



Application of support vector regression integrated with firefly optimization algorithm for predicting global solar radiation

Saeed Samadianfard

Department of Water Engineering, University of Tabriz, Tabriz, Iran, s.samadian@tabrizu.ac.ir
ORCID: 0000-0002-6876-7182

Salar Jarhan

Department of Water Resources Engineering, University of Tabriz, Tabriz, Iran, salar.jarhan@gmail.com
ORCID: 0000-0002-3432-9985

Hamed Sadri Nahand

Department of Water Engineering, University of Tabriz, Tabriz, Iran, hamed.sadri7@yahoo.com,
ORCID: 0000-0002-7091-505X

Arrived: 09.09.2018 Accepted: 13.11.2018 Published: 31.12.2018

Abstract: A fundamental factor for proficient designing of solar energy systems is providing precise estimations of the solar radiation. Global solar radiation (GSR) is a vital parameter for designing and operating solar energy systems. Because records of GSR are not available in many places, especially in developing countries, this research aims to model the GSR using support vector regression (SVR) in a hybrid manner that is integrated with the firefly Optimization algorithm (SVR-FFA). For this purpose, the daily meteorological parameters and GSR measured from beginning of 2011 to the end of 2013 at Tabriz synoptic station were utilized. For assessing the performance of the mentioned methods, different statistical indicators were implemented. For all of the defined predictive models with different combinations of meteorological parameters, the performance of the SVR-FFA hybrid model is better than the classical SVR, evidenced by the higher value of R (~0.892-0.982 relative to ~0.891-0.977) and lower values of RMSE and MAE (~1.551-3.725vs.1.748-4.067 and ~0.911-2.862vs.1.103-2.742). As a remarkable point studied empirical equations had higher prediction errors comparing with the developed SVR-FFA models. Conclusively, the obtained results proved the high proficiencies of SVR-FFA method for predicting global solar radiation.

Keywords: *El Fire fly optimization algorithm, Solar radiation, Statistical parameters, Support vector regression*

Cite this paper as: Samadianfard, S, Jarhan, S, Sadri Nahand, H. Application of support vector regression integrated with firefly optimization algorithm for predicting global solar radiation. *Journal of Energy Systems*, 2018; 2(4): 180-189, DOI: 10.30521/jes.458328

1. INTRODUCTION

The world is facing severe problems such as urban air pollution and global warming due to the rapid increasing of the consumption of fossil fuels. In the other hand, between the several accessible resources, solar energy has received a massive notice because it is sustainable, plentiful and environmental friendly [1]. Moreover, solar radiation as a renewable energy source provides the majority of the requiring heat energy for the most processes of the earth [2].

The awareness of solar data in a particular place is vital in designing and predicting the operations of solar systems. Also, the best way of obtaining the mentioned data is using specific instruments for measuring the solar radiation. But because these measurement instruments are expensive, different predictive models have been recommended for global solar radiation (GSR) estimation using different combinations of meteorological parameters [3]. The models for predicting GSR vary from simple empirical equations to the complex numerical procedures. So, another possibility which has been suggested by different researchers is using data-driven methods. These models have been implemented extensively for solving non-linear problems.

Support vector machines (SVMs), as one of the data driven techniques [4,5,6], has received noticeable attention in environmental and engineering problems [7]. Support vector classification (SVC) and support vector regression (SVR) are its two basic categories. SVM, as one of the learning frameworks, implements a high-dimensional space [8]. Additionally, SVR is concentrated on a quantifiable learning theory and the rule of minimizing the structural risk. Moreover, it has been used extensively for nonlinear frameworks [9]. The accuracy of an SVM model is based on reliable determination of the model parameters. So, optimizing methods for choosing the best and precise parameters are of the high importance. Furthermore, inspecting the literature exposes that a very few studies have been carried out to use the optimized SVM for the precise calculation of solar radiation [10,11,12].

