Journal of Artificial Intelligence and Metaheuristic (JAIM)

♦ ISSN: 2833-5597 Vol: 08, No. 02, 2024 ♦ PP. 01–06

Article DOI: https://doi.org/10.54216/JAIM.080201



FULL LENGTH ARTICLE

Comparative Analysis of Machine Learning Models for Daytime Power Generation Prediction

Nima Khodadadi[1,*] **Benyamin Abdollahzadeh**[2]

[1]Department of Civil and Architectural Engineering, University of Miami, Coral Gables, FL, USA

[2]Department of Mathematics, Faculty of Science, University of Hradec Králové, 500 03, Hradec Králové, Czech Republic

Emails: nima.khodadadi@miami.edu; benyamin.abdolahzade@gmail.com

Received: March 24, 2024 Revised: May 25, 2024 Accepted: November 1, 2024 ⋆ Corresponding author

ABSTRACT

This paper proposes to evaluate how different machine-learning techniques can be used to predict daytime power generation based on the "Daily Power Generation Data" data set. The six models include Random Forest Regressor, Decision Tree Regressor, Nearest Neighbors, Linear Regression, MLP Regressor, and SVR, and their performance is assessed using multiple metrics. The Random Forest Regressor achieves the lowest Mean Squared Error (MSE), 3.57E-06, among the three machine-learning models. The introduction highlights the role of precise planning in the power market, and the following sections describe the topic mathematically. The table of performance results explains why the Random Forest Regressor is the superior full-proof model, with the lowest MSE, highest explained variance, and great resistance to outlying samples. The paper thus gives useful evaluation criteria for choosing the best model because the Random Forest Regressor achieves the highest performance metrics.

Keywords: Power Generation ▪ Daily Power Generation ▪ Machine Learning ▪ Random Forest Regressor

1. INTRODUCTION

The escalating energy demand and environmental concerns are driving a significant transformation in the global energy landscape, necessitating a shift towards innovative and sustainable power generation solutions [1]. Power generation optimization, amidst technological advancements and resource constraints, is paramount in this endeavor [2]. Fossil fuels have long been the backbone of the global energy system, but their limited availability, environmental impacts, and geopolitical complexities underscore the urgency of transitioning to cleaner energy sources. This research focuses on optimizing power generation to balance rising energy needs with environmental preservation [3]. The guiding principle of this paper is to provide comprehensive insights into power generation phenomena and advocate for sustainable practices [4]. By informing policy decisions, facilitating technological advancements, and engaging diverse stakeholders, including scientists, engineers, policymakers, and environmentalists, we aim to contribute to transforming the global energy landscape and achieving a sustainable future. The study's research objectives serve as road markers, directing attention toward innovative and sustainable solutions to power-generating issues. We delve into three key research questions: firstly, we explore the principal challenges and limitations associated with current power generation methods, aiming to identify barriers to sustainability and effectiveness; secondly, we investigate cutting-edge technologies that could enhance electricity generation efficiency, driving our technological research forward; and finally, we examine strategies to minimize the environmental impact of power generation and promote sustainability, guided by our commitment to environmental stewardship and the preservation of ecological equilibrium [5, 6]. As we embark on this intellectual voyage toward innovation, our collective endeavor is to seek creative, sustainable, and inclusive

© 2024 The Author(s). Published by ASPG. This article is licensed under CC BY 4.0.

approaches to energy generation. This research paper focuses on current power-generating methods, explores technological advancements, and advocates for environmentally friendly practices [7]. We invite the reader to join us on this journey, which promises to be both enlightening and impactful, as we contribute to the development of innovative power generation sectors and work towards sustaining the Earth's oasis state for future generations.

2. LITERATURE REVIEW

Power generation is vital for modern societies, supporting industries, infrastructure, and daily life. With rising energy demand worldwide, ensuring efficient, sustainable, and reliable power generation is crucial. This review covers traditional and renewable energy sources, like solar and wind power, and investigates smart-grid technologies to improve system efficiency and resilience.

Within the sphere of solar energy, achieving optimal solar radiation levels is paramount for the effective design and operation of solar energy systems. Research endeavors such as [8] delve into innovative modeling techniques, like the hybrid PSO–ELM model, which significantly enhances the accuracy of daily solar radiation estimation, particularly in regions where comprehensive onsite data is lacking. Moreover, [9] focuses on advancing power-prediction methodologies for photovoltaic plants through the application of sophisticated machine-learning algorithms. Concurrently, [10] introduces a novel hybrid model, the SDA–GA–ELM, tailored for precise hourly estimation of PV power production, showcasing the relentless pursuit of accuracy in solar-energy forecasting.

