



# Improving generalisation capability of artificial intelligence-based solar radiation estimator models using a bio-inspired optimisation algorithm and multi-model approach

Roozbeh Moazenzadeh<sup>1</sup> · Babak Mohammadi<sup>2</sup> · Zheng Duan<sup>2</sup> · Mahdi Delghandi<sup>1</sup>

Received: 14 July 2021 / Accepted: 25 November 2021 / Published online: 4 January 2022  
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

## Abstract

One way of reducing environmental pollution is to reduce our dependence on fossil fuels by replacing them with solar radiation (Rs), which is one of the main sources of clean and renewable energy. In this study, daily Rs values at seven meteorological stations in Iran (Ahvaz, Isfahan, Kermanshah, Mashhad, Bandar Abbas, Kerman and Tabriz) over 2010–2019 were estimated using empirical models, support vector machine (SVM), SVM coupled with cuckoo search algorithm (SVM-CSA) and multi-model approach in the form of two structures. In structure 1, data from each station were divided into training and testing sets. In structure 2, data from the former four stations were used for model training, and those from the latter three stations were used to test the models. The results showed that using meteorological parameters improved estimation accuracy compared with the use of geographical parameters for both SVM and SVM-CSA models. Coupling the CSA to SVM did improve the accuracy of radiation estimates, reducing RMSE by up to 38% (Kermanshah station) and 36% (Tabriz station) for the first structure and about 42.4% (Tabriz station) for the second. Performance analysis of the models over three intervals including, the first, middle and last third of measured radiation values at each station showed that for both structures (except at Tabriz station), the best model performance in under- and over-estimation sets of radiation values was obtained, respectively, in the first third interval (first structure, Mashhad station,  $RMSE = 28.39 \text{ J.cm}^{-2}.\text{day}^{-1}$ ) and the last third interval (first structure, Bandar Abbas station,  $RMSE = 12.23 \text{ J.cm}^{-2}.\text{day}^{-1}$ ). Determining the effects of climate change on Rs estimation and using remotely sensed data as inputs of the models could be considered as future works.

**Keywords** Cuckoo search algorithm · Generalisation capability · Meteorological parameters · Renewable energy · Support vector machine

Communicated by Philippe Garrigues.

## Highlights

- Estimation of daily solar radiation in different climates of Iran.
  - Performance evaluation of empirical models, SVM, SVM-CSA and multi-model approach.
  - An attempt was made to examine the generalisation capability of the models.
  - Meteorological variables were more effective compared to geographical variables.
  - SVM-CSA and multi-model approach exhibited the best results in most stations.
- Responsible editor: Philippe Garrigues

✉ Roozbeh Moazenzadeh  
romo\_sci@shahroodut.ac.ir

<sup>1</sup> Department of Water Engineering, Faculty of Agriculture, Shahrood University of Technology, Shahrood, Iran

<sup>2</sup> Department of Physical Geography and Ecosystem Science, Lund University, Sölvegatan 12, SE-223 62 Lund, Sweden

## Introduction

Rapid economic growth in developing countries has highlighted the dependence on energy sources. Meanwhile, burning fossil fuels can have adverse environmental consequences, leading to air pollution and contamination of soil and water resources. It also leads to increased carbon dioxide emissions and eventually to global warming, which is particularly important in developing countries, including Iran. Rising sea levels, melting glaciers, sharper droughts and increasing heat waves, powerful storms and floods, changing ecosystems, growing and spreading agricultural pest populations and diseases and reduced food security are some of the most important sequels of climate change, accelerated by the use of fossil fuels. These threats have united governments towards taking management decisions regarding reduction and control of fossil fuel consumption and carbon dioxide

emissions, which are enforced through the annual Climate Change Conference.

Rs is one of the main sources of clean, renewable and accessible energy and can be an alternative to fossil fuels. Particularly, there is tremendous potential in such a country as Iran, with an arid to semi-arid climate and more than 13 h of sunshine per day in some areas. Rs is the primary driving force in agriculture and impacts many key processes such as photosynthesis, evapotranspiration and irrigation scheduling. It is also very useful from an industrial standpoint, playing a role in the design of solar panels and urban facilities and in photovoltaic power generation.

Rs can be directly measured at meteorological stations by pyranometers. In many developing countries, however, not all meteorological stations are equipped with these instruments, and incidentally they are costly to maintain. Therefore, either there are no recorded data for this parameter or the available data are not wholly credible. In recent years, accordingly, researchers have tried to estimate this parameter by various methods. Empirical models and artificial intelligence (AI) models are some of the main tools for Rs estimation.

### Empirical models

Empirical models can be divided into four categories: temperature-based (i.e. Hargreaves-Samani (1982); Annandale et al. (2002)), sunshine-based (Angstrom-Prescott (1940); Feng et al. (2018a, 2018b)), day of the year-based (Quej et al. (2017); Zang et al. (2018)) and hybrid models (Wu et al. (2007); Jahani et al. (2017)).

Adaramola (2012) investigated the performance of seven empirical models in estimating monthly Rs in Nigeria and showed that Angstrom-Prescott model exhibited the best performance (RMSE of  $0.257 \text{ kWh.m}^{-2}.\text{day}^{-1}$ ). Quej et al. (2017) studied the performance of five empirical day of the year-based models (four existing, and one proposed model) and reported RMSEs ranging between  $0.975 \text{ MJ.m}^{-2}.\text{day}^{-1}$  (proposed model) and  $2.197 \text{ MJ.m}^{-2}.\text{day}^{-1}$  (sine wave model). Jamil and Akhtar (2017) estimated monthly mean diffuse Rs using 16 proposed models categorised into two groups in a humid subtropical climatic region of India. RMSEs for the first group ranged from 1.29 to  $1.47 \text{ MJ.m}^{-2}.\text{day}^{-1}$ , and those of the second group from 1.29 to  $1.47 \text{ MJ.m}^{-2}.\text{day}^{-1}$ . Performance of two proposed models based on sunshine duration and relative humidity and 14 existing models in estimating daily Rs in Turkey was studied by Yildirim et al. (2018). Qin et al. (2018) used MODIS atmospheric and land products as well as daily meteorological data recorded at 837 meteorological stations in China as inputs of Yang's hybrid model (YHM), an efficient physically based model (EPP), an hourly solar radiation model (HSRM) and a neural network model (ANNM) for estimation of solar

radiation, and reported YHM as providing better estimates compared to EPP, ANNM and HSRM, with mean daily values of 2.414, 2.535, 2.855 and  $3.645 \text{ MJ.m}^{-2}.\text{day}^{-1}$ , respectively. Performance of 97 available models in literature and 5 newly established models for estimating diffuse Rs at 17 stations over China was evaluated by Wang et al. (2019). Their results showed that the proposed model with clearness index and relative sunshine duration as inputs, produced the highest accuracy. The results showed superiority of proposed models (with RMSEs of  $0.947 \text{ MJ.m}^{-2}$  at Adana station,  $1.086 \text{ MJ.m}^{-2}$  at Göksun station,  $1.074 \text{ MJ.m}^{-2}$  at Tarsus station) over existing models. Mohammadi and Moazenzadeh (2021) evaluated the performance of existing and proposed empirical models in estimating daily Rs at 13 weather stations of Peru. According to the RMSEs, the worst and best results are achieved at San Martin station (RMSE =  $509 \text{ J.cm}^{-2}.\text{day}^{-1}$ ) and Tacna station (RMSE =  $223 \text{ J.cm}^{-2}.\text{day}^{-1}$ ), respectively.

### AI-based models

Cao and Cao (2006) improved the accuracy of daily Rs estimates by combining wavelet analysis with a back propagation training algorithm, with the error rate reduced from 2.82 to approximately  $0.72 \text{ MJ.m}^{-2}.\text{day}^{-1}$ . Various combinations of day of the year, temperature and relative humidity were used by Rehman and Mohandes (2008) as input variables of ANN models for estimation Rs in Saudi Arabia, with MAPEs of 4.49 and 11.8 for different scenarios. Using main meteorological parameters, evaporation and soil temperature, Asl et al. (2011) estimated daily Rs in Dezful, Iran, by multi-layer perceptron (MLP) neural networks with an MAPE of 6.08. Wang et al. (2016a, 2016b) evaluated the performance of three neural network models including generalised regression neural network (GRNN), MLP and radial basis neural network (RBNN) in estimating daily Rs. The results of their study showed that MLP outperformed the other models, with RMSEs ranging from 1.94 to  $3.27 \text{ MJ.m}^{-2}.\text{day}^{-1}$ .

Three types of AI-based models including ANFIS-GP, ANFIS-SC and M5Tree were evaluated for estimating Rs at 21 stations over China by Wang et al. (2016a, 2016b) and a general tendency to under-estimation of high radiation values in some stations is reported. An experimental model named "Iqbal" and four AI-based models, including extreme learning machine (ELM), back-propagation neural networks optimised by genetic algorithm (GANN), random forests (RF) and GRNN were compared by Feng et al. (2017) in terms of estimating diffuse Rs. The results showed that GANN at Beijing station (RRMSE = 17.1%) and Zhengzhou station (RRMSE = 13.4%) had the best performance. Halabi et al. (2018) assessed the performance of four models, including ANFIS and its hybrids with three optimisation

algorithms — particle swarm optimisation (PSO), genetic algorithm (GA) and differential evolution (DE) — in estimating monthly  $R_s$  in Malaysia, concluding that ANFIS-PSO (RMSE=0.3065) has outperformed other models. Performance of back propagation (BP) and radial basis function (RBF) models and a new hybrid model, “ensemble empirical mode decomposition and self-organising map-back propagation hybrid neural networks” (EEMD-SOM-BP) was studied by Lan et al. (2018) in estimating seasonal  $R_s$  in China, with the proposed model leading to the best results in spring, summer, autumn and winter with RMSEs of 137.85, 123.58, 72.84 and 135.42  $W.m^{-2}$ , respectively. Among different models applied for modeling daily photosynthetically active radiation (Feng et al. (2018a, 2018b)), genetic model outperformed with lowest RMSE ( $0.5 MJ.m^{-2}.day^{-1}$ ). Kuhe et al. (2019) evaluated the performance of three ANN models including feed-forward back-propagation neural network (FFNN), radial basis function network (RBFN) and GRNN in estimating monthly  $R_s$  in Makurdi, Nigeria. All models had acceptable accuracies, with  $R^2=0.998$  and  $MSE=0.0142 (MJ. m^{-2}.day^{-1})$  on average. Performance of SVM-FA, Copula-based nonlinear quantile regression (CNQR) and empirical models in estimating daily diffuse  $R_s$  at four stations in China was studied by Liu et al. (2020). Their findings showed that CNQR and SVM-FA were much better than empirical models. SVM-FA results were slightly better than those of CNQR, resulting in a 0.67% decrease in MABE.

Given the vast surface area of Iran and its arid and semi-arid climate, and since many parts of the country receive long hours of sunshine, the present study aimed at (i) evaluating the effect of input parameters (geographical and meteorological) on estimating  $R_s$ , (ii) proposing a novel strategy based on multi-model approach by combining either support vector machine (SVM) or SVM boosted by the cuckoo search algorithm (SVM-CSA) with multi-layer perceptron and (iii) discussing the generalisation capability of proposed models using local and external analysis. Considering a data-fusion approach via multi-layer perceptron model for combining the advantages of empirical and AI models for estimating  $R_s$  and discussing the potential of proposed models in three different intervals are the main novelty aspects of the present study. Meteorological data and measured  $R_s$  values recorded at seven weather stations in Iran over 2010–2019 were used for this purpose.

