Enhancing solar radiation forecasting using machine learning algorithms

Mahesh Kumar K. M.[1], Uppuluri Lakshmi Soundharya[2], R. Hemalatha[3], Anjanappa C.[4], Suganya M. J.[5]
[1]Department of Electrical and Electronics Engineering, PES College of Engineering, Mandya, India
[2]Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, India
[3]Department of Computer Science and Engineering, St. Joseph's College of Engineering, OMR, Chennai, India
[4]Department of Electronics and Communication Engineering, The National Institute of Engineering, Mysore, India
[5]Department of Electrical and Electronics Engineering, Panimalar Engineering College, Chennai, India

Article Info

Article history:

Received Apr 16, 2024
Revised Nov 27, 2024
Accepted Feb 27, 2025

Keywords:

ARIMA
Forecasting
Machine learning
Performance metrics
Solar radiation

ABSTRACT

With the increasing amount of photovoltaic (PV) generation, accurate solar radiation forecasting is essential to the safe operation of power systems. This work examines many machines learning (ML) techniques that use both exogenous and endogenous inputs to forecast sun radiation. In order to find pertinent input parameters and their values based on previous observations, the forecasting models' performance is assessed using metrics like mean absolute error (MAE), mean squared error (MSE), R-squared (R2), and root mean squared error (RMSE). Accurate power output forecasting is becoming more and more necessary as the need to switch to renewable energy sources (RES) like solar and wind power grows. There is a clear demand for more reliable solutions because current models frequently struggle with temporal complexity and noise. A revolutionary deep learning-based technique designed especially for green energy power forecasting was developed in response. The study uses time series smoothing and the autoregressive integrated moving average (ARIMA) model for casing in order to create a solid basis for analysis and modeling that is free of noise and outliers. The proposed method aims to address the limitations of existing forecasting methods and promote the creation of more accurate and reliable forecasts in the field of renewable energy.

Corresponding Author:

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

1. INTRODUCTION

In the global shift to renewable energy sources (RES), solar energy consumption has grown in significance and sustainability. As of 2022, solar energy accounted for 12.8% of global energy capacity, according to figures from the International Energy Agency. By 2027, it is expected to have overtaken all other energy sources, accounting for 20% of global energy capacity [1]–[5]. Solar radiation has a lot of potential, but grid integration and energy management are severely hampered by its erratic and variable nature. For both short- and long-term planning, accurate photovoltaic (PV) electricity generation projection is essential. PV output is influenced by the amount of solar radiation reaching the PV cells, which in turn is affected by weather conditions such as temperature, wind speed, and cloud cover. These variables complicate energy system planning and management.

Accurate and trustworthy PV output projections are becoming more and more necessary as solar energy use rises. Market participation, grid integration, system stability, energy management, and research and development projects are just a few of the applications that depend on accurate forecasting. Recently, machine learning (ML) methods have gained prominence in PV forecasting due to advancements in processing power, offering the potential for more accurate and efficient forecasting models [5]–[10].

Historically, statistical time-series models like autoregressive integrated moving average (ARIMA), SARIMA, and exponential smoothing approaches have been used to forecast power generation from RES. These methods, which primarily address short-term forecasting needs, use historical power data to identify seasonal and temporal trends. However, they fall short in capturing the complex and nonlinear interactions between various factors, such as weather patterns and fluctuations in power supply, that influence power generation. Additionally, the assumptions of stationarity and linearity inherent in these methodologies may not apply to the dynamics of solar power generation. Despite these limitations, their simplicity and ease of use have maintained their appeal [10]–[15].

The limitations of traditional statistical methods raise concerns about their suitability for solar power forecasting. There is a pressing need for more sophisticated and accurate forecasting techniques that can account for the complexities inherent in renewable energy dynamics. Advancing beyond conventional approaches is imperative to meet the evolving demands for reliable and accurate power forecasting [16]–[18].

