Research Article

A Novel Model Based on Square Root Elastic Net and Artificial Neural Network for Forecasting Global Solar Radiation

He Jiang [1,2] **and Yao Dong** [1,2]

[1]*School of Statistics, Jiangxi University of Finance and Economics, Nanchang 330013, China*
[2]*Applied Statistics Research Center, Jiangxi University of Finance and Economics, Nanchang 330013, China*

Correspondence should be addressed to Yao Dong; dongyao20051987@126.com

In recent years, solar energy has attracted a great deal of attentions from scientific researchers because it is a clean and renewable form of energy. To make good use of solar energy, an effective way to forecast solar radiation is essential to guarantee the reliability of grid-connected photovoltaic installations. Although an artificial neural network (ANN) is of great importance, irrelevant variables are utilized which results in complex model and intractable computation cost. To remove these irrelevant variables, the combination of variable selection methods and ANN are applied. However, how to select the regularization parameters in these techniques is challenging. This paper successfully investigates a square root elastic net- (SREN-) based approach to tackle this challenge and selects all the important variables. An Elman neural network (ENN) is constructed with the important variables selected by SREN as inputs. Based on meteorological data, SRENENN has been developed for 1-year period in Xinjiang area of China. The present model delivers superior relationship between the estimated and measure values.

1. Introduction

Owing to the rapid development of the global economy, energy crisis and environmental pollution problems have threatened the sustainable development and human health. More and more countries pay much attention to green and renewable sources of energy, so it is essential to utilize sources of clean energy instead of fossil fuel [1, 2]. In fact, all kinds of energy sources derived from the sun have a diameter of 1.39×10^9 m and emit ferocious energy of 3.8×10^{20} MW, but the earth only obtains a small fraction of 1.7×10^{14} kW [3]. As one of the most significant forms of green energy, solar energy was used since the prehistoric times because it can be captured anywhere. Solar energy is a renewable and clean alternative for solving the worldwide energy shortage and environmental problems [4]. It can be applied in several fields including locating photovoltaic power plants, scheduling electrical load, and developing low-carbon economy [5]. It is significant to get the reliable global solar radiation data

for investigating, assessing, and utilizing solar energy resource. Although ground-based measurements can obtain the accurate global solar radiation, all the locations are not available [6]. In recent years, geostationary weather satellites can be applied to estimate global solar radiation at ground level, but it is worse than the forecasting models because it is an indirect approach [7]. In addition, the weather is intrinsically chaotic and instable which greatly affects the global solar radiation. These volatilities will threaten the stability and quality of the whole power system [8]. Therefore, it is vital to develop some models to improve the forecasting accuracy of global solar radiation through several atmospheric factors.

Many researchers attempt to study several soft computing methods to forecast solar radiation and evaluate their potential of solar energy. These models include time series regression models (ARMA, ARIMA, and GARCH), empirical models, and machine learning techniques (artificial neural networks, support vector machine, etc.) [9]. Sun et al. proposed ARMAX-GARCH model to forecast daily global

solar radiation using several meteorological variables. The results of experiment showed that global solar radiation depends more on sunshine duration than temperature difference at certain stations [10]. David et al. applied the combination of ARMA and GARCH model to provide probabilistic forecasts for solar irradiance. The proposed recursive ARMA-GARCH model was easier to estimate parameters and got a good accuracy [11]. Quej et al. developed a new empirical model to predict hourly global solar radiation applying meteorological factors such as rainfall, temperature, and humidity at six sites in Mexico. Through comparing with other models, the proposed model had the best forecasting precision [12]. Ouderni et al. utilized several empirical models including Benson model, Page model, and Angstrom-Prescott-Page model to assess the solar potential in the gulf of Tunis [13].