## Materials and methods

### Study area

With an area of 1,648,000  $km^2$ , Iran extends between longitudes  $44^\circ$  and  $63^\circ$  east and between latitudes  $25^\circ$  and  $40^\circ$

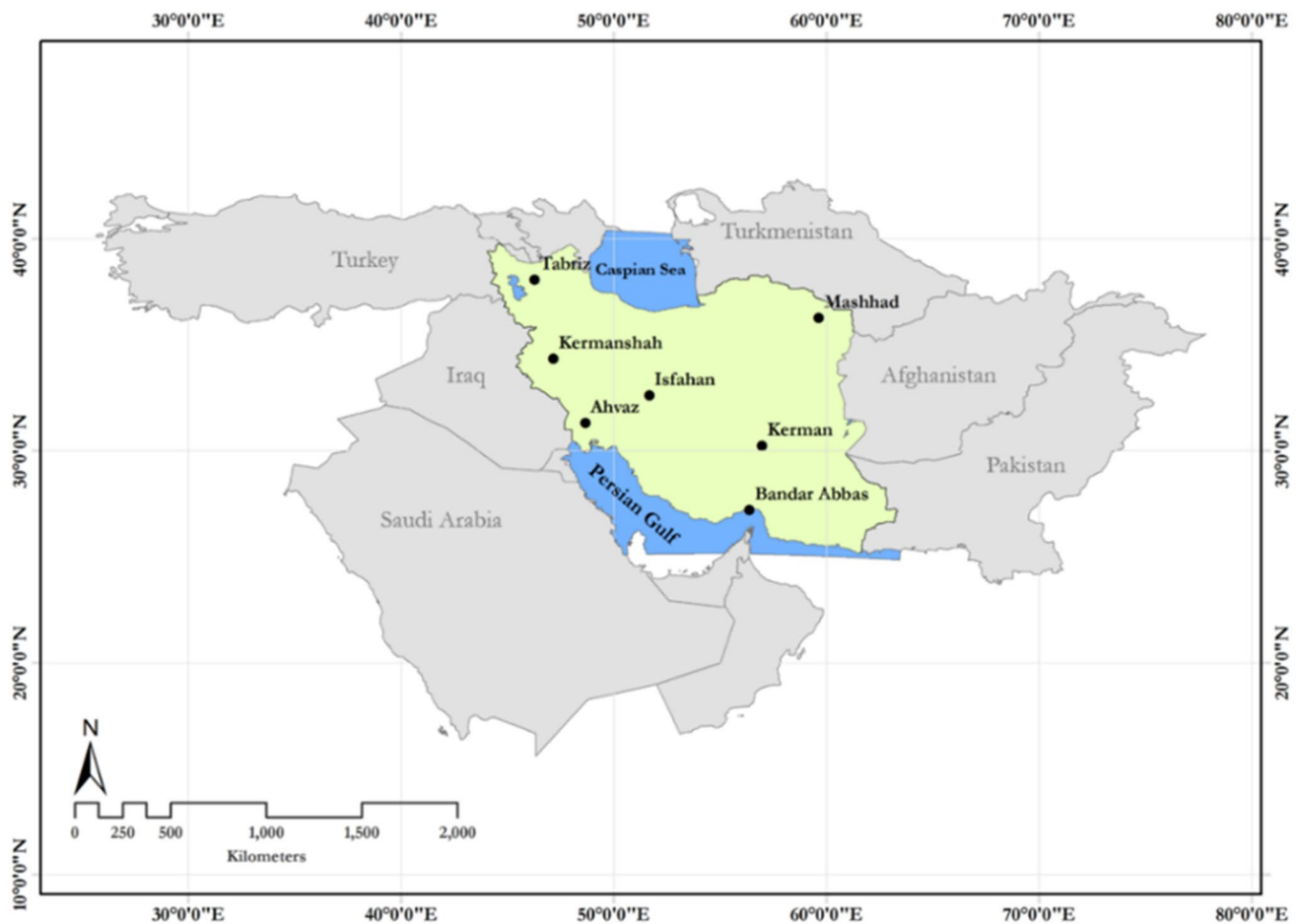
north (Fig. 1). The country's climate is generally arid or semi-arid (Aghelpour et al., 2019), with maximum sunshine hours of about 14 h per day in some regions, which is a high figure. Iran characterised by hot summers and cold winters and different type of climates, from arid (most part of Iran) to humid region (north of Iran). The average of annual precipitation is about 250 mm. Iran is a developing country and inevitably utilises various energy sources. However, fossil fuel consumption can cause irreparable damage to the environment in the long term. Studies on clean and renewable sources of energy such as  $R_s$ , and its proper estimation, can therefore be of great help in reducing the country's dependence on fossil fuels. In the present study, geographical and meteorological parameters recorded at seven weather stations in Iran were used for estimating daily  $R_s$ . Statistical indices of the meteorological parameters are given in Table 1.

### $R_s$ estimation

In the present study, two different structures were used for  $R_s$  estimation, flowcharts of which are shown in Fig. 2. In structure 1 (local analysis), data from each of the seven stations were divided into two sets (training and testing), and performance of each method at each station was evaluated separately. In structure 2 (external analysis), data from four stations (Ahvaz, Isfahan, Kermanshah, and Mashhad) were used for the training, and the data from three other stations (Bandar Abbas, Kerman, and Tabriz) were used for testing. In fact, in structure 2 an attempt was made to examine the generalisation capability of the models through rendering the data of the two sets (training and testing set) dissimilar, from two perspectives, including the distance between stations and climatic differences between stations. In this study, inputs of  $R_s$  simulator models were divided into two categories, geographical and meteorological inputs, so that in addition to evaluating model performance, the effect of input type could also be examined. A summary of model input data can be seen in Table 2. The quality of meteorological data (inputs) and solar radiation (output) was checked before applying in the modeling processes. For this aim, outliers and missing data were found and time series without noise data were considered for modeling.

### Empirical models

In this study, performance of empirical models based on temperature (6 models) or sunshine hours (2 models) as well as 3 hybrid models in estimating  $R_s$  over 2010–2019 was studied based on meteorological data recorded at seven weather stations (Table 3).



**Fig. 1** Iran (the study area), neighbouring countries and locations of weather stations

### Support vector machine (SVM)

SVM was introduced by Vapnik (1995) as a supervised learning algorithm and a statistical learning technique which can be used for solving classification, regression and forecasting problems. SVM employs kernel functions to transform the data from input space into a higher dimensional feature space in order to simplify classification problems. In addition, the “ $\epsilon$ ” insensitive loss function enables SVM to solve nonlinear regression problems. For non-separable classes, where an exact separating hyperplane cannot be found, the input space is mapped to a higher-dimensional feature space using nonlinear functions called feature functions ( $\emptyset$ ) and kernels, thus enabling SVM to form nonlinear boundaries and model highly complex problems (Raghavendra and Deka, 2014). SVM has been used for simulating evaporation (Goyal et al., 2014; Moazenzadeh et al., 2018), reference evapotranspiration (Kisi and Cimen, 2009; Mohammadi and Mehdizadeh, 2020), estimation of lake water level fluctuations (Cimen and Kisi, 2009), streamflow simulation

(Mohammadi et al., 2021), drought forecasting (Deo et al., 2017), soil temperature (Moazenzadeh and Mohammadi, 2019), velocity prediction (Ebtehaj and Bonakdari, 2016), prediction of discharge coefficient and depth around bridge piers (Sharafi et al., 2016; Azimi et al., 2019) and soil saturated hydraulic conductivity (Kashani et al., 2020) estimation.

### Cuckoo search algorithm (CSA)

This algorithm is based on the following assumptions:

- Each cuckoo randomly selects a nest and lays a single egg in it.
- The nests with the highest quality of eggs (that is, solutions to the problem) are carried over to the next generation.
- The number of the available nests is fixed. Original owners of the nests are able to recognise cuckoo eggs with a probability  $P_a \in [0, 1]$ .

**Table 1** Geographical locations of the stations and statistical indices of meteorological parameters used in this study

Station	Latitude	Longitude	Altitude	Parameter	Min	Max	Mean	SD	CV
Ahvaz	31° 20' N	48° 40' E	22.5 m	Tmin (°C)	-0.6	35.6	19.46	8.51	0.44
				Tmax (°C)	6.6	52.2	34	10.9	0.32
				RH (%)	7.6	95.3	38.79	19.71	0.51
				n/N	0.02	1.0	0.75	0.21	0.28
				Rs (J.cm <sup>-2</sup> .day <sup>-1</sup> )	220	3089	1837.7	612.6	0.33
Bandar Abbas	27° 13' N	56° 22' E	9.8 m	Tmin	3.9	33.2	21.9	6.9	0.31
				Tmax	16.2	46.5	32.56	5.83	0.18
				RH	7.6	93.4	63.39	12.18	0.19
				n/N	0.007	0.99	0.74	0.16	0.22
				Rs (J.cm <sup>-2</sup> .day <sup>-1</sup> )	646	2938	1895.2	459.5	0.24
Isfahan	32° 37' N	51° 40' E	1550.4 m	Tmin	-14.4	28.8	9.87	8.93	0.9
				Tmax	-3.0	43	25.2	10.0	0.39
				RH	8.5	96.6	32.49	16.64	0.51
				n/N	0.04	0.99	0.8	0.17	0.22
				Rs (J.cm <sup>-2</sup> .day <sup>-1</sup> )	712	3184	2077.8	639.7	0.31
Kerman	30° 15' N	56° 58' E	1753.8 m	Tmin	-15.2	26.6	8.27	8.18	0.99
				Tmax	-1.2	41.0	26.45	8.76	0.33
				RH	8.9	89.1	29.6	15.6	0.52
				n/N	0.01	1.0	0.79	0.19	0.25
				Rs (J.cm <sup>-2</sup> .day <sup>-1</sup> )	553	3424	2248.3	639.6	0.28
Kermanshah	34° 21' N	47° 09' E	1318.6 m	Tmin	-18.3	27.6	7.64	8.0	1.05
				Tmax	-4.0	43.1	25.04	11.07	0.44
				RH	9.8	95.6	41.24	21.5	0.52
				n/N	0.007	0.99	0.69	0.23	0.34
				Rs (J.cm <sup>-2</sup> .day <sup>-1</sup> )	461	3221	1829.5	673.2	0.36
Mashhad	36° 16' N	59° 38' E	999.2 m	Tmin	-18.2	27.4	10.33	8.33	0.8
				Tmax	-8.2	41.6	24.18	9.82	0.4
				RH	10.5	96.9	45.24	19.68	0.43
				n/N	0.01	0.97	0.74	0.2	0.27
				Rs (J.cm <sup>-2</sup> .day <sup>-1</sup> )	483	3523	2088	692.3	0.33
Tabriz	38° 05' N	46° 17' E	1361 m	Tmin	-16.8	27.6	8.15	9.43	1.15
				Tmax	-7.9	41	19.76	11.02	0.55
				RH	13.9	95.1	48.89	16.55	0.34
				n/N	0.007	0.99	0.67	0.25	0.36
				Rs (J.cm <sup>-2</sup> .day <sup>-1</sup> )	360	3088	1546.9	683	0.44

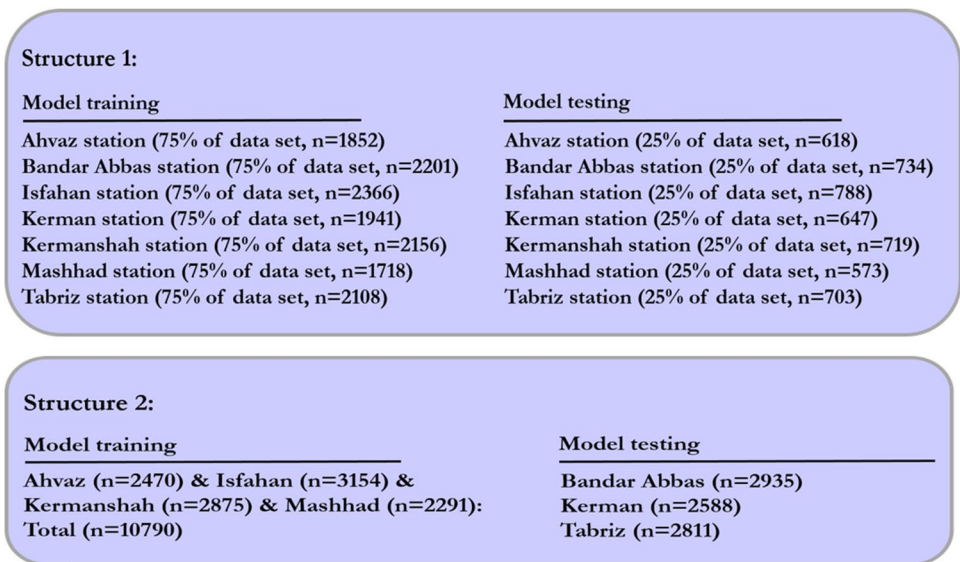
Upon learning that an alien egg is laid in its nest, the host bird will either discard that egg or desert the nest and build a new one. The last assumption can be approximated by a fraction Pa of the n nests being replaced by new nests (with new random solutions at new locations). For maximisation problems, the fitness of a solution can be proportional to the objective function. For details and background information about CSA, see Yang and Deb (2009) and Gandomi et al. (2013). In the present study, CSA was used to determine the best parameters and weights for SVM in the form of hybrid models. Previous studies have confirmed that such coupled CSA optimisation approach via AI models can produce a capable hybrid model (Liu and Fu 2014; He et al., 2018; Puspaningrum et al., 2020). Figure 3 shows the modelling

flowchart for the base SVM and SVM coupled with CSA (SVM-CSA) as used in this study.

**Multi-model approach**

A new approach based on the multi-model concept was developed, tested and compared with the empirical and AI models for estimation of daily Rs. In this approach, outputs of empirical and SVM/SVM-CSA models under the best scenarios were used as inputs to the multi-model, structure of which was based on MLP. Multi-model outputs were named SVM-MLP and SVM-CSA-MLP. Figure 4 outlines the multi-model approach employed in our study.

