A Review of Data Mining and Solar Power Prediction

Mehmet Yesilbudak[1], Medine Colak[2], Ramazan Bayindir[2]

[1]Nevsehir Haci Bektas Veli University, Faculty of Engineering and Architecture, Department of Electrical and Electronics Engineering, 50300, Nevsehir, Turkey.
[3]Gazi University, Faculty of Technology, Department of Electrical and Electronics Engineering, 06500, Ankara, Turkey.
myesilbudak@nevsehir.edu.tr, medinecolak@gmail.com, bayindir@gazi.edu.tr

***Abstract*—Solar energy is one of the clean and renewable energy sources that are mostly available in the world. As a result of this situation, there are many research studies done on the solar energy in order to get the maximum solar radiation during the day time, to estimate the solar power generation and to increase the efficiency of solar systems. In this paper, especially, a review of data mining methods employed for solar power prediction in the literature is introduced briefly. Input data, recording intervals, the number of training and test datasets of each study are also considered in the review process. It is shown that artificial neural networks are the most preferred methods in order to predict solar power generation.**

Keywords*—Data mining, solar power prediction, methods, review

I. INTRODUCTION

Nowadays, fossil energy sources are quickly running out in the world. Due to this reason, we are seeking new energy sources for the energy needs of next generations. In addition, we should make all energy sources to be more sustainable in an environmental, social and economic manner [1, 2]. The Sun serves as the main source of energy in the world and we can generate the electrical energy from solar energy systems efficiently. However, solar power prediction is so important in terms of installation and cost analyses and some meteorological data processed with data mining methods are required for more accurate prediction results. Among meteorological data, average temperature, average relative humidity, average wind velocity, total sunshine duration, total global solar radiation and total solar electricity production are mostly used in the literature [1, 3, 4].

In case of examining the knowledge discovery process in databases shown in Figure 1, it contains the stages of data cleaning, data integration, data selection, data transformation, data mining, pattern evaluation and knowledge presentation [1-4]. In data cleaning, integration, selection and transformation stages, inconsistent data are removed, multiple data sources are combined, the data relevant to the analysis task are retrieved and the convenient data form is performed, respectively. Data mining is based on searching the hidden patterns and it summarizes the meaningful and useful information [2-3]. The interesting patterns are identified and the knowledge mined is presented to users with visualization techniques in the stages of pattern evaluation and knowledge presentation, respectively. There are many data mining techniques in the literature, which are categorized as characterization and discrimination, evolution analysis, outlier analysis, association analysis, cluster analysis and classification [3, 5]. According to the [6], artificial neural networks, support vector machines, k-nearest neighbor algorithm, Naïve Bayes algorithm, regression analysis and decision trees are employed for the classification purposes. In addition to these, heuristic methods, grid-based methods, density-based methods, partitioning methods and hierarchical methods are utilized for the clustering purposes. Apriori algorithm is also considered for the association analysis, typically.

This study concentrates on reviewing the data mining methods along with their input data, recording intervals and the number of training and test datasets used for solar power prediction in the literature. As a result, many deficiencies in the literature are uncovered in an effective manner.

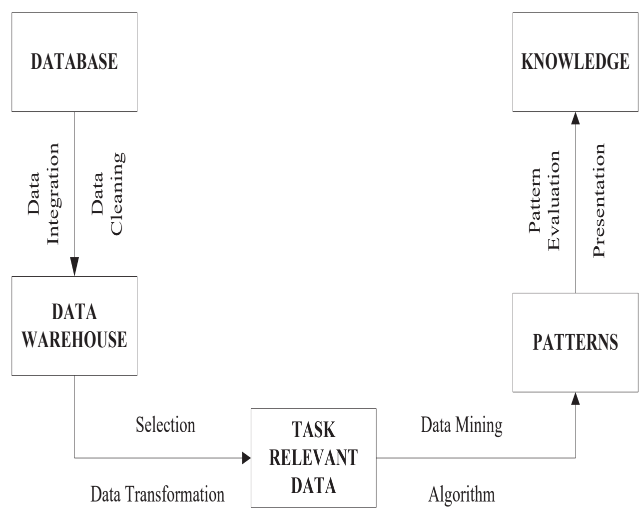

Fig. 1. Knowledge discovery process in databases [2-4]

