

Research Article

Forecasting Solar Energy Production Using Machine Learning

C. Vennila,¹ Anita Titus,² T. Sri Sudha,³ U. Sreenivasulu,⁴ N. Pandu Ranga Reddy,³ K. Jamal,⁵ Dayadi Lakshmaiah,⁶ P. Jagadeesh,⁷ and Assefa Belay⁸

¹Department of Electrical and Electronics Engineering, Alagappa Chettiar Government College of Engineering and Technology, Karaikudi, 630003 Tamil Nadu, India

²Department of Electronics and Communication Engineering, Jeppiaar Engineering College, Semmenchery Raghiv Gandhi Salai OMR, Chennai, 600119 Tamil Nadu, India

³Department of Electronics and Communication Engineering, Malla Reddy Engineering College, (A) Hyderabad 500100, India

⁴Department of Electronics and Communication Engineering, NBKR Institute of Science and Technology, Vidyannagar, Andhra Pradesh 524413, India

⁵Department of Electronics and Communication Engineering, GRIET Gokaraju Rangaraju Institute of Engineering and Technology, Bachupally, Hyderabad 500090, India

⁶Department of Electronics and Communication Engineering, Sri Indu Institute of Engineering and Technology, Hyderabad 501510, India

⁷Department of Electronics and Communication Engineering, Saveetha School of Engineering, (SIMATS) Chennai, 602105 Tamil Nadu, India

⁸Department of Mechanical Engineering, Mizan Tepi University, Ethiopia

Correspondence should be addressed to Assefa Belay; assefa@mtu.edu.et

Received 8 February 2022; Accepted 24 March 2022; Published 30 April 2022

Academic Editor: V. Mohanavel

Copyright © 2022 C. Vennila et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

When it comes to large-scale renewable energy plants, the future of solar power forecasting is vital to their success. For reliable predictions of solar electricity generation, one must take into consideration changes in weather patterns over time. In this paper, a hybrid model that integrates machine learning and statistical approaches is suggested for predicting future solar energy generation. In order to improve the accuracy of the suggested model, an ensemble of machine learning models was used in this study. The results of the simulation show that the proposed method has reduced placement cost, when compared with existing methods. When comparing the performance of an ensemble model that integrates all of the combination strategies to standard individual models, the suggested ensemble model outperformed the conventional individual models. According to the findings, a hybrid model that made use of both machine learning and statistics outperformed a model that made sole use of machine learning in its performance.

1. Introduction

Machine learning approaches have been increasingly popular across a wide range of businesses where data-driven difficulties have been common in recent decades. Machine learning encompasses a wide range of disciplines, including data mining, optimization, and artificial intelligence, to name a few of the more prominent. Machine learning approaches seek to discover connections between input data and output data, whether or not they make use of mathe-

matical models in the process. Following training with the training dataset, the forecasting input data can be fed into the well-trained machine learning models, which can then be used to make predictions [1, 2]. This stage is crucial to machine learning since it has the ability to improve the performance and speed of the algorithm.

Generalizations aside, machine learning relies on three forms of training: supervised training, unsupervised training, and reinforcement training. Clustering criteria are used, and the number of clusters can change depending on the situation.

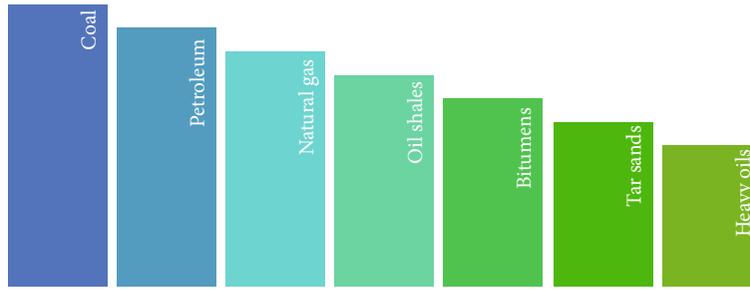


FIGURE 1: Fossil fuels.

