Proceedings of the 20th Conference on Computer Science and Intelligence Systems (FedCSIS) pp. 405–410 DOI: 10.15439/2025F9025 ISSN 2300-5963 ACSIS, Vol. 43

Unified Data-Driven Prediction of Photovoltaics Output from Weather and Geographic Data Across Diverse Systems

Joanna Wójcicka
Faculty of Electrical Engineering,
Automatics, Computer Science
and Biomedical Engineering,
AGH University of Krakow,
Al. Mickiewicza 30,
30-059 Krakow, Poland
Email: wojcickajoanna1@gmail.com,
joannawoj@student.agh.edu.pl

Tomasz Hachaj
ORCID: 0000-0003-1390-9021
Faculty of Electrical Engineering,
Automatics, Computer Science
and Biomedical Engineering,
AGH University of Krakow,
Al. Mickiewicza 30,
30-059 Krakow, Poland
Email: thachaj@agh.edu.pl

***Abstract*—As the adoption of renewable energy continues to rise, precise forecasting of solar power generation is essential for optimizing energy storage and distribution. This article explores the prediction of energy output in photovoltaic systems using machine learning models that leverage environmental and geographical factors. The study utilizes data from 9,182 private photovoltaic installations across Poland and publicly available weather records. Additionally, a data preprocessing method was introduced to filter out non-useful data, such as records indicating malfunctioning installations, ensuring that only relevant information is used for prediction. Key variables considered include temperature, cloud cover, wind speed, and solar panel efficiency. This paper studies the effectiveness of data-driven energy production forecasting methods, namely linear regression, polynomial regression, decision tree regression, random forest regression, and multilayer fully connected artificial neural network, designed to make predictions for various installations with different parameters and geographical locations, considering atmospheric conditions in contrast to frequently published articles in which predictions are fitted on data from a single photovoltaic installation. Due to this, our work has broad research value, explores the boundaries and limitations of such approaches, and can be considered a reference for energy engineers, computer scientists, and researchers.**

***Index Terms*—Photovoltaic Infrastructure, Machine Learning, Prediction, Feature Selection, Data-Driven**

I. Introduction

RENEWABLE energy has experienced substantial growth in recent decades, driven by the increasing global population and the recognition of the finite nature of fossil fuel resources. The transition to renewable energy sources is widely regarded as a viable strategy to reduce dependence on fossil fuels while mitigating the environmental impacts of their extraction and consumption. Technological advancements in renewable energy have been considerable, with numerous solutions already implemented in practical applications in private and industrial sectors. Additionally, reliable environmental and production data availability has improved significantly, caused by the growing adoption of monitoring systems in private photovoltaic installations.

Among renewable energy sources, solar and wind technologies have been recognized for their substantial potential to meet global energy demands. As noted by Gross et al. [1], these technologies were considered to have the most significant impact within the renewable energy sector. However, subsequent research by Olabi and Abdelkareem [2] emphasised that both solar and wind energy are highly dependent on environmental conditions, introducing variability and limiting their reliability as primary global energy sources.

One of the critical challenges in integrating renewable energy sources is the efficient management of energy storage and distribution to maintain a stable balance between generation and consumption. As discussed by Roberts et al. [3], the variability of renewable energy output, primarily influenced by environmental conditions, requires the implementation of adaptive and responsive strategies for storage and distribution. At a macro-level, particularly within centralised energy systems such as Nord Pool [4], this involves determining optimal periods for storing energy for future use versus distributing it to external sectors to meet immediate demand. In addition, these strategies should include dynamic pricing models that account for fluctuations in energy demand and the availability of generation resources, thereby enhancing the overall efficiency and reliability of the energy system.

Therefore, accurate forecasting of solar energy production has become a critical component in the efficient operation of modern energy systems. Reliable predictions enable better planning for energy storage, grid balancing, and market operations. Forecasting supports decision-making processes by providing information on expected energy availability, essential for maintaining supply stability, reducing reliance on backup fossil fuel generation, and optimising energy pricing and distribution strategies.

