Forecasting Solar Irradiance with Geographical Considerations: Integrating Feature Selection and Learning Algorithms

Saman Soleymani[1], Amirreza Talebi[2]

[1](Department of Electrical Engineering/ Semnan University, Iran)
[2](Industrial engineering, university of Tehran, Iran)

ABSTRACT: The increasing global demand for clean and environmentally friendly energy resources has caused increased interest in harnessing solar power through photovoltaic (PV) systems for smart grids and homes. However, the inherent unpredictability of PV generation poses problems associated with smart grid planning and management, energy trading and market participation, demand response, reliability, etc. Therefore, solar irradiance forecasting is essential for optimizing PV system utilization. This study proposes the next-generation machine learning algorithms such as random forests, Extreme Gradient Boosting (XGBoost), Light Gradient Boosted Machine (lightGBM) ensemble, CatBoost, and Multilayer Perceptron Artificial Neural Networks (MLP-ANNs) to forecast solar irradiance. Besides, Bayesian optimization is applied to hyperparameter tuning. Unlike tree-based ensemble algorithms that select the features intrinsically, MLP-ANN needs feature selection as a separate step. The simulation results indicate that the performance of the MLP-ANNs improves when feature selection is applied. Besides, the random forest outperforms the other learning algorithms.

Keywords– Smart grids, solar irradiance forecasting, MLP-ANNs, random forest, LightGBM, CatBoost, feature selection.
----

1.   INTRODUCTION

Solar energy has become a key source of sustainable electricity generation due to the growing need for renewable energy sources and concern over climate change [1]. Although solar energy is a clean and accessible resource, its stochastic nature and dependence on weather present distinct challenges to smart grids [2]. Solar radiation, which measures sun intensity reaching the surface of the Earth, plays an important role in determining solar energy production system effectiveness and predictability [3]. Solar irradiance forecasting has become a critical process in the smart grid sector [4]. This forecasting is essential for smart grid integration and stability [5], and reduces the negative impact of solar power instability on the grid, power electronic equipment [6], optimizing energy markets [7], energy storage management [8], transformers [9], environmental concerns [10]. So, a comprehensive understanding of solar irradiance is essential to effectively integrating solar power into smart grids. It is essential to develop advanced forecasting methods that can provide reliable insights into future irradiance patterns in order for this understanding to be achieved.

In recent years, machine learning has emerged as a powerful tool for addressing the various and complex problems in smart grids, such as environment explosiveness level evaluation [11], smart monitoring systems [12], diagnosing location [13], generation [14], demand-side cooperation [15], and occupant estimation [16].

One of the most important applications of machine learning algorithms is their ability to forecast. A machine learning model can be trained using historical solar irradiance data, such as sophisticated meteorological variables and sensor-based data. The authors assess the use of hybrid machine learning models for solar radiation forecasting [17]. To build a more precise forecasting model, they mix various machine learning techniques. Convolutional neural networks are utilized in [18] to forecast solar radiation by taking advantage of the spatial and temporal trends in the inputs and outputs to make better forecasts. An MLP-ANN algorithm is developed in [19] for solar irradiance forecasting based on meteorological data. In this work, feature importance is used to specify the most relevant features. Long Short-Term Memory is one of the various types of recurrent neural networks for sequential data, which are used in [20] for solar irradiance forecasting. [21] develops a support vector machine regression model to forecast solar irradiance.

Although significant studies have been conducted by using machine learning to predict solar irradiance, there are still significant aspects such as feature importance analysis, hyperparameter tuning, model validation, and testing on different unseen datasets that are crucial for developing accurate and robust forecasting models. Addressing these aspects can lead to more reliable predictions in solar irradiance forecasting that can improve the performance of the smart grids. The main contribution of this study can be summarized as follows:

- Advanced machine learning comparison: this study contributes by employing next-generation machine learning algorithms, including Random Forest, Extreme Gradient Boosting (XGBoost), Light Gradient Boosted Machine (LightGBM) ensemble, Multilayer Perceptron Artificial Neural Networks (MLP-ANNs) and CatBoost. By utilizing these next-generation techniques, the research not only explores their individual capabilities but also provides a comprehensive comparison of their performance.
- Enhanced forecasting through feature selection: another key contribution is the incorporation of feature selection techniques to improve forecasting results.
- Hyperparameter tuning via Bayesian optimization: the research advances the area by using Bayesian optimization to adjust the hyperparameters of the machine learning models.
- Cross-Validation: this work further contributes by precisely testing the developed forecasting algorithms on diverse unseen datasets.
- Spatial generalization through multilocation testing: the final contribution is to test the developed models on different datasets collected from various geographical locations.