Solmaz and Ozgoren [13] applied the artificial neural network (ANN) for determining the hourly GSR values of six selected locations in Turkey. According to their results, ANN produced proficient results in predicting solar radiation. Also, an ANN model for predicting the GSR values in China was established by Jiang [14]. He stated that the ANN model has more precision as compared to other regression models. In another research, several ANN models have been developed by Benghanem et al. [15] for GSR values in Saudi Arabia. The obtained results indicated that the most accurate model only uses sunshine duration and air temperature parameters. In another study, Mellit et al. [16] suggested a combined model of the neural network and fuzzy logic for predicting GSR values. Due to the obtained high correlations and fewer errors, they commented that the mentioned model is suitable for GSR estimation. Moreover, a specific approach based on artificial neural networks is developed by Bosch et al. [17] for estimation of daily global irradiation. The proposed ANN models produced sufficient results in tested phase. So, based on the adequate findings, this methodology has been recommended to be applied in other areas with a complex topography.

Furthermore, the use of support vector machines (SVMs), as one of the data driven techniques, has grown in the environmental subjects [7]. SVM is a learning framework, but support vector regression (SVR) is based on minimizing structural risk in a computable learning process in complex and nonlinear problems [9]. Inspecting the literature indicates that despite the increasing uses of SVM and SVR, only a few researches have been done using them for predicting GSR [10,11,12]. Ramedani et al. [18] implemented SVR for developing predictive models of GSR values in Tehran, Iran. In this research, the accuracy of radial basis and polynomial functions in SVR model has been compared and it has been found that SVR with radial basis function provided more accurate results of GSR than other studied function.

In the current research, we intend to implement a hybrid approach for predicting GSR values by using Firefly Algorithm (FFA) to optimize the architecture of SVR models. FFA is a nature-inspired optimizing algorithm which is based on the flashing behavior of the fireflies. Although FFA has been utilized productively in diverse fields, but to the best of our knowledge, there is no study in the literature investigating the capability of the hybrid SVR with FFA algorithm (SVR-FFA) in GSR prediction. For demonstrating the suitability of the hybrid SVR-FFA approach, the results are compared with the SVR model. So, the goal of the current study is exploring the capabilities of SVR-FFA and comparison with SVR model to predict GSR values in the north-west of Iran.

2. MATERIALS and METHODS

2.1. Support Vector Regression (SVR) and Hybrid SVR-FFA Model

SVM, which is based on statistical learning theory, has been extensively implemented for recognizing complex patterns of different environmental phenomena. Furthermore, SVR as one of the versions of the SVM is utilized for regression problems. Guzmán et al. [19] presented the mathematical details of SVR model in their research about investigating the ground water levels. Additionally, it is notable that the key merit of the SVR is its simple applicability in finding the relationships in nonlinear and complex problems using a hyperplane function and applying kernel function and the epsilon coefficient that is the limit of deviation between observed and predicted values. Fig. 1 designates schematic configuration of SVR model.

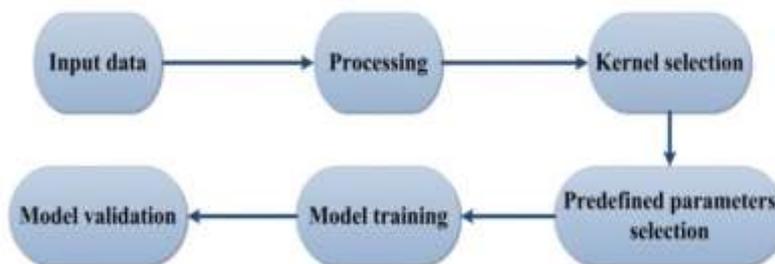


Figure 1. schematic configuration of SVR model.

SVR model permits tolerating errors up to ε in the training data sets. So, the SVR look for a linear function as follows:

$$P(x) = F^T x + L \tag{1}$$

where F and L represent the coefficients of the weight vector. This linear regression can be defined as the following:

$$\min \frac{1}{2} \|F\|^2 + C \sum_{i=1}^N (\xi_i + \xi_i^*) \quad \text{subject to} \quad \begin{cases} F^T x + L - y_i \leq \varepsilon + \xi_i^* \\ y_i - F^T x - L \leq \varepsilon + \xi_i \\ \xi_i, \xi_i^* \geq 0 \quad i = 1, 2, \dots, N \end{cases} \tag{2}$$

C is a predefined constant trade-off factor for the grade of the experimental error. The FFA, which is firstly established by Yang [20], is one of the intelligent optimization algorithms. The basis of the mentioned algorithm is the movement patterns of fireflies. Moreover, the FFA uses the brightness of the fireflies as their most elegant characteristics. In other words, the fireflies are moving toward the light

and in their group traveling, the fireflies are chasing the brightest firefly. So, the FFA optimizes the parameters of the SVR (ε and C) using the mentioned characteristics of the fireflies and the variation of light intensity in their swarms. This optimized method is called SVR-FFA [21].