Two primary methodologies are commonly used to forecast the energy production of solar installations. The first is a deterministic approach, which involves the development of a physical model that represents the electrical behaviour of photovoltaic cells. This method simulates the underlying physical processes to estimate the energy output based on the known specifications and characteristics of the solar panel. A wide range of physical models can be applied to predict the performance of photovoltaic systems with varying configurations, as examined by Dolara et al. [5], [6].

An alternative approach to forecasting solar energy production involves using statistical and machine learning models, which are developed based on historical data from photovoltaic installations and relevant environmental variables that influence energy generation. These methods aim to identify and model the underlying relationships between features in the dataset to enable accurate prediction. Linear regression and artificial neural networks (ANNs) are the most commonly used models in this domain. Both approaches have been extensively applied in various research contexts with notable success. For example, Gratidi et al. [7], Barrera et al. [8], and Zazoum [9] each developed accurate prediction models tailored to different datasets, photovoltaic system architectures, and methodological frameworks. These studies highlight the flexibility and effectiveness of machine learning techniques in capturing the complexity inherent in solar energy production forecasting.

Despite significant advancements and promising outcomes in solar energy forecasting, considerable room for improvement remains. Studies by Gratidi et al. [7] and Barrera et al. [8] commonly relied on data collected from small-scale installations located in research centres, which, while controlled and consistent, may limit the model's ability to generalise to real-world scenarios. For predictive models to be applicable globally, it is crucial to incorporate diverse datasets drawn from various geographic locations and installation types. Additionally, many existing studies utilize well-structured solar radiation data complemented by ambient temperature measurements. Although this data can enhance model accuracy, it often depends on specialized infrastructure that may not be available in all regions. This reliance poses a challenge to scalability, particularly in areas lacking comprehensive meteorological monitoring systems. It highlights the need for models capable of operating effectively with more readily accessible environmental data.

A. Novelty of this paper

This work examines the effectiveness of statistical and machine learning models for predicting energy output in photovoltaic systems that leverage environmental and geographical factors. New to previous work dealing with similar topics, the study was conducted on a large dataset containing installations located almost all over a medium-sized European country. Our goal was to create a predictive model that can effectively predict energy production based on data such as the power of an installation, its geographic location, and weather data to improve forecast accuracy [10]. Thus, in this work, we study the effectiveness of data-driven energy production forecasting methods designed to make predictions for various installations with different parameters and geographical locations, taking into account atmospheric conditions, in contrast to frequently published articles in which predictions are fitted on data from a single photovoltaic installation [11]–[15]. Due to this fact, our work has broad research value, explores the boundaries and limitations of such an approach, and can be considered a reference for energy engineers, computer scientists, and researchers.

II. Materials and Methods

The foundation of this study is a dataset that accurately represents real-world conditions of photovoltaic installations. The data were collected from multiple private solar installations and supplemented with commonly observed meteorological variables to ensure a comprehensive analysis. The selected methodology prioritizes data that are accessible to the average user, eliminating the need for specialized professional measuring equipment. The dataset from photovoltaic installations covers the year 2020 and, due to licensing and legal restrictions, cannot be publicly shared with other researchers.

A. Data set

The initial database comprised 145,068,684 records from 9,182 photovoltaic installations in Poland through the PVmonitor monitoring system. The measurements taken from these installations were aggregated every three minutes, capturing various parameters such as energy production, efficiency over specific time intervals, the cumulative sum of energy generated since monitoring began, and the corresponding timestamps. Each installation was assigned a unique ID and included localization details and the manufacturer's power specifications for the photovoltaic panels.

The weather dataset was gathered to use only commonly available environmental information like temperature, cloudiness, or wind speed. To achieve this, the data was collected from a publicly available AccuWeather archive using a data scraping technique. This source was selected due to its frequent measurements and broad coverage of features relevant to this study. A notable characteristic of this dataset is the cloudiness variable, which is provided as a percentage. This numerical representation is more precise and requires less preprocessing than categorical cloud cover classifications.

