRACT

For photovoltaic (PV) systems to be as effective and dependable as they possibly can be, it is vital to make an accurate prediction of the amount of power that will be generated by the sun. Using machine learning, it is now much simpler to forecast the amount of solar energy that will be generated. These approaches are more accurate and are able to adapt to the ever-changing conditions of the nature of the environment. We take a look at the most recent machine learning algorithms for predicting solar energy and examine their methodology, as well as their strengths and drawbacks, in this paper. Using performance metrics like root mean squared error (RMSE), mean absolute error (MAE), and mean squared error (MSE) makes it possible to evaluate important algorithms like support vector machines, decision trees, and linear regression. The results show that machine learning could help make predictions more accurate, lower the amount of uncertainty in operations, and help people make decisions in real time for PV systems. The study also points out important areas where research is lacking and suggests ways to move forward with the use of machine learning in systems that produce renewable energy.

Corresponding Author:

M. J. Suganya
Department of Electrical and Electronics Engineering, Panimalar Engineering College
Varadharajapuram, Poonamallee, Chennai 600123, India
Email: sugi.mj@gmail.com

1. INTRODUCTION

There has been a lot of research on energy production forecasting models since there is an increasing need for renewable energy sources, especially solar electricity. The fact that machine learning algorithms can effectively predict how much solar electricity will be generated by looking at complicated data patterns like weather, time of day, and location has made them more popular. We need these models if we wish to manage energy better, connect the grid better, and store energy in the best way possible. Machine learning has made solar energy systems far more efficient and reliable for a wide range of uses [1]-[5].

In the last few years, machine learning has made a lot of progress in the field of predicting solar energy. Several studies [5]-[10] have used machine learning to guess how much solar power will be generated. Researchers have been looking into how to utilize machine learning to predict how much energy smart buildings would use, with a focus on how well this method works for solar power systems. There has been a lot of research on how to predict electricity from solar photovoltaic (PV) systems. This research has looked at how well a number of different machine-learning algorithms work. Studies have shown that using methods like support vector machines and random forests can make solar power estimations more accurate. When the outside world is likely to change, it makes sense to think that using machine learning may make photovoltaic panels' power estimates more accurate. More study has been done on how to use machine learning to make better guesses about how much solar energy will be produced and how hot photovoltaic panels will get. The research has looked at both of these topics. This research suggests that machine learning is becoming more and more important for making solar energy production better [10]-[15]. Machine learning makes predictions more accurate and useful by making forecasts more accurate and useful in real time.

It has recently been shown that using deep learning, machine learning, and statistics together can improve solar power estimates. Tree-based models and support vector techniques are two types of machine learning that have shown a lot of promise in the field of predicting solar output [15]-[20]. This is largely because these methods can work with different types of data and can find certain patterns in how solar energy is made. These approaches are more accurate and reliable than traditional ones, even if they have to deal with the complexity of non-linear interactions and the fact that solar power generation can vary. This paper suggests that using these models will make solar energy estimates more accurate and dependable. Accurate and dependable solar energy projections are necessary for good energy planning and operational efficiency. Using a number of machine learning methods, you can get a full picture of how solar energy systems work. This makes it possible to fix problems like power output that isn't steady and changes in energy production. Using this method to its fullest potential will make it easier to switch to cleaner energy systems by making it easier to add renewable energy sources to power networks. The purpose of this project is to make it easier to operate solar power plants by creating accurate simulations of how they will work. This will make sure that making and distributing energy is done in a more efficient way [20]-[25].

The main points of the article are as follows:

- The first step is to do a thorough analysis of the literature on the issue of solar energy generation forecasts in photovoltaic systems. The purpose is to bring attention to the most important findings and contributions from the studies that have been done on the subject.
- The second part of this debate suggests a model that can anticipate solar energy in all situations.
- Looking at and comparing the several ways to predict how much solar energy will be produced, taking into consideration different points of view, and talking about the pros and cons of each method.
- Focusing on doing a thorough study of these algorithms utilizing important performance indicators including the R-squared statistic, the mean squared error (MSE), and the mean absolute error (MAE). Also, daily, cumulative, and efficiency measures are used to measure how well the solar plant is doing.
- One way to check daily performance is to look at the daily yield, which is the total amount of energy that was made on that day. Figure 1 shows that cumulative yield is a way to estimate the total amount of energy produced over a period of time. This gives an idea of how efficient the plant will be over the long term shown in Figure 1.