II. DATA MINING METHODS USED FOR SOLAR POWER PREDICTION

In the literature, many data mining methods have been implemented for different purposes such as the estimation of

978-1-5090-3388-1/16/$31.00 ©2016 IEEE

daily global solar radiation, the prediction of solar radiation intensity, the construction of hybrid intelligent solar power predictor, etc. [7-13]. Particularly, as a result of the reviewing process conducted in this study, a number of data mining methods used for solar power prediction in the literature are summarized in Table I.

As shown in Table I, adaptive neuro-fuzzy inference system, multilayer perceptron, radial basis functions, recurrent neural networks, back-propagation neural networks, wavelet-neural networks, support vector machines, relevance vector machine, extreme learning machine, self-organizing maps, fuzzy theory, genetic algorithm, particle swarm optimization, k-nearest neighbor classifier, learning vector quantization, nonlinear autoregressive with exogenous inputs, kernel ridge regression, least absolute shrinkage and selection operator, regression trees, bagging trees, boosted trees and principal component analysis have been used in the literature. It is clear that artificial neural networks are the most preferred method in the literature. In addition, air temperature, latitude, longitude, solar angle, solar radiation, relative humidity, wind speed and day length are among the commonly used input parameters in the literature. It should be noted that each study has its own recording interval and dataset number in the literature.

Table I. Different methods implemented for solar power prediction

Ref.	Input data	Recording intervals	Total dataset in		Methods implemented in the studies
			Training	Test	
[14]	Rainfall, humidity, maximum and minimum air temperature, elevation, longitude, latitude	1 month	N/A	N/A	Artificial neural networks
[15]	Relevant electrical quantities, local temperature, solar radiation intensity	10 minutes	27278	9275	Flexible neural tree
[16]	Solar energy	N/A	N/A	N/A	Knowledge-based neural network, multilayer perceptron
[17]	Air pressure, wind speed, humidity, temperature, DNI, PG	N/A	N/A	N/A	Principal component analysis
[18]	Height above sea level, hours, longitude, sunshine, latitude	N/A	N/A	N/A	Radial basis functions, neural network fitting tool
[19]	Solar power generation	5 minutes	N/A	N/A	Artificial neural networks
[20]	Solar hour angle and sun declination, solar azimuth angle, solar elevation angle	N/A	N/A	N/A	Recurrent neural networks
[21]	two different ground global horizontal radiation (a public weather station and a local station)	N/A	11480	5740	Support vector regression, extreme learning machine, k-nearest neighbor classifier, persistence model
[22]	Temperature, solar irradiance, cloud cover	N/A	N/A	N/A	Artificial neural networks
[23]	Sun position, humidity, cloud, air temperature, solar irradiance	N/A	N/A	N/A	Physical model, artificial neural networks
[24]	Satellite images, numerical weather data, ground measurement stations	10 minutes	9048	4524	Artificial neural networks
[25]	Meteorological data	N/A	N/A	N/A	Gaussian process, relevance vector machine, kernel ridge regression, support vector machine, extreme learning machine, boosted trees, bagging trees, regression trees, least absolute shrinkage and selection operator, regularized least squares linear regression
[26]	Solar irradiance, temperature, humidity, wind speed	5 to 60 min	N/A	N/A	Neural networks and support vector regression
[27]	Solar radiation, solar energy	N/A	N/A	N/A	Artificial neural networks
[28]	Daily total solar radiation	N/A	N/A	N/A	Wavelet-neural networks
[29]	Ultraviolet radiation index, wind speed, wind direction, precipitation probability, temperature	N/A	N/A	N/A	Support vector regression, self-organizing maps, learning vector quantization
[30]	Climatic data	N/A	2963	3052	Support vector regression
[31]	Meteorological data	N/A	N/A	N/A	Extreme learning machine, grouping genetic algorithm

