❏   900

Daily predictions of solar radiation utilizing genetic programming techniques

Rahima Ummi Kulsum Nadya[1], Ali. Najah Ahmed[2], Abdoulhdi A. Borhana[3], N. A. Mardhiah[4], Amr El-Shafie[5], Ahmed El-Shafie[6]

[1,2]Institute of Energy Infrastructure (IEI), Department of Civil Engineering, College of Engineering, Universiti Tenaga Nasional, Malaysia
[3,4]Department of Mechanical Engineering, College of Engineering, Universiti Tenaga Nasional, Malaysia
[3]Department of Mechanical Engineering, College of Engineering Science & Technology, Libya
[5]Civil Engineering Department, Giza High Institute for Engineering and Technology, Egypt
[6]Department of Civil Engineering, Faculty of Engineering, University of Malaya (UM), Malaysia

Article Info

ABSTRACT

The solar radiation prediction in Kuala Terengganu located in Terengganu, Malaysia was investigated in this study to improve the solar system design. Solar radiation data and number of parameters such as solar radiation, temperature, humidity, wind speed and sunshine hours were obtained from Malaysian Meteorological Malaysia MMD. In order to predict the solar radiation, Genetic Programming Techniques (GP) models were develop and tested. Two scenarios were considered in this study in order to validate the efficiency of the proposed model. Coefficients of determination (R^2) for the solar radiation during training and testing phases were ranged between 0.99402 to 0.98934 for all months of the year. This study confirms the ability of GP to predict solar radiation values precisely and accurately. The predictions from the GP models could enable scientists to locate and design solar energy systems in Malaysia.

Corresponding Author:

Rahima Ummi Kulsum Nadya,
Department of Civil Engineering,
College of Engineering,
Universiti Tenaga Nasional,
Jalan IKRAM-UNITEN, 43000 Kajang, Selangor, Malaysia.
Email: Mahfoodh@uniten.edu.my

1. INTRODUCTION

Energy is essential to economic and social development and improved quality of life of human being. There exists a direct correlation between the country development and its energy consumption. World reserve of conventional energy sources are limited and will be used up once. Therefore, the whole world is looking for non-exhaustible energy sources for their future [1]. Renewable energy resources have gained significant importance in the 21st century due to awareness of environmental pollution and depleting reservoirs of fossil fuels [2]. Among a list of renewable energy sources, solar energy is an efficient, flexible and one of the most important renewable energy source. Moreover solar energy conversion technology is environmentally sound. Because of the importance of solar energy as a renewable energy source, an accurate determination of significant covariates and their relationships with the amount of global solar radiation reaching the Earth is a critical research problem.

Solar energy is one of the most important renewable energy sources for world's energy demand. Assessment of solar energy through the available solar radiation data is the first step towards the solar energy applications. Accurate knowledge of solar radiation is necessary for different solar energy applications. In addition, solar power design is only achievable with accurate knowledge of solar radiation of the project

area. This is required so as to have an effective and efficient solar power design that will be able to supply the required load demand and also have proper knowledge of the energy that can be generated from the project area [3].

Forecasting of solar radiation is essential in solar energy power plants (SEPPs) development. The electrical energy generated from the sunlight relies on the weather and climate conditions in the area where the power plants are installed. The condition of solar irradiation will indirectly affect the electrical grid system into which the SEPPs are injected, i.e. the amount and direction of the power flow, voltage, frequency, and also the dynamic state of the system. Therefore, the prediction of solar radiation condition is very crucial to identify its impact into the system. There are several approaches in determining the solar radiation prediction, either by mathematical approach or by heuristic approach such as artificial intelligent method [4]. Some methods which can be used to predict the availability of solar sources are based on the probabilitic methods, network monitoring data, artificial neural network and linear regression methods, fuzzy logic method, mathematical approach using the atmospheric and geometric theory [5].