2. METHODOLOGY

Data collection and preprocessing

The data used to create the models used in this study comes from the National Solar Radiation Database [22]. The dataset contains 8 characteristics: solar irradiance, ambient temperature, pressure, humidity, wind speed wind direction, time, and length of day. Data are for 4 years from 2017 to 2021 at hourly intervals. After cleaning the datasets, it is a common practice to normalize the dataset. Normalization is a preprocessing technique that adjusts all the features to a similar scale, which can improve the performance and convergence of learning processes [23]. This study utilizes the standardization technique that transforms the data such that it has a mean of 0 and a standard deviation of 1 ensuring it follows the normal distribution [24].

Feature selection

Feature selection is a technique in machine learning where a subset of the most relevant features or inputs is selected from a larger set of features. The goal of feature selection is to improve model performance, reduce computational complexity, enhance interpretability, and mitigate the risk of overfitting by focusing on the most relevant features [25]. The simplest way to remove the less significant and select the most relevant features is to train the machine learning through all possible combinations of features and choose the best one. However, this is an extremely time-consuming procedure. Accordingly, to select the most important features of a learning model, the feature selection technique is used. It should be noted that machine learning models like random forest, XGBoost, LightGBM, and CatBoost are based on the inherent feature importance provided by the algorithms during their training process [26]. That is, there is no need for a separate step to

select the features. On the other hand, MLP-ANNs typically do not have the feature selection intrinsically. Feature selection techniques involve selecting features before feeding the data into the machine learning algorithms [27]. The selected features are then used for training. Among the most common feature selection techniques are wrapper and filter-based models. Many models are developed with all possible combinations of features, and the best model is selected based on a performance metric among all the models created by the wrapper feature selection models [28]. The present wrapper algorithms are forward, backward elimination, and recursive feature selection. Filter methods rely on statistical measures to rank and select features before applying a machine learning algorithm. Unlike wrapper methods that involve training and evaluating models, filter methods pre-process features independently of the chosen machine learning algorithm. Common filter methods include the Pearson Correlation Coefficient (PCC), Chi-Squared test [29], and ANOVA analysis. This study utilized the PCC to select the most important features in order to feed them into the models. PCC is the ratio between the covariance of two variables and the product of their standard deviations and measures the linear correlation between a feature and the output variable. Features with high correlation coefficients are considered more relevant. The PCC value ranges between -1 and +1. A value of -1 indicates a perfect negative correlation, 0 indicates no correlation, and 1 indicates a perfect positive correlation [19].

As mentioned tree-based algorithms like random forest, XGBoost, LightGBM, and CatBoost typically do not need feature selection. However, the contribution of each feature on the final fitted model can be demonstrated. To do that, Shapley's Additive explanation (SHAP) plot analysis is performed. SHAP plot is utilized to visualize the feature importance in the random forest, XGBoost, LightGBM, and CatBoost. This approach enhances the ability to interpret and validate the model decision-making process.

Learning algorithms

Learning algorithms are powerful tools that can be used for forecasting due to their learning ability. Random forest, MLP-ANNs, LightGBM, XGBoost, and CatBoost are utilized in this study. Random forest is an ensemble algorithm that uses a collection of decision trees for classification and regression. It uses a method called bagging (bootstrap aggregation) to reduce variance while maintaining low bias. LightGBM is a fast, distributed, high-performance gradient boosting framework based on the decision tree algorithm, used for ranking, classification, and many other machine learning tasks. XGBoots is another powerful machine learning model that constructs an ensemble of decision trees sequentially, optimizing a loss function and handling data imbalance, missing values, and overfitting. MLP-ANNs are a type of deep learning model composed of multiple interconnected layers and neurons [30]. It is designed to capture complex patterns in data. The main idea of CatBoost is to sequentially combine many weak models and, thus, create a solid competitive predictive model through greedy search.