This essay provides an in-depth analysis of improving solar radiation forecasting using ML algorithms. A detailed research methodology outlines the step-by-step development of the study. The initial stage involves exploring the principles of ARIMA modeling and conducting a thorough literature review to establish ARIMA as a benchmark model for assessing various strategies and predicting renewable energy generation. The second section examines the state of RES utilization in selected countries and their approaches to transitioning toward sustainable energy systems. This background sets the stage for the subsequent analysis. In the third stage, ARIMA models are constructed for two different time series hourly and monthly data at two different locations, followed by corresponding forecasts. The last phase entails a thorough examination of the information and the development of perceptive judgments [19]–[21]. This systematic approach guarantees a thorough examination of ARIMA modeling, its applicability for forecasting solar radiation, and its consequences in relation to the larger framework of policies for the development of renewable energy [22]–[26].

2. MACHINE LEARNING APPROACHES FOR SOLAR RADIATION FORECASTING

A number of important elements need to be taken into account when using ML techniques for solar radiation forecasting in order to improve the models' efficacy and accuracy. Key among these is the selection of relevant variables to measure, which can include meteorological data such as temperature, humidity, wind speed, and sunshine duration. Choosing the right variables ensures that the model captures all relevant influences on solar radiation. The duration for which data is recorded continuously also impacts model performance. Longer periods of data collection provide a more comprehensive understanding of seasonal and temporal variations in solar radiation, which can improve forecasting accuracy. Additionally, establishing the appropriate time resolution is crucial; data should be collected at intervals that balance detail and manageability. For example, high-frequency data may capture short-term fluctuations better but require more processing power.

Several important measures are frequently used to assess how well ML models anticipate solar radiation. The average absolute difference between expected and actual values is calculated using mean absolute error (MAE), which provides a straightforward indicator of prediction accuracy and interpretability. The average squared difference between expected and actual values is measured by mean squared error (MSE), which gives larger errors more weight and sheds light on the overall performance of the model. The coefficient of determination, or R-squared, measures how much of the variance in the dependent variable can be predicted from the independent variables, indicating how well the model explains the data. When combined, these indicators offer a thorough assessment of the model's performance, pointing out both its predicting strengths and shortcomings. They are crucial for refining models to ensure they deliver reliable and actionable insights in solar radiation forecasting shown in Figure 1.

3. ARIMA METHOD

Approach stochastic process models fall under the broad field of ARIMA models. These models are expansions of integrated ARMA (autoregressive moving average) models. Component (I) to deal with non-stationary problems in the data. Two main parts make up ARMA models for time series: regression analysis (AR): regression modeling based on the time series's lagged values is used in this component. Moving average (MA): this part combines the prior error terms linearly to describe the error term. Application of ARMA modeling approaches is made easier by the integration component in ARIMA models, which plays a crucial role in converting non-stationary time series data into stationary ones. This integration process helps in capturing the underlying patterns and trends within the data, making ARIMA models a powerful tool for time series analysis and forecasting.

$$y_t = \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{j=1}^{q} \theta_j e_{t-j} \tag{1}$$