Meanwhile, the transition to smart grids presents a host of complex challenges, including demand-side management, cybersecurity, and the optimization of grid infrastructure. The evaluation conducted in [11] provides invaluable insights into the effectiveness of machine-learning algorithms within smart-grid technologies, highlighting the need for further research to optimize data processing and enhance network-management capabilities. Additionally, [12] introduces LSTM-PC, an advanced model aimed at accurately forecasting PV plant energy production, addressing inherent challenges such as weather-dependent predictability and fluctuating energy demand.

Furthermore, [13] undertakes the crucial task of predicting greenhouse-gas emissions from Turkey's electricity sector, employing state-of-the-art deep-learning techniques alongside traditional methodologies. This research underscores the urgent need to curb emissions and transition to cleaner energy sources, given the alarming rise in fossil-fuel consumption and its detrimental environmental impact.

Navigating the complexities of power-system operation requires innovative solutions to effectively manage uncertainty. Reference [14] presents a pioneering method for generating statistically valid scenarios from probabilistic projections, facilitating improved decision-making in power-system management. Meanwhile, [15] proposes a data-driven framework for accurate monthly renewable-energy forecasts, leveraging advanced techniques like STL and LSTM to enhance prediction performance and ensure grid stability.

Lastly, ensuring grid observability remains paramount for efficient electricity-industry operation, as discussed in [16]. By employing state-estimation approaches and machine-learning techniques, this study showcases the potential for optimizing grid operation under varying conditions, paving the way for real-time status prediction and improved system resilience in the face of evolving energy demands and environmental challenges.

In summary, this literature review underscores the imperative of transitioning to clean and sustainable technologies within the power-generation sector. Collaboration among stakeholders, along with active interdisciplinary research, is essential for navigating the evolving energy landscape and fostering the development of a resilient, resourceful, and environmentally friendly power-generation system that meets the needs of society while minimizing environmental impact.

3. DATASET

3.1 Dataset Description

The "Daily Power Generation Data" dataset [17] is notable for its comprehensive information on power-generation activities, offering insights sorted by geographical stations. Covering a significant period from September 1, 2017, to January 19, 2023, this dataset provides a detailed account of electricity-production history, enabling a deep understanding of power-generation dynamics over time. However, it is essential to highlight significant data gaps on specific dates, including October 2, 2017, November 19, 2017, November 26, 2017, April 3, 2018, and April 4, 2018, as well as the prolonged gap from March 19, 2020, to May 31, 2020. These gaps obscure critical insights into power generation during these periods, presenting challenges for comprehensive analysis and decision-making processes.

Furthermore, the dataset is available in two distinct file formats, necessitating precise processing and analysis to ensure consistency and accuracy across both formats. Overcoming these challenges is paramount for reliable data preparation, particularly for predictive analysis and informed decision-making in energy management. Exploring strategic approaches to address data gaps and associate power-generation processes with level estimations will be essential to enhance the dataset's trustworthiness and usefulness for future analysis. By confronting these challenges head-on, researchers can unlock valuable insights and facilitate informed decision-making in the realm of power-generation dynamics.

3.2 Dataset Pre-processing Steps

Before commencing the analysis of daily power-generation data, a meticulous preparatory phase was undertaken to ensure the dataset's integrity and consistency. The following processes delineate the preprocessing methods employed [18]:

- **Handling Missing Data:** The identification and resolution of dates with missing data involved a thorough review to understand the underlying reasons. Imputation techniques were then applied to fill these gaps, offering researchers flexibility in choosing methods tailored to the data's nature and analytical objectives. Techniques such as mean imputation, interpolation, or predictive modeling were considered based on missing data and research objectives.

- **File Format Harmonization:** Given the disparity in file formats across the dataset, a harmonization method was implemented to ensure accurate representation of the dataset. This involved converting and aligning the storage format to establish a well-ordered structure conducive to future analyses. Additionally, efforts were made to reconcile discrepancies in data encoding or formatting conventions between the two files to ensure seamless integration and consistency.
- **Standardization of Units:** A systematic standardization procedure was executed on the dataset's units of measurement to enhance clarity and facilitate comparisons across variables. Specifically, a methodical transformation from mega units (MU) to megawatts (MW) was conducted, promoting a cohesive and uniform analysis.