Hyperparameter tuning

A practice to select the best hyperparameters in the learning models to improve performance is known as hyperparameter tuning. This study utilized Bayesian optimization which has been able to successfully find the optimal combination of hyperparameters of the learning models [31]. Bayesian optimization is a probabilistic model to predict which hyperparameters will lead to better performance and then tests those predictions. To implement Bayesian optimization in learning algorithms, a common practice involves utilizing the k-fold cross-validation technique. A piece of the training dataset is assigned for validation, and the model is trained iteratively over k-folds. During each iteration, one set is assigned for validation, while the remaining k-1 folds are used for training. Model performance is averaged across these k iterations to mitigate variability. In each iteration, Bayesian optimization seeks to determine hyperparameters that optimize the forecasting results.

Evaluation metrics

Evaluation metrics are used to measure the quality of the machine learning models. This study uses Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and coefficient of determination or R^2 [32] to assess the performance of the models.

3. SIMULATION AND RESULTS

The data are split into three sets: training, cross-validation, and testing datasets. 80% of the data are selected as the training set and 20% as the test. in addition, cross-validation is performed on 20% of the training datasets. In this work, the input data are first analyzed by the PCC algorithm, and the data with the highest correlation with solar radiation are identified. This data is then used as input for the MLP-ANN model. Fig. 1 demonstrates the PCC results.

Based on Fig. 1 variables such as ambient temperature, humidity, and wind direction show a substantial
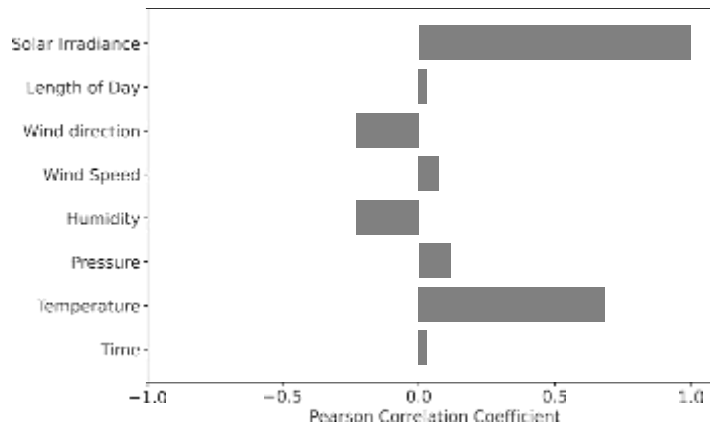

Fig. 1. PCC values calculated between solar irradiance and various meteorological parameters.

correlation with solar irradiance, indicating a higher impact on solar irradiance forecasting. While features like pressure and wind speed exhibited lower influence. Length and time of day could be considered redundant features. To assess the impact of the feature selection technique on the performance of the models, the MLP-ANNs are trained and tested with and without feature selection. The optimal hyperparameters of the fitted MLP-ANNs with PCC by utilizing Bayesian optimization are provided in Table 1. Besides, Table 2 shows solar irradiance forecasting results with and without PCC feature selection. MLP-ANN model performance is enhanced by PCC-based feature selection as MAE, RMSE decreased and R2 increased, indicating improved performance. Table 2 shows the results of predicting solar radiance using random forest, XGBoost, LightGBM, and CatBoost. The results given in Table II compare the performance of various algorithms for forecasting solar radiance. Random forest performs the lowest RMSE of 87.18 MAE of 34.82 and highest R2 of 0.95, indicating superior predictive accuracy. LightGBM and XGBoost algorithms also demonstrated competitive performance, with RMSE values of 111.49 and 114.01, MAE values of 50.16 and 51.57, and R2 values of 0.87 and 0.86, respectively. These results highlight the effectiveness of ensemble techniques like random forest, LightGBM, and XGBoost for accurate solar radiance prediction, while also emphasizing the significance of considering feature importance in the MLP-ANN model.