**Fig. 2** Two structures used for estimating daily Rs. Structure 1: dataset (*n*) from each station is divided into training and testing sets. Structure 2: datasets from four stations are used for training of the models, and the results are validated separately against data from each of the other three stations



**Table 2** Rs estimator models and their inputs categorised as geographical or meteorological parameters

Model inputs	Parameter	SVM	SVM-CSA	Multiple model
Geographical	DOY	SVM1G	SVM1G-CSA	(SVMG,SVMM,empirical) <sub>best</sub> -MLP; (SVMG-CSA,SVMM-CSA,empirical) <sub>best</sub> -MLP
	DOY,DOM	SVM2G	SVM2G-CSA	
	DOY,DOM,NOM	SVM3G	SVM3G-CSA	
	DOY,DOM,NOM,ELE	SVM4G	SVM4G-CSA	
	DOY,DOM,NOM,ELE,LON	SVM5G	SVM5G-CSA	
	DOY,DOM,NOM,ELE,LON,LAT	SVM6G	SVM6G-CSA	
Meteorological	Tmax	SVM1M	SVM1M-CSA	
	Tmax,Tmin	SVM2M	SVM2M-CSA	
	Tmax,Tmin,Tdew	SVM3M	SVM3M-CSA	
	Tmax,Tmin,Tdew,Twet	SVM4M	SVM4M-CSA	
	Tmax,Tmin,Tdew,Twet,Tdry	SVM5M	SVM5M-CSA	
	Tmax,Tmin,Tdew,Twet,Tdry,RH	SVM6M	SVM6M-CSA	
	Tmax,Tmin,Tdew,Twet,Tdry,RH,n	SVM7M	SVM7M-CSA	
	Tmax,Tmin,Tdew,Twet,Tdry,RH,n,N	SVM8M	SVM8M-CSA	
	Tmax,Tmin,Tdew,Twet,Tdry,RH,n,N,CC	SVM9M	SVM9M-CSA	
	Tmax,Tmin,Tdew,Twet,Tdry,RH,n,N,CC,Ra	SVM10M	SVM10M-CSA	

DOY day of the year; DOM day of the month; NOM number of months; ELE station’s elevation (m); LON station’s longitude; LAT station’s latitude; Tmax, Tmin, Tdew, Twet, and Tdry represent maximum, minimum, dew point, wet, and dry temperature (°C), respectively; RH relative humidity; n actual sunshine hours (hr); N maximum possible sunshine hours (hr); CC cloud cover; Ra extraterrestrial radiation (J.cm<sup>-2</sup>.day<sup>-1</sup>)

**Evaluation indices**

Performance of the developed Rs estimator models was evaluated using three statistical indices including RMSE, mean absolute percentage error (MAPE) and relative root mean square error (RRMSE), calculated using Eqs. 1, 2, 3, respectively. Ertekin and Yaldiz (2000) proposed the following categories for rating models according to accuracy: A model is excellent if its RRMSE is below 10%, good if

10% < RRMSE < 20%, fair if 20% < RRMSE < 30%, and poor if the RRMSE is higher than 30%.

$$RMSE = \left[ \sum_{i=1}^n (Rs_{(i,obs)} - Rs_{(i,est)})^2 \div n \right]^{0.5} \tag{1}$$

$$MAPE = \left[ \left( \sum_{i=1}^n \left| (Rs_{(i,est)} - Rs_{(i,obs)}) \div Rs_{(i,obs)} \right| \right) \div n \right] \tag{2}$$

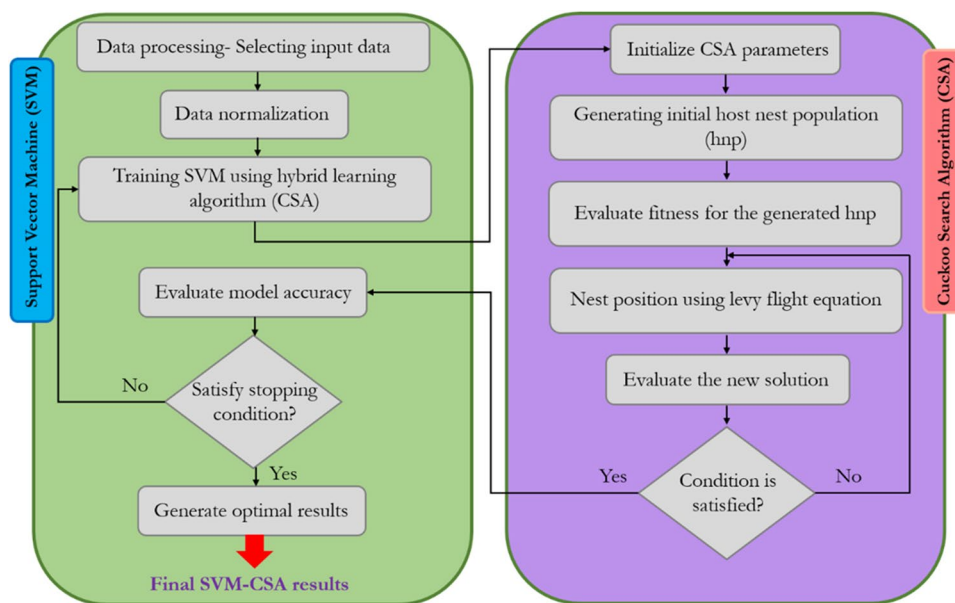
**Table 3** Empirical models used for estimating  $R_s$

Model type	Model's name	Equation
Temperature-based	Hargreaves-Samani	$R_s = R_a a (\Delta T)^{0.5}$
	Samani	$R_s = R_a [c_1 \Delta T^{2.5} + c_2 \Delta T^{1.5} + c_3 \Delta T^{0.5}]$
	Annandale et al	$R_s = R_a d (1 + 2.7 * 10^{-6} Z) \Delta T^{0.5}$
	Chen et al	$R_s = R_a (e_1 + e_2 \ln(\Delta T))$
	Jahani et al. 1	$R_s = R_a (\beta_1 + \beta_2 \Delta T + \beta_3 \Delta T^2 + \beta_4 \Delta T^3)$
	Jahani et al. 2	$R_s = R_a (\gamma_1 + \gamma_2 \Delta T^{0.5} + \gamma_3 \Delta T^{1.5} + \gamma_4 \Delta T^{2.5})$
Sunshine-based	Angstrom-Prescott	$R_s = R_a (f_1 + f_2 (n/N))$
	Bahel	$R_s = R_a (g_1 + g_2 (n/N) + g_3 (n/N)^2 + g_4 (n/N)^3)$
Hybrid	Jahani et al. 3	$R_s = R_a (\tau_1 + \tau_2 (n/N) + \tau_3 \Delta T^{0.5} + \tau_4 \Delta T_{dw})$
	Abdalla	$R_s = R_a (b_1 + b_2 (n/N) + b_3 (RH) + b_4 (T_{mean}))$
	Garj-Garj	$R_s = R_a (h_1 + h_2 (n/N) + h_3 (W_{at}))$

$$W_{at} = \left[ RH \left( 4.7923 + 0.3647 T_a + 0.0055 (T_a)^3 \right) \right]$$

where  $\Delta T$  is the difference between minimum and maximum temperatures ( $^{\circ}C$ ),  $Z$  is station's altitude (m),  $\Delta T_{dw}$  is the difference between dry and wet bulb air temperatures ( $^{\circ}C$ ),  $W_{at}$  is the amount of water vapour per unit volume of air, and  $T_a$  is air temperature ( $^{\circ}C$ ).

**Fig. 3** Flowchart of the hybrid model (SVM-CSA) used in the present study



$$RRMSE = \left[ RMSE \div \left( \frac{1}{n} \sum_{i=1}^n R_{s(i,obs)} \right) \right] \tag{3}$$

where  $R_{s_{i,obs}}$  and  $R_{s_{i,est}}$  are observed and estimated solar radiation values and  $n$  refers to the number of data points.

## Results

### Local performance of the models (structure 1)

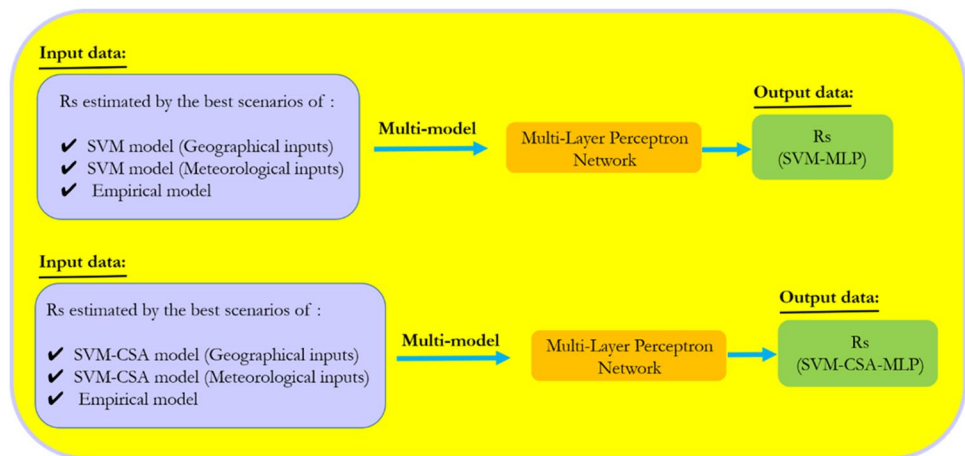
For this structure, we divided the data from each station into two parts (training and testing). Results of the best empirical

model, SVM, SVM-CSA and the multi-model approach under the best scenarios at each station are plotted in Fig. 5.

### Ahvaz station

The use of meteorological data as model inputs greatly improved the accuracy of radiation estimates, to the extent that RMSEs of the SVM10M and SVM10M-CSA models (best scenarios with meteorological inputs) were approximately 49% and 60% lower than those of SVM1G and SVM1G-CSA (best scenarios with geographical inputs), respectively. Coupling the CSA to the SVM model was effective and reduced RMSE from 153.21 to 114.22. Garj-Garj

**Fig. 4** Flowchart of the multi-model strategy used for estimating  $R_s$



model exhibited the best performance among the empirical models, with  $RMSE = 166.97 \text{ J.cm}^{-2}.\text{day}^{-1}$ . Examination of the results showed that the RMSE of the multi-model approach has been 4 and 2% lower compared to the base SVM and SVM-CSA, respectively. However, the difference is insignificant and the use of the multi-model approach is not recommended.

#### Bandar Abbas station

The results, under the best scenarios, of SVM, SVM-CSA, the best empirical model and the multi-model approach are represented in Fig. 5, indicating that the use of meteorological rather than geographical parameters as inputs of SVM and SVM-CSA has reduced RMSEs by about 21 and 28 per cent, respectively (Table 4). Among the empirical models, Garj-Garj ( $RMSE = 220.24 \text{ J.cm}^{-2}.\text{day}^{-1}$ ) outperformed SVM, but its performance was inferior to SVM-CSA with meteorological parameters. Also, the SVM8M-CSA scenario reduced estimation error by 31% compared to SVM8M, which highlights the importance of coupling the CSA to the base SVM model.

#### Isfahan station

Using meteorological rather than geographical parameters has greatly improved the performance of SVM and SVM-CSA, reducing their estimation errors by about 56 and 47%, respectively. Comparing the outputs of SVM ( $RMSE = 125.31 \text{ J.cm}^{-2}.\text{day}^{-1}$ ), SVM-CSA ( $RMSE = 125.86 \text{ J.cm}^{-2}.\text{day}^{-1}$ ) and Garj-Garj empirical model ( $RMSE = 125.87 \text{ J.cm}^{-2}.\text{day}^{-1}$ ) shows that application of the CSA to the base SVM model has failed to improve the results, and the empirical Garj-Garj model is recommended according to its ease of use compared to the more complex AI models. Performance of 42 different SVM structures in estimating daily  $R_s$  in Ghardaia, Algeria, was

studied by Belaid and Mellit (2016). Daily RMSE values for the four selected structures of SVM ranged between 2.777 and 2.807  $\text{MJ.m}^{-2}$ , whereas for the MLP model, the range of RMSEs increased to 2.788–3.047.

#### Kerman station

At this station, Abdalla's model outperformed all other empirical models, with an error rate similar to that of SVM. However, application of the CSA to the base SVM model reduced estimation error by 15%. Examination of the results also shows that application of the multi-model approach to outputs of SVM and SVM-CSA has only slightly improved the results, reducing RMSEs by about 7 and 2%, respectively.