This conversion facilitated a more intuitive understanding of power-generation metrics and eliminated potential confusion arising from disparate unit representations. Ensuring dataset integrity was crucial during preparation, with subsequent verifications conducted to address any inconsistencies. Through cross-variable consistency tests, researchers scrutinized relationships between variables, rectifying any discrepancies detected to ensure coherence and accuracy.

3.3 Descriptive Statistics

Descriptive statistics were instrumental in comprehensively understanding the traits and patterns within the "Daily Power Generation Data," offering a quantitative overview of key characteristics and scrutinizing the dataset's major tendencies and variation. Spanning September 1, 2017, to January 19, 2023, the dataset captures a rich tapestry of temporal dynamics in daily power generation. Across diverse power stations, each contributing uniquely to the overall landscape, computed descriptive statistics shed light on individual performance metrics. The dynamic heat map in Figure 1 illustrates the daily power-generating statistics. By utilizing color gradients, it reveals variations in power levels over time and across different stations, providing valuable information on geographical and temporal trends.

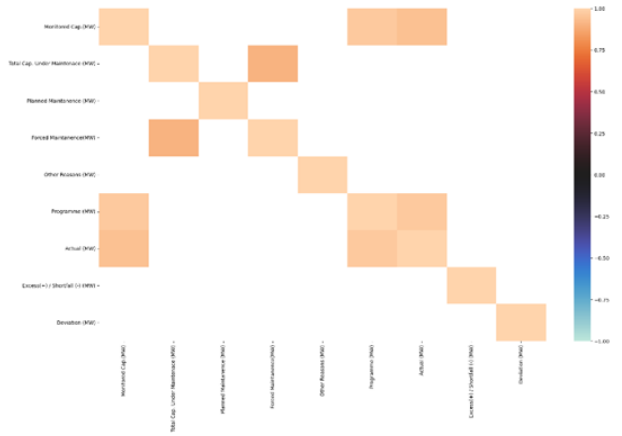


Figure 1. Heatmap of the dataset.

Figure 2 vividly displays variations between projected and actual power generation, using bars to represent real data.

4. RESULTS

This section is devoted to the outcomes of applying several machine-learning cases to the "Daily Power Output Data" dataset. The model was used in the back-end system to fore-cast power generation from different power stations, using features such as Expected Power Generation, Actual Power Generation, Deviation, and a few others. The objective of this study is to test the models and select the one that provides better power-generation prediction than the other. The following machine-learning models were evaluated:

1. Random Forest Regressor
2. Decision Tree Regressor
3. Nearest Neighbors
4. Linear Regression
5. MLP Regressor
6. SVMRegressor

For each model, we developed indicators to offer a comprehensive insight into the model's accuracy in forecasting power production. The metrics considered are Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Explained Variance Score (EVS), Max Error, Median Absolute Error (MedAE), Mean Absolute Percentage Error (MAPE), R-squared (R^2), and the Modified Tchebycheff Distance.

The following sections focus on the exact results concerning each design, supplying an understanding of their pros and cons in capturing the specialized details of the electricity-producing dataset. This analysis provides the basis for making more informed decisions about the most suitable model choice for forecasting power generation.

4.1 Machine Learning Models

This subsection describes the specific machine-learning models used to analyze the "Daily Power Generation Data" dataset. These predictive-analytics tools have different design features and mathematical foundations, leading to a wide range of predictive-analytics models. The choice of these models represents the highly interrelated nature of the data, which contributes to the overall picture of power-generation processes [19].

Random Forest Regressor: The Random Forest Regressor is an ensemble learning method that uses a collection of decision trees formed during training. This method works by combining the trees' forecasts. The forest delivers a robust and accurate model. Every tree is trained on a different slice of the data, which lowers the overall diversity of datasets, thus reducing the associated risk of overfitting. The final prediction averages the individual tree predictions [20]:

$$H(x) = \frac{1}{N}\sum_{i=1}^{N} h(x, \theta_i). \tag{1}$$

Decision Tree Regressor: A Decision Tree Regressor evenly splits its dataset into subsets using the most important feature at each node. A simple averaging of the relative target variables at the leaves of the decision tree achieves the final output. A decision tree is natural and complex enough to represent relationships in the data that are not linear. Let R_m represent the region represented by the m-th leaf node, and let $\bar{y}_m$ be the mean of the target variable in R_m. The prediction