In recent times the number of larger solar installations; both large scale photovoltaic and also concentrated solar thermal plants are considerably increased due to financial support of the governments. In order to first influence financial backers to participate in their development, and also to potentially compete in the electricity markets, better solar energy prediction models are required [6]. Reinforcement Learning (RL) is an attractive method to provide adaptation mechanisms in the dynamic environments through trial and error, where the rewards are given by the environments depending on the actions taken by the agent. The aim of the agent is to maximize the rewards, and RL learns a policy to maximize the accumulated rewards. State of the art can be seen that the integration of RL to Evolutionary Algorithms (EA), such as Genetic Algorithm (GA) [7], Genetic Programming (GP) [8, 9] and Genetic Network Programming (GNP) were studied in many researches [10], where the integration can improve the performance as shown in GNP with RL (GNP-RL) which was implemented to navigate the mobile robot.

Therefore, the predictions of solar radiation is useful for the purpose of planning, analyze, design, operation, optimization, utilizing the Genetic Programming Technique (GP). Even though there are many studies on solar radiation prediction were reported, very limited studies have been conducted using the GPT. It has been seen from many literature studies that there are not many studies using this methods to predict the solar radiation especially in Malaysia. End of this problem statement is to aim by using the Genetic Programming Technique (GP) to optimize and predict the solar radiation in Terengganu. The main objectives on this study is to develop model to predict solar radiation at Terengganu by using the Genetic Programming Technique (GP), to optimize the hyperparameters of GP and to study the impacts of input parameters on the accuracy of the model, and finally to validate the proposed model capable of predicting solar radiation in Malaysia.

2. METHODOLOGY

2.1. Data

One of the key functions of GP is to identify the model input parameters that could impact the output parameters considerably. The selection of input parameters depends on a priori knowledge regarding causal variables as well as statistical analysis pertaining to the potential outputs and inputs. On the basis of the literature [11-13], the following were chosen for GP modelling: Maximum Air Temperature, Minimum Air Temperature, Humidity, Wind Speed and Sunshine Hours. The basic statistical parameters, i.e. mean, minimum, maximum and standard deviation (S.D.) of the input and output parameters deployed in this study are depicted in Table 1.

Table 1. Input parameters for GP model

	Solar Radiation	Maximum Air Temperature	Minimum Air Temperature	Humidity	Wind Speed	Sunshine Hours
	MJ/m^2	C	C	%	m/s	h
Mean	46183.81638	31.51367742	23.86687711	45.26	2.84	6.541319508
Min	14316.69868	23.24	18.24	77.64886482	1.34	1.2
Max	90213.79489	34.06	25.88	90.54	2.84	8.42
SD	23665.2685	3.135051876	2.32872812	12.8729793	0.337841081	1.893465003

2.2. Geneteic programming – GP

Genetic programming (GP), proposed by [14] as a method to genetically develop populations of mathematical models for prediction of a system behavior with even high complexity [15]. The GP family

algorithms are the biologically-inspired techniques which attempt to find a solution (program) for a problem in a symbolic form through a number of genetic operations represented by tree structures [16-20].

2.3. Model performance evaluation

It is necessary to clearly recognise the criteria that are associated with judging the model's performance [21-23]. The criteria employed to assess the performance of the model in this study were clearly mentioned in the literature. In this study, Coefficients of Determination (CE) [24] as well as Root Mean Square Error (RMSE) [25] was employed to examine the model's performance.

$$RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^{n} ((Xm) - (Xp))^2} \tag{1}$$

$$CE = 1 - \frac{\sum_{i=1}^{n} (X_m - X_p)^2}{\sum_{i=1}^{n} (X_m - \overline{X}_m)^2} \tag{2}$$

3. RESULTS AND DISCUSSION

The proposed Genetic Programming models were developed by applying two different scenarios and cases where different input combinations have been investigated the combination of input and output in order to identify which of these models were more accurate. Table 2 shows the selection of generation of GP for each month.