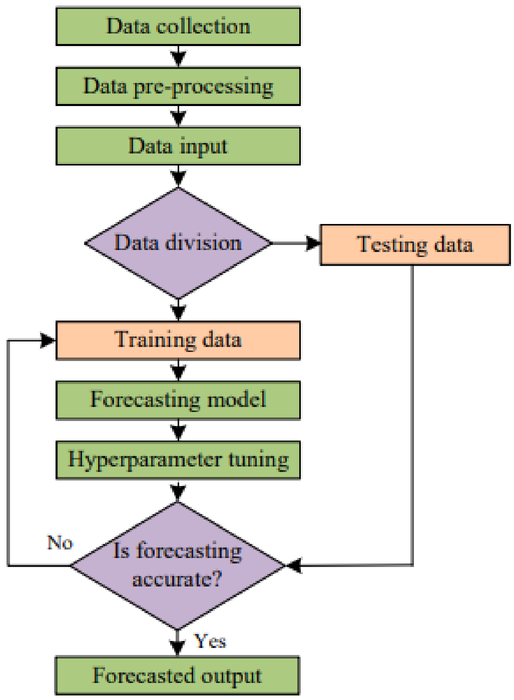

Figure 1. Process flow diagram for load forecasting

4. ERROR METRICS AND COMPUTATIONAL PERFORMANCE

Model accuracy is frequently assessed using metrics such as MAE and root mean squared error (RMSE). Nevertheless, there is a scale-dependency in these metrics, which makes the conclusions inconsistent between time series of different magnitudes. The mean absolute scaled error (MSE), an accuracy statistic, was established in order to alleviate this issue. By scaling the error in relation to a naive forecast, the MSE offers a more consistent assessment that facilitates insightful comparisons across various time series datasets.

$$MSE = \frac{MAE}{MAE_{in-sample,naive}} \tag{2}$$

where MAE, a frequently used accuracy statistic, stands for mean absolute error.

$$MAE = \frac{1}{N} \sum_{i=0}^{N-1} |\hat{y_i}[t] - y_i\ [t]| \tag{3}$$

A statistical indicator of how much of the variability in a dependent variable can be accounted for by the independent variables in a regression model is the coefficient of determination, also known as R-squared (R^2). It serves as a barometer for the model's adherence to the data. The R2 statistic, which provides information on how well the model captures the patterns in the data, is computed by dividing the explained variance by the overall variance.

$$R^2 = 1 - \frac{\sum_{i=1}^{k}(yi - \hat{yi})^2}{\sum_{i=1}^{k}(yi - \bar{y}i\,)^2} \tag{4}$$

Where, $\bar{y}$i -Actual values.

The code provided uses the Seaborn package to create a heatmap that shows the dataset Table 1 correlation matrix. The heatmap uses color to indicate the direction and intensity of correlations to show the associations between various variables. Lighter hues indicate significant positive associations, while darker hues indicate strong negative correlations. The correlation coefficients are shown in the annotations inside each cell. Through the identification of patterns and correlations between variables, this visualization technique facilitates feature selection, detects multicollinearity, and directs additional study. In data exploration and modeling procedures, the heat map provides a clear summary of the relationships within the dataset, enabling informed decision-making Figure 2.

Table 1. Metrics for performance evaluation

Sl.no	Metric	Value
1.	MAE	272.926791
2.	MSE	450,772.835729
3.	RMSE	671.396184
4.	R-squared	0.796661