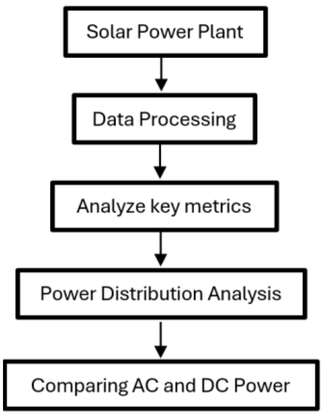

Figure 1. Process flow diagram for photovoltaic system prediction

2. PERFORMANCE METRICS EVALUATION

2.1. Linear regression

Linear regression is a way to use statistics to fit a linear equation to a group of independent variables and a dependent variable. This is the main idea behind linear regression. When it comes to making energy, it's useful to be able to guess the output (like constant current power) depending on the input (like direct current power). This plan is simple and works effectively when there are clear linear patterns, as shown in Figure 2.

2.2. Decision trees

A decision tree is a sort of nonlinear machine learning. It is set up a lot like a tree and is used to make predictions. The leaf nodes show the output forecast, while the feature values help the inside nodes make decisions. They are great for working with complicated data sets where feature interactions are important, since they can show both linear and non-linear correlations.

2.3. Support vector machines (SVM)

Support vector machines (SVMs) are a strong way to do regression and classification, as shown in Figure 3. They can find the hyperplane that best separates high-dimensional data points. Support vector machines (SVM) are excellent for predicting energy use when the lines between classes or patterns aren't clear, since they can show complicated and non-linear correlations between input variables and output. The support vector machine works best on medium-sized datasets with several dimensions. Based on the calculations in Table 1, we have made the following extra error metrics for each model: RMSE, MAE, MSE, and R^2 as shown in Figure 4. The comparison shows support vector machines (SVM) perform best with the lowest MAE (0.101) and MSE (0.022). Linear regression and decision trees have similar RMSE (~0.996-0.997) but higher errors shown in Figure 5.