#### Kermanshah station

Coupling the CSA to SVM base model has improved radiation estimates, with the SVM9-M-CSA scenario reducing RMSE by about 30% and 38% compared to the best empirical model (third model of Jahani et al.) and the best SVM scenario (SVM9-M), respectively. Application of the multi-model approach to SVM outputs reduced the RMSE by 9%. Kim et al. (2018) evaluated the performance of single soft computing models, including MLP, SVM, ANFIS and MARS (multivariate adaptive regression spline) in estimating daily  $R_s$ . With various inputs, the best performance was obtained for SVM ( $RMSE = 4.399$ ) and MARS ( $RMSE = 4.207$ ) at Big Bend and Incheon stations, respectively.

#### Mashhad station

The results showed that replacing geographical parameters with meteorological parameters as inputs of SVM and SVM-CSA reduced RMSEs by 45 and 25%, respectively. Both the base SVM and SVM coupled with CSA effectively estimated



Rs, with RMSEs reduced by about 11% and 37%, respectively, compared to the best empirical model (Abdalla's model).

#### Tabriz station

The results indicate that although the use of SVM model has failed to improve radiation estimation in comparison with the best empirical model (the first model of Jahani et al.,  $RMSE = 366.88 \text{ J.cm}^{-2}.\text{day}^{-1}$ ), SVM coupled with CSA has reduced the RMSE by 14% compared to the best empirical model. Application of the multi-model approach to outputs of the best scenarios of SVM and SVM-CSA was effective, reducing error rates by 7% and 16%, respectively.

#### External performance of the models (structure 2)

We defined this structure in order to examine the possibility of generalising the performance of radiation estimator models at stations which had not played a role in model training process. For this purpose, radiation data from four stations (Ahvaz, Isfahan, Kermanshah and Mashhad) were used to train the models, and performance of the models was tested on radiation data from the other three stations (Bandar Abbas, Kerman and Tabriz). Figure 6 depicts the results of the best empirical model, SVM, SVM-CSA and the multi-model approach under the best scenarios.

#### Bandar Abbas station

According to the results, although SVM has not led to lower estimation errors in comparison with the best empirical model (Abdalla model), application of CSA has reduced the error rate by 10%. The results also show that application of the multi-model approach to outputs of SVM and SVM-CSA has failed to improve radiation estimates.

#### Kerman station

At this station, application of meteorological rather than geographical parameters greatly improved Rs estimates, to the extent that RMSEs were reduced by 49 and 48% for SVM and SVM-CSA, respectively. Although the best empirical model (Garj-Garj) slightly outperformed the best scenario of SVM, coupling the CSA to the base SVM reduced RMSEs by 29% and 18% compared to SVM and Garj-Garj models, respectively. Application of the multi-model approach to both SVM and SVM-CSA led to better radiation estimates; but this improvement was negligible in both cases, especially for SVM-CSA.

#### Tabriz station

Application of CSA to the base SVM model has reduced the error rate by 25% compared to the best empirical model, testifying to the importance of coupling the optimisation algorithm to SVM. Unlike other stations, application of the multi-model approach to outputs of both SVM and SVM-CSA was effective, reducing error rates by about 24% (SVM-MLP) and 27% (SVM-CSA-MLP), respectively.

## Discussion

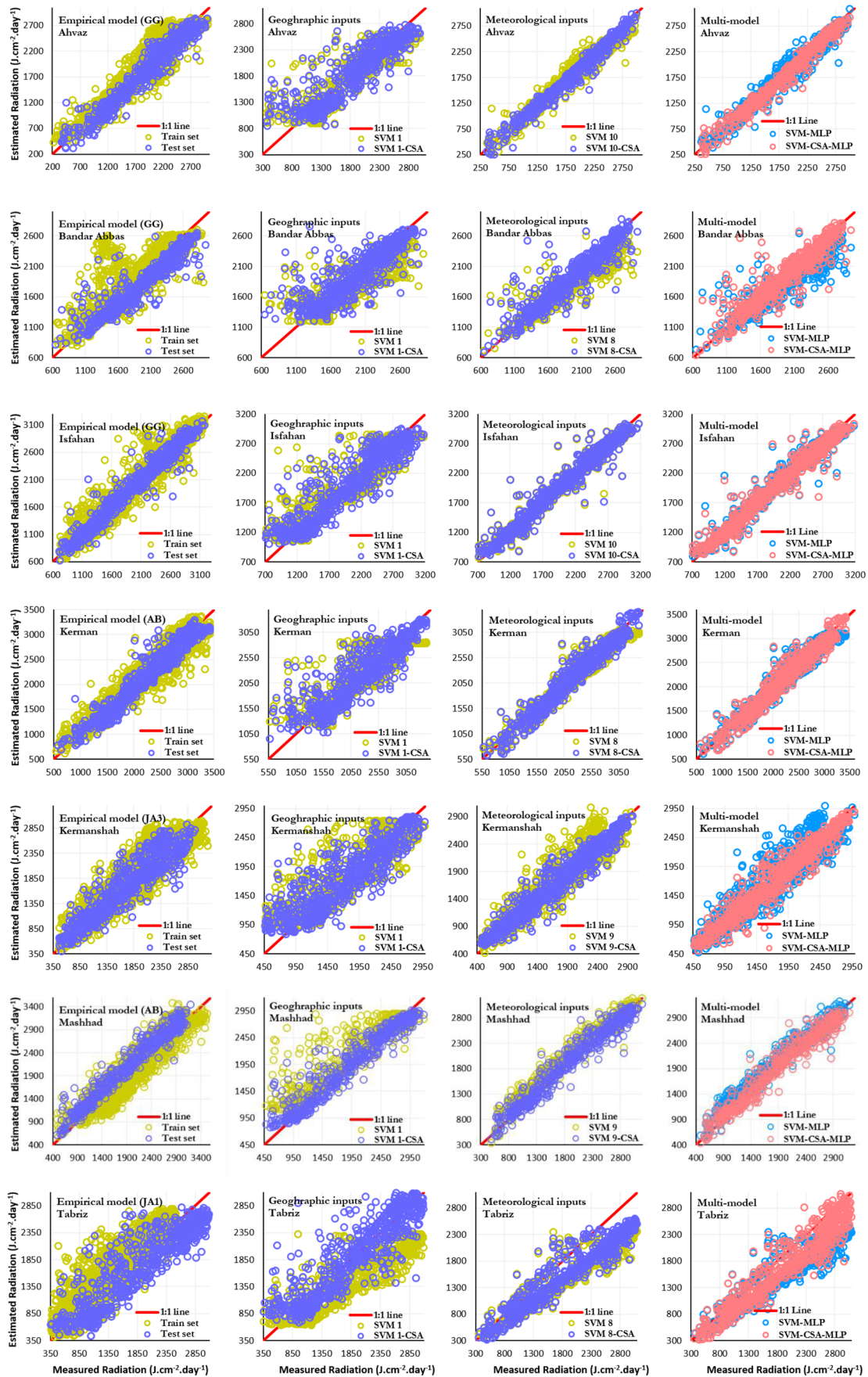
### Local performance of the models (structure 1)

Results of the best scenarios — in terms of under- or over-estimation of radiation amounts — and model accuracies, respectively, over the three discussed intervals (first, middle and last third of measured radiation values) are depicted in Figs. 7 and 8. Statistical indices including RMSE, MAPE and RRMSE are also listed in Table 4.

#### Ahvaz station

Apart from SVM-G and SVM-G-CSA, whose performances were “good” according to their RRMSE indices, performance of all other models was excellent, indicating the reliable performance of Garj-Garj model, superiority of meteorological variables over geographical variables and the importance of coupling the CSA to the base SVM model (Table 4).

Hassan et al. (2017) evaluated the performance of three different machine-learning algorithms as well as a proposed algorithm titled “decision trees” in estimating Rs in Cairo, Egypt. Among day of the year-based models, the proposed decision trees model exhibited the best performance ( $RMSE = 2.0489$ ). Zang et al. (2018) studied the performance of 14 empirical models, five ANFIS-based models and GPR and SVR models, in estimating daily Rs in China. RMSEs varied from 1.39 to 3.065  $\text{MJ.m}^{-2}.\text{day}^{-1}$  for the empirical models, from 1.379 to 2.976 for SVR, from 1.287 to 2.711 for GPR and between 1.203 and 2.721 for ANFIS-based models. Application of the multi-model approach to the SVM-CSA (SVM-CSA-MLP) led to more accurate results in over-estimation set. The results showed that SVM-CSA-MLP has estimated Rs with the lowest error rates in under- and over-estimation sets in the first third interval ( $RMSE = 103.86$ ) and the last third interval ( $RMSE = 41.64$ ), respectively.



◀**Fig. 5** Measured versus estimated radiation values under the best empirical model, the best scenarios of SVM and SVM-CSA (with either geographical or meteorological inputs) and the multi-model approach at each station (local analysis: structure 1)

### Bandar Abbas station

The results presented in Table 4 show that except SVM-M-CSA and SVM-CSA-MLP, which have RRMSEs in the “excellent” category and highlight the role of coupling the optimisation algorithm to the base SVM model and the importance of employing the multi-model approach in radiation estimation, respectively; other models are “good” with RRMSEs ranging between 10 and 20%. Performance of two empirical models as well as eight existing hybrid models and four proposed hybrid models in China was evaluated by Fan et al. (2018). According to their findings, Bahel model was recommended for cases where only sunshine hours data were available. Sanz et al. (2018) investigated the performance of four models including extreme learning machine (ELM), SVR, multiple linear regression (MLR) and multivariate adaptive regression spline (MARS), either alone or coupled with two optimisation algorithms, CRO (coral reefs optimisation) and GGA (grouping genetic algorithm), in Australia. The four base models had error rates ( $\text{MJ.m}^{-2}$ ) varying from 4.224 (ELM) to 4.289 (MARS), whereas coupling GGA and CRO algorithms to them led to error rates ranging from 4.246 (GGA-ELM) to 4.533 (GGA-MLR) and from 4.21 (CRO-ELM) to 4.468 (CRO-SVR), respectively.

As shown in Fig. 8, the best performance of all models in under- and over-estimation sets is obtained in the first third interval (measured radiation values below  $1675 \text{ J.cm}^{-2} \cdot \text{day}^{-1}$ ) and the last third interval (measured radiation values above  $2241 \text{ J.cm}^{-2} \cdot \text{day}^{-1}$ ), respectively. In under-estimation set, SVM-CSA-MLP in the first third ( $\text{RMSE} = 102.84 \text{ J.cm}^{-2} \cdot \text{day}^{-1}$ ) and SVM-G in the last third ( $\text{RMSE} = 375.25 \text{ J.cm}^{-2} \cdot \text{day}^{-1}$ ) had the best and the poorest performance, respectively. The best and poorest models in over-estimation set were 12.23 (last third, the best empirical model) and 412.88 (first third, SVM-G-CSA), respectively. Similar to what happened to most models at this station in the last third interval, Wang et al. (2016a, 2016b) showed in a study on 12 Chinese stations that ANN models under-estimated high radiation amounts at some stations.

### Isfahan station

With the exception of SVM-G-CSA, which has performed better in over-estimation set with an  $\text{RMSE}$  of  $190.97 \text{ J.cm}^{-2} \cdot \text{day}^{-1}$ , accuracy of the other models has been higher in under-estimation set. Similar to the previous two stations, although all models have performed best in under- and over-estimation sets in the first third (measured radiation values

below  $1675 \text{ J.cm}^{-2} \cdot \text{day}^{-1}$ ) and the last third (measured radiation values above  $2415 \text{ J.cm}^{-2} \cdot \text{day}^{-1}$ ), respectively, the difference between estimation errors in the first third (under-estimation set) and the last third (over-estimation set) is much less compared to the previous two stations. In under-estimation set, SVM-M-CSA in the first third interval ( $\text{RMSE} = 96.66 \text{ J.cm}^{-2} \cdot \text{day}^{-1}$ ) and SVM-G in the last third interval ( $\text{RMSE} = 231.8 \text{ J.cm}^{-2} \cdot \text{day}^{-1}$ ) demonstrated the best and the poorest performance, respectively. For the over-estimation set, corresponding values were 76.06 (last third, the SVM-MLP model) and 425.31 (first third, the SVM-G model), respectively.