B. Preprocessing

Since the installations dataset was collected from private users, it relied heavily on user input, with no control over the accuracy of the solar panel setup or the calibration of the monitoring devices. This resulted in a necessity to validate and prune data strictly. Several conditions must be satisfied to determine whether or not an installation is active and whether it should be considered.

- Power of Installation provided by a user cannot be equal to zero,

- Installation should at least once in a month reach over 40% efficiency,
- Installation's mean efficiency in winter (October through March) should be over 5%,
- Installation's mean efficiency in summer (April through September) should be over 10%,
- Installation has to be monitored throughout the entire year.

Applying these conditions reduced the number of installations to 32. Although this significantly decreased the amount of data available for energy production predictions, ensuring that the dataset contained only high-quality records was necessary. Despite the reduction, the remaining installations still contributed 181,306 measurement points, providing sufficient data for accurate predictions.

The next step was to combine both datasets into a consistent database. Since installation measurements were made every three minutes and weather data was aggregated hourly, it was necessary to aggregate the measurements to match this time frame. Aggregation was achieved by calculating the mean value of each efficiency of each Installation and the sum of power (P) and power increment (DPV). Finally, the data were aligned based on geographical location.

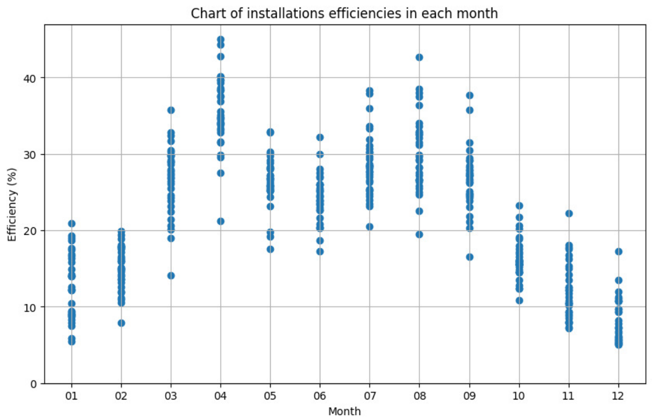

Fig. 1. Chart of mean installations' efficiencies aggregated monthly

After proper data filtering, it was possible to distinguish a desired trend in efficiency throughout the year presented in Figure 1. This trend is a direct result of yearly changes in the solar azimuth angle because photovoltaic installations are more efficient when the radiation reaches solar panels at around 90° angle [16], [17] (as a combination of azimuth and tilt angles) that occurs at the spring and autumnal equinox in Poland.

C. Experimental setup

The dataset was divided into training and testing sets in an 80%-20% ratio based on installation indexes, ensuring that instances from the same installation did not appear in both sets, thus maintaining a clear separation for evaluation. Due to the data being inconsistent regarding values range, the action taken to ensure good performance of models was scaling both training and test data using the Equation 1.

$$Z = \frac{x - \rho}{\sigma} \tag{1}$$

where:
x is the value of sample,
ρ is a mean values of all samples and
σ is the standard derivation.

The linear regression model was proposed as a baseline model for comparison, as it is a simple machine learning model that does not require additional hyperparameters tuning. Another models we also evaluated were polynomial regression, K-nearest neighbour regression (KNN), decision tree regression, random Forest regression (RFR) and artificial neural network (ANN). The architecture of implemented ANN is detailed in Table I.