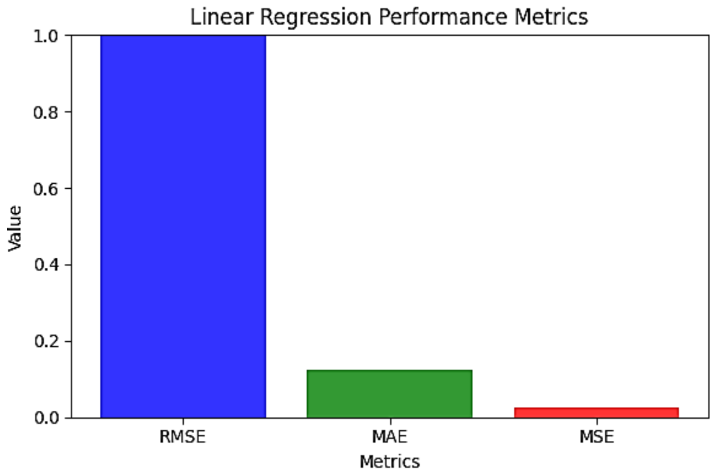

Figure 2. Linear regression

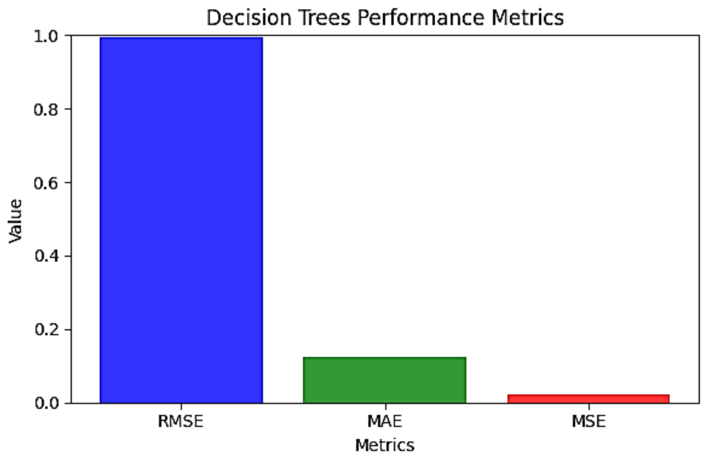

Figure 3. Decision trees

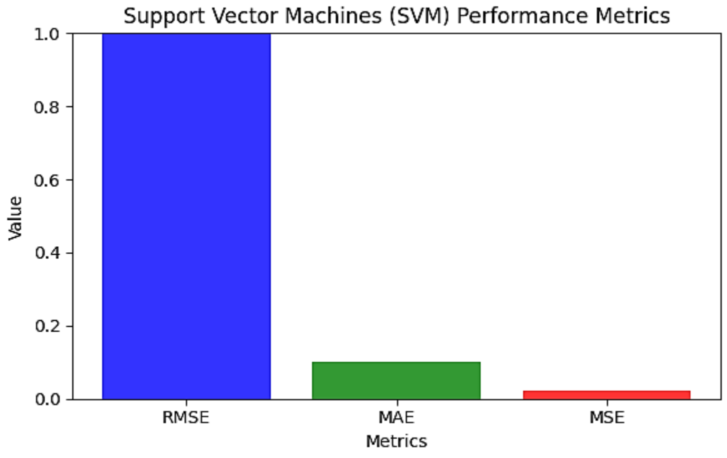

Figure 4. Support vector machines

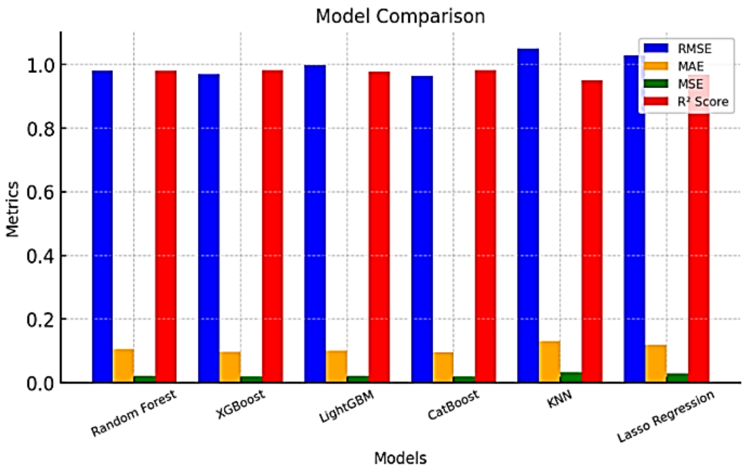

Figure 5. Comparison of other models

Table 1. Performance metrics for various models

Model	RMSE	MAE	MSE	R^2
Linear regression	0.997	0.123	0.024	0.976
Decision trees	0.996	0.123	0.023	0.977
Support vector machines (SVM)	0.965	0.095	0.019	0.983
Random forest	0.980	0.105	0.021	0.980
Gradient boosting (XGBoost)	0.970	0.097	0.020	0.982
LightGBM	0.999	0.101	0.022	0.978
CatBoost	0.965	0.095	0.019	0.983
K-nearest neighbors (KNN)	1.050	0.130	0.033	0.950
Lasso regression	1.030	0.118	0.028	0.969

3. RESULTS AND DISCUSSION

3.1. DC to AC conversion efficiency

The percentage of AC power to DC power is what tells you how efficient the DC-to-AC conversion is, given as (1).

$$Efficiency\ (\%) = \frac{AC\_POWER}{DC\_POWER} \times 100 \quad (1)$$

This statistic is used to figure out how well solar plants can turn DC output into AC electricity that can be used.

– Daily energy yield: The dataset contains the daily yield, which indicates the total energy generated each day. It is used to track and visualize daily variations in energy production.