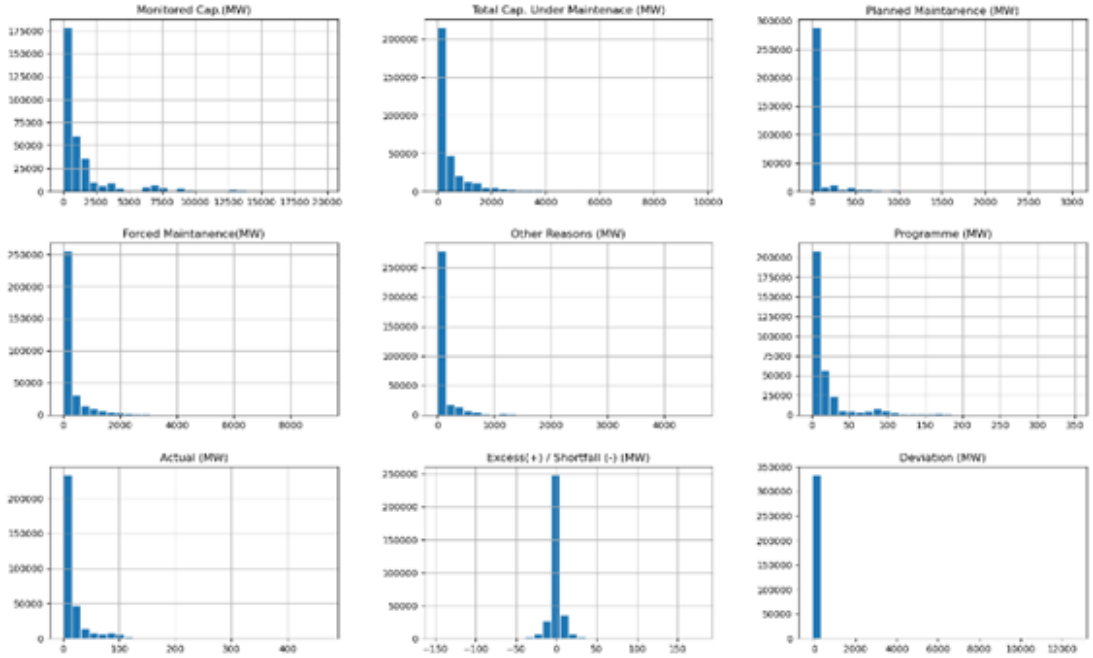


Figure 2. Histograms of the dataset.

is:

$$H(x) = \sum_{m=1}^{M} \bar{y}_m I(x \in R_m). \tag{2}$$

Nearest Neighbors: The k-nearest-neighbor algorithm predicts a desired variable based on the average of the k nearest neighbors in the feature space. It hypothesizes that the same series of inputs results in the same output values, so it will be weighted. For a given point x, the prediction is the average of the k nearest neighbors' target values:

$$H(x) = \frac{1}{k} \sum_{i=1}^{k} y_i. \tag{3}$$

Linear Regression: Linear regression assumes a straight-line association between the target variable and the independent variables through a linear function. The linear relationship between the independent variables and the outcome variable is also assumed. Given features X, weights θ, and bias b, the prediction is:

$$H(X) = X \cdot \theta + b. \tag{4}$$

MLP Regressor (Multi-Layer Perceptron): The MLP Regressor is a class of multilayer neural networks based on the structure of two-layer networks with input, intermediary, and output layers. The network has the power to unveil hidden links in the data by propagating the activating neuron signals throughout it. For a given input X, the output is calculated through activation of neurons in each layer:

$$H(X) = \sigma\left(W^{(2)} \cdot \sigma\left(W^{(1)} \cdot X + b^{(1)}\right) + b^{(2)}\right). \tag{5}$$

SVR (Support Vector Regressor): SVR utilizes support vectors to determine a nonlinear hyperplane with the fewest total deviations from the given data for the best projections in an artificially created space. It provides a useful function for capturing non-respective relations between data. Given input features X, target values y, a set of support vectors, and a kernel function K, the prediction is:

$$H(X) = \sum_{i=1}^{n} \alpha_i y_i K(X, X_i) + b, \tag{6}$$

where α_i are support-vector weights, y_i are target values, X_i are support vectors, and b is a bias term.