[32]	Average temperature, clearness index, pressure, humidity, longitude, latitude, altitude	N/A	N/A	N/A	Artificial neural networks
[33]	Solar radiation on the horizontal, ambient air temperature, etc.	N/A	N/A	N/A	Adaptive neuro-fuzzy inference system
[34]	Altitude, latitude, average rainfall, number of rainy days, day length, top of atmosphere radiation,	N/A	17	28	Artificial neural networks
[35]	Weather data	N/A	N/A	N/A	Artificial neural networks, fuzzy theory
[36]	Solar radiation, relative humidity, mean temperature, mean sunshine duration, altitude, longitude, latitude	N/A	900	450	Artificial neural networks
[37]	Weather data	N/A	N/A	N/A	Multilayer perceptron
[38]	Aerosol index data	N/A	N/A	N/A	Back-propagation artificial neural networks
[39]	Temperature, solar radiation, power generation	N/A	N/A	N/A	Back propagation artificial neural networks, particle swarm optimization
[40]	Wind speed, velocity of the molten salt, inlet temperature, solar flux	N/A	N/A	N/A	Back-propagation artificial neural networks
[41]	Meteorological data	N/A	N/A	N/A	Recurrent neural networks
[42]	Humidity, solar radiation, temperature	N/A	N/A	N/A	Artificial neural networks
[43]	Air pressure, air humidity, wind speed, air temperature, irradiance, day	N/A	N/A	N/A	Artificial neural networks
[44]	Cloud coverage, wind speed, atmospheric pressure, humidity, temperature, irradiance, day	N/A	N/A	N/A	Artificial neural networks
[45]	Temperature, humidity, wind speed	10 minutes	N/A	N/A	Wavelet recurrent neural networks
[46]	Humidity, precipitation, sky cover, wind speed, dew point, temperature, day	N/A	N/A	N/A	Support vector machines
[47]	Incident angles, air mass ratio, solar azimuth, solar altitude, solar declination angle, hour angle	1 minute	N/A	N/A	Artificial neural networks, modified Meinel model, modified Ashrae model
[48]	Solar radiation, temperature, humidity, sun hour	4 months	528	48	Recurrent simple addressing structure for cerebella model articulation controller with general basis function
[49]	Sunshine duration, relative humidity, daily mean air temperature, day of the year	N/A	1461	366	Artificial neural networks
[50]	Global horizontal irradiation, sunshine duration, wind speed, relative humidity, air temperature	N/A	2479	531	Nonlinear autoregressive with exogenous inputs, adaptive neuro-fuzzy inference system, artificial neural networks,
[51]	Atmospheric pressure, cloud coverage, elevation, humidity, precipitation, wind speed, sunshine duration, average temperature	N/A	67	14	Artificial neural networks
[52]	Mean sunshine hours, mean wind speed, mean relative humidity, mean vapor pressure, mean temperature, mean pressure, month, location	N/A	N/A	N/A	Nonlinear autoregressive with exogenous inputs

III. Discussion and Conclusion

In this study, data mining methods used for solar power prediction in the literature are reviewed briefly. Especially, many comparisons in terms of the models employed, the input data utilized, and the recording interval and the total dataset considered are made efficiently. The deficiencies and the key points determined for the current literature are given below:

❖ Data mining is a good way used for improving the solar power prediction.

❖ Researchers use a number of different input parameters in their own studies.

❖ The information about the recording intervals of input parameters and the number of training and test datasets is not enough in most of the reviewed studies.

❖ Seasonal performance of the employed models should be considered for better solar power predictions.

❖ Other classification, clustering and association techniques utilized in data mining should be focused in addition to artificial neural networks for benchmark tests.

In future studies, the mentioned deficiencies should be considered and new hybrid methods should be used for addressing the different studies on solar power prediction.

References

[1] W.A.M. Ghoneim, A.A. Helal and M.G.A. Wahab, "Renewable energy resources and recovery opportunities in wastewater treatment plants", 3rd International Conference on Renewable Energies for Developing Countries, pp. 1-8, 2016.