In order to maximize the intended benefits of reinforcement learning, the learner must interact with their environment in order to obtain feedback from it. This is known as interactivity.

There have been a variety of theories and implementations presented that are based on the three fundamental learning principles [3]. As a result, deep learning is capable of achieving characteristic nonlinear features and invariant high-level data configurations, and as a result, it has been applied in a variety of diverse fields with good results.

According to some studies, a single machine learning model has also been used to anticipate the availability of renewable energy sources [4]. Because of the large range of datasets and time steps, prediction ranges, settings, and performance measurements, a single machine learning model cannot improve forecasting performance on a single dataset or time step. There have been a number of studies in renewable energy forecasting that have resulted in hybrid machine learning models or overall prediction methodologies that are intended to improve prediction performance. Significant attention has lately been drawn to support vector machines (SVMs) and deep learning algorithms [5].

In addition to hastening the depletion of fossil fuel reserves as in Figure 1, overconsumption of fossil fuels has a negative influence on the environment as a whole. Increased health risks as well as climate change threats will result as a result of these issues. Renewable energy, which includes both nuclear power and fossil fuels, is the energy source, surpassing both.

Renewable energy has recently received a great deal of attention as a result of its long-term viability and minimal impact on the surrounding environment. In the near future, the provision of renewable energy will be one of the most critical challenges to be addressed. Alternative terminology: the inclusion of renewable energy sources into current or future electric power generation systems.

Existing energy challenges, such as increasing supply stability and alleviating regional power shortages, can be solved through the evolution of renewable energy technologies. This creation of diverse energy sources, on the other hand, is interrupted and chaotic as a result of the volatility of the energy market as well as the unpredictable and intermittent renewable energy. Dealing with renewable energy fluctuation in an accurate way prevails as a challenge. The energy system efficiency is improved via energy monitoring with high precision. The application of energy forecasting technologies can assist in the creation, management, and formulation of energy policy at all levels of government. As renewable energy sources

become more generally available, it is vital to create cutting-edge technologies for storing this energy [6]–[7].

Several studies have discovered that a variety of machine learning algorithms have been employed to estimate the output of renewable energy resources. With the help of data-driven models, it is possible to make more accurate predictions about renewable energy. With the use of hybrid machine learning algorithms, projections for renewable energy sources have also been enhanced. In order to effectively predict the availability of renewable energy sources, it was required to use a number of time intervals. When it comes to renewable energy forecasting, these criteria have been widely used to evaluate the accuracy and efficiency of machine learning algorithms [8].

2. Related Works

Wang et al. [9] investigated deep learning-based renewable energy forecast algorithms in their research. There were four kinds of approaches: stack autoencoder, deep belief networks, recurrent neural networks, and other approaches are lumped together. Certain data processing is employed in order for performance improvement of the predicted results even further.

When it comes to dependability and estimating energy, Bermejo et al. [10] proposed a model using ANN. In this inquiry, potential sources of energy such as solar, hydraulic, and wind power were all taken into consideration. A number of examples are developed to demonstrate the advantages of ANN in the prediction of energy and trustworthiness, among other things. In their study, Mosavi et al. [11] looked at the analysis and classification of ML algorithms used in energy systems. According to the findings of the study, hybrid models outperform standard ML models in energy systems.

According to Ahmed and Khalid [12], they investigated the reliability of renewable power generation systems and optimal reserve capacity in order to better understand forecasting models for renewable power production systems. According to the power industry, this review gave current trends and forecasts for future improvements in system design and operation. In the field of solar and wind energy forecasting, Zendejboudi et al. [13] discovered that support vector machine (SVM) outperformed others. Furthermore, when it comes to forecasting accuracy, hybrid SVM models outperforms single SVM models.