These mathematical details provide insight into how the models work and support interpretation of the results.

4.2 Performance Metrics

The performance metrics employed for assessing machine-learning models on the "Daily Power Generation Data" dataset have been designed for a particular reason. The metrics are used to assess model predictive performance and define different aspects of performance.

Mean Squared Error (MSE): MSE calculates the squared deviation of actual values from predicted values and their average. It scores larger deviations harder than small errors, which, in turn, provides a measure of overall model accuracy. Lower MSE values indicate optimal model performance based on small prediction errors.

Root Mean Squared Error (RMSE): RMSE is an algebraic formula of MSE, which makes it more comprehensible because RMSE tells errors in the same unit as the target variable. As with MSE, smaller RMSE values primarily point to higher model accuracy.

Mean Absolute Error (MAE): MAE presents the average deviation between the real value and the predicted value. It is a rigorous test of the accuracy of the model's prediction capability. Lower MAE values indicate higher accuracy since they more suitably ignore outliers.

Explained Variance Score (EVS): EVS indicates model accuracy by the ratio of residual variance among the explained variance of the target variable. A score of 1 means perfect prediction, and a lower score indicates less accurate predictions. Higher EVS observations display better concordance, while

Table 1. Regression Models Results

Models	MSE	RMSE	MAE	EVS	Max error	MedAE	MAPE	R^2	MTD	RRMSE	Willmott	MBE	SD
Random Forest Regressor	3.57×10^{-6}	0.00189	0.000714	0.780449	0.14408	0.000348	4.3×10^{9}	0.780449	3.57×10^{-6}	0.008007	0.000346	-9.9×10^{-7}	0.00189
Decision Tree Regressor	6.16×10^{-6}	0.002482	0.001056	0.621495	0.182195	0.000552	2.28×10^{10}	0.621422	6.16×10^{-6}	0.010515	0.989277	3.45×10^{-5}	0.002481
Nearest Neighbors	8.55×10^{-6}	0.002924	0.001408	0.475002	0.176012	2.82×10^{-5}	1.16×10^{11}	0.474369	8.55×10^{-6}	0.01239	0.960413	0.000102	0.002922
Linear Regression	1.33×10^{-5}	0.003651	0.001833	0.181297	0.226556	0.000782	1.63×10^{12}	0.18041	1.33×10^{-5}	0.015471	0.030117	−0.00012	0.003649
MLP Regressor	1.46×10^{-5}	0.003822	0.002061	0.110707	0.2254	0.000895	1.68×10^{12}	0.101939	1.46×10^{-5}	0.016195	0.089541	−0.000038	0.003803
SVR	0.008185	0.090469	0.090336	−0.53908	0.133703	0.0091572	2.5×10^{13}	−0.50215	0.008185	0.383326	0.002338	0.090331	0.005004

a score of 1 stands for the perfect fit, reflecting the actual estimation.

Max Error: Max Error expresses the largest possible difference in absolute values of the predicted and actual values. It determines the largest error between actual observations and forecasts produced by machine learning. Smaller Max Error values mean less severe prediction-error indices.

Median Absolute Error (MedAE): MedAE stands for the median absolute difference between expected and resulting outcomes. It provides a means to assess prediction error. For the data, smaller values indicate more accuracy, especially in datasets with outliers.

MAPE (Mean Absolute Percentage Error): MAPE gives the average percentage error that occurred between actual and predicted values, which enables us to estimate the size of errors. Larger MAPE values indicate worse accuracy, and the metric has the advantage of indicating errors as a percentage.

R-squared (R^2): R^2 explains the model in terms of the percentage of variance in the target variables that the model explains. A value of 1 means the model perfectly fits the data, while 0 means there is no connection modeled between predictors and the dependent variable. Stronger R^2 values reflect a better model fit, with 1 ideal and representing perfect variance explained.

Modified Tchebycheff Distance (MTD): MTD compares the similarity between two probability distributions, giving a measure of how close and productive the models are when aiming to predict the true distribution of the target variable. Lower MTD (mean target deviation) represents better tuning between forecasts and the real phenomenon.

Relative Root Mean Squared Error: RMMSE restructures RMSE to take the mean of the observed values, facilitating comparison across a range of datasets. Lower RRMSE values indicate higher model performance when the dataset is large enough.