[2] S. Agarwal, "Data mining concepts and techniques", International Conference on Machine Intelligence Research and Advancement, pp. 203-207, 2013.

[3] B.M. Ramageri, "Data mining techniques and applications", Indian Journal of Computer Science and Engineering, vol. 1, no.4, pp. 301-305, 2010.

[4] H. Karacan and M. Yesilbudak, "User-centered interactive data mining: A literature review", International Journal of Informatics Technologies, vol. 3, no. 1, pp. 17-21, 2010.

[5] E.B. Mohamed, H. Ltifi and M.B. Ayed, "Using visualization techniques in knowledge discovery process for decision making", 13th International Conference on Hybrid Intelligent Systems, pp. 93-98, 2013.

[6] I. Colak, S. Sagiroglu and M. Yesilbudak, "Data mining and wind power prediction: A literature review", Renewable Energy, vol. 46, pp. 241-247, October 2012.

[7] N. Kumar, S.P. Sharma, U.K. Sinha and Y. Nayak, "Prediction of solar energy based on intelligent and modeling", International Journal of Renewable Energy Research, vol. 6, no. 1, pp. 183-188, 2016.

[8] M. Khademi, M. Moadel and A. Khosravi, "Power prediction and techno-economic analysis of a solar PV power plant by MLP-ABC and COMFAR III, considering cloudy weather conditions", International Journal of Chemical Engineering, vol. 6, pp.1-8, 2016.

[9] S.S. Priya and M.H. Iqbal, "Solar radiation prediction using artificial neural network", International Journal of Computer Applications, vol. 116, pp. 28-31, 2015.

[10] H.H. Abbas and A. Khashman, "Evaluation of geographical information system and intelligent prediction of solar energy", International Journal of Control Systems and Robotics, vol. 1, pp. 106-112, 2016.

[11] G. Zhang, X. Wang and Z. Du, "Research on the prediction of solar energy generation based on measured environmental data international journal of u- and e- service", Science and Technology, vol. 8, pp. 385-402, 2015.

[12] J. Zeng and W. Qiao, "Short-Term solar power prediction using an RBF neural network", IEEE Power and Energy Society General Meeting, pp.1-8, 2011.

[13] M.R. Hossain, A.M.T. Oo and A.B.M.S. Ali, "Hybrid prediction method of solar power using different computational intelligence algorithms", 22nd Australasian Universities Power Engineering Conference, pp.1-8, 2012.

[14] A. Prastawa and R. Dalimi, "Hybrid prediction method of solar power using different computational intelligence algorithms", International Conference on QiR (Quality in Research), pp. 166-169, 2013.

[15] S. Basterrech, L. Prokop, T. Burianek and S. Misak, "Optimal design of neural tree for solar power prediction", 15th International Scientific Conference on Electric Power Engineering, pp.273-278, 2014.

[16] Q. Liu and Q.J. Zhang, "Accuracy improvement of energy prediction for solar-energy-powered embedded systems", IEEE Transactions on Very Large Scale Integration Systems, vol. 24, no. 6, pp. 2062-31, June 2016.

[17] J. Xiu, C. Zhu and Z. Yang, "Prediction of solar power generation based on the principal components analysis and the BP neural network", IEEE 3rd International Conference on Cloud Computing and Intelligence Systems, pp. 366-374, 2014.

[18] A.K. Yadav and H. Malik, "Comparison of different artificial neural network techniques in prediction of solar radiation for power generation using different combinations of meterological variables", IEEE International Conference on Power Electronics, Drives and Energy Systems, pp.1-5, 2014.

[19] R.M. Ehsan, S.P. Simon and P.R. Venkateswaran, "Day-Ahead prediction of solar power output for grid connected solar photovoltaic installations using artificial neural networks", IEEE 2nd International Conference on Emerging Electronics, pp.1-4, 2014.

[20] S. Cao, "Total daily solar ırradiance prediction using recurrent neural networks with determinants", Asia-Pacific Power and Energy Engineering Conference, pp. 1-4, 2010.