TABLE I
ARCHITECTURE OF IMPLEMENTED ARTIFICIAL NEURAL NETWORK

Layer	Layer Type	Number of Neurons	Activation Function
Input	Input	7 (number of features)	
Hidden n°1	Dense	10	ReLU
Hidden n°2	Dense	10	ReLU
Hidden n°3	Dense	10	ReLU
Output	Dense	1	Linear

D. Metrics used to evaluate models' performance

Evaluating machine learning models using standardised metrics is essential to ensure comparability. The performance of the models was assessed on the following metrics:

- Normalized Root Mean Squared Error

$$NRMSE_{minmax} = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}}{y_{max} - y_{min}}, \; y_{max} \neq y_{min} \tag{2}$$

- Mean Squared Error

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{3}$$

- R-squared

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2}, \exists i, j \text{ such that } y_i \neq y_j \text{ for } i \neq j, \tag{4}$$

where:
N is a number of testing samples,
y_i are the predicted values,
$\hat{y}_i$ are the actual values,
$y_{\max}$ and $y_{\min}$ are the maximum and minimum values in the dataset,
$\bar{y}$ is mean of the observed values.

III. RESULTS

Table II presents the evaluation metrics calculated for the training and test datasets. The NRMSE$_{minmax}$ metric is the primary indicator of predictive accuracy, as it assesses the performance of the models relative to the data scale. In addition,

TABLE II
EVALUATION OF MODELS ON TRAINING AND TEST DATA

Model	NRMSE$_{minmax}$	MAE	R^2
Test Set			
Linear Regression	0.1809	17.4593	0.2390
Regression with Polynomial Features	0.1210	11.1261	0.6591
K-Nearest Neighbors (KNN)	0.1304	10.5962	0.6045
Decision Tree Regressor (DTR)	0.1237	9.8932	0.6445
Random Forest Regressor (RFR)	**0.1113**	**9.1079**	**0.7122**
Artificial Neural Network (ANN)	**0.1137**	**9.4136**	**0.6994**
Training Set			
Model	**NRMSE$_{minmax}$**	**MAE**	**R^2**
Linear Regression	0.2019	17.4593	0.2290
Regression with Polynomial Features	0.1260	10.1974	0.6996
K-Nearest Neighbors (KNN)	**0.0637**	**4.0101**	**0.9231**
Decision Tree Regressor (DTR)	0.1179	7.9926	0.7228
Random Forest Regressor (RFR)	**0.0254**	**1.5648**	**0.9877**
Artificial Neural Network (ANN)	0.1189	8.7776	0.7322

MSE and R^2 serve as complementary metrics, offering further insight into the overall effectiveness of the models.

In the test set, the NRMSE$_{minmax}$ values indicate that the best-performing models are RFR and ANN, with RFR showing a slight advantage. This suggests that RFR is particularly effective at capturing the dataset's complexity while maintaining high predictive accuracy. In contrast, models such as DTR and polynomial regression obtained satisfactory results but did not achieve the same level of accuracy as RFR and ANN.

Comparison with training data provides insight into the generalisability of the models. Although RFR and KNN achieve exceptionally high performance on the training data, their performance on the test dataset is noticeably lower. Among the evaluated models, ANN exhibits the most consistent performance in both training and test datasets, indicating a strong balance between predictive accuracy and generalisability.

There is a notable improvement between the baseline model (linear regression) and the more advanced models, emphasizing the effectiveness of non-linear approaches. Among the models tested, RFR demonstrates the best overall performance, closely followed by ANN. Additionally, based solely on the evaluation metrics, regression with polynomial features and KNN exhibit strong performance for this dataset. Further insights into the models can be obtained by analysing Figures 2 and 3, which compares the real and predicted values for each model across training and test datasets. These visualizations offer a clear representation of the models' generalization capabilities and enable a direct comparison of their performance on training versus test data.