Table 2. Statistical properties of multigene model during training

Month	Scenario 1				Scenario 2			
	Case1		Case 2		Case1		Case 2	
	Best Fitness	Generation	Best Fitness	Generation	Best Fitness	Generation	Best Fitness	Generation
February	0.0096054	3	0.0093124	33	0.0096054	3	0.0083907	9
March	0.0098545	15	0.0098333	14	0.009829	30	0.0093372	44
April	0.0063636	4	0.0098306	53	0.0099261	45	0.0099617	70
May	0.0075041	53	0.0097948	51	0.0089579	5	0.009968	63
June	0.0099184	33	0.0097631	36	0.0097613	13	0.0097954	17

Table 3 presents the results of GP techniques errors for training phase for month of February, March, April, May and June. The table shows that the R-Squared and Adj. R-Squared for training set were minimal where the values are close to one. This shows that there were very little difference between the actual data and the analyzed data.

Table 3. Statistical properties of the proposed models during training phase

Month	Scenario 1				Scenario 2			
	Case1		Case 2		Case1		Case 2	
	R-Squared	Adj. R-Squared	R-Squared	Adj. R-Squared	R-Squared	Adj. R-Squared	R-Squared	Adj. R-Squared
February	0.98934	0.9874	0.99673	0.99614	0.98934	0.9874	0.99735	0.99686
March	0.9941	0.99345	0.99531	0.99479	0.99413	0.99348	0.99735	0.9953
April	0.99754	0.99727	0.99553	0.99501	0.99402	0.99335	0.99541	0.99487
May	0.99658	0.9962	0.99535	0.99483	0.99513	0.99458	0.99518	0.99465
June	0.99402	0.99336	0.99538	0.99487	0.99421	0.99357	0.99417	0.99352

Figure 1 depicts the measured and estimated parameters of solar radiation for the most excellent network, which provided the most precise estimation. On the whole, the predictive capability of this model was fairly good for each of the parameters of the solar radiation in the training and testing duration. The findings showed the proposed GP model is consistent in the prediction despite of high variations and intrinsic non-linear correlation among the parameters of the solar radiation.
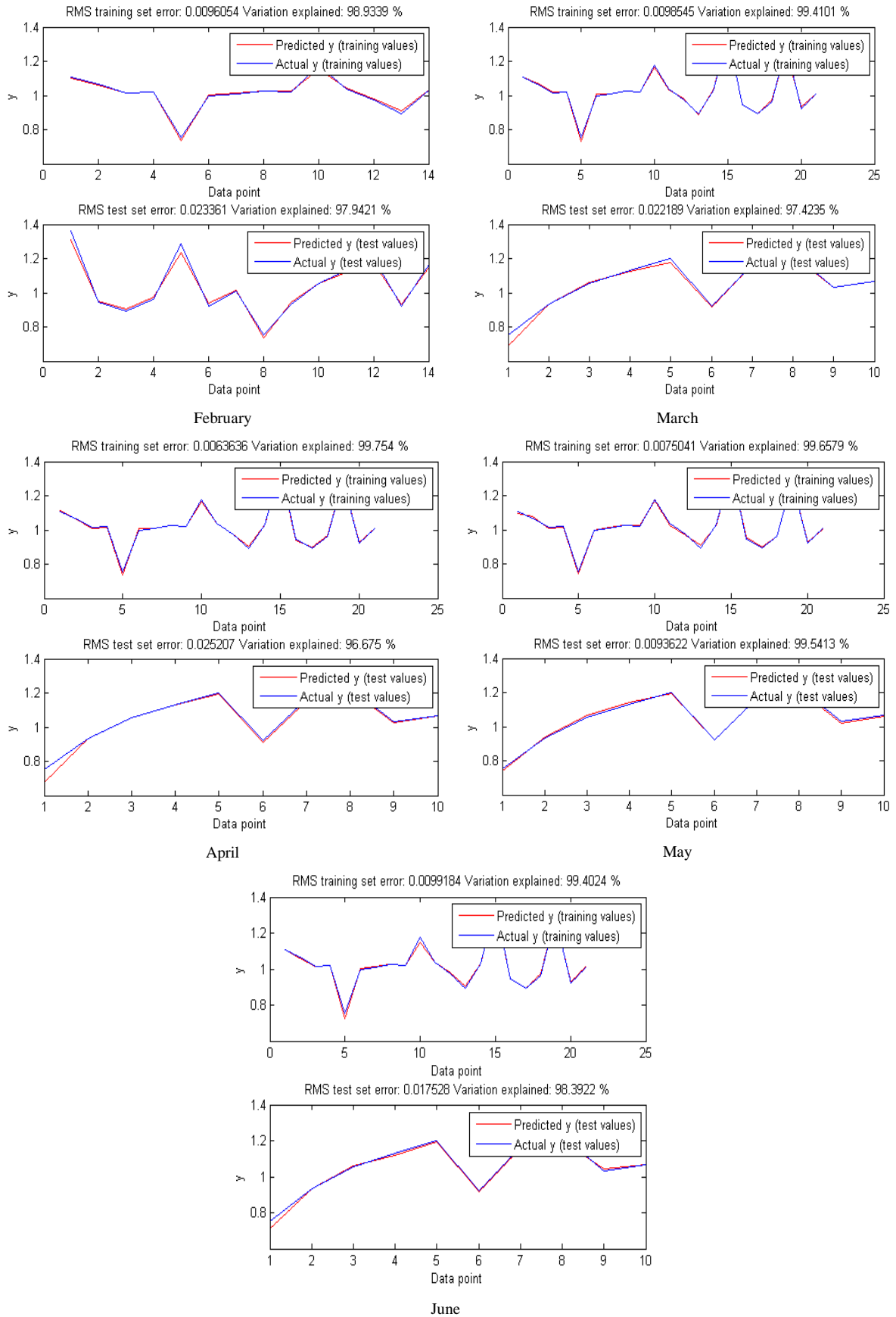
Figure 1. The measured and estimated parameters of solar radiation