[21] S. Ferrari, M. Lazzaroni, V. Piuri, A. Salman, L. Cristaldi and M. Faifer, "Computational intelligence models for solar radiation prediction", IEEE International Instrumentation and Measurement Technology Conference, pp. 757-762, 2013.

[22] S. Sperati, S. Alessandrini and L. D. Monache, "An application of the ECMWF Ensemble Prediction System for short-term solar power forecasting", Solar Energy, vol. 133, pp. 437-450, August 2016.

[23] Y. Huang, J. Lu, C. Liu, X. Xu, W. Wang and X. Zhou, "Comparative study of power forecasting methods for PV stations", International Conference on Power System Technology, pp.1-6, 2010.

[24] S. Ferrari, M. Lazzaroni, V. Piuri, A. Salman, L. Cristaldi, M. Rossi and T. Poli, "Illuminance prediction through extreme learning machines", IEEE Workshop on Environmental Energy and Structural Monitoring Systems, pp. 97-103, 2012.

[25] S. Salcedo-Sanz, C. Casanova-Mateo, J. Muñoz-Marí and G. Camps-Valls, "Prediction of daily global solar irradiation using temporal Gaussian processes", IEEE Geoscience and Remote Sensing Letters, vol. 11, no. 11, pp. 1936-1940, November 2014.

[26] M. Rana, I. Koprinska and V.G. Agelidis, "Univariate and multivariate methods for very short-term solar photovoltaic power forecasting", Energy Conversion and Management, vol. 121, pp. 380-390, August 2016.

[27] N. Zhang and P.K. Behera, "Solar radiation prediction based on recurrent neural networks trained by Levenberg-Marquardt backpropagation learning algorithm", IEEE PES Innovative Smart Grid Technologies, pp.1-7, 2012.

[28] H. Zhou, W. Sun, D. Liu, J. Zhao and N. Yang, "The research of daily total solar-radiation and prediction method of photovoltaic generation based on wavelet-neural network", Asia-Pacific Power and Energy Engineering Conference, pp.1-5, 2011.

[29] H.T. Yang, C.M. Huang, Y.C. Huang and Y.-S. Pai, "A weather-based hybrid method for 1-day ahead hourly forecasting of PV power output", IEEE Transactions on Sustainable Energy, vol. 5, no. 3, pp. 917-926, July 2014.

[30] F. Bellocchio, S. Ferrari, M. Lazzaroni, L. Cristaldi, M. Rossi, T. Poli, and R. Paolini, "Illuminance prediction through SVM regression", IEEE Workshop on Environmental Energy and Structural Monitoring Systems, pp. 1-5, 2011.

[31] A. Aybar-Ruiz, S. Jimenez-Fernandez, L. Cornejo-Bueno, C. Casanova-Mateo, J. Sanz-Justo, P. Salvador-Gonzalez and S. Salcedo-Sanz, "A

novel grouping genetic algorithm–extreme learning machine approach for global solar radiation prediction from numerical weather models inputs", Solar Energy, vol. 132, pp. 129-142, 2016.

[32] K. Chiteka and C.C. Enweremadu, "Prediction of global horizontal solar irradiance in Zimbabwe using artificial neural networks", Journal of Cleaner Production, vol. 135, pp. 701-711, 2016.

[33] W. Yaïci and E. Entchev, "Adaptive neuro-fuzzy inference system modelling for performance prediction of solar thermal energy system", Renewable Energy, vol. 86, pp. 302-315, 2016.

[34] E.F. Alsina, M. Bortolini, M. Gamberi and A. Regattieri, "Artificial neural network optimisation for monthly average daily global solar radiation prediction", Energy Conversion and Management, vol. 120, pp. 320-329, 2016.

[35] A. Yona, T. Senjyu, T. Funabashi and C.H. Kim, "Determination method of insolation prediction with fuzzy and applying neural network for long-term ahead PV power output correction", IEEE Transactions on Sustainable Energy, vol. 4, no. 2, pp. 527-533, April 2013.