As one of the most popular forecasting models, machine learning techniques including artificial neural networks (ANNs), intelligent optimization algorithms, and support vector machines (SVMs) own self-adaptiveness and robustness and have already been successfully applied to forecast global solar radiation. ANN techniques include backpropagation (BP), radial basis function (RBF), multi-layer perceptron (MLP), and extreme learning machine (ELM) [14–17]. Benmouiza and Cheknane used k-means method to find the input samples and took advantage of nonlinear autoregressive (NAR) neural networks to forecast hourly global horizontal solar radiation [18]. Chen et al. presented a model based on fuzzy rules and neural network to forecast solar radiation; the case study revealed that the proposed technique achieve excellent forecasting accuracy [19]. Renno et al. developed two ANN models to estimate hourly direct normal irradiance and global radiation [5]. Salcedo-Sanz et al. proposed a novel approach Coral Reefs Optimization-Extreme Learning Machine (CRO-ELM) to predict daily global solar radiation and achieved satisfactory results [20]. Gairaa et al. adopt a new hybrid technique combining the linear ARMA and the nonlinear ANN to forecast daily global solar radiation in Algeria. The experimental results revealed that the hybrid model is superior to the single one [21].

Although ANNs have been widely exploited because of their nonlinear mapping ability, prediction capabilities, and robustness, the optimal parameters in the network such as weights, bias, and the number of the hidden layer nodes are not easy to determine, and the training of the network is likely to converge to a local minimum [22]. Furthermore, its structure of the network would be quite intricate if all the variables are applied as inputs. This will cause the following two problems: (1) the complex structure makes critically trouble for forecasting and selection performance and (2) the complex structure needs much computation time. The weights between the nodes in an ANN are going to be estimated, and it would spend a lot of time if ANN has excessive number of nodes. Based on the above discussion, investigating an effective method to establish a simple neural network is essential. Since its structure relies much on the number of input sets, variable selection techniques are needed to choose the significant variables which are considered as inputs of an ANN.

Some researchers focus on selecting some important variables as inputs of the forecasting models including but not limited to ANN and SVM. Benghanem et al. applied Levenberg-Marquardt learning algorithm to construct ANN to study daily global irradiation of Saudi Arabia. Air temperature, sunshine duration, relative humidity, and day of the year are used as the input variables which achieved good forecasting accuracy [23]. Rahimikhoob used temperature including the highest temperature and the lowest temperature to forecast global solar radiation in Southwest of Iran [24]. Qing and Niu developed a new technique long short-term memory (LSTM) networks to predict hourly solar irradiance and used the weather data (temperature, wind speed, dew point, etc.) to enter the networks [25]. Vakili et al. established MLP neural network to estimate daily solar irradiance using temperature, wind speed, relative humidity, and particulate matter 10 [26]. Rohani et al. proposed a Gaussian process with K-fold cross-validation model to forecast daily and monthly solar radiation using temperature, humidity, pressure, and sunshine hours as input variables [27]. It is found that the above hybrid approaches combine the advantages of several single models and perform better. Variable selection algorithms can be used to reduce high-dimensional data that select the optimal input variables or model [28, 29]. Jović et al. studied the solar radiation and used adaptive neuro-fuzzy inference system (ANFIS) to select the most relevant factors from temperature, mean sea level, and relative humidity as the predictors [30]. Almaraashi applied four different feature selection methods to determine the input space and forecast daily solar radiation in Saudi Arabia based on a multilayer neural network [31]. Aybar-Ruiz et al. adopted a grouping genetic method to select the relevant atmospherical features in extreme learning machine model for predicting global solar radiation [32]. Mori chose meteorological variables using graphical modelling to estimate solar radiation [33]. Jiang and Dong developed penalized kernel SVM approaches to select structural variables and forecast global horizontal radiation [34].