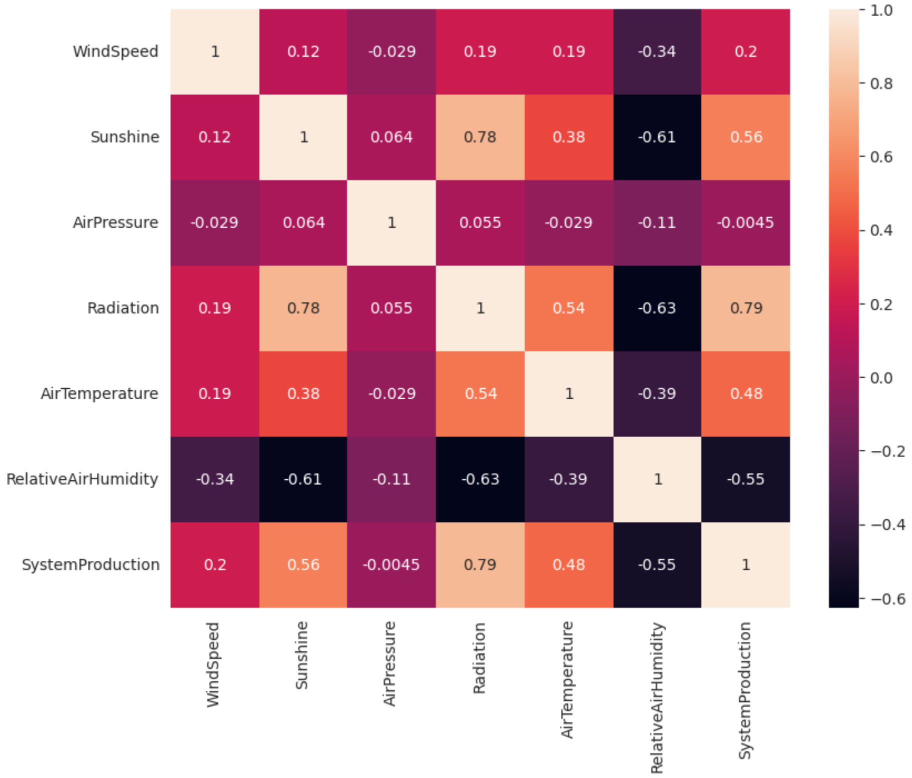

Figure 2. Heatmap for data correlation

5. RESULT

5.1. Energy prediction every month

The code generates two insightful line plots. Figure 3 illustrates the average monthly energy production, with each point reflecting the mean energy output for that month. This plot effectively highlights seasonal trends and variations, revealing peak production months and fluctuations in energy generation. Such analysis is essential for understanding seasonal demand patterns, optimizing resource allocation, and refining energy management strategies.

Figure 4 compares energy production with meteorological factors wind speed, sunshine duration, air pressure, and humidity across the dataset. Each line represents a variable's values over time, allowing for an in-depth analysis of how these environmental conditions influence energy output. This multi-variable plot helps identify correlations and dependencies, such as how increased sunshine duration may boost energy production or how higher humidity might reduce output. The inclusion of a legend aids in distinguishing between variables, enhancing interpretability.

Figure 4 compares energy production with meteorological factors wind speed, sunshine duration, air pressure, and humidity across the dataset. Each line represents a variable's values over time, allowing for an in-depth analysis of how these environmental conditions influence energy output. This multi-variable plot helps identify correlations and dependencies, such as how increased sunshine duration may boost energy production or how higher humidity might reduce output. The inclusion of a legend aids in distinguishing between variables, enhancing interpretability. Figure 4 compares energy production with meteorological factors wind speed, sunshine duration, air pressure, and humidity across the dataset. Each line represents a variable's values over time, allowing for an in-depth analysis of how these environmental conditions influence energy output. This multi-variable plot helps identify correlations and dependencies, such as how increased sunshine duration may boost energy production or how higher humidity might reduce output. The inclusion of a legend aids in distinguishing between variables, enhancing interpretability.

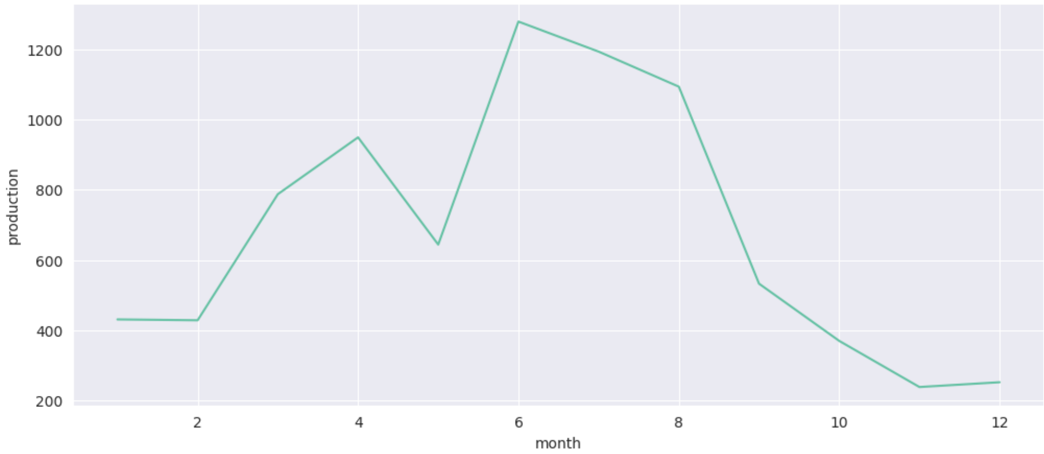

Figure 3. Predition of energy production every month

Figure 4. Predicted production VS wind speed, sunshine, air pressure and humidity energy month