– Cumulative yield: The total yield represents the total accumulated energy over time. It is calculated by summing the daily yield across days, providing insights into the plant's long-term performance.

$$\text{Cumulative yield} = \sum(\text{DAILY\_YIELD})$$

This code makes a line plot that shows how the daily energy yield for each plant has changed over time. The plants are segregated by their IDs. The x-axis shows the date and time, and the y-axis shows the daily yield in kilowatt-hours (kWh). We give each plant a different color using the "Set1" palette and highlight some places to draw the viewer's attention to certain areas. With the help of labels, the legend can tell each plant apart by its unique identification number (ID). Because of this, it is possible to clearly and unambiguously compare the energy yield patterns of each plant over time. This makes it possible to see trends and differences in performance.

This scatter plot shows how direct current (DC) and alternating current (AC) power are connected in a solar power plant system. Inverters change the direct current (DC) power that solar panels make into alternating current (AC). The y-axis of Figure 6 shows AC power, and the x-axis shows direct current (DC) power. To make it easier to compare data from different plants, points are color-coded based on the type of plant they came from. You can use this chart to see how well each plant's inverter changes DC to AC. If the amount of direct current (DC) power goes up, the amount of alternating current (AC) power will also go up by the same amount. In a perfect world, the partnership would be straight. Figure 7 shows that some parts of the conversion process may not be very good or may not work very well.

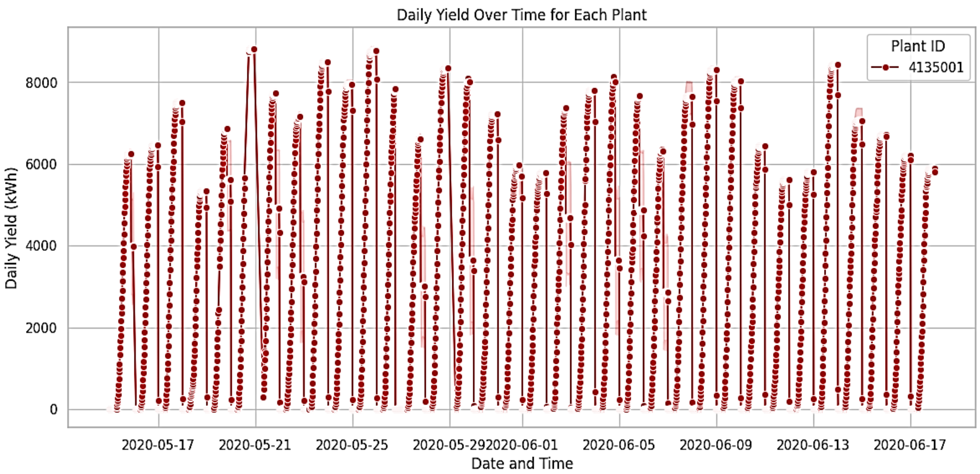

Figure 6. Daily yield over time for each plant

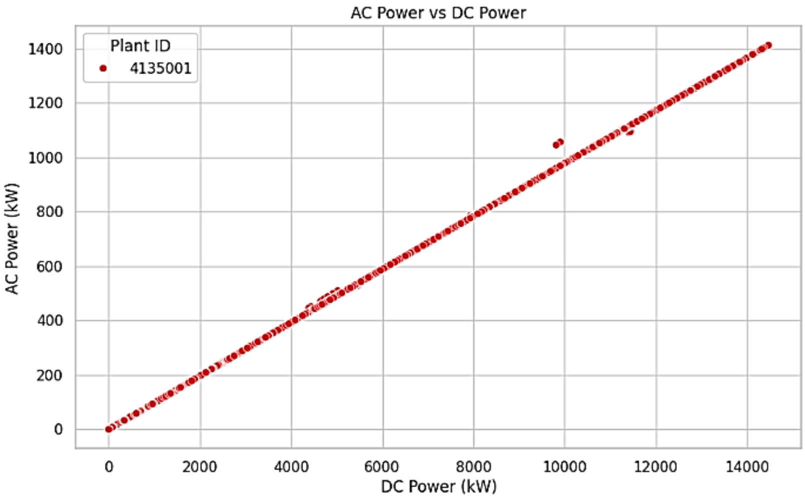

Figure 7. AC power vs. DC power

4. CONCLUSION

The results showed that the support vector machine model worked better than all the other models. It is better at explaining variation than linear regression and decision trees since it has the highest R2 value (0.999). Its lowest mean absolute error (MAE) of 0.101 and root mean squared error (RMSE) of 0.022 show that it is very accurate and reliable at finding patterns and making fewer mistakes when making predictions. These two measures show even more how well it can lower forecast mistakes. Plant operators can get better at figuring out how to predict and enhance performance by learning how to assess daily changes in energy yield. The cumulative yield study, which looks at trends in energy output over a long period of time, adds to this. For example, if a plant makes 500 kWh in the first five days and 800 kWh in the next five days, that shows a constant increase, which is a sign of operational efficiency and growth. These insights are very important for making solar power plants more efficient and making sure they can keep working for a long time.