As far as we know, the current research papers focus on the way to select variables by trying some specific combinations or groups. However, there is no theoretical guarantee of the way to determine these combinations and considering all the possible combination of variables is time-consuming. Penalized variable selection methods are advocate to select the important variables directly without trying possible combinations, and they are more straightforward to use. Furthermore, compared with the conventional ANNs and SVMs, Elman neural network (ENN) is a local recurrent neural network with a single hidden layer, which owns fast learning rate, good dynamic characteristics, and high global stability [35, 36]. In this paper, an ENN structure can be selected as the forecasting technique for global solar radiation forecasting. This work advocates square root elastic net variable selection procedure in the Elman neural network (SRENENN) approach to forecast the global solar radiation in the Xinjiang area of China. The primary
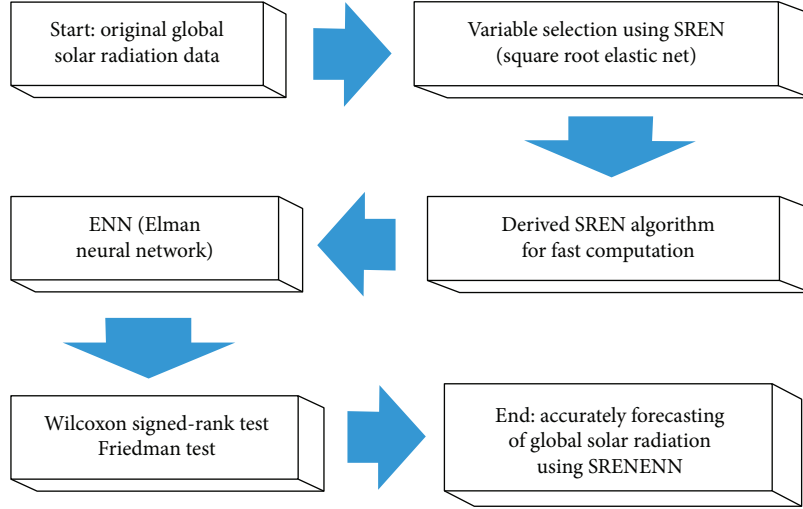
FIGURE 1: A schematic overview of the whole paper.

novelty and contributions of this study are provided in the following list:

(1) An ENN is applied to forecast global solar radiation with meteorological variables.

(2) Square root elastic net is used to effectively extract the meteorological variables which are applied as inputs of ENN, and the optimal model is determined by the 10-fold cross-validation to improve forecasting precision.

(3) A novel square root elastic net variable selection procedure in the Elman neural network (SRENENN) algorithm is proposed, and the corresponding forecasting results are compared systematically using Wilcoxon signed-rank test and Friedman test.

The structure of this study is given: Section 2 describes the square root elastic net variable selection procedure and Elman neural network; Section 3 investigates the case study based on real data analysis; Section 4 provides the forecasting accuracies and corresponding experimental results; the conclusions are presented in Section 5. The schematic overview of the whole paper is given in Figure 1.

2. Materials and Methods

2.1. Square Root Elastic Net. Based on the dataset $\mathbf{Z} = (\mathbf{X}, \mathbf{y})$, the following linear regression model is considered after centering $\mathbf{X}$ and $\mathbf{y}$:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta}^* + \boldsymbol{\varepsilon}, \tag{1}$$

where $\mathbf{y} \in R^n$ denotes the target response which is going to be studied. $\mathbf{X} \in R^{n \times p}$ represents the data matrix with n samples and p variables, and $\boldsymbol{\beta}^* \in R^p$ is the coefficient for the true model. Let $\mathbf{I}$ be the identity matrix, the error term $\boldsymbol{\varepsilon}$ follows Gaussian distribution $N(\mathbf{0}, \sigma^2 \mathbf{I})$ with $\sigma^2 > 0$.

To obtain an interpretable model, the following optimization problem is considered:

$$\min_{\boldsymbol{\beta} \in R^p} \left\{ \frac{\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2}{2n} + \Omega(\boldsymbol{\beta}; \lambda) \right\}, \tag{2}$$

where $\Omega(\boldsymbol{\beta}; \lambda)$ denotes the penalty function with λ representing the tuning parameter. When $\Omega(\boldsymbol{\beta}; \lambda) = \lambda \|\boldsymbol{\beta}\|_1$ which is convex penalty, (2) becomes a well-known LASSO [37] problem given in (3). LASSO is more easy to compute in big data because of its convex form.

$$\text{LASSO} = \min_{\boldsymbol{\beta} \in R^p} \left\{ \frac{\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2}{2n} + \lambda \|\boldsymbol{\beta}\|_1 \right\}. \tag{3}$$

In addition to convex penalty function, nonconvex penalty function is also proposed to perform variable selection. For instance, [38] investigate the SCAD penalty which is given below.