3.1. Scenarios and cases comparison

Case 1 and Case 2 for both scenarios show the most effective combinations of the training and testing data sets to predict the global solar radiation. To evaluate the performance between these cases RMSE is used as indicator. Table 4 shows the accuracy of the proposed models during training and testing phases for both scenarios and cases. It can be seen from the table that Case 1 for Scenario 1 shows the least RMSE which is an indicator that the model is performed well. That the observed values are the most similar to the actual data compared to other scenario and cases. The model developed for scenario 1 case 1 showed excellent efficiency in predicting solar radiation.

Table 4. Comparison between Scenario 1 and Scenario 2 with their cases

	SCENARIO 1 Case 1					SCENARIO 2 Case 1			
Months	RMSE		Variation %		Months	RMSE		Variation %	
	Training	Testing	Training	Testing	Testing	Testing	Testing	Training	Testing
February	0.0096054	0.023361	98.9339	97.9421	February	0.0096054	0.023361	98.9339	97.9421
March	0.0098545	0.022189	99.4101	97.4235	March	0.009829	0.013669	99.4132	99.0223
April	0.0063636	0.025207	99.754	96.675	April	0.0099261	0.024088	99.4015	96.9638
May	0.0075041	0.0093622	99.6579	99.5413	May	0.0089579	0.033579	995126	94.0994
June	0.0099184	0.017528	99.4024	98.3922	June	0.0097613	0.015229	99.4212	98.7863
	Case 2					Case 2			
Months	RMSE		Variation %		Months	RMSE		Variation %	
	Training	Testing	Training	Testing		Training	Testing	Training	Testing
February	0.0093124	0.03635	99.673	84.7316	February	0.0083907	0.016171	99.7345	96.9783
March	0.0098333	0.026688	99.5313	93.2105	March	0.0093372	0.011388	99.5774	98.7637
April	0.0098306	0.016228	99.5531	97.4896	April	0.0099617	0.012549	99.5411	98.4989
May	0.0097948	0.020463	99.5349	96.0085	May	0.009968	0.014155	99.5183	98.09
June	0.0097631	0.026687	99.5379	93.211	June	0.0097954	0.021428	99.4172	97.5973