2.2. Study Area and Data Collection

The daily climatic data of a weather station located in Tabriz, Iran, were utilized in the current study (Fig. 2). Tabriz, which is located at northwest of Iran, has a semi-arid climate and cold winters. Meteorological parameters which were used for the current research are: minimum temperature ($^{\circ}\text{C}$) (T_{\min}), maximum temperature ($^{\circ}\text{C}$) (T_{\max}), relative humidity (%) (RH), sunshine hours (hr) (n), maximum sunshine hours (hr) (N), corrected clear-sky solar irradiation (-) (ICSKY), day of year (day) (DOY) and extra-terrestrial radiation ($\text{MJ m}^{-2} \text{d}^{-1}$) (R_a) and global solar radiation ($\text{MJ m}^{-2} \text{d}^{-1}$) (GSR) with the time period of 2011 to 2013. Table 1 represents the daily statistical parameters of the applied meteorological variables for Tabriz Station in both training and test phases. Most of the variables indicate normal distributions because they have low skewness values, except n/N and GSR which they show negative and positive skewed distributions, respectively. Computing correlation coefficients reveal that the R_a , T_{\max} and T_{\min} have the first, second and third highest correlations with GSR values, respectively. Additionally, there is a high inverse correlation between the RH and GSR. The observed meteorological data are shown in Fig. 3.



Figure 2. The location of Tabriz station [22]

Table 1. Statistical analysis of meteorological parameters

	Variable	X_{mean}	X_{min}	X_{max}	S_x	C_V	C_{sx}	Correlation with GSR	with GSR
Train	T_{\min}	7.42	-13.20	28.20	9.94	1.34	-0.03		0.74
	T_{\max}	19.46	-3.20	40.00	11.51	0.59	-0.08		0.81
	RH	50.10	15.00	93.75	17.73	0.35	0.27		-0.71
	n/N	0.65	0.00	0.96	0.28	0.43	-1.04		0.60
	ICSKY	4.92	0.00	9.00	2.50	0.00	-0.44		-0.49
	DOY	182.64	1.00	365.00	110	0.00	0.00		-0.01
	R_a	28.97	14.71	41.82	9.66	0.33	-0.11		0.85
	GSR	18.46	1.97	91.52	9.36	0.51	1.20		1.00

Test	T_{min}	8.90	-10.40	25.00	8.87	1.00	-0.08	0.70
	T_{max}	20.35	-2.20	38.80	10.48	0.52	-0.17	0.80
	RH	51.15	15.00	89.63	17.20	0.34	0.15	-0.76
	n/N	0.64	0.00	0.95	0.29	0.45	-1.07	0.59
	ICSKY	5.21	0.00	9.00	2.373	0.50	-0.58	-0.48
	DOY	183.00	1.00	365.00	105.51	0.58	0.00	-0.08
	R_a	28.93	14.71	41.82	9.67	0.33	-0.10	0.89
	GSR	18.30	1.09	86.59	9.08	0.50	1.09	1.00

Note: the terms X_{mean} , X_{min} , X_{max} , S_x , C_v and C_{sx} denote the mean, minimum, maximum, standard deviation, coefficient of variation and skewness, respectively.