Das et al. [14] conducted an investigation and evaluation of the forecasting methodologies utilized in solar photovoltaic

electricity generation. According to the findings of this study, artificial neural networks and the support vector machine model were found to be particularly prevalent in this field. In their paper, the scientists noted that weather conditions have an impact on the accuracy of solar power forecasts.

Due to the fact that solar radiation is a key source of solar energy, Voyant et al. [15] investigated the application of machine learning algorithms in forecasting solar radiation. Several strategies for forecasting solar radiation have been described. Perez-Ortiz et al. [16] conducted an evaluation of categorization approaches for problems of renewable energy and provided insights for both academics and industry practitioners in this field. In this study, we employed evolutionary approaches and game theory to investigate the feasibility of hybrid renewable energy systems and the obstacles associated with them.

To give a comprehensive survey [17]–[18] of current developments in the field, this paper examines data pretreatment methodologies, machine learning algorithms, parameter selection, and performance assessments of machine learning models in renewable energy projections.

3. Proposed Method

In this section, we validate the forecasting made by the ensemble model for optimal prediction of power generation using PV plants. The study considers two case studies, where the former is simulated for smaller PV farms of 1000 PV cells and larger PV farms of 100000 PV cells. The illustration of training the ensemble model is given in Figure 2. In these proposed methods, the data is classified into a single classifier and another set will act as a training data from which we will classify the samples, classifier, Aggregation was done once it was completed. It will be moved to performance metric evaluation with several comparison techniques.

3.1. Feature Selection. The sun delivers solar energy in the form of solar radiation, which is produced by the photovoltaic effect. Sunlight intensity is the most important factor influencing the output of photovoltaic (PV) solar panels. A PV system output can be affected by a variety of different environmental variables among others. Identifying which parts of PV are valuable and which aspects are not is also essential so that a suitable feature subset may be selected as an input to the model. We propose a hybrid method for feature variable selection that comprises two basic processes, namely, the filter stage and the wrapper stage, as depicted in Figure 3.

Prior to begin the learning process, the filter technique analyzes features based on the inherent attributes of each one of the features. Filter criteria are used to select a subset of features from a dataset based on their relevance.

Because of the characteristics of PV data, the Pearson correlation coefficient (PCC) is employed to assess the relationship between input factors and the target variable. A PCC is a statistical metric that is used to determine the linear correlation between two variables, X and Y , in a dataset. Data from time series analysis captures the degree to which a target variable Y correlates with an input variable X over the course of an observation period. When calculating the correlation between two variables, the time series data at

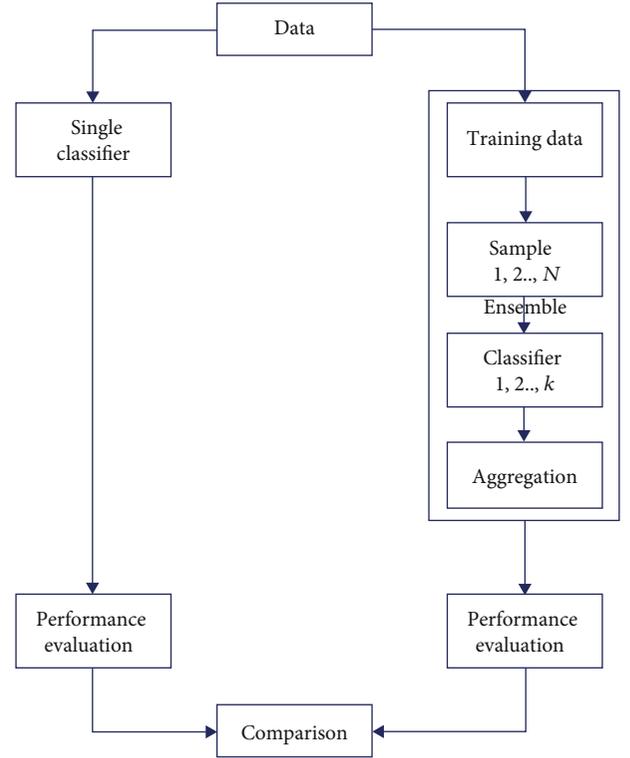


FIGURE 2: Proposed model.