4. CONCLUSION

The study attempted to develop a robust artificial intelligence model such as genetic programing to predict a daily solar radiation. Three statistical criteria were adopted, namely absolute error, r-squared and correlation of determination. Based on the results, case 1 from scenario 1 was chosen as the best model compared to the other scenario and cases. The proposed GP model can predict the solar radiation fairly accurately only by using 6 parameters as input. It has shown that the solar radiation, temperature, humidity, wind speed and sunshine hours affect the solar radiation prediction significantly. Coefficients of determination (R^2) for the solar radiation during training and testing phases were found to be 0.98934 for February, 0.9941 for March, 0.99754 for April, 0.99658 for May and 0.99402 for June. These results confirmed that GP is the suitable technique to predict solar radiation values precisely. Further improvements can be achieved by utilizing more input parameters and also by augment optimization techniques with GP model.

REFERENCES

[1]   M. Arif, S. Bhuiyan, and M. J. A. Patwary, "Emperical Computation of Solar Radiation and Determination of Regression Emperical Computation of Solar Radiation and Determination of Regression Coefficients for Khulna City," *TELEKOMNIKA Indonesian Journal of Electrical Engineering*, vol. 12, no. 12, pp. 8015–8021, 2014.
[2]   A. Qazi, H. Fayaz, A. Wadi, R. Gopal, N. A. Rahim, and W. Ahmed, "The artificial neural network for solar radiation prediction and designing solar systems : a systematic literature review," *Jouornal of Cleaner Production*, vol. 104, pp. 1–12, 2015.
[3]   S. Salisu, M. W. Mustafa, and M. Mustapha, "A Wavelet Based Solar Radiation Prediction in Nigeria Using Adaptive Neuro-Fuzzy Approach," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 12, no. 3, pp. 907–915, 2018.
[4]   H. Suyono, R. N. Hasanah, R. A. Setyawan, P. Mudjirahardjo, and A. Wijoyo, "Comparison of Solar Radiation Intensity Forecasting Using ANFIS and Multiple Linear Regression Methods," *Bulletin Electrical Engineering. Informatics (BEEI)*, vol. 7, no. 2, pp. 191–198, 2018.

[5]   M. Elyaqouti, L. Bouhouch, and A. Ihlal, "Modelling and Predicting of the Characteristics of a Photovoltaic Generator on a Horizontal and Tilted Surface," *International Journal of Electrical and Computer Engineering(IJECE)*, vol. 6, no. 6, pp. 11740, 2016.

[6]   G. Landeras, J. Javier, O. Kisi, and J. Shiri, "Comparison of Gene Expression Programming with neuro-fuzzy and neural network computing techniques in estimating daily incoming solar radiation in the Basque Country (Northern Spain)," *Energy Conversion and Management*, vol. 62, pp. 1–13, 2012.

[7]   S. N. Anual, M. F. Ibrahim, N. Ibrahim, and A. Hussain, "GA-based Optimisation of a LiDAR Feedback Autonomous Mobile Robot Navigation System," *Bulletin Electrical Engineering. Informatics (BEEI)*, vol. 7, no. 3, pp. 433–441, 2018.

[8]   J. R. Koza, "Genetic Programming II: Automatic Discovery of Reusable Programs," *Cambridge, MA, USA: MIT Press*, 1994.

[9]   Y. Xiaoyu, C. A. I. Meng, and L. I. Jianxun, "Path Planning for Unmanned Aerial Vehicles Based on Genetic Programming," in *2016 28th Chinese Control and Decision Conference (CCDC)*, pp. 717–722, 2016.

[10] F. Liu, S. Liang, and X. Xian, "Optimal Path Planning for Mobile Robot Using Tailored Genetic Algorithm," *TELEKOMNIKA Indonesian Journal of Electrical Engineering*, vol. 12, no. 1, pp. 1–9, 2014.

[11] Muzathik, "Reference Solar Radiation Year and Some Climatology Aspects of East Coast of West Malaysia," *Am. J. Eng. Appl. Sci.*, vol. 3, no. 2, pp. 293–299, 2010.