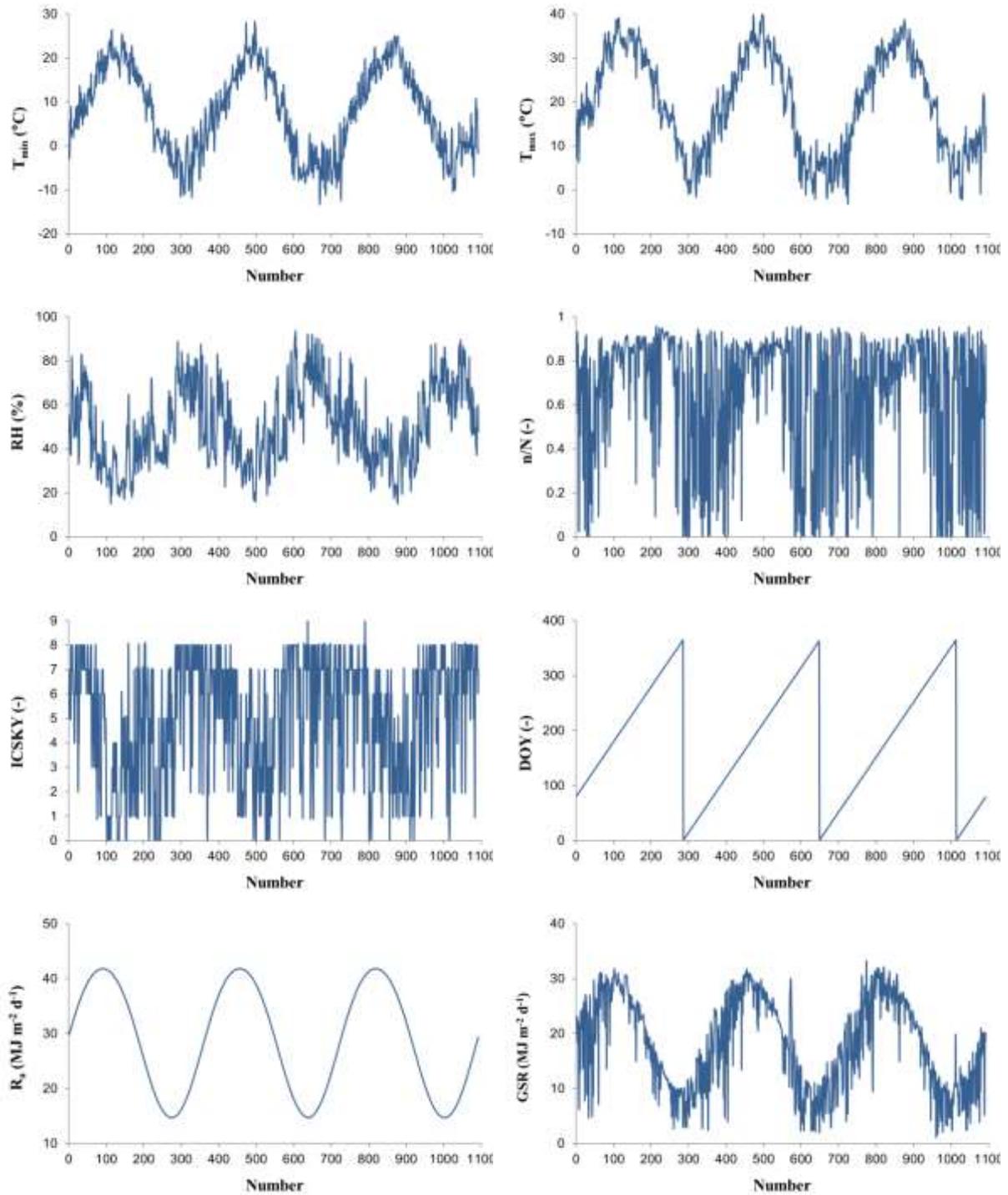


Figure 3. Time variation graphs of the meteorological parameters.

2.3. Empirical Equations

Empirical equations, which utilize meteorological parameters, are favorable for GSR estimation. Angstrom [23] and Prescott [24] recommended Angstrom-Prescott method for predicting GSR using n/N by applying linear relationship. Bristow and Campbell [25] estimated GSR employing minimum and maximum temperatures and applying nonlinear equations. Elagib and Mansell [26] used an exponential function using n/N parameter and Chen et al. [27] offered accurate predictions of GSR values by adding minimum and maximum temperature difference. The selected empirical equations for comparing their accuracies with data-driven techniques are demonstrated in Table 2.