### Kerman station

According to the results from Isfahan and Kerman stations (Table 4) and in confirmation of high RMSEs for SVM-G and SVM-G-CSA, RRMSEs of these two models are above 10% compared to the other models, placing them in “good” category; and this denotes the subordinate role of geographical parameters in estimating radiation, even in case of using an AI model solely or its coupling with the CSA. This finding is important in that employing the AI model will not necessarily lead to a proper estimation of radiation, and care must be taken when selecting model inputs. Meenal and Slavakumar (2018) studied the performance of 16 empirical models, 16 different structures of the SVM, and 3 structures of the ANN, at four stations in India. According to the results, the lowest RMSEs for empirical models were 0.638, 1.15 and  $0.744 \text{ MJ.m}^{-2} \cdot \text{day}^{-1}$ , for sunshine-based, temperature-based and hybrid models, respectively, and the lowest RMSEs for ANN and SVM were 0.581 and  $0.42 \text{ MJ.m}^{-2} \cdot \text{day}^{-1}$ , respectively. Zou et al. (2017) introduced improved forms of two empirical models (Bristow-Campbell’s and Yang’s hybrid model) and an ANFIS-based model for estimating daily  $R_s$  in China. Their results indicated that ANFIS had the lowest RMSEs and MAEs, ranging from 0.59 to 1.6 and from 0.42 to  $1.21 \text{ MJ.m}^{-2} \cdot \text{day}^{-1}$ , respectively.

The lowest and highest differences in error rates between under- and over-estimation sets were 7.38 and  $164.72 \text{ J.cm}^{-2} \cdot \text{day}^{-1}$  for Abdalla’s empirical model and SVM-G-CSA, respectively (Fig. 8). In under-estimation set, except for SVM-M-CSA and SVM-CSA-MLP which had their lowest error rates in the middle third interval (measured radiation values above 1887 and below  $2590 \text{ J.cm}^{-2} \cdot \text{day}^{-1}$ ), the best performance of the other models occurred in the first third interval (measured radiation values below  $1887 \text{ J.cm}^{-2} \cdot \text{day}^{-1}$ ). In over-estimation set, all models performed noticeably better in the last third compared to the other two intervals. According to the results, the minimum and maximum estimation errors were  $69.44 \text{ J.cm}^{-2} \cdot \text{day}^{-1}$  (last third of over-estimation set, SVM-M model) and  $540.06 \text{ J.cm}^{-2} \cdot \text{day}^{-1}$  (first third of over-estimation set, SVM-G model).

**Table 4** Performance evaluation indices for the best empirical model, SVM, SVM-CSA, and the multi-model approach for structure 1 (local analysis). *G* and *M* denote geographical and meteorological inputs, respectively

Station: Ahvaz			
Models	RMSE (J.cm <sup>-2</sup> .day <sup>-1</sup> )	MAPE	RRMSE (%)
Best empirical model	166.97	0.09	9.05
SVM-G (M)	301.09 (153.21)	0.175 (0.07)	16.32 (8.3)
SVM-G-CSA (M)	287.56 (114.22)	0.16 (0.05)	15.59 (6.18)
SVM-MLP	146.49	0.07	7.94
SVM-CSA-MLP	111.75	0.05	6.06
Station: Bandar Abbas			
Best empirical model	220.24	0.1	11.29
SVM-G (M)	318.12 (250.62)	0.14 (0.11)	16.3 (12.93)
SVM-G-CSA (M)	241.76 (173.42)	0.1 (0.06)	12.39 (8.89)
SVM-MLP	230.84	0.1	11.83
SVM-CSA-MLP	171.08	0.06	8.77
Station: Isfahan			
Best empirical model	125.87	0.05	6.24
SVM-G (M)	282.85 (125.31)	0.13 (0.05)	14.01 (6.21)
SVM-G-CSA (M)	236.29 (125.86)	0.1 (0.05)	11.71 (6.24)
SVM-MLP	125.72	0.05	6.23
SVM-CSA-MLP	126.59	0.05	6.27
Station: Kerman			
Best empirical model	178.35	0.07	8.03
SVM-G (M)	348.69 (175.91)	0.15 (0.07)	15.7 (7.92)
SVM-G-CSA (M)	302.64 (148.75)	0.13 (0.06)	13.63 (6.7)
SVM-MLP	163.8	0.06	7.38
SVM-CSA-MLP	145.97	0.06	6.57
Station: Kermanshah			
Best empirical model	237.97	0.12	13.89
SVM-G (M)	364.45 (267.5)	0.22 (0.14)	21.28 (15.62)
SVM-G-CSA (M)	306.53 (165.76)	0.18 (0.09)	17.9 (9.68)
SVM-MLP	243.02	0.13	14.19
SVM-CSA-MLP	165.21	0.08	9.65
Station: Mashhad			
Best empirical model	241	0.13	12.38
SVM-G (M)	386.38 (214.2)	0.21 (0.12)	19.85 (11.01)
SVM-G-CSA (M)	202.48 (151.76)	0.1 (0.07)	10.4 (7.8)
SVM-MLP	207.25	0.12	10.65
SVM-CSA-MLP	153.14	0.07	7.87
Station: Tabriz			
Best empirical model	366.88	0.2	19.61
SVM-G (M)	539.22 (493.45)	0.27 (0.22)	28.82 (26.37)
SVM-G-CSA (M)	319.93 (316.05)	0.17 (0.14)	17.1 (16.89)
SVM-MLP	458.68	0.22	24.51

**Table 4** (continued)

Station: Ahvaz			
Models	RMSE (J.cm <sup>-2</sup> .day <sup>-1</sup> )	MAPE	RRMSE (%)
SVM-CSA-MLP	264.93	0.13	14.16

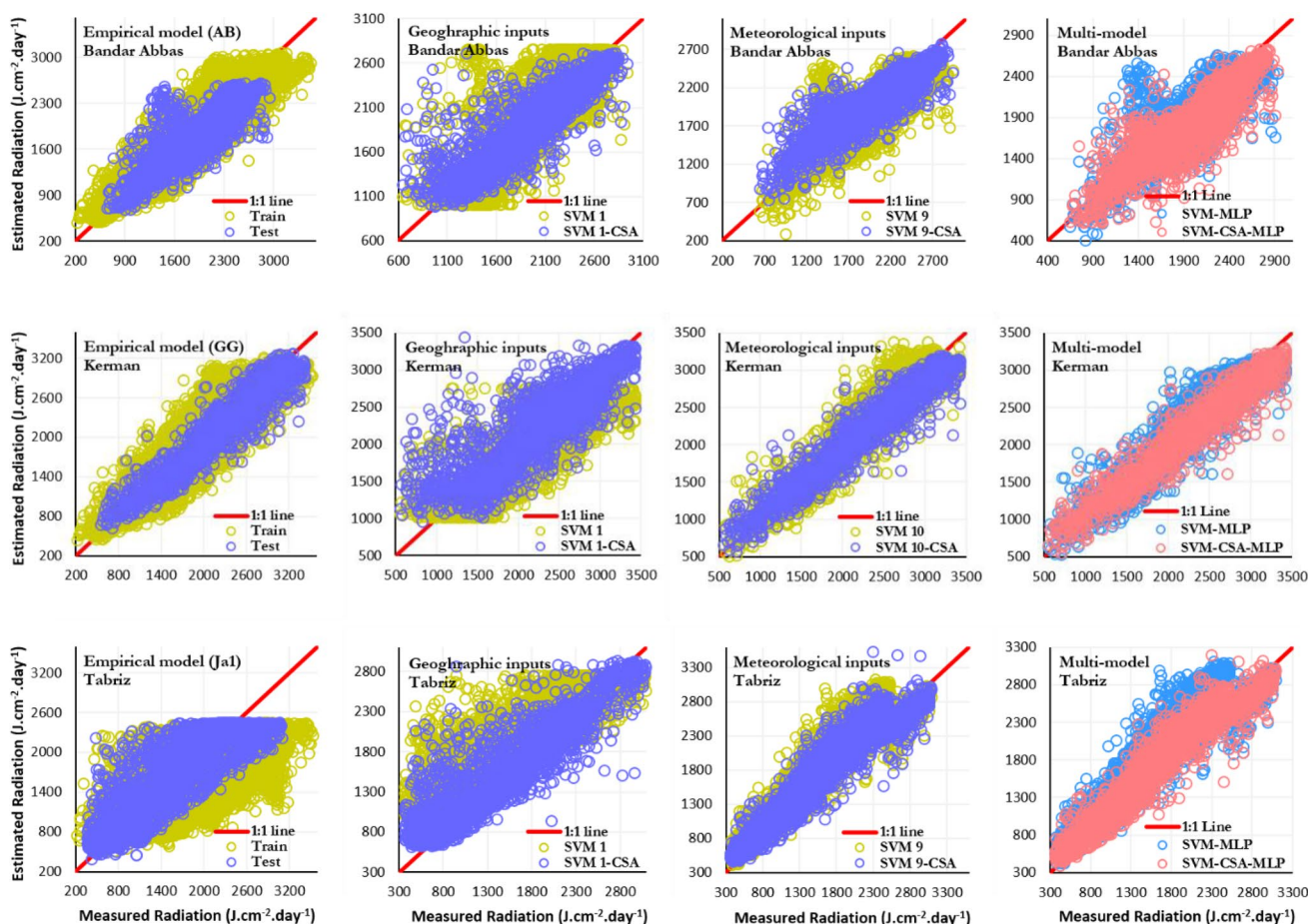
### Kermanshah station

Aside from SVM-M-CSA and SVM-CSA-MLP, which demonstrated a slightly better performance in over-estimation set (RMSE = 161.64 and RMSE = 161.47, respectively), all other models exhibited their best performance in under-estimation set. In under-estimation set, all seven models exhibited their best performance in the first third interval (measured radiation values below 1352 J.cm<sup>-2</sup>.day<sup>-1</sup>), with the lowest and highest differences in error rates being  $\Delta$ RMSE = 36.11 J.cm<sup>-2</sup>.day<sup>-1</sup> (between the first and the last third intervals, SVM-CSA-MLP model) and  $\Delta$ RMSE = 185.56 J.cm<sup>-2</sup>.day<sup>-1</sup> (between the first and the last third intervals, the best empirical model), respectively (Fig. 7). The minimum and maximum estimation errors in over-estimation set were RMSE = 76.34 and RMSE = 523.65 J.cm<sup>-2</sup>.day<sup>-1</sup>, for the last third interval of SVM-M-CSA and the first third interval of SVM-G, respectively.

### Mashhad station

According to the results presented in Table 4, only RRMSEs of SVM-M-CSA and SVM-CSA-MLP are below 10% (“excellent”), which, in confirmation of RMSEs, point to the importance of using meteorological parameters, application of the optimisation algorithm and employment of the multi-model approach in estimating *R<sub>s</sub>* at Kermanshah and Mashhad stations. Using air temperature as the sole input data, Feng et al. (2019) reported RMSEs in the ranges (3.309–3.375), (3.834–4.021), (3.379–3.406) and (3.811–4.053) MJ.m<sup>-2</sup> for empirical models and (2.814–3.103), (3.715–3.939), (3.35–3.491) and (3.54–3.866) for machine learning models at Turpan, Yinchuan, Dunhuang and Xilingol stations in China, respectively. Performance of an empirical model and four AI models in estimating daily *R<sub>s</sub>* in Zhengzhou region, China, was investigated by Xue and Zhou (2019). According to RMSE values, which ranged from 0.7524 to 1.9632 MJ.m<sup>-2</sup>.day<sup>-1</sup> for the 5 models, PSO-LSSVM and the empirical model had the best and the poorest performance, respectively.