Willmott Index: The Willmott Index employs preferred risk, negativity, and correlation to deduce whether the model functioned properly. It ranges from −1 to 1, with 1 standing for excellent harmony. Values closer to 1 mean high model accuracy if the weighted average comprising both associations and bias is not high enough.

Mean Bias Error (MBE): MBE figures out the average difference between observed and estimated values, which is useful for identifying systematic errors. MBE values close to 0 indicate that a machine-learning model is highly unbiased and its predictions are free of any biases.

Standard Deviation (SD): SD indicates the level of spread among predicted values by indicating the amount of variation or dispersion in the value set. It provides reliable information and a route to consistency. Smaller SDs imply smaller spreads around the mean.

Knowing what the performance metric measures and how to interpret it unites all the machine-learning models' contributions to predicting power generation based on the given dataset.

4.3 Regression Results

Here, we further examine the results of our machine-learning model experiments using a variety of measures. The main goal is to conduct a thorough examination of the models' abilities to accurately predict power-generation values within the "Daily Power Generation Data" dataset. As indicated above, a series of different metrics is used to highlight different aspects of predictive performance individually. Table 1 showcases the performance metrics for each evaluated machine-learning model.

Discussion of Key Findings

1. **Random Forest Regressor:** Laboratory experimental results revealed the lowest MSE, RMSE, and MAE, consequently showing highly precise predictions. High EVS and R^2 values indicate great modeling ability for the data. It demonstrates a lower maximum absolute error and MedAE, suggesting good removal of outliers.
2. **Decision Tree Regressor:** It allows the ability to compete across performance metrics. High Willmott Index and R^2 values suggest a good fit for the case data. Only 0.4 and 2.15 units of MAE and Max Error are higher than those of Random Forest, respectively.
3. **Nearest Neighbors:** Lower MAE and MedAE are evident, suggesting high accuracy and precision when predicting real energy-generation values. Diagnostic R^2 and EVS values are significantly lower than those of Random Forest and Decision Tree; hence, they reflect a smaller portion of the response variability.
4. **Linear Regression:** It does reasonably well for moderate values under most of the metrics. Ideally, the R-squared and adjusted R-squared values (R^2 and EVS) of tree-based models should be higher when predicting nonlinear relationships.
5. **MLP Regressor:** This model illustrates complex relationships with competitive performance metrics. Its Max Error is also slightly higher than that of the Random Forest and Decision Tree models.
6. **SVR:** Negative R^2 and EVS imply a lack of suitable representation of the underlying pattern in the data.

The Random Forest Regressor [20] achieves the best performance results among the considered models according to the complete assessment for this dataset, suggesting it as the most probable model for power-production prediction. Its com-

bination of low error metrics, high explained variance, and robustness to outliers makes it a top method for this problem.

5. CONCLUSION

In summary, this research aims to achieve accurate forecasts of power generation using a data-driven approach, accomplished through in-depth analysis of multiple machine-learning models on the "Daily Power Generation Data" dataset. The exploration covers six models: Random Forest Regressor, Decision Tree Regressor, Nearest Neighbors, Linear Regression, MLP Regressor, and SVR, evaluated by performance metrics on all fronts. The Random Forest Regressor was judged the most excellent model in terms of the lowest Mean Squared Error (MSE), 3.57E-06, indicating its effectiveness. Evaluation using RMSE, MAE, EVS, and Max Error demonstrates its performance in handling outliers and external variations in power-data analysis. The results are significant for those involved in power-company management, providing useful information about the method best suited to implement predictive analytics successfully. The Random Forest Regressor demonstrated dominant competence among the models, while the others performed adequately but less well than the RF Regressor. This supports the RF Regressor as the best model for accurate and reliable power-generation predictions. Nevertheless, developing an ultimate framework to meet all target aspects and limitations remains challenging.

REFERENCES

[1] E. M. Almetwally and M. A. Meraou, "Application of environmental data with new extension of Nadarajah-Haghighi distribution," *Computational Journal of Mathematical and Statistical Sciences*, vol. 1, no. 1, pp. 26–41, 2022.

[2] C. Scott, M. Ahsan, and A. Albarbar, "Machine learning for forecasting a photovoltaic (PV) generation system," *Energy*, vol. 278, p. 127807, 2023.

[3] S. K. Towfek, "Navigating the storm: Cutting-edge risk mitigation and analysis for volatile markets," *Journal of Artificial Intelligence and Metaheuristics*, vol. 4, no. 2, pp. 36–44, 2023.