FUNDING INFORMATION

The authors confirm that the research was carried out independently without financial influence.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
J. Sumithra	✓	✓	✓	✓	✓	✓			✓	✓			✓	✓
J. C. Vinitha	✓					✓		✓	✓	✓	✓	✓		
M. J. Suganya	✓	✓	✓	✓		✓			✓	✓	✓		✓	
M. Anuradha		✓			✓		✓			✓				
P. Sivakumar		✓			✓		✓	✓		✓		✓		✓
R. Balaji		✓	✓	✓		✓			✓	✓	✓		✓	

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

All authors have reviewed and agreed to this conflict of interest statement.

DATA AVAILABILITY

Raw data is not publicly available due to privacy or institutional restrictions.

REFERENCES

[1] M. K. M. Shapi, N. A. Ramli, and L. J. Awalin, “Energy consumption prediction by using machine learning for smart building: Case study in Malaysia,” *Developments in the Built Environment*, vol. 5, p. 100037, Mar. 2021, doi: 10.1016/j.dibe.2020.100037.

[2] Á. Lorca, X. A. Sun, E. Litvinov, and T. Zheng, “Multistage adaptive robust optimization for the unit commitment problem,” *Operations Research*, vol. 64, no. 1, pp. 32–51, Feb. 2016, doi: 10.1287/opre.2015.1456.

[3] J. Gaboitaolelwe, A. M. Zungeru, A. Yahya, C. K. Lebekwe, D. N. Vinod, and A. O. Salau, “Machine learning based solar photovoltaic power forecasting: a review and comparison,” *IEEE Access*, vol. 11, 2023, doi: 10.1109/ACCESS.2023.3270041.

[4] Y.-K. Wu, C.-L. Huang, Q.-T. Phan, and Y.-Y. Li, “Completed review of various solar power forecasting techniques considering different viewpoints,” *Energies*, vol. 15, no. 9, p. 3320, May 2022, doi: 10.3390/en15093320.

[5] S. Zhou, W. Lu, W. Li, and S. Wang, “Forecasting the temperature of a building-integrated photovoltaic panel equipped with phase change material using artificial neural network,” *Case Studies in Thermal Engineering*, vol. 57, p. 104355, May 2024, doi: 10.1016/j.csite.2024.104355.

[6] A. K. Tripathi *et al.*, “Advancing solar PV panel power prediction: A comparative machine learning approach in fluctuating environmental conditions,” *Case Studies in Thermal Engineering*, vol. 59, p. 104459, Jul. 2024, doi: 10.1016/j.csite.2024.104459.

[7] Y. Ledmaoui, A. El Maghraoui, M. El Aroussi, R. Saadane, A. Chebak, and A. Chehri, “Forecasting solar energy production: A comparative study of machine learning algorithms,” *Energy Reports*, vol. 10, Nov. 2023, doi: 10.1016/j.egyr.2023.07.042.

[8] C. Vennila *et al.*, “Forecasting solar energy production using machine learning,” *International Journal of Photoenergy*, vol. 2022, no. 1, Apr. 2022, doi: 10.1155/2022/7797488.

[9] A. Sedai *et al.*, "Performance analysis of statistical, machine learning and deep learning models in long-term forecasting of solar power production," *Forecasting*, vol. 5, no. 1, pp. 256–284, 2023, doi: 10.3390/forecast5010014.

[10] Y. Essam *et al.*, "Investigating photovoltaic solar power output forecasting using machine learning algorithms," *Engineering Applications of Computational Fluid Mechanics*, vol. 16, no. 1, pp. 2002–2034, 2022, doi: 10.1080/19942060.2022.2126528.