TABLE 1. Characterization of MLP-ANN parameters with PCC.

Hyperparameter	Value	Hyperparameter	Value
Solver	adam	Loss function	Cross-Entropy
alpha	0.0001	Activation function	Relu
batch_size	Auto	Neurons of input layers	15
learning_rate	0.001	Neurons of the first hidden layers	10
power_t	0.5	Neurons of second hidden layers	5
max_iter	5000	Neurons of third hidden layers	5
shuffle	true	Neurons of output layers	1
Random state	42	Hyperparameter optimizer	Bayesian

TABLE 2. Characterization of MLP-ANN parameters with PCC.

Algorithm	RMSE	MAE	R^2
MLP without PCC	124.50	55.41	0.83
MLP with PCC	95.16	41.45	0.91
Random Forest	87.18	34.82	0.95
LightGBM	111.49	50.16	0.87
XGBoost	114.01	51.57	0.86
CatBoost	130.42	60.12	0.79

Because random forest produces favorable results, further investigation of its generalization capabilities is warranted. Through a five-fold cross-validation process, potential overfitting problems are addressed. As a result of this analysis, we can observe whether the observed performance is consistent for different training set sizes and their corresponding validation scores.

Fig. 2 demonstrates the learning curve associated with the fitted random Forest algorithm. The training MAE values range from 5 to 32 as the training set size increases. Furthermore, the validation MAE values are decreasing from 230 to 34.2. There is a considerable gap between the training and validation MAE values at the beginning of the curve (when training sets are smaller), which may indicate overfitting, in which the model performs remarkably well on the training data but struggles to generalize to unseen data (higher validation MAE). However, as the training set size increases the validation MAE decreases and the gap
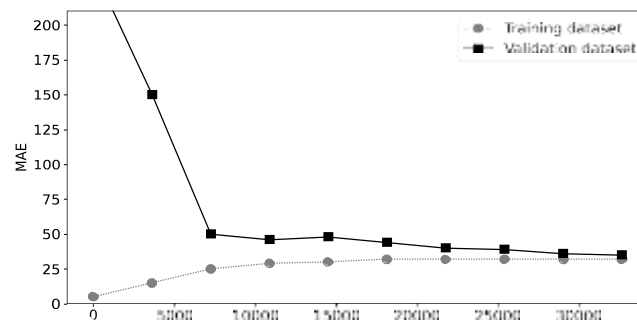

Fig. 2. Learning curve associated with the fitted random forest.

narrows, indicating that the generalization ability of the model is improved. Additionally, as training and validation MAE values converge, it indicates improved generalization and reduced overfitting concerns. As a result, it is evident from the learning curve that the random forest algorithm provides accurate and consistent solar irradiance forecasts. Besides, the optimal hyperparameters of the fitted random forests by utilizing Bayesian optimization are provided in Table 3.

TABLE 3. Characterization of the random forest parameters.

hyperparameter	value
criterion	Gini
Random-state	0
n_estimators	400
min_impurity_decrease	0
max_depth	none
bootstrap	true
max_features	*auto*

Further, a SHAP plot is also created to assess the importance and contribution of each feature on the trained model in Fig 3. The SHAP plot demonstrates the distribution of each feature over all samples in the dataset. The light gray and gray areas represent features with low and high values, respectively. The Y-axis
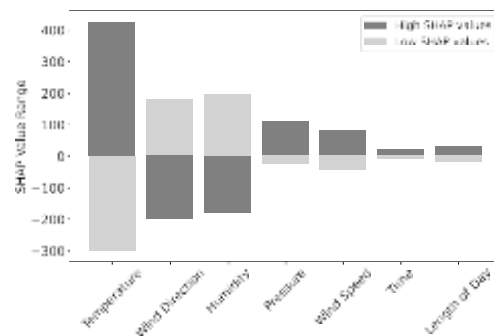

Fig. 3. SHAP plot associated with the fitted Random forest

represents the importance of the features on solar irradiance forecasting. Besides, the features on the left X-axis are ordered based on their mean absolute SHAP values. These orders are consistent with the PCC results indicating that the random forest captures the relevant features. The ambient temperature has the highest SHAP value, indicating its high contribution to solar irradiance forecasting. In addition, the higher the ambient temperature values, the higher the impact on model output. Following the ambient temperature, the second and third significant features are wind direction and humidity which have a negative correlation to the target values. Features like wind speed and air pressure play intermediate roles, contributing to solar irradiance prediction. As the values of wind speed and air pressure increase (from blue to red), their corresponding SHAP values increase, indicating a positive correlation with target values. Finally, the time and length of the day have minimal impact on solar irradiance forecasting.