To objectively evaluate the performance of the models, it is

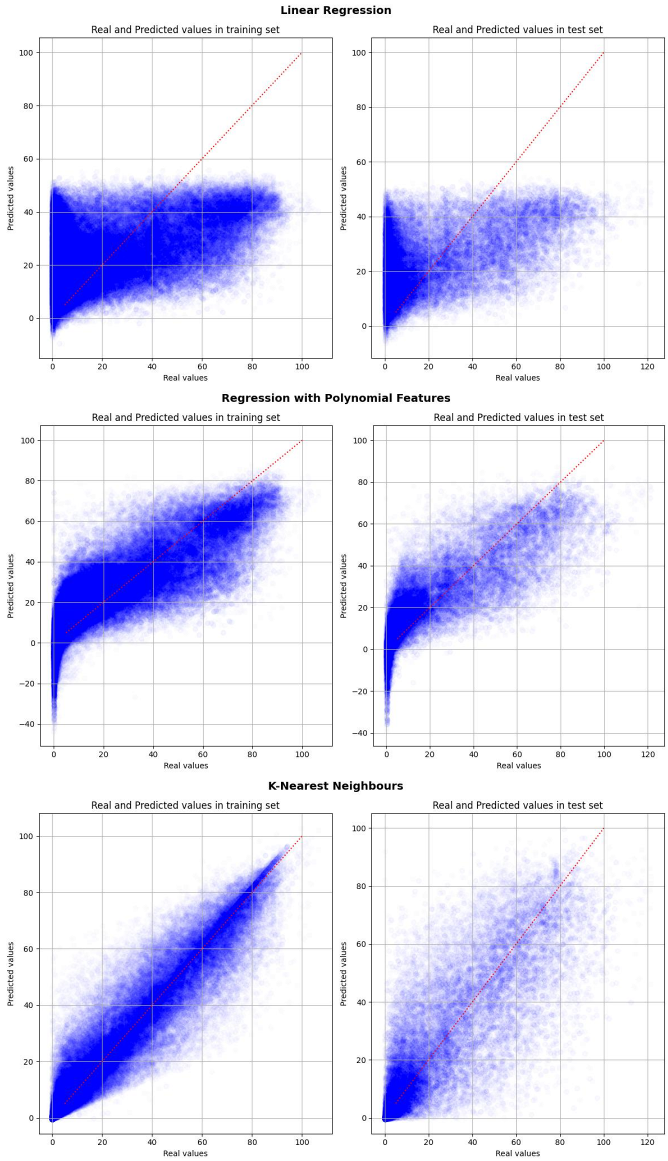

Fig. 2. Comparison of training and test predicted values vs actual values - linear regression, regression with polynomial features and K-nearest neighbour

crucial to compare the results with those reported in other studies. Table III presents a performance summary of the selected models in various data sets and methodologies. The comparison is based on NRMSE values, with linear regression serving as the baseline, normalised to 100% using the min-max method in this study and different normalisation approaches reported in the studies [7], [18], [19]. The relative performance of other models is then expressed as a percentage improvement or decline relative to this baseline. This approach ensures the comparability of the metric between different methodologies.

IV. DISCUSSION

The comparison between the training and test results (Table II) provides insights into the generalisation capabilities of the