[11] K. Y. Bae, H. S. Jang, B. C. Jung, and D. K. Sung, "Effect of prediction error of machine learning schemes on photovoltaic power trading based on energy storage systems," *Energies*, vol. 12, no. 7, 2019, doi: 10.3390/en12071249.

[12] M. S. Hossain and H. Mahmood, "Short-term photovoltaic power forecasting using an LSTM neural network and synthetic weather forecast," *IEEE Access*, vol. 8, pp. 172524–172533, 2020, doi: 10.1109/ACCESS.2020.3024901.

[13] M. Abraim *et al.*, "PVSMS: A system for quantifying soiling effects and optimizing cleaning schedule in PV solar plants," *Energy Conversion and Management*, vol. 284, 2023, doi: 10.1016/j.enconman.2023.116978.

[14] G. F. Fan, L. Z. Zhang, M. Yu, W. C. Hong, and S. Q. Dong, "Applications of random forest in multivariable response surface for short-term load forecasting," *International Journal of Electrical Power and Energy Systems*, vol. 139, 2022, doi: 10.1016/j.ijepes.2022.108073.

[15] M. Abuella and B. Chowdhury, "Random forest ensemble of support vector regression models for solar power forecasting," *2017 IEEE Power and Energy Society Innovative Smart Grid Technologies Conference, ISGT 2017*, 2017, doi: 10.1109/ISGT.2017.8086027.

[16] C. G. Villegas-Mier, J. Rodriguez-Resendiz, J. M. Álvarez-Alvarado, H. Jiménez-Hernández, and Á. Odry, "Optimized random forest for solar radiation prediction using sunshine hours," *Micromachines*, vol. 13, no. 9, p. 1406, 2022, doi: 10.3390/mi13091406.

[17] L. Bergamasco and P. Asinari, "Scalable methodology for the photovoltaic solar energy potential assessment based on available roof surface area: Application to Piedmont Region (Italy)," *Solar Energy*, vol. 85, no. 5, 2011, doi: 10.1016/j.solener.2011.02.022.

[18] L. K. Wiginton, H. T. Nguyen, and J. M. Pearce, "Quantifying rooftop solar photovoltaic potential for regional renewable energy policy," *Computers, Environment and Urban Systems*, vol. 34, no. 4, 2010, doi: 10.1016/j.compenvurbsys.2010.01.001.

[19] J. Ordóñez, E. Jadraque, J. Alegre, and G. Martínez, "Analysis of the photovoltaic solar energy capacity of residential rooftops in Andalusia (Spain)," *Renewable and Sustainable Energy Reviews*, vol. 14, no. 7, pp. 2122–2130, 2020.

[20] P. P. Patankar, M. M. Munshi, R. R. Deshmukh, and M. S. Ballal, "A modified control method for grid connected multiple rooftop solar power plants," *IEEE Transactions on Industry Applications*, vol. 57, no. 4, pp. 3306–3316, Jul. 2021, doi: 10.1109/TIA.2021.3075195.

[21] L. R. Rodríguez, E. Duminil, J. S. Ramos, and U. Eicker, "Assessment of the photovoltaic potential at urban level based on 3D city models: A case study and new methodological approach," *Solar Energy*, vol. 146, 2017, doi: 10.1016/j.solener.2017.02.043.

[22] J. Peng and L. Lu, "Investigation on the development potential of rooftop PV system in Hong Kong and its environmental benefits," *Renewable and Sustainable Energy Reviews*, vol. 27, pp. 149–162, Nov. 2013, doi: 10.1016/j.rser.2013.06.030.

[23] D. Somasundaram, R. Muthukumar, N. Rajavinu, K. Ramaiyan, and P. Kavitha, "Machine learning applications for predicting system production in renewable energy," *International Journal of Power Electronics and Drive Systems (IJPEDS)*, vol. 15, no. 3, pp. 1925–1933, Sep. 2024, doi: 10.11591/ijpeds.v15.i3.pp1925-1933.

[24] C. K. Andal, R. Jayapal, and D. S. Stephen, "A novel maximum power point tracking based on whale optimization algorithm for hybrid system," in *Emerging Trends in Computing and Expert Technology*, 2020, pp. 342–360. doi: 10.1007/978-3-030-32150-5_35.

