

Machine Learning Technologies for a Solar Plant's System

Subjects: [Energy & Fuels](#)

Contributor: Ekaterina Engel , Nikita Engel

A solar plant system has complex nonlinear dynamics with uncertainties due to variation of the system parameters and insolation. Thereby, it is sophisticated to approximate these complex dynamics by conventional algorithms whereas Machine Learning (ML) methods yield the essential performance. ML models are key units in recent sensor systems for the solar plant's design, forecasting, maintenance, and control to provide best safety, reliability, robustness and performance, as compared to classical methods which are usually employed in hardware and software of the solar plants.

machine learning

neural networks

DL

PV

1. Introduction

Solar plant systems have complex nonlinear dynamics with uncertainties since the system's parameters and insolation fluctuate ^[1]. Thereby, it is complicated to approximate these complex dynamics with classical methods, while ML methods provide the required performance ^[2]. In modern sensor systems, ML methods are crucial units to increase the quality of big dataset processing for solar plant design, forecasting, maintenance, and control ^{[1][2]}. Within the EU COVID-19 strategic reply, the smart energy standards define a cloud platform specification for a distributed solar big data ecosystem that will provide the creation of effective ML technologies for smart solar energy ^[3]. The long-term contribution of solar energy is dependent on overcoming the remaining issues of grid integration, high costs, and low efficiency, mainly through the research and development of a smart solar plant system based on ML methods on account of traditional methods' ineffectiveness. Within breakthrough studies, ML technologies collected, analyzed, and converted a huge number of sensory datasets into ML knowledge. These big data sets are collected by supervisory control and data acquisition (SCADA) systems ^[4]. The SCADA system is able to integrate the sensor system and ML technologies into an ML sensor system based on software that implements ML sensor models and integrates with SCADA through API. Further, the application of ML technologies for the digital transformation of solar plant systems has a massive potential to increase their stability, reliability, dynamic response, cost-effectiveness, and other essential advancements, easing their integration into electric grids.

The contribution of this research is threefold. First, researchers reviewed more than 100 research papers devoted to state-of-the-art ML technologies of solar plant systems, most of the articles were published in the last five years. Second, researchers reviewed resources where researchers can find open datasets, source code, and ML framework and simulation environments to create ML technologies for a solar plant system. Third, in contrast with

other review articles, this research proposes a simple but effective pipeline scheme for an ML sensor solar plant system's implementation and outlines its future digital transformation into a smart solar plant based on integrated, cutting-edge technologies; estimates the impact of the ML technologies based on the proposed scheme on a solar plant value chain; sets benchmarks for performance comparison of the reviewed ML models for a solar plant's system based on the comparative studies' results summaries; and briefly summarizes researchers' self-adaptive models for sizing, forecasting, maintenance, and control of a solar plant based on a modified fuzzy neural net (MFNN) that is automatically created with regard to tasks' complexities and overfitting problems [5][6][7][8].

2. Machine Learning Technologies for a Solar Plant's System

Real-life solar plant systems have complex, nonlinear dynamics due to variations in system parameters and insolation. Thus, ML methods have been proposed to approximate this complex dynamic. The recent studies [1][2][5][6][7][8][9][10][11][12][13] prove that ML technologies for a solar plant's design, forecasting, maintenance, and control increase the effectiveness and reliability of the solar plant as compared to conventional methods. In smart sensor systems of solar plants, ML methods are crucial units to increase the quality of datasets processing the solar plant's design, forecasting, maintenance, and control. SCADA is a control system architecture that uses sensors, programmable logic, and discrete PID controllers to control the processes of a solar plant system. The solar plant's system includes advanced sensors. Big data from SCADA are collected 24/7. Combined with weather big data, this enables the creation of ML technologies to solve complex tasks of a solar plant's design, forecasting, maintenance, and control.

2.1. ML Sensor System of a Solar Plant

Smart models based on ML technologies have the advantage of parallel computation through modern graphical processing units, which significantly decreases the time cost in SCADA datasets processing for solar plant design, forecasting, maintenance, and control [12].