6. CONCLUSION

With a MAE of 272.93, MSE of 450,772.84, RMSE of 671.40, and an R-squared value of 0.7967, the evaluation of solar radiation forecasting algorithms demonstrates respectable predictive accuracy. These results highlight the potential of combining deep learning with time series smoothing and ARIMA modeling to enhance forecast accuracy. This hybrid approach addresses limitations of traditional models, contributing to better energy management and grid stability, essential for the integration of solar power into energy systems. Future research should focus on noise reduction, improving temporal dynamics, real-time forecasting, integration with other renewable sources, and scalability across diverse regions and climates. By advancing these fields, forecasting methods will be further improved, promoting the shift to sustainable energy and facilitating the smooth integration of solar energy into global energy systems.

FUNDING INFORMATION

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

AUTHOR CONTRIBUTIONS STATEMENT

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Mahesh Kumar K. M.		✓	✓	✓			✓	✓					✓	✓
Uppuluri Lakshmi Soundharya	✓							✓			✓	✓		
R. Hemalatha			✓	✓		✓					✓		✓	
Anjanappa C.							✓		✓			✓	✓	
Suganya M. J.					✓		✓	✓		✓		✓		✓

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] F. Rodríguez, I. Azcárate, J. Vadillo, and A. Galarza, "Forecasting intra-hour solar photovoltaic energy by assembling wavelet based time-frequency analysis with deep learning neural networks," *International Journal of Electrical Power and Energy Systems*, vol. 137, 2022, doi: 10.1016/j.ijepes.2021.107777.

[2] C. Voyant *et al.*, "Machine learning methods for solar radiation forecasting: a review," *Renewable Energy*, vol. 105, pp. 569–582, May 2017, doi: 10.1016/j.renene.2016.12.095.

[3] D. P. Mishra, S. Jena, R. Senapati, A. Panigrahi, and S. R. Salkuti, "Global solar radiation forecast using an ensemble learning approach," *International Journal of Power Electronics and Drive Systems (IJPEDS)*, vol. 14, no. 1, pp. 496–505, 2023, doi: 10.11591/ijpeds.v14.i1.pp496-505.

[4] C. K. Andal, R. Jayapal, and D. S. Stephen, "A novel maximum power point tracking based on whale optimization algorithm for hybrid system," *Emerging Trends in Computing and Expert Technology*, vol. 35, pp. 342–360, 2020, doi: 10.1007/978-3-030-32150-5_35.

[5] 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, 2015, doi: 10.1016/j.rser.2015.07.173.

[6] A. Aybar-Ruiz *et al.*, "A novel grouping genetic algorithm-extreme learning machine approach for global solar radiation prediction from numerical weather models inputs," *Solar Energy*, vol. 132, pp. 129–142, 2016, doi: 10.1016/j.solener.2016.03.015.

[7] Y. Wu and J. Wang, "A novel hybrid model based on artificial neural networks for solar radiation prediction," *Renewable Energy*, vol. 89, pp. 268–284, 2016, doi: 10.1016/j.renene.2015.11.070.

[8] R. Kumar, R. K. Aggarwal, and J. D. Sharma, "Comparison of regression and artificial neural network models for estimation of global solar radiations," *Renewable and Sustainable Energy Reviews*, vol. 52, pp. 1294–1299, 2015, doi: 10.1016/j.rser.2015.08.021.

[9] X. Lü, T. Lu, C. J. Kibert, and M. Viljanen, "Modeling and forecasting energy consumption for heterogeneous buildings using a physical-statistical approach," *Applied Energy*, vol. 144, pp. 261–275, 2015, doi: 10.1016/j.apenergy.2014.12.019.