Table 2. Empirical equations employed in the current study

Method	Notation	Equation	Source
Angstrom and Prescott	A-P	$R_s = R_a(a + b \times (n/N))$	[21]
Bristow and Campbell	B-C	$R_s = R_a \times a \times (1 - \exp(-b \times (T_{\max} - T_{\min})^c))$	[22]
Elagib and Mansell	E-M	$R_s = R_a \times a \times \exp(b \times (n/N))$	[23]
Chen et al.	C	$R_s = R_a \times (a \times \ln(T_{\max} - T_{\min}) + b \times (n/N)^c + d)$	[24]

2.4. Evaluation Parameters

For comprehensive validation of the considered models and empirical equations, some statistical parameters including correlation coefficient (CC), root mean square error (RMSE) and Mean absolute error (MAE) were employed in model evaluation [28]. These statistical metrics can be described as follows:

$$CC = \frac{\left(\sum_{i=1}^n O_i P_i - \frac{1}{n} \sum_{i=1}^n O_i \sum_{i=1}^n P_i \right)}{\left(\sum_{i=1}^n O_i^2 - \frac{1}{n} \left(\sum_{i=1}^n O_i \right)^2 \right) \left(\sum_{i=1}^n P_i^2 - \frac{1}{n} \left(\sum_{i=1}^n P_i \right)^2 \right)} \quad (3)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (P_i - O_i)^2} \quad (4)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |P_i - O_i| \quad (5)$$

where O_i and P_i are the observed and predicted the i^{th} value of GSR and n is the number of the GSR values.

3. RESULTS and DISCUSSION

In this study nine input combinations based on seven meteorological parameters (Table 3) were assembled and evaluated to discover whether the proposed SVR-FFA hybrid model was a capable data-driven tool for modelling the GSR values. In all modeling scenarios, we followed the notion that there is no rule of thumb for that the universal way the training and testing data are partitioned. For example, the study of Kurup and Dudani [29] used a total of their 63% of data for training of the model while Pal [30] applied 69% and Samadianfard et al. [28,31] used 67% of total data. Considering these studies, the data in the present study were separated into two distinct parts; training (67%) and testing (33%) for all of the chosen input combinations. In other words, meteorological parameters of the years of 2011 and 2012 were selected for training and the data of the year of 2013 was chosen to testing. Tables 4 and 5 illustrate the designated models and the correspondent statistical parameters and the correspondent time series plots of observed and predicted GSR values are demonstrated in Figure 4. Additionally, Figure 5 illustrates the observed and predicted values of GSR in the test period using the best models of SVR and SVR-FFA models and the best empirical equation.

Table 3. Input parameters for implementation of studied models and empirical equations

Number of input combinations	Input parameters
1	R _a
2	R _a , n/N
3	R _a , RH
4	R _a , T _{min} , T _{max}
5	R _a , RH, T _{min} , T _{max}
6	R _a , RH, T _{min} , T _{max} , n/N
7	R _a , RH, T _{min} , T _{max} , n/N, DOY
8	R _a , RH, T _{min} , T _{max} , n/N, ICSKY
9	R _a , RH, T _{min} , T _{max} , n/N, DOY, ICSKY

Table 4. Statistical parameters of the various scenarios of data-driven techniques in the test period

Method	Parameter	Number of model								
		1	2	3	4	5	6	7	8	9
SVR	CC	0.891	0.975	0.913	0.913	0.921	0.977	0.976	0.977	0.977
	MAE (MJ m ⁻² d ⁻¹)	2.742	1.177	2.510	2.408	2.380	1.130	1.139	1.103	1.106
	RMSE (MJ m ⁻² d ⁻¹)	4.067	1.846	3.412	3.440	3.244	1.771	1.803	1.748	1.772
SVR-FFA	CC	0.892	0.980	0.920	0.936	0.941	0.982	0.982	0.982	0.982
	MAE (MJ m ⁻² d ⁻¹)	2.882	1.010	2.314	2.069	1.972	0.916	0.911	0.919	0.923
	RMSE (MJ m ⁻² d ⁻¹)	3.725	1.644	3.293	2.929	2.803	1.551	1.568	1.564	1.573

Table 5. Statistical parameters of the considered empirical equations in the test period

Parameter	A-P	E-M	B-C	C
CC	0.977	0.975	0.932	0.955
MAE (MJ m ⁻² d ⁻¹)	1.156	1.254	2.231	1.788
RMSE (MJ m ⁻² d ⁻¹)	1.786	1.869	2.999	2.481