The reliability, accuracy, and other demanded quality parameters must be composed as the performance of an ML model. This model must be created effectively with high-quality datasets to have optimal performance [14]. **Figure 1** shows the basic life cycle of an ML sensor model. Smart model creation has two phases: data preparation (DP phase) and model creation (MC phase). They should be elaborated by the Cross-Industry Standard Process for Data Mining cycle (CRISP-DM) [15] and Open Neural Network Exchange (ONNX) format [16]. The CRISP-DM cycle [15] provides a pipeline for the implementation of smart models in real-time scenarios.

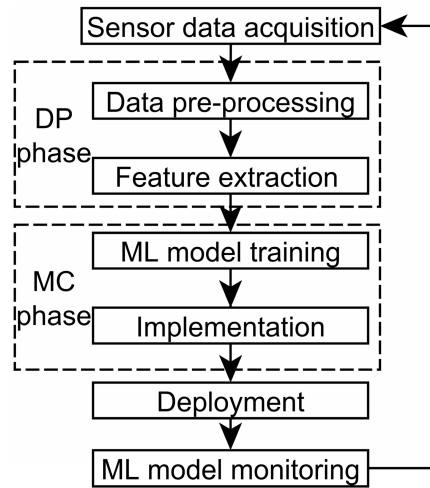


Figure 1. A basic life cycle of an ML sensor model.

The sensor data of a solar plant are compiled into raw SCADA datasets. Then, these datasets are preprocessed (**Figure 1**) in a simple way (standardization or encoding). Data preparation methods include dimensionality reduction (principal component analysis (PCA)), sampling (subsampling, oversampling), transformation, encoding, feature extraction, and selection [14].

Feature extraction is a crucial step in a smart sensor system's creation because it provides knowledge for ML model creation [14]. The DM methods generate features. The most relevant data are further separated into train, validating, and test datasets (**Figure 1**). An ML model to solve either classification or regression tasks is trained based on a train dataset. When a smart model provides the demanded performance, its weights are frozen. The ML frameworks, which researchers review in Section 2.3, provide an automatic MC phase, including validating (**Figure 1**). The trained ML model is deployed. If a monitored ML model does not provide optimal performance, then it is retrained based on updated datasets.

2.2. ML Methods for Smart Sensor Creation

An ML sensor model can be developed based on neural network (NN) or non-NN algorithms [14]. The last ones include PCA, Random Forest (RF), support vector machine (SVM), and Decision Tree (DT). In contrast with non-NN methods, NN architectures can include various neurons which are specified by ONNX [17], highly effective learning, and extracting features. A deep neural learning/network (DL/DNN), such as a recurrent neural network (RNN), convolutional neural network (CNN), and transformers, is part of the ML methods with feature learning that use multiple layers, complex connectivity architectures, and different transfer operators to automatically mine meta features from the input. NNs, such as artificial neural networks (ANNs), radial basis function neural networks (RBF-NNs), generative adversarial networks (GANs), RNNs, and CNNs have recently made major progress in practical applications of solar energy [1].

Figure 2 shows two NN methods' classes and the ML method groups according to the task they solved for a solar plant system [2].