[10] L. Sehovac, C. Nesen, and K. Grolinger, "Forecasting building energy consumption with deep learning: a sequence to sequence approach," in *Proceedings - 2019 IEEE International Congress on Internet of Things, ICIOT 2019 - Part of the 2019 IEEE World Congress on Services*, Jul. 2019, pp. 108–116, doi: 10.1109/ICIOT.2019.00029.

[11] F. Rodríguez, I. Azcárate, J. Vadillo, and A. Galarza, "Forecasting intra-hour solar photovoltaic energy by assembling wavelet based time-frequency analysis with deep learning neural networks," *International Journal of Electrical Power and Energy Systems*, vol. 137, 2022, doi: 10.1016/j.ijepes.2021.107777.

[12] S. Gupta *et al.*, "Long-term power prediction of photovoltaic panels based on meteorological parameters and support vector machine," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 33, no. 2, pp. 687–695, 2024, doi: 10.11591/ijeecs.v33.i2.pp687-695.
[13] L. Zou, M. S. Munir, K. Kim, and C. S. Hong, "Day-ahead energy sharing schedule for the P2P prosumer community using LSTM and swarm intelligence," in *2020 International Conference on Information Networking (ICOIN)*, Jan. 2020, vol. 2020-Janua, pp. 396–401, doi: 10.1109/ICOIN48656.2020.9016520.
[14] B. Singh and D. Pozo, "A guide to solar power forecasting using ARMA models," in *2019 IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe)*, Sep. 2019, pp. 1–4, doi: 10.1109/ISGTEurope.2019.8905430.
[15] R. Basmadjian, A. Shaafieyoun, and S. Julka, "Day-ahead forecasting of the percentage of renewables based on time-series statistical methods," *Energies*, vol. 14, no. 21, p. 7443, Nov. 2021, doi: 10.3390/en14217443.
[16] V. A. Natarajan and P. Karatampati, "Survey on renewable energy forecasting using different techniques," in *2019 2nd International Conference on Power and Embedded Drive Control (ICPEDC)*, Aug. 2019, pp. 349–354, doi: 10.1109/ICPEDC47771.2019.9036569.
[17] G. S. Thirunavukkarasu *et al.*, "Very short-term solar irradiance forecasting using multilayered long-short term memory," in *2022 7th International Conference on Smart and Sustainable Technologies (SpliTech)*, Jul. 2022, pp. 1–8, doi: 10.23919/SpliTech55088.2022.9854244.
[18] M. Mohamed, F. E. Mahmood, M. A. Abd, A. Chandra, and B. Singh, "Dynamic forecasting of solar energy microgrid systems using feature engineering," *IEEE Transactions on Industry Applications*, vol. 58, no. 6, pp. 7857–7869, 2022, doi: 10.1109/TIA.2022.3199182.
[19] V. H. Wentz, J. N. Maciel, J. J. G. Ledesma, and O. H. A. Junior, "Solar irradiance forecasting to short-term PV power: accuracy comparison of ANN and LSTM models," *Energies*, vol. 15, no. 7, 2022, doi: 10.3390/en15072457.
[20] S. M. Maleki, H. Hizam, and C. Gomes, "Estimation of hourly, daily and monthly global solar radiation on inclined surfaces: models re-visited," *Energies*, vol. 10, no. 1, p. 134, Jan. 2017, doi: 10.3390/en10010134.
[21] M. Bai, Y. Chen, X. Zhao, J. Liu, and D. Yu, "Deep attention ConvLSTM-based adaptive fusion of clear-sky physical prior knowledge and multivariable historical information for probabilistic prediction of photovoltaic power," *Expert Systems with Applications*, vol. 202, p. 117335, Sep. 2022, doi: 10.1016/j.eswa.2022.117335.
[22] A. Andreotti, B. Caiazzo, A. Di Pasquale, and M. Pagano, "On comparing regressive and artificial neural network methods for power system forecast," in *2021 AEIT International Annual Conference (AEIT)*, Oct. 2021, pp. 1–6, doi: 10.23919/AEIT53387.2021.9626938.
[23] P. Bendiek, A. Taha, Q. H. Abbasi, and B. Barakat, "Solar irradiance forecasting using a data-driven algorithm and contextual optimisation," *Applied Sciences*, vol. 12, no. 1, p. 134, Dec. 2021, doi: 10.3390/app12010134.
[24] R. Vinayakumar, M. Alazab, K. P. Soman, P. Poornachandran, A. Al-Nemrat, and S. Venkatraman, "Deep learning approach for intelligent intrusion detection system," *IEEE Access*, vol. 7, pp. 41525–41550, 2019, doi: 10.1109/ACCESS.2019.2895334.
[25] V. B. Kumar, V. M. Nookesh, B. S. Saketh, S. Syama, and J. Ramprabhakar, "Wind speed prediction using deep learning-LSTM and GRU," in *2021 2nd International Conference on Smart Electronics and Communication (ICOSEC)*, Oct. 2021, pp. 602–607, doi: 10.1109/ICOSEC51865.2021.9591886.
[26] Z. Qian *et al.*, "Deep learning-based short-term wind power prediction considering various factors," in *2022 17th International Conference on Control, Automation, Robotics and Vision (ICARCV)*, Dec. 2022, pp. 529–533, doi: 10.1109/ICARCV57592.2022.10004261.