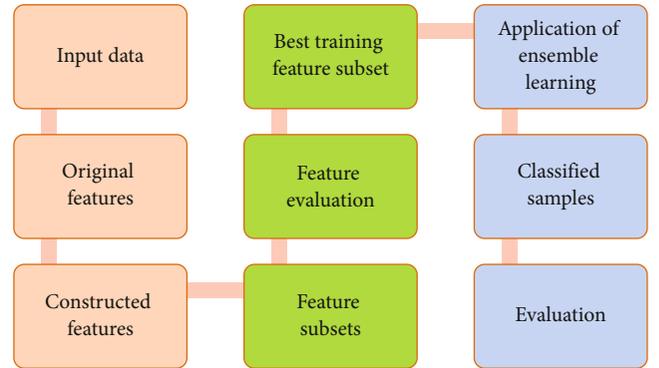


FIGURE 3: Feature selection.

the points t and $t1$ of the variable are not used. In our example, the meteorological factors that have an impact on PVPG are represented by the letters Y and X , respectively. As a result, the PCC may be expressed mathematically as

$$\rho(X, Y) = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^N (y_i - \bar{y})^2}}, \quad (1)$$

where PCC is the value that lies between +1 and -1. PCC is one of the most commonly used criteria for describing the relationship between variables in practice, and it is also one of the most widely studied.

The wrapper approach is utilized to analyze each subset that has been selected. The learning algorithm is integrated into the feature selection process, which in turn makes use

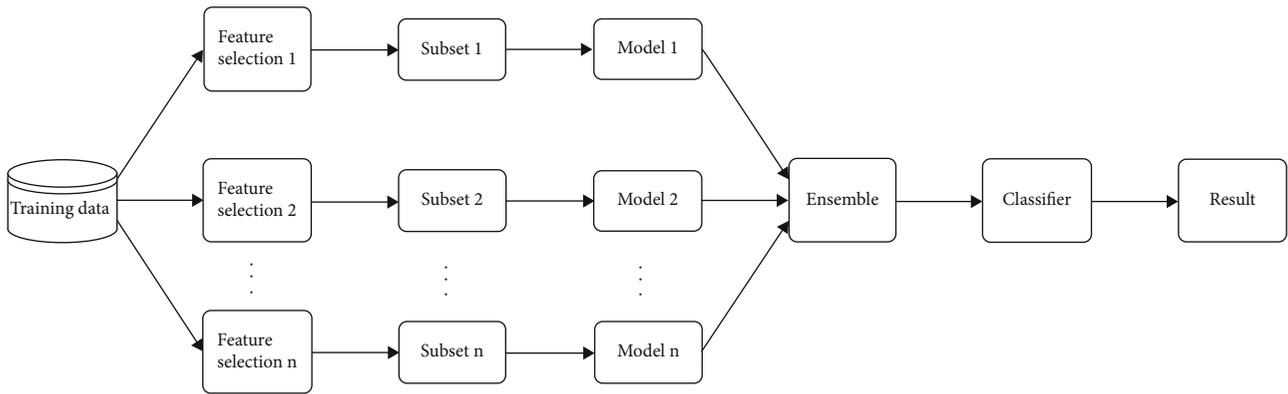


FIGURE 4: Classification [19].