Overall, all the seven models performed better in under-estimation compared to over-estimation set; and the lowest and highest differences in error rates between the two sets were observed for the SVM-CSA-MLP model



**Fig. 6** Measured versus estimated radiation values under the best empirical model, the best scenarios of SVM and SVM-CSA (with either geographical or meteorological inputs) and the multi-model approach at each station (external analysis: structure 2)

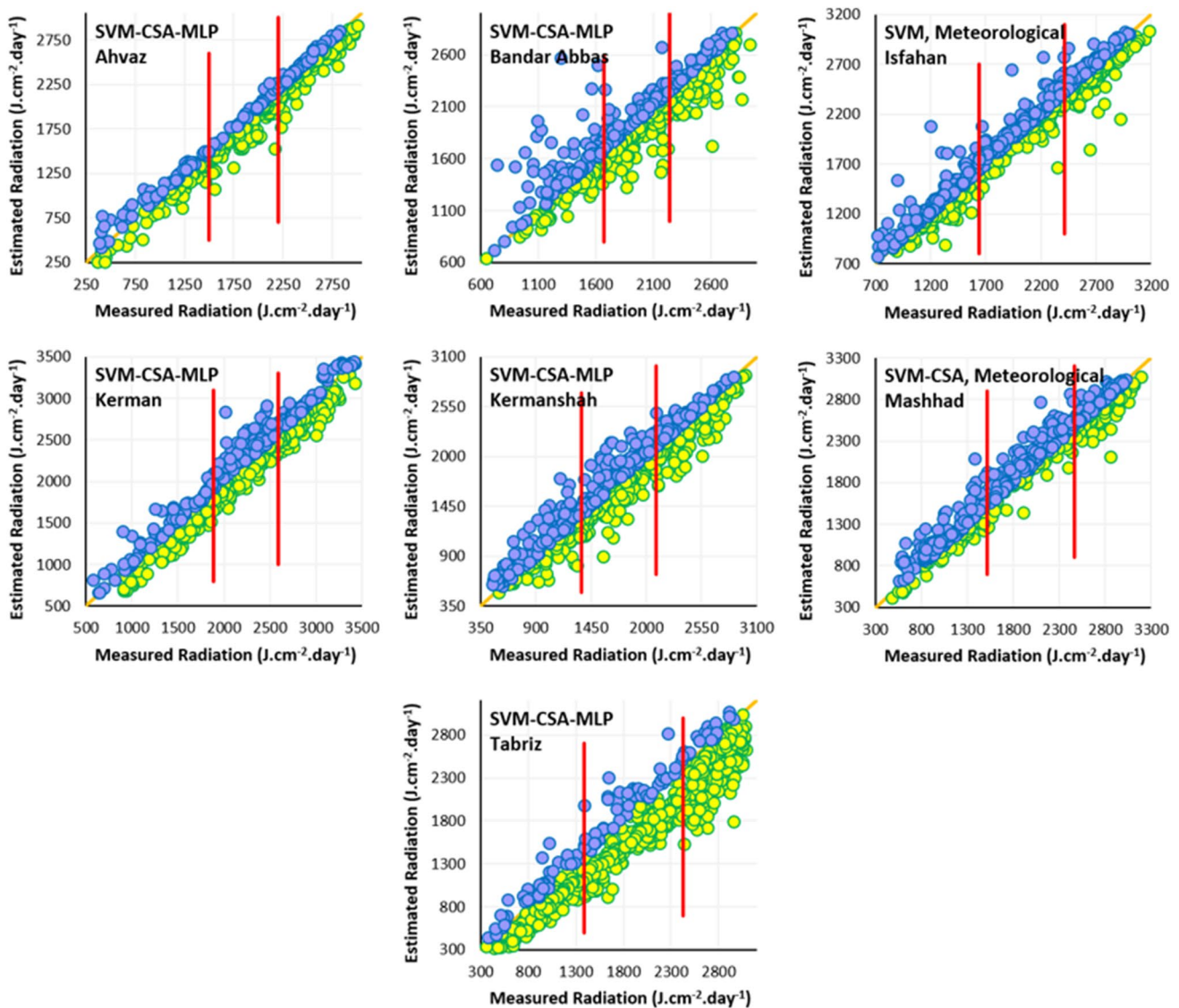
( $\Delta RMSE = 29.06 \text{ J.cm}^{-2}.\text{day}^{-1}$ ) and the SVM-G model ( $\Delta RMSE = 277.76 \text{ J.cm}^{-2}.\text{day}^{-1}$ ), respectively. The best performances of all models in under- and over-estimation sets were obtained in the first third interval (measured radiation values below  $1518 \text{ J.cm}^{-2}.\text{day}^{-1}$ ) and the last third interval (measured radiation values above  $2466 \text{ J.cm}^{-2}.\text{day}^{-1}$ ), respectively (Fig. 8). Abdalla’s empirical model in the first third of under-estimation set ( $RMSE = 28.39 \text{ J.cm}^{-2}.\text{day}^{-1}$ ) and SVM-G in the first third of over-estimation set ( $RMSE = 573.39 \text{ J.cm}^{-2}.\text{day}^{-1}$ ) exhibited the best and the poorest performance in estimating solar radiation, respectively.

**Tabriz station**

In confirmation of the higher RMSEs compared to the other stations, performance of the best empirical model, SVM-CSA and SVM-CSA-MLP were “good” (10–20%) and that of the other models were “fair” (20–30%), according to the RRMSE. These findings are indicative of the fine performance of the first model of Jahani et al. and the noticeable

superiority of SVM-CSA over the stand-alone SVM model. A critical review of the various types of solar radiation estimator models was undertaken by Zhang et al. (2017). Their results showed that RMSEs have been ranging from 1.11 to  $4.5 \text{ MJ.m}^{-2}$  for sunshine-based models, from 2.05 to  $4.7 \text{ MJ.m}^{-2}$  for non-sunshine-based models, and from 1.24 to  $4.2 \text{ MJ.m}^{-2}$  for ANN models. Ghimire et al. (2019) compared the performance of various models at 5 sites in Australia and reported that ANN (with a mean  $RMSE = (1.715\text{--}2.27) \text{ MJ.m}^{-2}.\text{day}^{-1}$ ) has outperformed the other models ( $RMSE = 2.14\text{--}5.9$ ). Hou et al. (2018) compared the performance of ELM integrated with variable forgetting factor (FOS-ELM) and classical ELM in estimating  $R_s$  in Burkina Faso. The results showed that FOS-ELM has reduced RMSE and MAE by (68.8–79.8)% compared to ELM.

The smallest differences in error rates between under- and over-estimation sets were observed for the first model of Jahani et al. ( $\Delta RMSE = 43.09 \text{ J.cm}^{-2}.\text{day}^{-1}$ ) and the SVM-G-CSA model ( $\Delta RMSE = 86.8 \text{ J.cm}^{-2}.\text{day}^{-1}$ ), respectively. In under-estimation set, all seven models exhibited their best performance in the first third interval,



**Fig. 7** Under-estimated (yellow circles) and over-estimated (purple circles) radiation values under the best model at each station. Vertical lines divide measured radiation data points (ranked by magnitude) into three numerically equal groups (local analysis: structure 1)

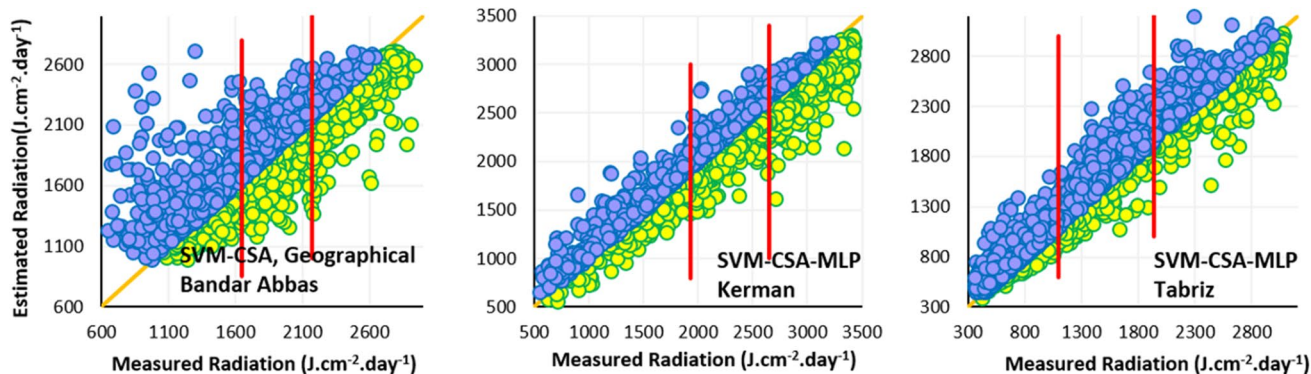
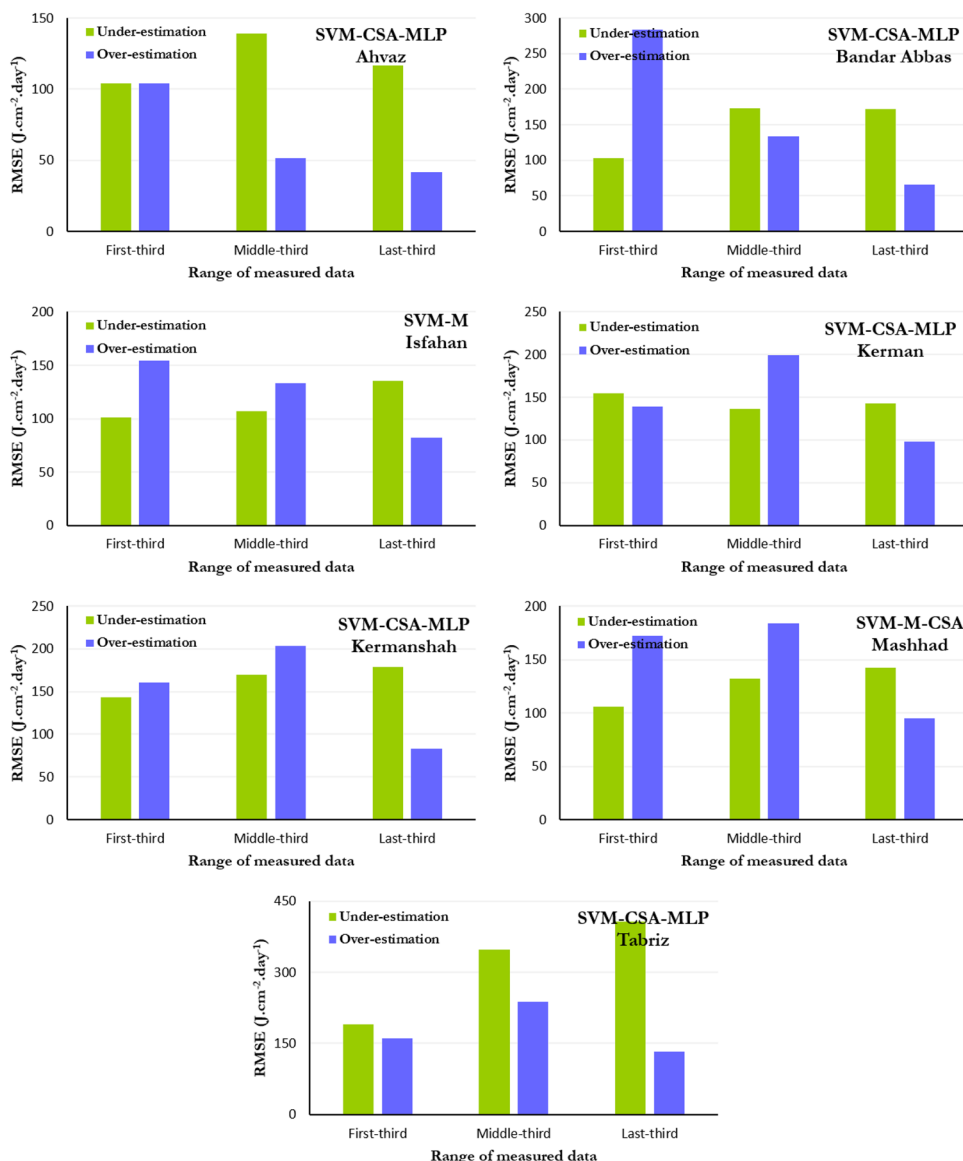
and error rates were rising from the first third to the last third interval (Fig. 7). Accordingly, the minimum and maximum error rates in under-estimation set were 163.99  $\text{J.cm}^{-2}.\text{day}^{-1}$  (first third interval, SVM-G-CSA model) and 725.28  $\text{J.cm}^{-2}.\text{day}^{-1}$  (last third interval, SVM-G). Unlike under-estimation set, the lowest error rates in over-estimation set did not occur in a single interval: SVM-M, SVM-M-CSA and SVM-MLP exhibited their best performance in the first third interval (measurement radiation values below 1386  $\text{J.cm}^{-2}.\text{day}^{-1}$ ); SVM-G performed best in the middle third interval ( $1386 \leq R_s \leq 2431$ ); and the best performance of SVM-G-CSA, the first model of Jahani et al., and SVM-CSA-MLP was obtained in the last third interval (measured radiation values above 2431  $\text{J.cm}^{-2}.\text{day}^{-1}$ ). In

over-estimation set, the lowest and highest error rates were approximately 106.85 (last third interval, the first model of Jahani et al.) and 497.04 (last third interval, SVM-G-CSA), respectively.

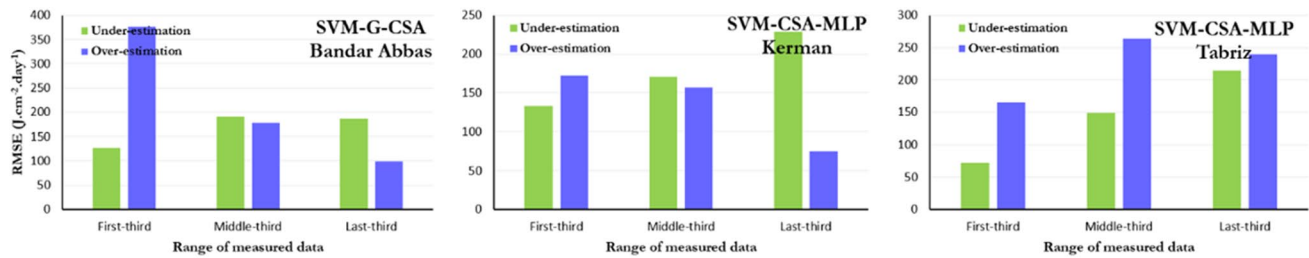
### External performance of the models (structure 2)

Results of the best scenarios — in terms of under- or over-estimation of radiation amounts — and model accuracies, respectively, over the three discussed intervals (first, middle, and last third of measured radiation values) are shown in Figs. 9 and 10. Statistical indices including RMSE, MAPE and RRMSE are also listed in Table 5.