BIOGRAPHIES OF AUTHORS

Dr. Mahesh Kumar K. M. working as an associate professor and head in the Department of Electrical and Electronics Engineering PES College of Engineering, Mandya has about 13 years of teaching experience. He received his B.E. degree in electrical and electronics engineering and M.Tech. degree in electronics engineering from Visvesvaraya Technological University, Belagavi, Karnataka. He received Ph.D. degree in electronics engineering from University of Mysore, Karnataka state. He has published 11 research papers in refereed international journals and in the proceedings of various international conferences. His areas of research include high voltage engineering, programmable logic controllers, renewable energy systems, and analog electronics circuits. He is an active member of IEEE and IEI. He can be contacted at email: maheshkm@pesce.ac.in.

Mrs. Uppuluri Lakshmi Soundharya is a highly accomplished professional with a diverse educational background and a profound expertise in computer science and engineering. Pursuing Ph.D. in computer science and engineering from SRM University, Chennai, established as solid academic foundation in Tamil Nadu. Post graduation in computer science and engineering from NRI Institute of Technology and under graduation in computer science and engineering from NRI Institute of Technology affiliated to JNTUK. Her commitment to professional development is evident through an impressive array of certifications, such as PCAP 31-03, Microsoft certifications, Oracle Cloud Infrastructure. Served as teaching associate at SRM University AP. Currently working as an assistant professor at Koneru Lakshmaiah Educational Foundation, adding wealth of academic and industry experience to his role. She can be contacted at email: mail2usoundharya@gmail.com.

Dr. R. Hemalatha is presently associate professor, Department of Computer Science and Engineering, St Joseph's College of Engineering, Chennai. She has done B.E in computer science and engineering from Bharathidasan University, M.E in computer science and engineering from Annamalai University and Ph.D. in information and communication engineering from Anna University. She has done her specialization in mobile ad hoc networks. Her other area of interest includes mobile computing, ML, AI, and wireless networks. She can be contacted at email: hemes.kumar@gmail.com.

Dr. Anjanappa C. is presently working as an Associate Professor in the department of electronics and communication engineering, The National Institute of engineering,mysuru, Karnataka, india. He received his B.E. degree in electronics and communication engineering and M.Tech. degree in electronics and communication engineering from Bangalore University, Bangalore, Karnataka. He received Ph.D. degree in electronics engineering from University of Mysore, Karnataka state. He has published more than 20 research papers in refereed international journals and in the proceedings of various international conferences. His areas of research interest include VLSI, signal processing medical imaging, AI, and ML application to VLSI. He can be contacted at email: anjanappagayathri@gmail.com.

Suganya M. J. is an assistant professor in 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 Government College of Engineering, Tirunelveli in 2008 and 2010, respectively. She has been an assistant professor in 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 library. She can be contacted at email: sugi.mj@gmail.com.