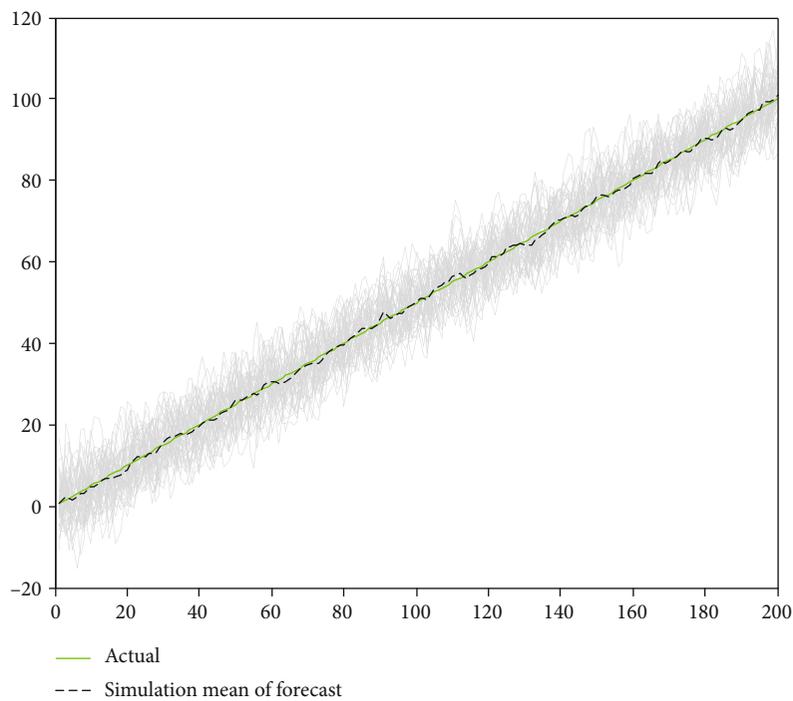


FIGURE 5: Observations from ensemble model for smaller PV farms.

of the error of a given model to determine which feature subset is most important to the user. It was decided to employ the traditional LSTM model for the analysis of feature subsets in this study because of its capacity to address time series forecasting issues. As a result, the optimal subset of training characteristics may be determined from among all of the subsets that have been investigated.

Through the use of a hybrid approach, the proposed feature selection attempts to integrate the best aspects of wrapper and filter methods into a single method. After examining the correlations between variables using filter criteria, appropriate thresholds are determined in order to reduce the number of feature variables that can be evaluated. The filter technique, in contrast to other learning algorithms, is univariate in nature.

This results in it being significantly more efficient and faster to compute than the wrapper technique, and it is capable of dealing with massive datasets with ease. No consideration is

currently given to how features interact with one another or with the learning algorithms, which is a problem. In this case, the wrapper technique is required because coupled features in the single feature evaluation. In order to effectively use an individual wrapper strategy, a significant amount of computer power is required. This is owing to the learning methods used and the enormous number of feature subsets that must be analyzed. Despite this, when the correlation results of the filter approach are utilized as a guide, fewer feature subsets are generated and analyzed than would otherwise be the case. As a result, the hybrid method that has been proposed has the potential to improve the effectiveness of the feature selection.

3.2. Ensemble Feature Classification. In this study, we employed an approach known as multimodel ensemble feature selection, which is an alternative to the methods that have been previously explored. Initially, the training data is