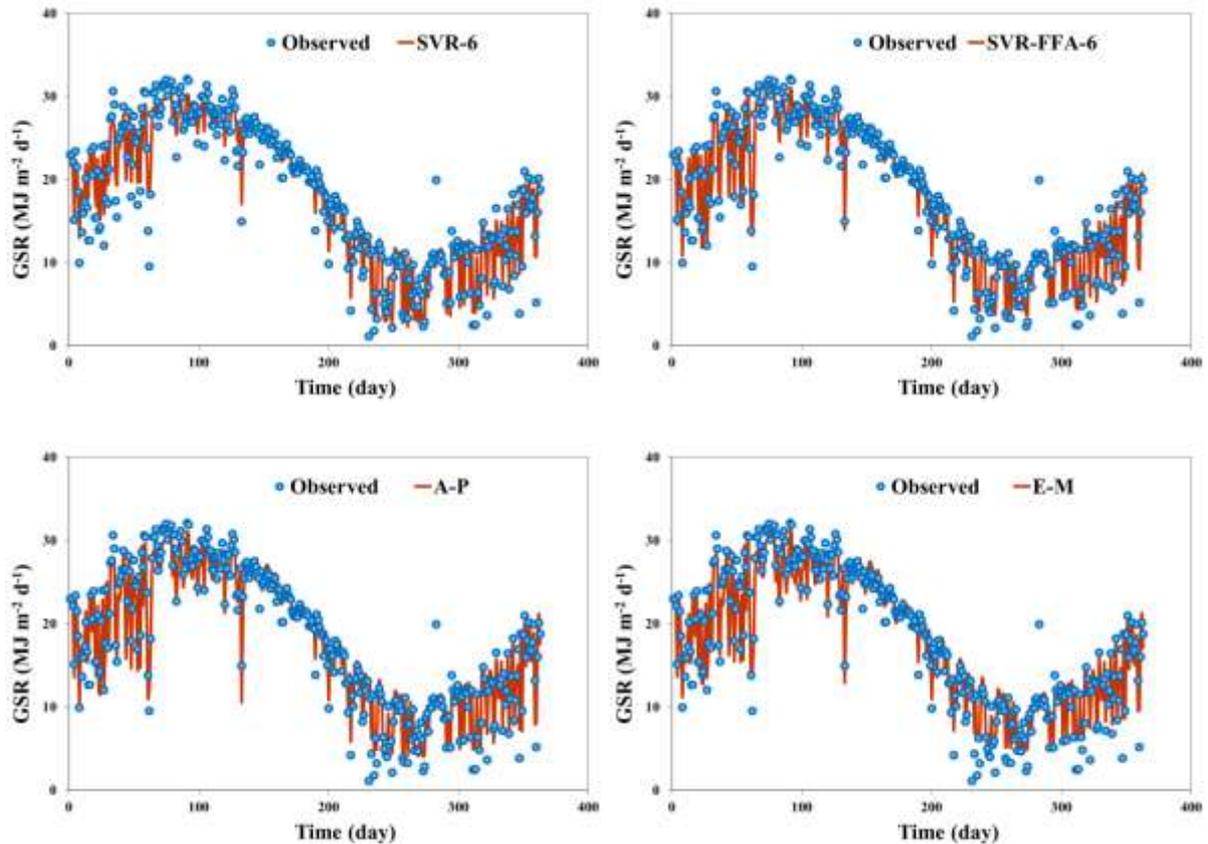


Figure 4. Time series plot of observed and estimated GSR in the test period

After checking the impact of the FFA algorithm on the SVR models, it is evident that the predictive performance regarding the RMSE value directed to an expressively lower value of the error for all model input combinations. This implied that the employment of the FFA into SVR models led to an improvement in the overall model accuracy, specifying that the hybrid SVR-FFA models were optimized more precisely and are likely to lead to better performance for the prediction of GSR values in the testing dataset.