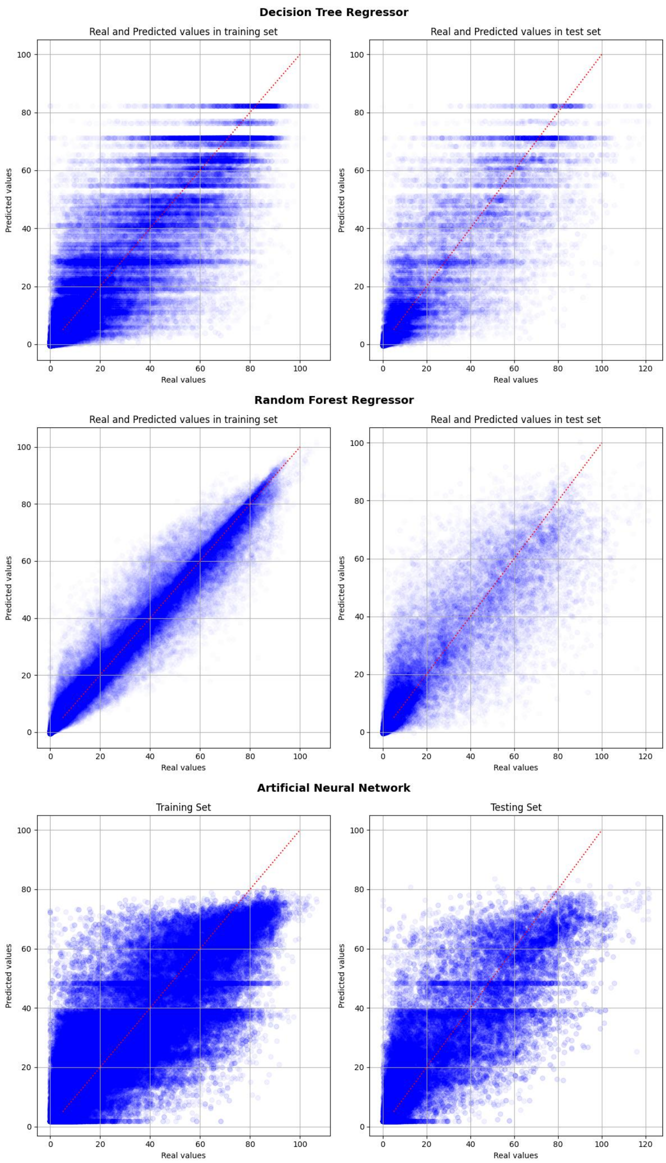

Fig. 3. Comparison of training and test predicted values vs actual values - decision tree regressor, random forest regression and artificial neural network

models. Although RFR and KNN achieve high performance on the training set, their test performance declines, indicating overfitting. This suggests that these models capture patterns specific to the training data but struggle to generalise to unseen data. In contrast, ANN exhibits the most consistent performance across both datasets, highlighting its robust generalisation ability. This suggests that ANN effectively learns underlying relationships in the data without overfitting, making it a more reliable model for photovoltaic energy prediction.

The observed differences in model performance may be attributed to the nonlinear nature of photovoltaic energy generation, which is influenced by multiple environmental factors [20], [21]. Although decision tree-based models like RFR effectively handle complex interactions, their sensitivity to

TABLE III
COMPARISON OF MODEL PERFORMANCE BASED ON NRMSE VALUES, WITH LINEAR REGRESSION SERVING AS THE BASELINE (100%), AND THE PERFORMANCE OF OTHER MODELS EXPRESSED AS A PERCENTAGE IMPROVEMENT OR DECLINE RELATIVE TO THIS BASELINE

Model	relative performance (in %)			
	S	S_1 [7]	S_2 [18]	S_3 [19]
LR	100	100	100	100
KNN	138.7			173.33
RFR	162.53		160.8	288.89
ANN	158.10	71.38	350.33	236.36

training data can lead to overfitting. On the other hand, the ANN model, with its ability to capture nonlinear dependencies, demonstrates a better balance between flexibility and generalisation.

In Figures 2–3, there is a noticeable increase in point density for lower efficiency values in all models. This can be attributed to the dataset's characteristics, where higher efficiency values are less frequent than lower ones. Power generation is typically most efficient during specific periods, particularly in summer and around midday, if weather conditions are favourable. This distribution reflects the natural variability of photovoltaic systems, which is strongly influenced by seasonal and temporal factors. The effect is even more pronounced in the training set, probably because of the smaller number of samples available. The uneven distribution of training samples in the dataset is an unfavorable phenomenon for machine learning. It could have impacted the training and final performance of the models considered.

Table III compares the findings of this study with related works. Graditi et al. [7] analysed a single photovoltaic installation under controlled conditions, where linear regression performed well due to simple variables. Jebli et al. [18] used historical weather and radiation data and found ANN to be the best performer, reinforcing the advantage of nonlinear models. Similarly, Chaaban and Alfadl [19] studied solar power plant data, with RFR and KNN outperforming linear regression, aligning closely with the findings of this research.

The presented metric provides valuable information on the dataset's characteristics and underscores the performance improvements of RFR and ANN models in this specific problem setting. Compared to other studies, this research's outcomes are satisfactory, as the observed trends across different models align well with the findings reported in the literature.