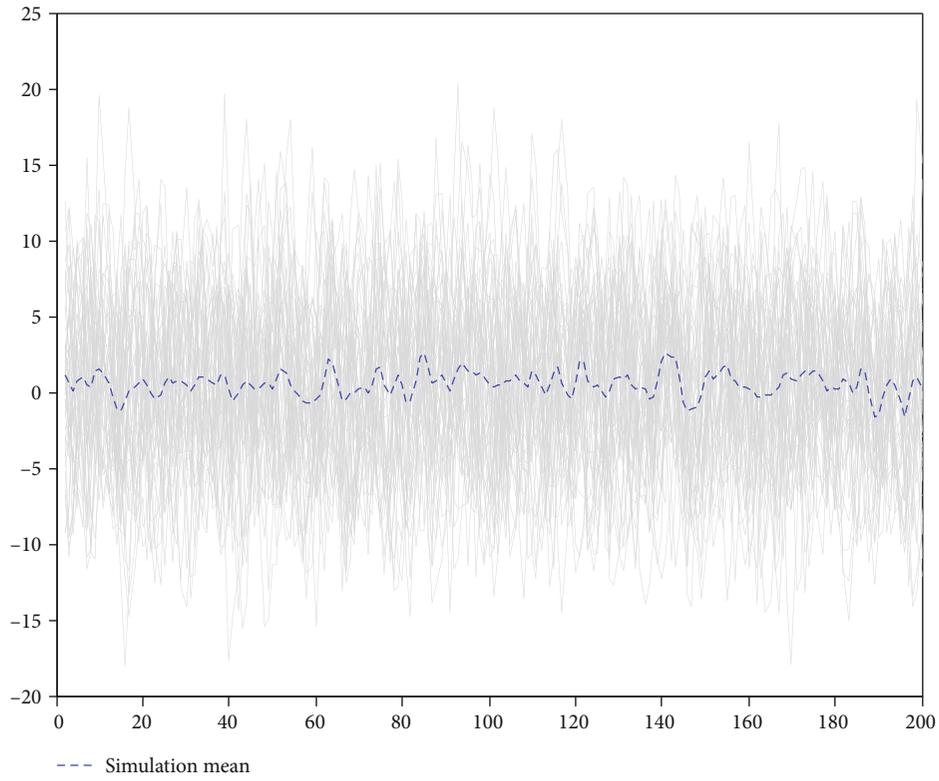


FIGURE 6: Difference between the smaller and larger PV farms using ensemble model.

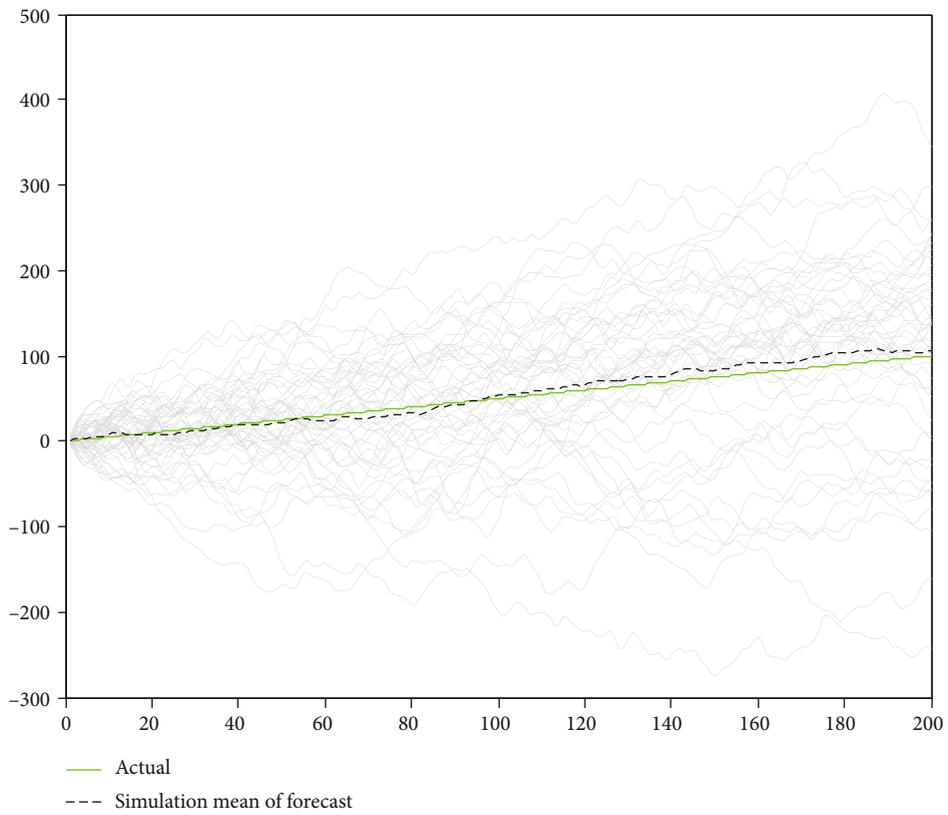


FIGURE 7: Observations from ensemble model for larger PV farms.

analyzed using a number of feature selection algorithms, each of which generates a subset of the total data. Second, a model is trained using a single subset of data that was previously acquired, as described above. Finally, we compile the results of all of the simulations that we have conducted. Figure 4 depicts the process of selecting ensemble features from a large number of candidates.