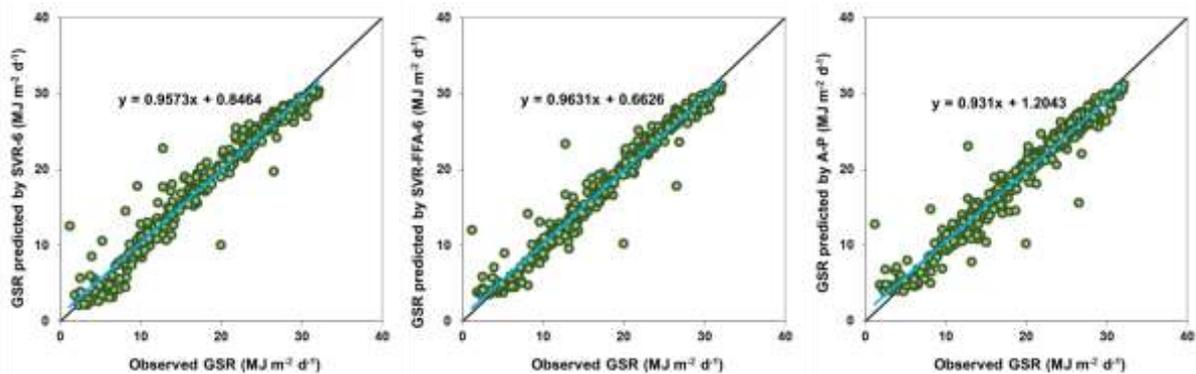


Figure 5. Scatterplots of observed and predicted GSR in the test period

Finally, it can be comprehended that the predictions of SVR-FFA-6 are in better agreement with observed GSR values comparing other models, this maybe was due to the fact that the input parameters of R_a , RH , T_{min} , T_{max} , n/N have the most correlations with GSR values.

4. CONCLUSION

In the present research, we tried to estimate daily GSR values using the SVR-FFA and SVR models. The meteorological parameters of Tabriz station have been gathered, and nine diverse groups were prepared. The results attained exposed that the sixth input combination (utilizing meteorological parameters of R_a , RH, T_{\min} , T_{\max} , n/N) generated the most precise results in the independent testing phase. Additionally, the hybrid SVR-FFA model produced a substantial decrease in MAE and RMSE values compared to the standalone SVR model by approximately 18.9% and 12.4%, respectively, and therefore, is nominated as the optimal model for of the prediction of GSR values. As a conclusion, the acquired results endorsed the adequacy of SVR-FFA hybrid model and pointed out its efficiency for solar energy estimations.

REFERENCES

- [1] Akikur, R, Saidur R, Ping, H, Ullah, K. Comparative study of stand-alone and hybrid solar energy systems suitable for off-grid rural electrification: a review. *Renewable Sustainable Energy Reviews* 2013; 27(3):738-52.
- [2] Kalogirou, SA. Artificial neural networks and genetic algorithms for the modeling, simulation and performance prediction of solar energy systems. *Assessment and Simulation Tools for Sustainable Energy Systems*. Green Energy and Technology 2013; 225–245.
- [3] Fadare, DA. Modelling of solar energy potential in Nigeria using an artificial neural network model. *Applied Energy* 2009; 86:1410–22.
- [4] Celik, E, Gor, H, Ozturk, N, Kurt, E. Application of artificial neural network to estimate power generation and efficiency of a new axial flux permanent magnet synchronous generator. *International Journal of Hydrogen Energy*, 2017; 42(28): 17692-17699.
- [5] Celik, E, Uzun, Y, Kurt, E, Ozturk, N, Topaloglu, N. A neural network design for the estimation of nonlinear behavior of a magnetically-excited piezoelectric harvester. *Journal of Electronic Materials*, 2018; 47(8): 4412-4420.
- [6] Celik, E, Cavusoglu, O, Gurun, H, E, Ozturk, N, Topaloglu, N. Estimation of the clearance effect in the blanking process of CuZn30 sheet metal using neural network—A comparative study. *Bilişim Teknolojileri Dergisi*, 2018; 11(2), 187-193.
- [7] Ornella, L, Tapia, E. Supervised machine learning and heterotic classification of maize (*Zea mays* L.) using molecular marker data. *Computers and Electronics in Agriculture* 2010; 74: 250–257.
- [8] Ananthakrishnan S, Prasad R, Stallard D, Natarajan P. Batch-mode semisupervised active learning for statistical machine translation, *Comput. Speech Lang.* 2013; 17: 397–406.
- [9] Wei Z, Tao T, ZhuoShu D, Zio E, A dynamic particle filter-support vector regression method for reliability prediction, *Reliab. Eng. Syst. Safe.* 2013; 11(9): 109–116.
- [10] Chen, JL, Liu, HB, Wu, W, Xie, DT. Estimation of monthly solar radiation from measured temperatures using support vector machines—a case study. *Renewable Energy* 2011; 36: 413–420.
- [11] Zeng, J, Qiao, W. Short-term solar power prediction using a support vector machine. *Renewable Energy* 2013; 52:118–127.
- [12] Ekici, BB. A least squares support vector machine model for prediction of the next day solar insolation for effective use of PV systems. *Measurement* 2014; 50: 255–262.
- [13] Solmaz, O, Ozgoren, M. Prediction of hourly solar radiation in six provinces in turkey by artificial neural networks. *Journal of Energy Engineering* 2012; 138, 194–204.
- [14] Jiang, Y. Computation of monthly mean daily global solar radiation in China using artificial neural networks and comparison with other empirical models. *Energy* 2009; 1276–1283.
- [15] Benganem, M, Mellit, A, Alamri, S. ANN-based modelling and estimation of daily global solar radiation data: a case study. *Energy Conversion and Management* 2009; 50(7):1644–55.
- [16] Mellit, A, Hadj, Arab, A, Khorissi, N, Salhi, H. An ANFIS-based forecasting for solar radiation data from sunshine duration and ambient temperature. In: *IEEE power engineering society general meeting*; 24–28 June 2007, Florida (USA). p. 1–6