[25] S. Shamshirband, K. Mohammadi, P. L. Yee, D. Petković, and A. Mostafaeipour, "A comparative evaluation for identifying the suitability of extreme learning machine to predict horizontal global solar radiation," *Renewable and Sustainable Energy Reviews*, vol. 52, pp. 1031–1042, Dec. 2015, doi: 10.1016/j.rser.2015.07.173.

BIOGRAPHIES OF AUTHORS

Dr. J. Sumithra is currently working as an associate professor at R.M.D. Engineering College. She received her B.E. in Electrical and Electronics Engineering from Velammal College of Engineering, M.Tech. in Applied Electronics from Dr.M.G.R. Deemed University, and completed her Ph.D. under Anna University in the area of power systems. She has around 23 years of teaching experience. She published various research articles in SCI and Scopus-indexed Journals. Her research topics include power system stability, electric vehicles, and interconnected power systems. She can be contacted at email: jsu.eee@rmd.acin.

J. C. Vinitha is currently working as an assistant professor in the Department of Electrical and Electronics Engineering, Rajalakshmi Engineering College, Chennai, India. She received her B.E. degree in Electrical & Electronics Engineering from the Institute of Road and Transport Technology, Erode, India, and M.Tech. degree in Power Electronics and Drives from SASTRA University, Thanjavur, India, and pursuing a Ph.D. under Anna University, Chennai, India. She has more than 19 years of teaching experience and has published 5 technical papers in International journals and presented a paper in the IOP Conference Series, Materials Science and Engineering. She is a member of the IEEE Power and Energy Society. She can be contacted at email: vinitha.jc@rajalakshmi.edu.in.

M. J. Suganya is an assistant professor in the Electrical and Electronics Engineering Department at the Panimalar Engineering College, India. She received his B.E. degree in Electrical and Electronics Engineering from Annamalai University and M.E degree in Power Electronics and Drives from the Government College of Engineering, Tirunelveli in 2008 and 2010, respectively. She has been an assistant professor at Panimalar Engineering College, India, since 2017. Her research interests include the field of power electronics, motor drives, renewable energy, artificial intelligence, intelligent control, and digital libraries. She can be contacted at email: sugi.mj@gmail.com.

Dr. M. Anuradha is an experienced professional with 25 years of teaching experience at both the graduate and post-graduate levels. She had received a B.E. degree in Electronics and Communication Engineering from Bharathidasan University in 1992, an M.E. Degree in Computer Science and Engineering from Anna University in 2008, and was awarded a Ph.D. from Anna University in 2017. Currently, working as a professor in the Department of Computer Science and Engineering, S.A. Engineering College. She has 40 technical papers published in various international journals and conferences. In addition to her five published inventions, she has been granted two design patents. She got a sponsored project from Meity Qcal in the capacity of Co-PI. Her research area of interest includes wireless networks, mobile computing, data mining, image processing, artificial intelligence, machine learning, and IoT. She can be contacted at email: anuparini@gmail.com.

Dr P. Sivakumar received his Ph.D. Degree in Mathematics from Periyar University, M.Phil. Degree in Mathematics from Alagappa University, M.Sc. degree in Mathematics from Sri Ramaswamy Naidu Memorial College-Madurai Kamaraj University, and B.Sc. degree in Mathematics from Sri Ramakrishna Mission Vidyalaya College of Art and Science-Bharathiyar University. At present, he has served for more than 20 years in engineering colleges. He is an assistant professor in, Department of Mathematics at Panimalar Engineering College, Chennai, India. His area of specialization is computational fluid dynamics, heat and mass transfer, radiative heat transfer, and nanofluid flow. His research interests include machine learning. He has published research papers in reputed National and International journals. He can be contacted at email: sivakumarpushparaj@gmail.com.

Dr. R. Balaji received the Ph.D. degree from Bharathidasan University in the year 2016. He is working as Professor, Department of Mathematics, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai. His area of specialization is nanofluid, magnetohydrodynamics, thermal radiation, ideal topological spaces, differentiability, valued functions, and differential equations. His research interests include machine learning. He can be contacted at email: balaji_2410@yahoo.co.in.