[36] H. Kutucu and A. Almryad, "Modeling of solar energy potential in Libya using an artificial neural network model", IEEE 1st International Conference on Data Stream Mining & Processing, pp. 356-359, 2016.

[37] Q. Liu, T. Mak, T. Zhang, X. Niu, , W. Luk and A. Yakovlev, "Power-Adaptive computing system design for solar-energy-powered embedded systems", IEEE Transactions on Very Large Scale Integration Systems, vol. 23, no. 8, pp. 1402-1414, August 2015.

[38] J. Liu, W. Fang, X. Zhang and C. Yang, "An improved photovoltaic power forecasting model with the assistance of aerosol index data", IEEE Transactions on Sustainable Energy, vol. 6, no. 2, pp. 434-442, April 2015.

[39] Z. Zhifeng, T. Jianjun, Z. Tianjina and Z. Linlin, "PV power short-term forecasting model based on the data gathered from monitoring network", China Communications, vol. 11, no.14. pp. 61-69, 2015.

[40] T. Ren, S. Liu, G. Yan and H. Mu, "Temperature prediction of the molten salt collector tube using BP neural network", IET Renewable Power Generation, vol. 10, no. 2, pp. 212-220, 2016.

[41] G. Capizzi, F. Bonanno and C. Napoli, "A wavelet based prediction of wind and solar energy for long-term simulation of integrated generation systems", International Symposium on Power Electronics, Electrical Drives, Automation and Motion, pp. 586-592, 2010.

[42] A. Asrari, T. Wu and B. Ramos, "PA hybrid algorithm for short-term solar power prediction- sunshine state case study", IEEE Transactions on Sustainable Energy, pp. 1-10, 2016.

[43] F. Grimaccia, M. Mussetta and R. Zich, "Neuro-Fuzzy predictive model for pv energy production based on weather forecast", IEEE International Conference on Fuzzy Systems, pp. 2454-2457, 2011.

[44] D. M. Quan, E. Ogliari, F. Grimaccia, S. Leva, M. Mussetta, "Hybrid model for hourly forecast of photovoltaic and wind power", IEEE International Conference on Fuzzy Systems, pp. 1-6, 2013.

[45] G. Capizzi, C. Napoli and F. Bonanno, "Innovative second-generation wavelets construction with recurrent neural networks for solar radiation forecasting", IEEE Transactions on Neural Networks and Learning Systems, vol. 23, no. 11, pp. 1805-1815, November 2012.

[46] N. Sharma, P. Sharma, D. Irwin and P. Shenoy, "Predicting solar generation from weather forecasts using machine learning", IEEE International Conference on Smart Grid Communications, pp. 528-533, 2011.

[47] L.P. Naing and D. Srinivasan, "Estimation of solar power generating capacity", IEEE 11th International Conference on Probabilistic Methods Applied to Power Systems, pp. 95-100, 2010.

[48] C.W. Lu, C.Y. Hsieh and C.T. Chiang, "A Recurrent S_CMAC_GBF based estimation for global solar radiation from environmental information", International Joint Conference on Neural Networks, pp. 1-5, 2010.

[49] M.A. Khan, S. Huque and A. Mohammad, "A neural network model for estimating global solar radiation on horizontal surface", International Conference on Electrical Information and Communication Technology, pp. 1-4, 2013.

[50] S. Hussain and A.A. Alili, "Soft computing approach for solar radiation prediction over Abu Dhabi, UAE: A comparative analysis", IEEE International Conference on Smart Energy Grid Engineering, pp. 1-6, 2015.

[51] K.M. Rabbi, I. Nandi, A.S. Saleh, F. Faisal and S. Mojumder, "Prediction of solar irradiation in Bangladesh using artificial neural network (ann) and data mapping using GIS technology", 4th International Conference on the Development in the Renewable Energy Technology, pp. 1-6, 2016.

[52] S. Mohanty and P.K. Patra, "Prediction of global solar radiation using nonlinear auto regressive network with exogenous inputs (NARX)", 39th National Systems Conference, pp. 1-6, 2015.