[4] S. K. Towfek, "A semantic approach for extracting the medical association rules," *Journal of Artificial Intelligence and Metaheuristics*, vol. 5, no. 1, pp. 46–52, 2023.

[5] A. Sedai, R. Dhakal, S. Gautam, A. Dhamala, A. Bilbao, Q. Wang, A. Wigington, and S. Pol, "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.

[6] R. Wazirali, E. Yaghoubi, M. S. S. Abujazar, R. Ahmad, and A. H. Vakili, "State-of-the-art review on energy and load forecasting in microgrids using artificial neural networks, machine learning, and deep learning techniques," *Electric Power Systems Research*, vol. 225, p. 109792, 2023.

[7] E.-S. El-Kenawy, A. Ibrahim, S. Mirjalili, Y.-D. Zhang, S. Elnazer *et al.*, "Optimized ensemble algorithm for predicting metamaterial antenna parameters," *Computers, Materials & Continua*, vol. 71, no. 3, pp. 4989–5003, 2022.

[8] D. S. Khafaga, A. Ibrahim, E.-S. M. El-Kenawy, A. A. Abdelhamid, F. K. Karim *et al.*, "An Al-Biruni earth radius optimization-based deep convolutional neural network for classifying monkeypox disease," *Diagnostics*, vol. 12, no. 11, p. Article 11, 2022.

[9] A. Djaafari, A. Ibrahim, N. Bailek, K. Bouchouicha, M. A. Hassan *et al.*, "Hourly predictions of direct normal irradiation using an innovative hybrid LSTM model for concentrating solar power projects in hyper-arid regions," *Energy Reports*, vol. 8, pp. 15 548–15 562, 2022.

[10] H. AlEisa, E.-S. El-Kenawy, A. Alhussan, M. Saber, A. Abdelhamid *et al.*, "Transfer learning for chest X-rays diagnosis using dipper throated algorithm," *Computers, Materials & Continua*, vol. 73, no. 2, pp. 2371–2387, 2022.

[11] E.-S. M. El-Kenawy, N. Khodadadi, S. Mirjalili, T. Makarovskikh, M. Abotaleb *et al.*, "Metaheuristic optimization for improving weed detection in wheat images captured by drones," *Mathematics*, vol. 10, no. 23, p. Article 23, 2022.

[12] A. A. Abdelhamid, S. K. Towfek, N. Khodadadi, A. A. Alhussan, D. S. Khafaga *et al.*, "Waterwheel plant algorithm: A novel metaheuristic optimization method," *Processes*, vol. 11, no. 5, p. Article 5, 2023.

[13] D. Khafaga, E.-S. El-Kenawy, F. Karim, S. Alshetewi, A. Ibrahim *et al.*, "Optimized weighted ensemble using dipper throated optimization algorithm in metamaterial antenna," *Computers, Materials & Continua*, vol. 73, no. 3, pp. 5771–5788, 2022.

[14] S. Bhavsar, R. Pitchumani, and M. A. Ortega-Vazquez, "Machine learning enabled reduced-order scenario generation for stochastic analysis of solar power forecasts," *Applied Energy*, vol. 293, p. 116964, 2021.

[15] S. Ding, H. Zhang, Z. Tao, and R. Li, "Integrating data decomposition and machine learning methods: An empirical proposition and analysis for renewable energy generation forecasting," *Expert Systems with Applications*, vol. 204, p. 117635, 2022.

[16] D. Mukherjee, S. Chakraborty, and S. Ghosh, "Power system state forecasting using machine learning techniques," *Electrical Engineering*, vol. 104, no. 1, pp. 283–305, 2022.

[17] Kaggle, "Daily power generation data," Kaggle dataset.

[18] S. Alam and N. Yao, "The impact of preprocessing steps on the accuracy of machine learning algorithms in sentiment analysis," *Computational and Mathematical Organization Theory*, vol. 25, no. 3, pp. 319–335, 2019.

[19] C. Janiesch, P. Zschech, and K. Heinrich, "Machine learning and deep learning," *Electronic Markets*, vol. 31, no. 3, pp. 685–695, 2021.

[20] L. Xue, Y. Liu, Y. Xiong, Y. Liu, X. Cui *et al.*, "A data-driven shale gas production forecasting method based on the multi-objective random forest regression," *Journal of Petroleum Science and Engineering*, vol. 196, p. 107801, 2021.