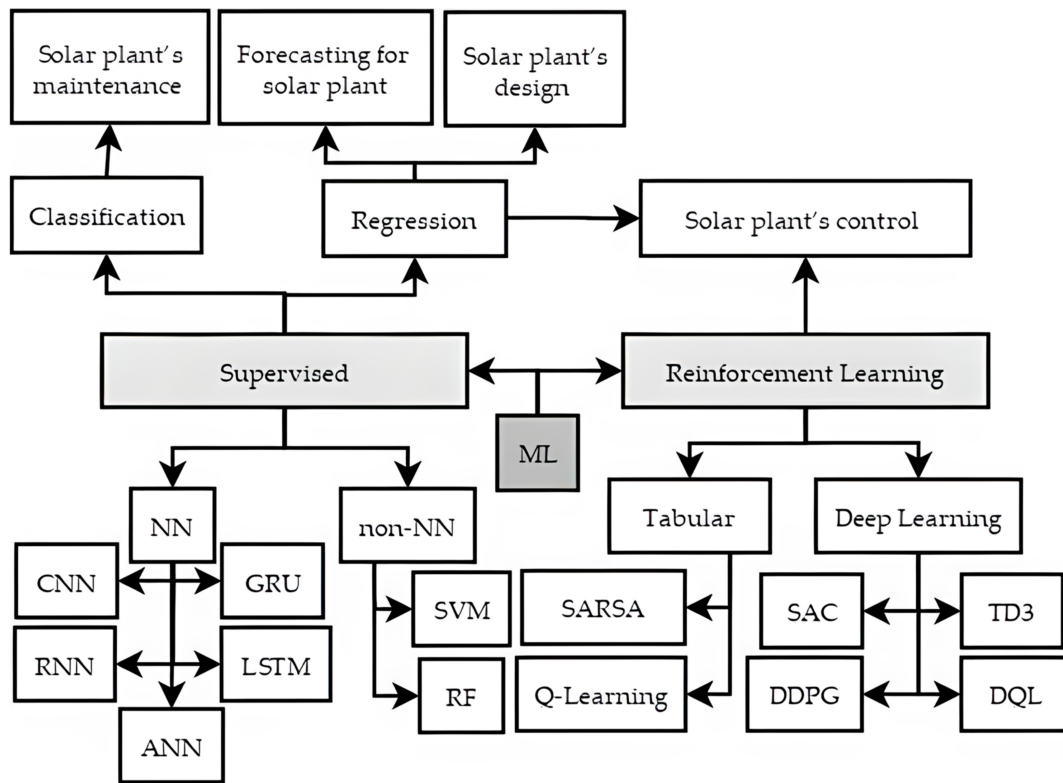


Figure 2. Classification of tasks that are solved based on ML methods.

The ensemble's types are bagging, boosting, and stacking/blending [18][19]. **Table 1** presents the comparison of ensemble techniques [18]. There are constant and dynamic weighting ensemble approaches. In recent studies, the most used ensemble methods are RF, Extreme Gradient Boosting (XGBoost), Extreme Learning Machine (ELM), etc.

Model training methods that optimize performance include quasi-Newton, stochastic gradient descent (SGD), evolutionary computation, genetic programming, etc. [15]. The creation of the ML model is the most complex and important task which includes the creation of an optimal ML model's architecture and requires a multidimensional global optimization (GO).

The bias and variance estimate the effectiveness of a model. The improvement of a model's bias always makes gains at the expense of variance and vice versa. The performance of ML models highly correlates with the representativeness of a dataset. A lot of techniques provide a model's evaluation, including cross-validation, kfold, holdout with a different performance including accuracy (ACC), mean squared error (MSE), precision, receiver operating characteristics (ROC), recall, Matthew's correlation coefficient (MCC), F1, area under the curve (AUC), mean absolute error (MAE), and root-MSE (RMSE). The relative errors, such as normalized RMSE (nRMSE), normalized MAE (nMAE), etc., facilitate the comparison between models that are tuned based on datasets with different scales.

With the goal to develop intelligent models for sizing, forecasting, and control of a solar plant system and to make an RNN more adaptive with regard to a task's complexity and overfitting problem, researchers developed an MFNN

[5][6][7][8]. The MFNN includes RNNs with fuzzy units and/or a convolutional block to process images. An RNN approximates a membership function in contrast to an Adaptive Network-Based Fuzzy Inference System (ANFIS). Researchers combined the modified multidimensional quantum-behaved particle swarm optimization (PSO) with the Levenberg–Marquardt algorithm (MD QPSO) and developed a hierarchical encoder of the particle's dimension component [5][6][7][8] to automatically create an optimal architecture of an MFNN and improve the convergence.

Table 1. Comparison of ensemble techniques.