To further evaluate the robustness of the fitted random forest model, it is tested on different locations to assess its capability to generalize across altering environmental conditions and geographical areas. Fig. 4 shows the RMSE and MAE of the fitted random forest for the original dataset location and two different locations that are 100 and 500 miles away from the original location. The model performance decreases as the distance from the original location increases. Both RMSE and MAE are slightly increased by a 100-mile displacement. Further, a distance of 500 miles shows more increase in model performance parameters. The slight decrease in model performance can be attributed to varying geographical and climatic conditions. Despite the decreasing performance, random forests still outperform LightGBM, XGBoost, CatBoost, and MLP- ANN in this case. Random forest models are particularly robust to varying geographical and climate conditions because they are intrinsically adaptable to complicated relationships between input features and solar
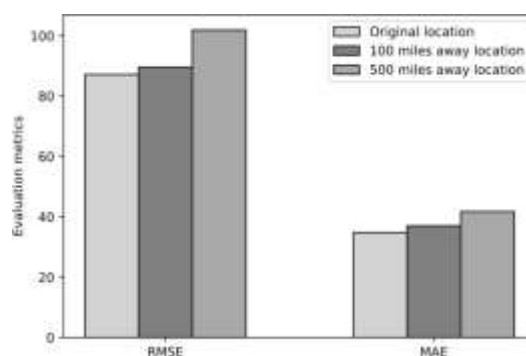

Fig. 4. RMSE and MAE of the random forest for different locations.

radiation.

4. CONCLUSION

In this study, we explored the use of advanced machine learning algorithms for forecasting solar irradiance, a key factor for efficient smart grid management. We applied the PCC for feature selection and Bayesian optimization for fine-tuning the learning algorithms. The findings reveal that the MLP-ANNs perform better when coupled with PCC-based feature selection. The random forest algorithm emerged as the most accurate in our tests, even in different geographical conditions. Future work could explore different feature selection methods, such as wrapper and filter methods, and other ensemble algorithms for potentially better accuracy. Additionally, alternate optimization techniques can be examined to improve the model's performance. Furthermore, using more parameters such as the temperature of the earth's crust, pollution index, and cloud index can improve forecasting.

5. REFERENCES

1. Keeble, Brian R. "The Brundtland report: 'Our common future'." *Medicine and war* 4, no. 1 (1988): 17-25.

2. H. Niakan, A. P. Sirat and B. Parkhideh, "Current Distribution Monitoring of Paralleled GaN HEMTs," 2020 IEEE Energy Conversion Congress and Exposition (ECCE), Detroit, MI, USA, 2020, pp. 2865-2870, doi: 10.1109/ECCE44975.2020.9235770.

3. Nematirad, R., Pahwa, A., Natarajan, B., & Wu, H. Optimal Sizing of Photovoltaic-Battery System for Peak Demand Reduction Using Statistical Models. Frontiers in Energy Research, 11, 1297356, doi.org/10.3389/fenrg.2023.1297356.

4. P. Sirat, H. Mehdipourpicha, N. Zendehdel and H. Mozafari, "Sizing and Allocation of Distributed Energy Resources for Loss Reduction using Heuristic Algorithms," 2020 IEEE Power and Energy Conference at Illinois (PECI), Champaign, IL, USA, 2020, pp. 1-6, doi: 10.1109/PECI48348.2020.9064636.

5. F. Jamaatisomarin, R. Chen, S. Hosseini-Zavareh, and S. Lei, "Laser Scribing of Photovoltaic Solar Thin Films: A Review," Journal of Manufacturing and Materials Processing, vol. 7, no. 3, p. 94, May 2023, doi: 10.3390/JMMP7030094.