4. Results and Discussions

During the experiment, a total of 10 MW of capacity was achieved through the use of polycrystalline solar panels (Poly-SI) and thin-film solar cells (TFSC). The time series dataset can be accessed in a number of different formats. For our research, we used data with a five-minute resolution.

Low-bias machine learning models also reveal that estimates for energy generation are pretty close to the reality. According to the example studies depicted in Figure 5, clear weather conditions outperformed overcast and partly cloudy conditions for each ML model. When the weather is clear, the shift in cloud cover is gradual, which allows for more accurate power predictions. The RMSE of the power forecast for a gloomy day is slightly higher than that of the power forecast for a clear day. It is the actual and forecasted of the predicted power for the partly cloudy day that are particularly bad. The error in a forecast is influenced by the variety of the forecast. It is mostly owing to the greater unpredictability associated with partially cloudy conditions that the abovementioned discrepancies in forecast errors exist. In contrast to the findings in Figure 6, the results in Figure 7 appear to be contradictory.

It is obvious from this dataset that the ensemble model is capable of producing accurate estimates. It is possible that a generalized ML model that can be applied to any PV plant will never be developed. It is possible that the ensemble model, despite its excellent performance in this study, will perform even better on a different dataset. The location and construction of the power plant may have an impact on the weather conditions and the amount of electricity that is generated. Clear data is more abundant in the dataset than cloudy data, which indicates that the data is more reliable. Because the dataset has a higher proportion of clear data than other models for polar regions, the ML model overall performance is superior to that of other models for polar regions.

5. Conclusions

An integrated machine learning model and the statistical approach are used to anticipate future solar power generation from renewable energy plants. This hybrid model improves accuracy by integrating machine learning methods and the statistical method. In order to improve the accuracy of the suggested model, an ensemble of machine learning models was used in this study. When comparing the performance of an ensemble model that integrates all of the combination strategies to standard individual models, the suggested ensemble model outperformed the conventional individual models. According to the findings, a hybrid model that made use of both machine learning and statistics

outperformed a model that made sole use of machine learning in its performance. In future work, the proposed method can improvise the performance, accuracy, and the other metrics using several deep learning mechanisms.

Data Availability

The data used to support the findings of this study are included within the article. Further data or information is available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

The authors thanks to Mizan Tepi University, Ethiopia, for providing help during the research and preparation of the manuscript.