**Fig. 8** RMSEs in the first, middle and last third intervals of estimated radiation amounts under the best model, in under- and over-estimation sets; at each station (local analysis: structure 1)



**Fig. 9** Under-estimated (yellow circles) and over-estimated (purple circles) radiation values under the best model at each station. Vertical lines divide measured radiation data points (ranked by magnitude) into three numerically equal groups (structure 2)



**Fig. 10** RMSEs in the first, middle and last third intervals of estimated radiation amounts under the best model, in under- and over-estimation sets, at each station (structure 2)

**Table 5** Performance evaluation indices for the best empirical model, SVM, SVM-CSA, and the multi-model approach for structure 2 (external analysis). *G* and *M* denote geographical and meteorological inputs, respectively

Station: Bandar Abbas			
Models	RMSE ( $\text{J.cm}^{-2}.\text{day}^{-1}$ )	MAPE	RRMSE (%)
Best empirical model	250.83	0.1	13.23
SVM-G (M)	365.45 (267.89)	0.17 (0.11)	19.28 (14.14)
SVM-G-CSA (M)	215.83 (225.5)	0.09 (0.11)	11.39 (11.9)
SVM-MLP	268.16	0.11	14.15
SVM-CSA-MLP	230.97	0.11	12.19
Station: Kerman			
Best empirical model	203.96	0.07	9.07
SVM-G (M)	459.13 (234.18)	0.2 (0.09)	20.42 (10.42)
SVM-G-CSA (M)	322.39 (167.25)	0.13 (0.06)	14.34 (7.44)
SVM-MLP	219.32	0.09	9.75
SVM-CSA-MLP	165.12	0.06	7.34
Station: Tabriz			
Best empirical model	336.86	0.2	21.78
SVM-G (M)	529.69 (436.2)	0.43 (0.27)	34.24 (28.2)
SVM-G-CSA (M)	286.18 (251.26)	0.19 (0.15)	18.5 (16.24)
SVM-MLP	331.22	0.2	21.41
SVM-CSA-MLP	184.31	0.11	11.91

### Bandar Abbas station

The ranges of RRMSE variations are given in Table 5, showing that all models have had a “good” performance; the lowest and highest RRMSEs (11.39 and 19.28%) are those of SVM-G-CSA and SVM-G, respectively. This finding underlines the importance of coupling the CSA with the base SVM and is confirmed by comparison of RMSEs of these two models. Marzo et al. (2017) evaluated the performance of the ANN in estimating daily Rs in desert areas, using data from several stations in Chile to train the network and data from other stations to validate the results (RRMSD = 6.6%).

Also, data from two other stations in Chile and four in Israel, South Africa, Saudi Arabia and Australia were used to test the generalisation capability of the proposed model to other desert regions of the world, with RRMSD values ranging from 8.1% (one of the stations in Chile) to 22.9% (the Australian station).

Overall, of the seven models studied, only SVM-M-CSA and SVM-CSA-MLP performed better in over-estimation set compared to the under-estimation set. However, the difference in error rates between under- and over-estimation sets for Abdalla’s empirical model, SVM-M and SVM-MLP were about 7, 23 and 31  $\text{J.cm}^{-2}.\text{day}^{-1}$ , respectively, which are negligible and indicative of comparable performance of these models in the two sets. In under- and over-estimation sets, the best performance of all seven models occurred in the first third interval (measured radiation values below  $1646 \text{ J.cm}^{-2}.\text{day}^{-1}$ ) and the last third interval (measured values above  $2171 \text{ J.cm}^{-2}.\text{day}^{-1}$ ), respectively. The minimum and maximum estimation errors in under-estimation set were 126.51 and  $315.02 \text{ J.cm}^{-2}.\text{day}^{-1}$  for SVM-G-CSA and SVM-G models, respectively; whereas corresponding values in over-estimation set were 77.94 and 675.34 for the SVM-M-CSA and SVM-G models, respectively, indicating that the differences in performance of all models between the one-third intervals have been greater in over-estimation set.

### Kerman station

Comparison of RRMSEs (Table 5) shows that Garj-Garj model, SVM-M-CSA and the multi-model approach have had an “excellent” performance, with RRMSEs below 10%. Hassan et al. (2016) assessed the generalisation capability of their proposed empirical models with application on the ten separate stations and showed that RMSEs were in the range (0.7035–2.447).

From the viewpoint of model performance over different one-third intervals, all models have estimated radiation amounts with the lowest error rates in the first third interval (measured radiation values below  $1928 \text{ J.cm}^{-2}.\text{day}^{-1}$ ) in the under-estimation set and in the last third interval (measured radiation values above  $2652 \text{ J.cm}^{-2}.\text{day}^{-1}$ ) in the



over-estimation set (Figs. 9 and 10). According to the results, only SVM-G-CSA and SVM-M have had lower error rates in under-estimation set. All other models have estimated radiation values more accurately in over-estimation set, although the differences in error rates between the two sets are negligible for SVM-G and SVM-MLP (approximately 3.2 and 1.6  $\text{J} \cdot \text{cm}^{-2} \cdot \text{day}^{-1}$ , respectively), indicating consistent performance of these models in under- and over-estimation sets. In under-estimation set, SVM-M-CSA and SVM-CSA-MLP were the most accurate models, with RMSEs of 131.27 and 133.26  $\text{J} \cdot \text{cm}^{-2} \cdot \text{day}^{-1}$ , respectively (both in the first third interval); and SVM-G had the lowest accuracy with  $\text{RMSE} = 500.71 \text{ J} \cdot \text{cm}^{-2} \cdot \text{day}^{-1}$  (in the last third interval). In over-estimation set, SVM-G ( $\text{RMSE} = 56.38 \text{ J} \cdot \text{cm}^{-2} \cdot \text{day}^{-1}$  in the last third interval) and SVM-G-CSA ( $\text{RMSE} = 567.5 \text{ J} \cdot \text{cm}^{-2} \cdot \text{day}^{-1}$  in the first third interval) had the highest and lowest accuracies in estimating  $R_s$ , respectively. Maximal variations in estimation error over the three one-third intervals occurred in over-estimation set and for the SVM-G model, from first interval to middle interval (46% error reduction) and from middle interval to last interval (81% error reduction), and for the SVM-G-CSA model, first to middle interval (42% error reduction) and middle to last interval (62% error reduction).

#### Tabriz station

According to the results (Table 5), the most challenging attempt at generalisation of the models to stations with no role in model training has been the one at Tabriz station. Although SVM-G (a “poor” model) and SVM-M, SVM-MLP and the first model of Jahani et al. (“fair”) did not perform well in estimating radiation at this station, SVM-CSA with meteorological or geographical inputs had “good” RRMSEs of 16.24 and 18.5%, respectively. The results show that the multi-model approach has been effective and efficient in generalising radiation estimation at a station that has played no role in its training, with  $\text{RRMSE} = 11.91\%$ , which is very close to the borderline value separating “excellent” and “good” categories (10%). Almost all models have performed noticeably better in under-estimation set compared to over-estimation set, with the minimum and maximum differences in error rates between the two sets being 49.39  $\text{J} \cdot \text{cm}^{-2} \cdot \text{day}^{-1}$  (the first model of Jahani et al.) and 314.17  $\text{J} \cdot \text{cm}^{-2} \cdot \text{day}^{-1}$  (SVM-G model). In under-estimation set, the lowest and highest error rates were 28.93  $\text{J} \cdot \text{cm}^{-2} \cdot \text{day}^{-1}$  (SVM-G, first third interval) and 420.69  $\text{J} \cdot \text{cm}^{-2} \cdot \text{day}^{-1}$  (the first model of Jahani et al., last third interval), respectively. An important point worth mentioning about the under-estimation set is the relatively low error rates in the first third interval, with all models except the first model of Jahani et al. having estimation errors below 90  $\text{J} \cdot \text{cm}^{-2} \cdot \text{day}^{-1}$ . In over-estimation set, there is almost no consistency in error rates

between the three intervals. The lowest errors rates of SVM-G, SVM-G-CSA and the first model of Jahani et al. occurred in the last third interval; whereas for SVM-M, SVM-M-CSA, SVM-MLP and SVM-CSA-MLP the lowest error rates occurred in the first third interval. SVM-G-CSA ( $\text{RMSE} = 160.4$ ), SVM-CSA-MLP ( $\text{RMSE} = 164.74$ ) and the first model of Jahani et al. ( $\text{RMSE} = 165.99$ ) exhibited the best performance, and SVM-G ( $\text{RMSE} = 666.17$ ) had the poorest performance in over-estimation set.

## Conclusions

In the current research, we evaluated the performance of empirical models, two AI models (SVM and SVM-CSA), and the novel “multi-model” approach, in estimating daily  $R_s$  values at seven Iranian meteorological stations over 2010–2019. For the first structure, model performances were examined separately at each station and in training and testing sets (local analysis). For the second structure (external analysis), an attempt was made to examine the generalisation capability of the models by separating the data used for model training (Ahvaz, Isfahan, Kermanshah and Mashhad stations) from those used for testing (Bandar Abbas, Kerman and Tabriz stations). The results showed that overall, meteorological parameters have played a more effective role in estimating radiation compared to geographical parameters. Considering the atmospheric conditions including energy transferring and sunshine duration, is one of the main advantages of meteorological compared to geographical parameters. SVM-CSA significantly improved radiation estimates at all stations except Isfahan, where Garj-Garj empirical model performed equally well and was comparable to SVM-CSA. All models except SVM-G at Bandar Abbas station, Garj-Garj model, SVM-M-CSA and the multi-model approach at Kerman station and SVM-CSA-MLP at Tabriz station could be effectively and efficiently generalised to stations that played no role in training those models (second structure). Although the multi-model approach demonstrated a much better performance under both structures and at most stations compared with empirical models and the base SVM, it is not preferable to the SVM-CSA model given that its superiority over SVM-CSA is negligible at most stations (except Tabriz) on the one hand, and it requires longer and more complex computations on the other hand. Using meteorological as well as geographical inputs and considering the ability of multi-model approach beside AI-based models are the main advantages of the proposed models in the present study. It appears that further studies at climatically diverse stations are needed before recommending the use of the multi-model approach for radiation estimation.

**Acknowledgements** The authors would like to thank the Islamic Republic of Iran's Meteorological Organization (IRIMO) for providing the meteorological dataset.

**Author's contribution** Conceptualisation; writing, review and editing; supervision; software; formal analysis: Roozbeh Moazenzadeh. Data curation, methodology, writing original draft, software, formal analysis: Babak Mohammadi. Writing, review and editing; supervision; formal analysis: Zheng Duan. Data curation: Mahdi Delghandi.