6. Amini, S., Omidi, A.H., Afkhami, H., Sabati, H., Mohsenzadeh, A., Soleymani, A., Zonobian, M.A., Ghanbarnejad,

7. N. and Mohammadi, M.R., 2024. Frequency of aac (6′)-le-aph (2′) gene and resistance to Aminoglycoside antibiotics in Staphylococcus aureus isolates. Cellular, Molecular and Biomedical Reports, 4(2), pp.120-128.

8. R. A. Ghaderloo, Y. Shen, C. Singhabahu, R. Resalayyan, and A. Khaligh, "Dead-time Compensation Method for Bus-clamping Modulated Voltage Source Inverter," Jul. 2023, Accessed: Sep. 09, 2023. [Online]. Available: http://arxiv.org/abs/2307.11868.

9. G. Wang, "Hybrid Energy Storage System Supporting Solar PV Farms and Their Dispatch in Australia s National Electricity Market," 2014, Accessed: Sep. 09, 2023. [Online]. Available: https://unsworks.unsw.edu.au/entities/publication/eed76694-f652-470d-b3c3-b5cf64095b74.

10. R. A. Ghaderloo, A. P. Sirat, and A. Shoulaie, "A High-Frequency Active Clamp Forward Converter with Coreless.

11. Transformer," Jul. 2023, Accessed: Sep. 09, 2023. [Online]. Available: http://arxiv.org/abs/2307.12804.

12. P. Gupta, R. S.-I. J. of Sustainable, and undefined 2021, "PV power forecasting based on data-driven models: a review," Taylor & FrancisP Gupta, R SinghInternational Journal of Sustainable Engineering, 2021•Taylor & Francis, vol. 14, no. 6, pp. 1733–1755, 2021, doi: 10.1080/19397038.2021.1986590.

13. R. Arjomandnia, M. Ilbeigi, M. Kazemidemneh, AN. Hashemi, "Renovating buildings by modelling energy–CO2 emissions using particle swarm optimization and artificial neural network (case study: Iran)." Indoor and Built Environment (2023): doi:10.1177/1420326X231151244.

14. Kahrizi, D., Mohammadi, M.R. and Amini, S., 2023. Micro and nano formulas of phyto-drugs naringenin and naringin with antineoplastic activity: Cellular and molecular aspects. Micro Nano Bio Aspects, 2(2), pp.27-33.

15. Yousefpour, A., Foumani, Z. Z., Shishehbor, M., Mora, C., & Bostanabad, R. (2023). GP+: A Python Library for Kernel-based learning via Gaussian Processes. arXiv preprint arXiv:2312.07694.

16. H. Torkaman, E. Zeraatkar, N. Deyhimi, H. H. Alhelou and P. Siano, "Rearrangement Method of Reducing Fault Location Error in Tied Uncompleted Parallel Lines," in IEEE Access, vol. 10, pp. 51862-51872, 2022, doi: 10.1109/ACCESS.2022.3174117.

17. M. Jabbari Zideh and S. S. Mohtavipour, "Two-sided tacit collusion: Another step towards the role of demand-side," Energies, vol. 10, no. 12, p. 2045, Dec 2017, https://doi.org/10.3390/en10122045.

18. S. S. Mohtavipour and M. Jabbari Zideh, "An iterative method for detection of the collusive strategy in prisoner's dilemma game of electricity market," IEEJ Transactions on Electrical and Electronic Engineering, vol. 14, no. 2, pp. 252–260, Feb 2019, https://doi.org/10.1002/tee.22804.

19. R. Nematirad, M. M. Ardehali, and A. Khorsandi, "Optimization of Residential Demand Response Program Cost with Consideration for Occupants Thermal Comfort and Privacy," arXiv.org, May 14, 2023. https://arxiv.org/abs/2305.08077v1 (accessed Jul. 19, 2023).

20. B. Sri Revathi, "A survey on advanced machine learning and deep learning techniques assisting in renewable.

21. energy generation," Environ Sci Pollut Res Int, vol. 30, no. 41, 2023, doi: 10.1007/S11356-023-29064-W.