References

- [1] I. Jebli, F. Z. Belouadha, M. I. Kabbaj, and A. Tilioua, "Prediction of solar energy guided by Pearson correlation using machine learning," *Energy*, vol. 224, article 120109, 2021.
- [2] G. Narvaez, L. F. Giraldo, M. Bressan, and A. Pantoja, "Machine learning for site-adaptation and solar radiation forecasting," *Renewable Energy*, vol. 167, pp. 333–342, 2021.
- [3] S. Park, Y. Kim, N. J. Ferrier, S. M. Collis, R. Sankaran, and P. H. Beckman, "Prediction of solar irradiance and photovoltaic solar energy product based on cloud coverage estimation using machine learning methods," *Atmosphere*, vol. 12, no. 3, p. 395, 2021.
- [4] K. Mahmud, S. Azam, A. Karim, S. Zobaed, B. Shanmugam, and D. Mathur, "Machine learning based PV power generation forecasting in Alice Springs," *IEEE Access*, vol. 9, pp. 46117–46128, 2021.
- [5] A. Nespoli, A. Niccolai, E. Ogliari, G. Perego, E. Collino, and D. Ronzio, "Machine learning techniques for solar irradiation nowcasting: cloud type classification forecast through satellite data and imagery," *Applied Energy*, vol. 305, article 117834, 2022.
- [6] F. Rodríguez, F. Martín, L. Fontán, and A. Galarza, "Ensemble of machine learning and spatiotemporal parameters to forecast very short term solar irradiation to compute photovoltaic generators output power," *Energy*, vol. 229, article 120647, 2021.
- [7] H. Musbah, H. H. Aly, and T. A. Little, "Energy management of hybrid energy system sources based on machine learning classification algorithms," *Electric Power Systems Research*, vol. 199, article 107436, 2021.
- [8] U. Singh, M. Rizwan, M. Alaraj, and I. Alsaidan, "A machine learning-based gradient boosting regression approach for wind power production forecasting: a step towards smart grid environments," *Energies*, vol. 14, no. 16, article 5196, 2021.
- [9] H. Wang, Z. Lei, X. Zhang, B. Zhou, and J. Peng, "A review of deep learning for renewable energy forecasting," *Energy Conversion and Management*, vol. 198, article 111799, 2019.
- [10] J. Ferrero Bermejo, J. F. Gomez Fernandez, F. Olivencia Polo, and A. Crespo Márquez, "A review of the use of artificial neural network models for energy and reliability prediction A

- study of the solar PV hydraulic and wind energy sources,” *Applied Sciences*, vol. 9, no. 9, article 1844, 2019.
- [11] A. Mosavi, M. Salimi, S. Faizollahzadeh Ardabili, T. Rabczuk, S. Shamshirband, and A. R. Varkonyi-Koczy, “State of the art of machine learning models in energy systems, a systematic review,” *Energies*, vol. 12, no. 7, article 1301, 2019.
- [12] A. Ahmed and M. Khalid, “A review on the selected applications of forecasting models in renewable power systems,” *Renewable and Sustainable Energy Reviews*, vol. 100, pp. 9–21, 2019.
- [13] A. Zendeboudi, M. A. Baseer, and R. Saidur, “Application of support vector machine models for forecasting solar and wind energy resources: a review,” *Journal of Cleaner Production*, vol. 199, pp. 272–285, 2018.
- [14] U. K. Das, K. S. Tey, M. Seyedmahmoudian et al., “Forecasting of photovoltaic power generation and model optimization: a review,” *Renewable and Sustainable Energy Reviews*, vol. 81, pp. 912–928, 2018.
- [15] C. Voyant, G. Notton, S. Kalogirou et al., “Machine learning methods for solar radiation forecasting: a review,” *Renewable Energy*, vol. 105, pp. 569–582, 2017.
- [16] M. Pérez-Ortiz, S. Jiménez-Fernández, P. A. Gutiérrez, E. Alexandre, C. Hervás-Martínez, and S. Salcedo-Sanz, “A review of classification problems and algorithms in renewable energy applications,” *Energies*, vol. 9, no. 8, article 607, 2016.
- [17] S. F. Stefenon, M. H. D. M. Ribeiro, A. Nied et al., “Time series forecasting using ensemble learning methods for emergency prevention in hydroelectric power plants with dam,” *Electric Power Systems Research*, vol. 202, article 107584, 2022.
- [18] K. Lingelbach, Y. Lingelbach, S. Otte, M. Bui, T. Künzell, and M. Peissner, “Demand forecasting using ensemble learning for effective scheduling of logistic orders,” in *In International Conference on Applied Human Factors and Ergonomics*, pp. 313–321, Springer, Cham, 2021.
- [19] S. Zhao, Y. Zhang, H. Xu, and T. Han, “Ensemble classification based on feature selection for environmental sound recognition,” *Mathematical Problems in Engineering*, vol. 2019, 7 pages, 2019.