V. CONCLUSION

The aim of this study was to create machine learning models to predict energy production in photovoltaic infrastructures with different parameters, locations, and under the influence of variable weather conditions. It is an area of study of growing importance, given the increasing adoption of renewable energy as a primary energy source across a large number of sectors. Its significance is particularly evident for energy management systems, which depend on accurate energy production predictions to adjust distribution strategies accordingly to resource availability.

We have shown that a single data-driven predictive model can accurately predict the output of many different installations. Among the evaluated models, the ANN showed the most consistent performance across both training and test datasets, demonstrating strong generalisability. This indicates that ANN-based approaches are particularly suitable for solar energy forecasting. In contrast, the Random Forest and K-Nearest Neighbors models exhibited indications of overfitting, as evidenced by a notable difference in performance between training and test sets. However, the RFR remains a highly accurate model in general, showcasing strong predictive capabilities.

There is still room for improvement in models' performance that was not addressed in this study. For future work, it would be beneficial to incorporate data from installations that span a broader geographical area. This would enhance the generalisation capability of the models, enabling them to make accurate predictions across diverse contexts and conditions. In their current form, the models developed in this project are not well-suited for deployment in locations beyond Poland and nearby areas.

References

[1] R. Gross, M. Leach, and A. Bauen, "Progress in renewable energy," *Environment international*, vol. 29, no. 1, pp. 105–122, 2003. doi: https://doi.org/10.1016/S0160-4120(02)00130-7

[2] A. Olabi and M. A. Abdelkareem, "Renewable energy and climate change," *Renewable and Sustainable Energy Reviews*, vol. 158, p. 112111, 2022. doi: https://doi.org/10.1016/j.rser.2022.112111

[3] B. P. Roberts and C. Sandberg, "The role of energy storage in development of smart grids," *Proceedings of the IEEE*, vol. 99, no. 6, pp. 1139–1144, 2011. doi: 10.1109/JPROC.2011.2116752

[4] E. Serban, M. Ordonez, C. Pondiche, K. Feng, M. Anun, and P. Servati, "Power management control strategy in photovoltaic and energy storage for off-grid power systems," in *2016 IEEE 7th International Symposium on Power Electronics for Distributed Generation Systems (PEDG)*, 2016. doi: 10.1109/PEDG.2016.7527070 pp. 1–8.

[5] A. Dolara, S. Leva, and G. Manzolini, "Comparison of different physical models for pv power output prediction," *Solar energy*, vol. 119, pp. 83–99, 2015. doi: https://doi.org/10.1016/j.solener.2015.06.017

[6] A. Bączkiewicz and J. Wątróbski, "Selection of floating photovoltaic system considering strong sustainability paradigm using ssp-copras method," in *Proceedings of the 18th Conference on Computer Science and Intelligence Systems, M. Ganzha, L. Maciaszek, M. Paprzycki, D. Ślęzak (eds). ACSIS*, 09 2023. doi: 10.15439/2023F492 pp. 901–905.

[7] G. Graditi, S. Ferlito, and G. Adinolfi, "Comparison of photovoltaic plant power production prediction methods using a large measured dataset," *Renewable energy*, vol. 90, pp. 513–519, 2016. doi: https://doi.org/10.1016/j.renene.2016.01.027

[8] J. M. Barrera, A. Reina, A. Maté, and J. C. Trujillo, "Solar energy prediction model based on artificial neural networks and open data," *Sustainability*, vol. 12, no. 17, p. 6915, 2020. doi: https://doi.org/10.3390/su12176915

[9] B. Zazoum, "Solar photovoltaic power prediction using different machine learning methods," *Energy Reports*, vol. 8, pp. 19–25, 2022. doi: https://doi.org/10.1016/j.egyr.2021.11.183

[10] F. Koutenský, J. Pihrt, M. Čepek, V. Rybar, P. Šimánek, M. Kepka, K. Jedlička, and K. Charvat, "Combining local and global weather data to improve forecast accuracy for agriculture," in *Communication Papers of the 19th Conference on Computer Science and Intelligence Systems (FedCSIS), M. Bolanowski, M. Ganzha, L. Maciaszek, M. Paprzycki, D. Ślęzak (eds). ACSIS*, 11 2024. doi: 10.15439/2024F5990 pp. 77–82.