22. S. Ghimire, R. C. Deo, N. Raj, and J. Mi, "Deep solar radiation forecasting with convolutional neural network and long short-term memory network algorithms," Applied Energy, vol. 253, p. 113541, Nov. 2019, doi: https://doi.org/10.1016/j.apenergy.2019.113541.

23. R. Nematirad and A. Pahwa, "Solar Radiation Forecasting Using Artificial Neural Networks Considering Feature Selection," 2022 IEEE Kansas Power and Energy Conference, KPEC 2022, 2022, doi: 10.1109/KPEC54747.2022.9814765.

24. Z. Pang, F. Niu, and Z. O'Neill, "Solar radiation prediction using recurrent neural network and artificial neural network: A case study with comparisons," Renewable Energy, Apr. 2020, doi: https://doi.org/10.1016/j.renene.2020.04.042.

25. M. Guermoui, K. Gairaa, J. Boland, and T. Arrif, "A Novel Hybrid Model for Solar Radiation Forecasting Using Support Vector Machine and Bee Colony Optimization Algorithm: Review and Case Study," Journal of Solar Energy Engineering, vol. 143, no. 2, Aug. 2020, doi: https://doi.org/10.1115/1.4047852. https://nsrdb.nrel.gov/data-sets/tmy.

26. L. Huang, J. Qin, Y. Zhou, F. Zhu, L. Liu and L. Shao, "Normalization Techniques in Training DNNs: Methodology, Analysis and Application," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 45, no. 8, pp. 10173-10196, Aug. 2023, doi: 10.1109/TPAMI.2023.3250241.

27. K. Cabello-Solorzano, I. Ortigosa de Araujo, M. Peña, L. Correia, and A. J. Tallón-Ballesteros, "The Impact of Data Normalization on the Accuracy of Machine Learning Algorithms: A Comparative Analysis," pp. 344–353, 2023, doi: 10.1007/978-3-031-42536-3_33.

28. Jovic, K. Brkic, and N. Bogunovic, "A review of feature selection methods with applications," 2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), May 2015, doi: https://doi.org/10.1109/mipro.2015.7160458.

29. H. Zhou, J. Zhang, Y. Zhou, X. Guo, and Y. Ma, "A feature selection algorithm of decision tree based on feature weight," Expert Systems with Applications, vol. 164, p. 113842, Feb. 2021, doi: https://doi.org/10.1016/j.eswa.2020.113842.

A. Hajnayeb, A. Ghasemloonia, S. E. Khadem, and M. H. Moradi, "Application and comparison of an ANN-based feature selection method and the genetic algorithm in gearbox fault diagnosis," Expert Systems with Applications, vol. 38, no. 8, pp. 10205–10209, Aug. 2011, doi: https://doi.org/10.1016/j.eswa.2011.02.065.

30. G. Chandrashekar and F. Sahin, "A survey on feature selection methods," Computers & Electrical Engineering, vol. 40, no. 1, pp. 16–28, Jan. 2014, doi: 10.1016/J.COMPELECENG.2013.11.024.

31. S. Ray, Khaldoon Alshouiliy, A. Roy, G. Murtaza, and D. P. Agrawal, "Chi-Squared Based Feature Selection for Stroke Prediction using AzureML," Oct. 2020, doi: https://doi.org/10.1109/ietc47856.2020.9249117.

32. R. J. Hyndman and A. B. Koehler, "Another look at measures of forecast accuracy," Int J Forecast, vol. 22, no. 4, pp. 679–688, Oct. 2006, doi: 10.1016/J.IJFORECAST.2006.03.001.

33. D. Chicco, M. J. Warrens, and G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," PeerJ Computer Science, vol. 7, no. 5, p. e623, Jul. 2021, doi: https://doi.org/10.7717/peerj-cs.623.

34. R. G. Pontius, O. Thontteh, and H. Chen, "Components of information for multiple resolution comparison between maps that share a real variable," Environ Ecol Stat, vol. 15, no. 2, pp. 111–142, Jun. 2008, doi: 10.1007/S10651-007-0043-Y.