Name of Method	Advantages	Disadvantages
Bagging	Tends to reduce variance more than bias	Does not work well with relatively simple models
Boosting	Reduces bias and variance	Sensory to noise and outliers in data. Susceptible of overfitting
Stacking/blending	Provides the optimal combination of base learners, reduces variance, and bias [18]	In the case of huge datasets, the computational time increases sufficiently as each classifier is working independently on the huge dataset.

Researchers implemented an MFNN and its life cycle, which includes automatic creation and self-adaptation as an intelligent framework based on the researchers' software [20]. This intelligent framework provides the automatic creation of the optimum architecture of an MFNN with regard to a task's complexity.

All the above-mentioned ML methods and algorithms were implemented as software by an ML Framework, which represents a tool to create a smart sensor system.

2.3. ML Frameworks

ML frameworks implement many ML methods [15]. **Table 2** shows the comparison of popular ML frameworks.

Big data ecosystems, namely Apache Flink, Apache Spark, and Cloudera Oryx 2, include built-in ML libraries for large-scale DM. These ML libraries evolve presently, but the potency of the entire ecosystem is significant.

Google, Facebook, and Microsoft developed most of the DL frameworks that support ONNX, namely PyTorch, TensorFlow, Caffe2, Microsoft CNTK, and MXNet.

Chainer, Theano, Deeplearning4, and H2O are also appropriate DL libraries and frameworks for smart sensor system creation.

The high-level DL wrapper libraries such as Keras, TensorLayer, and Gluon are developed on top of the DL frameworks. They provide a simpler but more computationally expensive way for smart sensor system creation.

The ML frameworks provide an automatic MC phase of an ML model, including validating (**Figure 1**). An ML sensor system can be implemented as software based on an ML framework that supports ONNX. Such implementation will provide flexibility and all an ML framework's advantages for a developed ML sensor system.

Table 2. Comparison of ML frameworks.

Name	Advantages	Disadvantages
TensorFlow	Open-source, API-oriented, cross-platform, ML/DL toolbox implements many ML methods.	The code is not flexible. Lack of documentation. Toolbox oriented for academic usage.
Microsoft CNTK	Open-source, fast-evolving, supports ONNX, supported by Microsoft.	Limited facilities for mobile platforms.
Caffe2	Cross-platform, supports mobile platforms, supports ONNX.	Complex as compared to PyTorch. Without dynamic graph computation.
PyTorch	Dynamic computational graph. Automatic implementation of ML models. Supports ONNX.	Absence of monitoring and visualization tools like a tensor board
Keras	Open-source, provides backend tools from Google and Microsoft. Detailed specification. API for DL. Quick implementation of DL models (e.g., TensorFlow, Theano, CNTK).	Modularity and simplicity make gains at the expense of flexibility. Limited facilities to create a new architecture.

2.4. Open Resources for ML Research in a Solar Plant System

The open solar energy data sources, including big data, provide the development of cutting-edge ML technologies in solar energy.

The GitHub repositories [\[21\]\[22\]](#) are implementations of maximum power point tracking (MPPT) systems [\[21\]](#) and management of cities' demand/load [\[22\]](#) based on an open-source Gym toolkit [\[23\]](#). An open-source tool pymgrid [\[24\]](#) provides the creation and simulation of various microgrids. Octave [\[25\]](#) and Scilab [\[26\]](#) are open sources that are compatible with MATLAB.

Table 3 presents a brief description of the open datasets to implement and validate ML solar plant systems.

Table 3. The open datasets to implement and validate ML solar plant systems.