[11] M. Ceci, R. Corizzo, F. Fumarola, D. Malerba, and A. Rashkovska, "Predictive modeling of pv energy production: How to set up the learning task for a better prediction?" *IEEE Transactions on Industrial Informatics*, vol. 13, no. 3, pp. 956–966, 2017. doi: 10.1109/TII.2016.2604758

[12] Y. Ledmaoui, A. El Maghraoui, M. El Aroussi, R. Saadane, A. Chebak, and A. Chehri, "Forecasting solar energy production: A comparative study of machine learning algorithms," *Energy Reports*, vol. 10, pp. 1004–1012, 2023. doi: https://doi.org/10.1016/j.egyr.2023.07.042. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2352484723011228

[13] N. Dimitropoulos, N. Sofias, P. Kapsalis, Z. Mylona, V. Marinakis, N. Primo, and H. Doukas, "Forecasting of short-term pv production in energy communities through machine learning and deep learning algorithms," in *2021 12th International Conference on Information, Intelligence, Systems & Applications (IISA)*, 2021. doi: 10.1109/IISA52424.2021.9555544 pp. 1–6.

[14] M. Pikus and J. Wąs, "Predictive modeling of renewable energy purchase prices using deep learning based on polish power grid data for small hybrid pv microinstallations," *Energies*, vol. 17, no. 3, 2024. doi: 10.3390/en17030628. [Online]. Available: https://www.mdpi.com/1996-1073/17/3/628

[15] M. Pikus and J. Was, "Using deep neural network methods for forecasting energy productivity based on comparison of simulation and dnn results for central poland—swietokrzyskie voivodeship," *Energies*, vol. 16, no. 18, 2023. doi: 10.3390/en16186632. [Online]. Available: https://www.mdpi.com/1996-1073/16/18/6632

[16] X. Chen, Y. Li, B. Zhao, and R. Wang, "Are the optimum angles of photovoltaic systems so important?" *Renewable and Sustainable Energy Reviews*, vol. 124, p. 109791, 2020. doi: https://doi.org/10.1016/j.rser.2020.109791

[17] M. Z. Jacobson and V. Jadhav, "World estimates of pv optimal tilt angles and ratios of sunlight incident upon tilted and tracked pv panels relative to horizontal panels," *Solar energy*, vol. 169, pp. 55–66, 2018. doi: https://doi.org/10.1016/j.solener.2018.04.030

[18] 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, p. 120109, 2021. doi: https://doi.org/10.1016/j.energy.2021.120109

[19] A. K. Chaaban and N. Alfadl, "A comparative study of machine learning approaches for an accurate predictive modeling of solar energy generation," *Energy Reports*, vol. 12, pp. 1293–1302, 2024. doi: https://doi.org/10.1016/j.egyr.2024.07.010

[20] M. A. Hassan, N. Bailek, K. Bouchouicha, A. Ibrahim, B. Jamil, A. Kuriqi, S. C. Nwokolo, and E.-S. M. El-kenawy, "Evaluation of energy extraction of pv systems affected by environmental factors under real outdoor conditions," *Theoretical and Applied Climatology*, vol. 150, no. 1, pp. 715–729, 2022. doi: https://doi.org/10.1007/s00704-022-04166-6

[21] T. Hai, M. Aksoy, and K. Nishihara, "Optimized mppt model for different environmental conditions to improve efficacy of a photovoltaic system," *Soft Computing*, vol. 28, no. 3, pp. 2161–2179, 2024. doi: https://doi.org/10.1007/s00500-023-09195-5