**Data Availability** The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

## Declarations

**Ethics approval and consent to participate** Not applicable.

**Consent for publication** Not applicable.

**Conflict of interest** The authors declare no competing interests.

## References

- Adaramola MS (2012) Estimating global solar radiation using common meteorological data in Akure, Nigeria. *Renew Energy* 47:38–44. <https://doi.org/10.1016/j.renene.2012.04.005>
- Aghelpour P, Mohammadi B, Biazar SM (2019) Long-term monthly average temperature forecasting in some climate types of Iran, using the models SARIMA, SVR, and SVR-FA. *Theor Appl Climatol* 138(3–4):1471–1480. <https://doi.org/10.1007/s00704-019-02905-w>
- Annandale J, Jovanovic N, Benade N, Allen R (2002) Software for missing data error analysis of Penman-Monteith reference evapotranspiration. *Irrigation Science* 21(2):57–67. <https://doi.org/10.1007/s002710100047>
- Asl SFZ, Karami A, Ashari G, Behrang A, Assareh A, Hedayat N (2011) Daily global solar radiation modelling using multi-layer perceptron (MLP) neural networks. *World Acad Sci Eng Technol* 79:740–742
- Azimi H, Bonakdari H, Ebtehaj I (2019) Design of radial basis function-based support vector regression in predicting the discharge coefficient of a side weir in a trapezoidal channel. *Appl Water Sci* 9(4):1–12. <https://doi.org/10.1007/s13201-019-0961-5>
- Belaid S, Mellit A (2016) Prediction of daily and mean monthly global solar radiation using support vector machine in an arid climate. *Energy Convers Manage* 118:105–118. <https://doi.org/10.1016/j.enconman.2016.03.082>
- Cao JC, Cao SH (2006) Study of forecasting solar irradiance using neural networks with preprocessing sample data by wavelet analysis. *Energy* 31:3435–3445. <https://doi.org/10.1016/j.energy.2006.04.001>
- Cimen M, Kisi O (2009) Comparison of two different data-driven techniques in modeling lake level fluctuations in Turkey. *J Hydrol* 378:253–262. <https://doi.org/10.1016/j.jhydrol.2009.09.029>
- Deo RC, Kisi O, Singh VP (2017) Drought forecasting in eastern Australia using multivariate adaptive regression spline, least square support vector machine and M5Tree model. *Atmos Res* 184:149–175. <https://doi.org/10.1016/j.atmosres.2016.10.004>
- Ebtehaj I, Bonakdari H (2016) A support vector regression-firefly algorithm-based model for limiting velocity prediction in sewer pipes. *Water Sci Technol* 73(9):2244–2250. <https://doi.org/10.2166/wst.2016.064>
- Ertekin C, Yaldiz O (2000) Comparison of some existing models for estimating global solar radiation for Antalya (Turkey). *Energy Convers Manage* 41:311–330. [https://doi.org/10.1016/S0196-8904\(99\)00127-2](https://doi.org/10.1016/S0196-8904(99)00127-2)
- Fan J, Wang X, Wu L, Zhang F, Bai H, Lu X, Xiang Y (2018) New combined models for estimating daily global solar radiation based on sunshine duration in humid regions: A case study in South China. *Energy Convers Manage* 156:618–625. <https://doi.org/10.1016/j.enconman.2017.11.085>
- Feng L, Lin A, Wang L, Qin W, Gong W (2018a) Evaluation of sunshine-based models for predicting diffuse solar radiation in China. *Renew Sust Energ Rev* 94:168–182. <https://doi.org/10.1016/j.rser.2018.06.009>
- Feng L, Qin W, Wang L, Lin A, Zhang M (2018b) Comparison of artificial intelligence and physical models for forecasting photosynthetically-active radiation. *Remote Sens* 10:1855. <https://doi.org/10.3390/rs10111855>
- Feng Y, Cui N, Zhang Q, Zhao L, Gong D (2017) Comparison of artificial intelligence and empirical models for estimation of daily diffuse solar radiation in North China Plain. *Int J Hydrogen Energy* 42(21):14418–14428. <https://doi.org/10.1016/j.ijhydene.2017.04.084>
- Feng Y, Gong D, Zhang Q, Jiang S, Zhao L, Cui N (2019) Evaluation of temperature-based machine learning and empirical models for predicting daily global solar radiation. *Energy Convers Manage* 198:111780. <https://doi.org/10.1016/j.enconman.2019.111780>
- Gandomi AH, Yang XS, Alavi AH (2013) Cuckoo search algorithm: a metaheuristic approach to solve structural optimization problems. *Eng Comput* 29:17–35. <https://doi.org/10.1007/s00366-011-0241-y>
- Ghimire S, Deo RC, Downs NJ, Raj N (2019) Global solar radiation prediction by ANN integrated with European centre for medium range weather forecast fields in solar rich cities of Queensland Australia. *J Clean Prod* 216:288–310. <https://doi.org/10.1016/j.jclepro.2019.01.158>
- Goyal MK, Bharti B, Quilty J, Adamowski J, Pandey A (2014) Modeling of daily pan evaporation in sub-tropical climates using ANN, LS-SVR, Fuzzy Logic, and ANFIS. *Expert Syst Appl* 41:5267–5276. <https://doi.org/10.1016/j.eswa.2014.02.047>
- Halabi LM, Mekhilef S, Hossain M (2018) Performance evaluation of hybrid adaptive neuro-fuzzy inference system models for predicting monthly global solar radiation. *Appl Energy* 213:247–261. <https://doi.org/10.1016/j.apenergy.2018.01.035>
- Hassan GE, Youssef ME, Mohamed ZE, Ali MA, Hanafy AA (2016) New temperature-based models for predicting global solar radiation. *Appl Energy* 179:437–450. <https://doi.org/10.1016/j.apenergy.2016.07.006>
- Hassan MA, Khalil A, Kaseb S, Kassem MA (2017) Potential of four different machine-learning algorithms in modelling daily global solar radiation. *Renew Energy* 111:52–62. <https://doi.org/10.1016/j.renene.2017.03.083>
- He Z, Xia K, Niu W, Aslam N, Hou J (2018) Semisupervised SVM based on cuckoo search algorithm and its application. *Math Probl Eng*. <https://doi.org/10.1155/2018/8243764>
- Hou M, Zhang T, Weng F, Ali M, Al-Ansari N, Yaseen ZM (2018) Global solar radiation prediction using hybrid online sequential extreme learning machine model. *Energies* 11(3415):1–19. <https://doi.org/10.3390/en11123415>
- Jamil B, Akhtar N (2017) Estimation of diffuse solar radiation in humid-subtropical climatic region of India: comparison of diffuse fraction and diffusion coefficient models. *Energy* 131:149–164. <https://doi.org/10.1016/j.energy.2017.05.018>
- Kashani MH, Ghorbani MA, Shahabi M, Raghavendra S, Diop L (2020) Multiple AI model integration strategy-application to

- saturated hydraulic conductivity prediction from easily available soil properties. *Soil Tillage Res* 196:104449. <https://doi.org/10.1016/j.still.2019.104449>
- Kim S, Seo Y, Rezaie-Balf M, Kisi O, Ghorbani MA, Singh VP (2018) Evaluation of daily solar radiation flux using soft computing approaches based on different meteorological information: peninsula vs continent. *Theor Appl Climatol* 137:693–712. <https://doi.org/10.1007/s00704-018-2627-x>
- Kisi O, Cimen M (2009) Evapotranspiration modelling using support vector machines. *Hydrol Sci J* 54:918–928. <https://doi.org/10.1623/hysj.54.5.918>
- Kuhe A, Achirgibenda VT, Agada M (2019) Global solar radiation prediction for Makurdi, Nigeria, using neural networks ensemble. *Energy Sources, Part A Recover. Util. Environ. Eff* 13 <https://doi.org/10.1080/15567036.2019.1637481>
- Lan H, Yin H, Hong YY, Wen S, Yu DC, Cheng P (2018) Day-ahead spatio-temporal forecasting of solar irradiation along a navigation route. *Appl Energy* 211:15–27. <https://doi.org/10.1016/j.apenergy.2017.11.014>
- Liu Y, Zhou Y, Chen Y, Wang D, Wang Y, Zhu Y (2020) Comparison of support vector machine and copula-based nonlinear quantile regression for estimating the daily diffuse solar radiation: A case study in China. *Renew Energy* 146:1101–1112. <https://doi.org/10.1016/j.renene.2019.07.053>
- Marzo A, Trigo-Gonzalez M, Alonso-Montesinos J, Martinez-Durban M, Lopez G, Ferrada P, Fuentealba E, Cortes M, Batlles FJ (2017) Daily global solar radiation estimation in desert areas using daily extreme temperatures and extraterrestrial radiation. *Renew Energy* 113:303–311. <https://doi.org/10.1016/j.renene.2017.01.061>
- Meenal R, Selvakumar AI (2017) Assessment of SVM, empirical and ANN based solar radiation prediction models with most influencing input parameters. *Renew Energy* 121:324–343. <https://doi.org/10.1016/j.renene.2017.12.005>
- Moazenzadeh R, Mohammadi B (2019) Assessment of bio-inspired metaheuristic optimization algorithms for estimating soil temperature. *Geoderma* 353:152–171. <https://doi.org/10.1016/j.geoderma.2019.06.028>
- Moazenzadeh R, Mohammadi B, Shamshirband S, Chau KW (2018) Coupling a firefly algorithm with support vector regression to predict evaporation in northern Iran. *Eng Appl Comp Fluid* 12(1):584–597. <https://doi.org/10.1080/19942060.2018.1482476>
- B Mohammadi S, Mehdizadeh (2020) Modeling daily reference evapotranspiration via a novel approach based on support vector regression coupled with whale optimization algorithm *Agric Water Manag* 106:145 <https://doi.org/10.1016/j.agwat.2020.106145>
- Mohammadi B, Moazenzadeh R (2021) Performance analysis of daily global solar radiation models in Peru by regression analysis. *Atmosphere* 12:389. <https://doi.org/10.3390/atmos12030389>
- B Mohammadi R, Moazenzadeh K, Christian Z, Duan (2021) Improving streamflow simulation by combining hydrological process-driven and artificial intelligence-based models *Environ SciPollut Res* <https://doi.org/10.1007/s11356-021-15563-1>
- Puspaningrum A, Suheryadi A, Sumarudin A (2020) Implementation of cuckoo search algorithm for support vector machine parameters optimization in pre collision warning. In *IOP Conference Series: Materials Science and Engineering* (Vol. 850, No. 1, p. 012027). IOP Publishing. <https://doi.org/10.1088/1757-899X/850/1/012027>
- Qin W, Wang L, Lin A, Zhang M, Xia X, Hu B, Niu Z (2018) Comparison of deterministic and data-driven models for solar radiation estimation in China. *Renew Sust Energy Rev* 81:579–594. <https://doi.org/10.1016/j.rser.2017.08.037>
- Quej VH, Almorox J, Ibrakhimov M, Saito L (2017) Estimating daily global solar radiation by day of the year in six cities located in the Yucatan Peninsula, Mexico. *J Clean Prod* 14:75–82. <https://doi.org/10.1016/j.jclepro.2016.09.062>
- Raghavendra S, Deka PC (2014) Support vector machine applications in the field of hydrology: A review. *Appl Soft Comput* 19:372–386. <https://doi.org/10.1016/j.asoc.2014.02.002>
- Rehman S, Mohandes M (2008) Artificial neural network estimation of global solar radiation using air temperature and relative humidity. *Energy Policy* 36:571–576. <https://doi.org/10.1016/j.enpol.2007.09.033>
- Sanz S, Deo RC, Cornejo-Bueno L, Camacho-Gomez C, Ghimire S (2018) An efficient neuro evolutionary hybrid modelling mechanism for the estimation of daily global solar radiation in the Sunshine State of Australia. *Appl Energy* 209:79–94. <https://doi.org/10.1016/j.apenergy.2017.10.076>
- Sharafi H, Ebtehaj I, Bonakdari H, Zaji AH (2016) Design of a support vector machine with different kernel functions to predict scour depth around bridge piers. *Nat Hazards* 84(3):2145–2162. <https://doi.org/10.1007/s11069-016-2540-5>
- Wang L, Kisi O, Zounemat-Kermani M, Salazar GA, Zhu Z, Gong W (2016a) Solar radiation prediction using different techniques: model evaluation and comparison. *Renew Sust Energy Rev* 61:384–397. <https://doi.org/10.1016/j.rser.2016.04.024>
- L Wang O, Kisi M, Zounemat-Kermani Z, Zhu W, Gong Z, Niu H, Liu Z, Liu (2016b) Prediction of solar radiation in China using different adaptive neuro-fuzzy methods and M5 model tree *Int J Climatol* <https://doi.org/10.1002/joc.4762>
- Wang L, Lu Y, Zou L, Feng L, Wei J, Qin W, Niu Z (2019) Prediction of diffuse solar radiation based on multiple variables in China. *Renew Sust Energy Rev* 103:151–216. <https://doi.org/10.1016/j.rser.2018.12.029>
- Xue X, Zhou H (2019) Soft computing methods for predicting daily global solar radiation. *Numer. Heat Transf., Part B: Fundam* 1–14. <https://doi.org/10.1080/10407790.2019.1637629>
- Yang XS, Deb S (2009) Cuckoo search via levy flights. *World Congress on Nature & Biologically Inspired Computing* 210–214.
- Yıldırım HB, Teke A, Antonanzas-Torres F (2018) Evaluation of classical parametric models for estimating solar radiation in the Eastern Mediterranean region of Turkey. *Renew Sust Energy Rev* 82:2053–2065. <https://doi.org/10.1016/j.rser.2017.08.033>
- Zang H, Cheng L, Ding T, Cheung KW, Wang M, Wei Z, Sun G (2018) Estimation and validation of daily global solar radiation by day of the year-based models for different climates in China. *Renew Energy* 135:984–1003. <https://doi.org/10.1016/j.renene.2018.12.065>
- Zhang J, Zhao L, Deng S, Xu W, Zhang YA (2017) critical review of the models used to estimate solar radiation. *Renew Sust Energy Rev* 70:314–329. <https://doi.org/10.1016/j.rser.2016.11.124>
- L Zou L, Wang L, Xia A, Lin B, Hu H, Zhu (2017) Prediction and comparison of solar radiation using improved empirical models and Adaptive Neuro-Fuzzy Inference Systems *Renew Energy* 106:343–353 <https://doi.org/10.1016/j.renene.2017.01.042>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.