Open Dataset	Data Source Location	Description
Duke California Solar Array Dataset [27]	-	Over 400 km ² of imagery and 16,000 hand-labeled solar arrays
SOLETE [28]	City: Roskilde, Denmark. Latitude and longitude:	Meteorological and active power 15 months dataset from PV array

Open Dataset	Data Source Location	Description
	55.6867, 12.0985	
Desert Knowledge Australia Center Dataset [29]	-	Data of solar technologies spanning multiple types, ages, models, and configurations
Girasol [30]	Albuquerque, USA	A meteorological (10 min sampling interval), insolation (a sampling rate ranging from 4 to 6 samples per second), and images (sampling interval of the cameras is 15 s) 242 days (of 3 years) dataset
ESOLMET-IER Dataset [31]	Institute of Renewable Energies UNAM, station "ESOLMET-IER"	Solar metric and meteorological dataset
The National Solar Radiation Data Base (NSRDB) [32]	The USA and neighboring countries	Solar insolation and meteorological 23 years dataset
Photovoltaic Thermal Images Dataset [33]	66 MW PV plant in Tomboruk	Thermal images of PV arrays with the presence of one or more anomaly cells and their respective masks
Pecan Street Dataset [34]	-	1300 customer loads one-year dataset

References

1. Kurukuru, V.S.B.; Haque, A.; Khan, M.A.; Sahoo, S.; Malik, A.; Blaabjerg, F. A Review on Artificial Intelligence Applications for Grid-Connected Solar Photovoltaic Systems. *Energies* 2021, 14, 4690.
2. Forootan, M.M.; Larki, I.; Zahedi, R.; Ahmadi, A. Machine Learning and Deep Learning in Energy Systems: A Review. *Sustainability* 2022, 14, 4832.
3. EITCI Institute. Smart Energy Standards Group. Available online: <https://eitci.org/sesg> (accessed on 10 September 2022).
4. Massaro, A.; Starace, G. Advanced and Complex Energy Systems Monitoring and Control: A Review on Available Technologies and Their Application Criteria. *Sensors* 2022, 22, 4929.
5. Engel, E. A Photovoltaic Applications on the Basis of Modified Fuzzy Neural Net. In *Solar Irradiance: Types and Applications*; Welsh, D.M., Ed.; Nova Science Publishers: Hauppauge, NY, USA, 2020; pp. 7–87.
6. Engel, E.; Engel, N. Photovoltaic System Control Model on the Basis of a Modified Fuzzy Neural Net. *Stud. Comput. Intell.* 2020, 856, 45–52.

7. Engel, E.A.; Engel, N.E.; Degtyarev, A.S.; Kosenko, V.I.; Savelyeva, M.V. Power forecasting for a photovoltaic system based on the multi-agent adaptive fuzzy neuronet. *IOP Conf. Ser. Mater. Sci. Eng.* 2018, 450, 072012.
8. Engel, E.; Engel, N. Automatic adaptation of a Solar Plant Intelligent Control System. *J. Phys. Conf. Ser.* 2021, 1889, 022029.
9. Engel, E.A.; Engel, N.E. Maximum photovoltaic array power point tracking algorithm based on modified particle swarm optimization under non-uniform irradiances. *IOP Conf. Ser. Mater. Sci. Eng.* 2020, 734, 012112.
10. Gaviria, J.F.; Narváez, G.; Guillen, C.; Giraldo, L.F.; Bressan, M. Machine learning in photovoltaic systems: A review. *Renew. Energy* 2022, 196, 298–318.
11. Youssef, A.; El-Telbany, M.; Zekry, A. The role of artificial intelligence in photo-voltaic systems design and control: A review. *Renew. Sustain. Energy Rev.* 2017, 78, 72–79.
12. De Oliveira, A.K.V.; Aghaei, M.; Rüther, R. Automatic Inspection of Photovoltaic Power Plants Using Aerial Infrared Thermography: A Review. *Energies* 2022, 15, 2055.
13. Berghout, T.; Benbouzid, M.; Bentrchia, T.; Ma, X.; Djurović, S.; Mouss, L.-H. Machine Learning-Based Condition Monitoring for PV Systems: State of the Art and Future Prospects. *Energies* 2021, 14, 6316.
14. Ha, N.; Xu, K.; Ren, G.; Mitchell, A.; Ou, J.Z. Machine Learning-Enabled Smart Sensor Systems. *Adv. Intell. Syst.* 2020, 2, 2000063.
15. Nguyen, G.; Dlugolinsky, S.; Bobák, M.; Tran, V.; García, L.; Heredia, I.; Malík, P.; Hluchý, L. Machine Learning and Deep Learning frameworks and libraries for large-scale data mining: A survey. *Artif. Intell. Rev.* 2019, 52, 77–124.
16. Open Neural Network Exchange. Available online: <https://onnx.ai/> (accessed on 11 September 2022).
17. Khan, A.; Bhatnagar, R.; Masrani, V.; Lobo, V.B. A Comparative Study on Solar Power Forecasting using Ensemble Learning. In *Proceedings of the 4th International E3S Web of Conferences*, Tirunelveli, India, 15–17 June 2020; Volume 309, p. 01163.
18. Brown, G.; Ensemble Sammut, C.; Webb, G.I. *Encyclopedia of Machine Learning and Data Mining*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 393–402.
19. Shahhosseini, M.; Hu, G.; Pham, H. Optimizing ensemble weights and hyperparameters of machine learning models for regression problems. *Mach. Learn. Appl.* 2022, 7, 100251.
20. The Module of the Modified Fuzzy Neural Net. M.: Federal Service for Intellectual Property (Rospatent), Certificate about State Registration of Software NO. 2021681065. 2021.