- [17] Bosch JL, Lopez G, Batllea FJ. Daily solar irradiation estimation over a mountainous area using artificial neural networks. *Renew Energy* 2008;33(7):1622–8.
- [18] Ramedani Z, Omid M, Keyhani A, Shamsirband S, Khoshnevisan B. Potential of radial basis function based support vector regression for global solar radiation prediction. *Renewable Sustainable Energy Rev* 2014;39(1):1005–11.
- [19] Guzmán, SM, Paz, JO, Tagert, MLM, Mercer, AE, Pote, JW. An integrated SVR and crop model to estimate the impacts of irrigation on daily groundwater levels. *Agricultural Systems* 2018; 159: 248-259.
- [20] Yang, XS. Firefly algorithm, stochastic test functions and design optimization. *Journal Bio-Inspired Computation* 2010; 2(2): 78–84.
- [21] Kayarvizhy, N, Kanmani, S, Uthariaraj R. ANN models optimized using swarm intelligence algorithms. *WSEAS Transactions on Computers* 2014; 13: 501-519.
- [22] URL1 < <https://www.google.com/maps/@34.3808769,50.7895208,2502254m/data=!3m1!1e3?hl=en> >
- [23] Angstrom, A. Solar and terrestrial radiation. *Quarterly Journal of the Royal Meteorological Society* 1924; 50: 121-125.
- [24] Prescott, JA. Evaporation from water surface in relation to solar radiation. *Transactions of the Royal Society of South Australia*. 1940; 64: 114–125.
- [25] Bristow, KL, Campbell, GS. On the relationship between incoming solar radiation and daily maximum and minimum temperature. *Agricultural and Forest Meteorology*. 1984; 31: 159-166.
- [26] Elagib, N, Mansell, MG. New approaches for estimating global solar radiation across Sudan. *Energy Conversion and Management* 2000; 41: 419-434.
- [27] Chen, RS, Ersi, K, Yang, JP, Lu, SH, Zhao, WZ. Validation of five global radiation models with measured daily data in China. *Energy Conversion and Management* 2004; 45: 1759-1769.
- [28] Samadianfard, S, Sattari, MT, Kisi, O, Kazemi, H. Determining flow friction factor in irrigation pipes using data mining and artificial intelligence approaches. *Applied Artificial Intelligence* 2014; 28: 793-813.
- [29] Kurup, PU, Dudani, NK. Neural networks for profiling stress history of clays from PCPT data. *Journal of Geotechnical and Geoenvironmental Engineering* 2014; 128(7): 569-579.
- [30] Pal, M. Support vector machines-based modelling of seismic liquefaction potential. *International Journal for Numerical and Analytical Methods in Geomechanics* 2006; 30(10): 983-996.
- [31] Samadianfard, S, Delirhasannia, R, Kisi, O, Agirre-Basurko, E. Comparative analysis of ozone level prediction models using gene expression programming and multiple linear regression. *GEOFIZIKA* 2013, 30, 43-74.