[12] A. M. Muzathik, M. Z. Ibrahim, K. B. Samo, and W. B. Wan Nik, "Estimation of global solar irradiation on horizontal and inclined surfaces based on the horizontal measurements," *Energy*, vol. 36, no. 2, pp. 812–818, 2011.

[13] S. Mohammad, E. S. Mostafavi, A. Jaafari, and A. Jaafari, "Using measured daily meteorological parameters to predict daily solar radiation," *Measurement*, vol. 76, pp. 148–155, 2015.

[14] J. R. Koza, "Genetic programming as a means for programming computers by natural selection," *Stat. Comput.*, vol. 4, no. 2, pp. 87–112, 1994.

[15] A. Asadi Tashvigh, F. Zokaee Ashtiani, M. Karimi, and A. Okhovat, "A novel approach for estimation of solvent activity in polymer solutions using genetic programming," *Calphad*, vol. 51, pp. 35–41, 2015.

[16] D. E. Goldberg and J. H. Holland, "Genetic Algorithms in Search, Optimization and Machine Learning," *Mach. Learn.*, vol. 3, pp. 95–99, 1988.

[17] F. Rahdari, M. Eftekhari, and R. Mousavi, "A two-level multi-gene genetic programming model for speech quality prediction in Voice over Internet Protocol systems R," *Comput. Electr. Eng.*, vol. 49, pp. 9–24, 2016.

[18] Escobar, Carlos A., Diana M. Wegner, Abhinav Gaur, and Ruben Morales-Menendez. "Process-Monitoring-for-Quality—A Model Selection Criterion for Genetic Programming." In *International Conference on Evolutionary Multi-Criterion Optimization*, pp. 151-164, 2019.

[19] Gao, Wei, Hossein Moayedi, and Amin Shahsavar. "The feasibility of genetic programming and ANFIS in prediction energetic performance of a building integrated photovoltaic thermal (BIPVT) system." *Solar Energy* vol. 183, pp. 293-305, 2019.

[20] Riahi-Madvar, Hossien, Majid Dehghani, Akram Seifi, and Vijay P. Singh. "Pareto Optimal Multigene Genetic Programming for Prediction of Longitudinal Dispersion Coefficient." *Water resources management,* vol. 33, no. 3,: pp. 905-921, 2019.

[21] Tikhamarine, Yazid, Doudja Souag-Gamane, Ali Najah Ahmed, Ozgur Kisi, and Ahmed El-Shafie. "Improving artificial intelligence models accuracy for monthly streamflow forecasting using grey Wolf optimization (GWO) algorithm." *Journal of Hydrology*, vol. 582, pp. 124435, 2020.

[22] Banadkooki, Fatemeh Barzegari, Mohammad Ehteram, Ali Najah Ahmed, Fang Yenn Teo, Chow Ming Fai, Haitham Abdulmohsin Afan, Michelle Sapitang, and Ahmed El-Shafie. "Enhancement of Groundwater-Level Prediction Using an Integrated Machine Learning Model Optimized by Whale Algorithm." *Natural Resources Research*, pp. 1-20, 2020.

[23] Ibrahim, Rusul Khaleel, Seef Saadi Fiyadh, Mohammed Abdulhakim AlSaadi, Lai Sai Hin, Nuruol Syuhadaa Mohd, Shaliza Ibrahim, Haitham Abdulmohsin Afan, Chow Ming Fai, Ali Najah Ahmed, and Ahmed Elshafie. "Feedforward Artificial Neural Network-Based Model for Predicting the Removal of Phenolic Compounds from Water by Using Deep Eutectic Solvent-Functionalized CNTs." *Molecules*, vol. 25, no. 7, pp. 1511, 2020.

[24] Rodríguez Sánchez, Ainara, Román Salmerón Gómez, and Catalina García. "The coefficient of determination in the ridge regression." *Communications in Statistics-Simulation and Computation*, pp. 1-19, 2019.

[25] Pasquier, Philippe, Angelo Zarrella, and Denis Marcotte. "A multi-objective optimization strategy to reduce correlation and uncertainty for thermal response test analysis." *Geothermics*, vol. 79, pp. 176-187, 2019.