21. Avila, L.; de Paula, M.; Trimboli, M.; Carlucho, I. Loavila/Mpvt-Gym. Available online: <https://github.com/loavila/mpvt-gym> (accessed on 11 September 2022).
22. Official Reinforcement Learning Environment for Demand Response and Load Shaping. Available online: <https://github.com/intelligent-environments-lab/CityLearn> (accessed on 5 October 2022).
23. Open AI. Gym A Toolkit for Developing and Comparing Reinforcement Learning Algorithms. Available online: <https://gym.openai.com> (accessed on 13 September 2022).
24. Pymgrid Library. Available online: <https://github.com/Total-RD/pymgrid> (accessed on 15 October 2022).
25. GNU Octave 7.2.0. Available online: <https://octave.org/download> (accessed on 15 October 2022).
26. Scilab 6.1.1. Available online: <https://www.scilab.org/download/scilab-6.1.1> (accessed on 15 October 2022).
27. Bradbury, K.; Saboo, R.; Malof, J.; Johnson, T.; Devarajan, A.; Zhang, W.; Collins, L.; Newell, R.; Streltsov, A.; Hu, W. Distributed Solar Photovoltaic Array Location and Extent Data Set for Remote Sensing Object Identification. *Sci. Data* 2016, 3, 160106.
28. The SOLETE Dataset. Available online: https://data.dtu.dk/articles/dataset/The_SOLETE_dataset/17040767 (accessed on 26 October 2022).
29. DKASC, Alice Springs. DKA Solar Centre. Available online: <http://dkasolarcentre.com.au/locations/alice-springs> (accessed on 12 September 2022).
30. Terrén-Serrano, G.; Bashir, A.; Estrada, T.; Martínez-Ramón, M. Girasol, a Sky Imaging and Global Solar Irradiance Dataset. *Data Brief* 2021, 35, 106914.
31. ESOLMET-IER. Instituto de Energías Renovables. Available online: http://esolmet.ier.unam.mx/Tipos_consulta.php (accessed on 12 September 2022).
32. Sengupta, M.; Xie, Y.; Lopez, A.; Habte, A.; Maclaurin, G.; James, S. The national solar radiation data base (NSRDB). *Renew. Sustain. Energy Rev.* 2018, 89, 51–60.
33. Pierdicca, R.; Paolanti, M.; Felicetti, A.; Piccinini, F.; Zingaretti, P. Photovoltaic Thermal Images Dataset. Available online: <https://vrai.dii.univpm.it/content/photovoltaic-thermal-images-dataset> (accessed on 12 September 2022).
34. Miranda, E.; Fierro, J.F.G.; Gabriel, N.A.; Giraldo, L.F.; Bressan, M. Prediction of site-specific solar diffuse horizontal irradiance from two input variables in Colombia. *Heliyon* 2021, 7, 08602.

Retrieved from <https://encyclopedia.pub/entry/history/show/85736>