$$\text{SCAD} = \Omega(\boldsymbol{\beta}; \lambda) = \sum_{j=1}^{p} \omega\left(\boldsymbol{\beta}_j; \lambda\right), \tag{4}$$

where $\omega(t; \lambda) = \int_0^{|t|} \{\lambda 1_{\{z \leq \lambda\}} + (a\lambda - z/a - 1)1_{\{z > \lambda\}}\}, a = 3.7$ is selected by general cross-validation. The elastic net (EN) [39] penalty is given as

$$\text{EN} = \Omega(\boldsymbol{\beta}; \lambda, \eta) = \lambda \|\boldsymbol{\beta}\|_1 + \frac{\eta}{2} \|\boldsymbol{\beta}\|_2^2. \tag{5}$$

In this paper, we are going to fulfill the following two tasks: (G1) model interpretation and (G2) forecasting accuracy. Elastic net can be used to achieve these goals because its penalty function consists of both LASSO and ridge penalty. However, its forecasting performance is still affected negatively by the noise level which is difficult to estimate. To solve this problem, square root regularization is considered in our work by using square root error loss function $\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2$ instead of square error

loss function $\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2$. Therefore, we combine the benefits of square root error loss and EN penalty by proposing square root elastic net (SREN) which considers the following optimization problem.

$$\text{SREN} = \min_{\boldsymbol{\beta} \in R^p} \left\{ \frac{\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2}{\sqrt{n}} + \lambda \|\boldsymbol{\beta}\|_1 + \frac{\eta}{2} \|\boldsymbol{\beta}\|_2^2 \right\}. \tag{6}$$

Comparing with EN which takes (5) into account, SREN has the following advantage: two tuning parameters (λ and η), which are determined by $\sqrt{(\log p)/n}$, can be selected properly since they do not involve σ that is extremely difficult to estimate in data analysis. Specifically, it is known that $\sigma^2 = \text{RSS}/(n - p)$, where RSS represents residual sum of square. When $p > n$, σ cannot be estimated. Even when $n > p$, the high coherence causes a large RSS value which results in large σ value. SREN avoids estimating σ in the parameter tuning work which boosts the model forecasting accuracy.

Square Root LASSO (SRL) [40] considers the optimization problem as follows:

$$\text{SRL} = \min_{\boldsymbol{\beta} \in R^p} \left\{ \frac{\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2}{\sqrt{n}} + \lambda \|\boldsymbol{\beta}\|_1 \right\}. \tag{7}$$

Although both SRL and LASSO use the same L_1 penalty, SRL applies square root error loss function which can facilitate the parameter tuning work. Comparing with SRL, SREN adds ridge penalty $(\eta/2)\|\boldsymbol{\beta}\|_2^2$ which is a L_2 type penalty to handle the high coherence between variables and enforce more shrinkage to the model. Although they both apply square root error loss function, SREN is able to get more accurate result in a model with high coherence. Furthermore, SREN applies two tuning parameters (λ and η) to adjust the model performance while SRL just use one tuning parameter λ.

Two novel plans are proposed to design the algorithm for solving (6), which are denoted by Plan A and Plan B, respectively.

(i) Plan A: denote the following:

$$\mathbf{y}_{\text{new}} = \begin{bmatrix} \mathbf{y} \\ \mathbf{0} \end{bmatrix},$$

$$\mathbf{X}_{\text{new}} = \begin{bmatrix} \mathbf{X} \\ \sqrt{\eta/2} \left\| \mathbf{y} - \mathbf{X}\boldsymbol{\beta}^{(j)} \right\|_2^{1/2} \mathbf{I} \end{bmatrix}. \tag{8}$$

The algorithm is designed based on the following iterations:

$$\text{SREN-A} = \boldsymbol{\beta}^{(j+1)} = \arg \min_{\boldsymbol{\beta}^{(j)} \in R^p} \left\{ \left\| \mathbf{y}_{\text{new}} - \mathbf{X}_{\text{new}} \boldsymbol{\beta}^{(j)} \right\|_2^2 \right.$$
$$\left. + \lambda \left\| \mathbf{y} - \mathbf{X}\boldsymbol{\beta}^{(j)} \right\|_2 \left\| \boldsymbol{\beta}^{(j)} \right\|_1 \right\}. \tag{9}$$

Notice that soft thresholding operator is able to be applied to solve (9).

(ii) Plan B: the algorithm is derived based on the following iterations:

$$\text{SREN-B} = \boldsymbol{\beta}^{(j+1)} = \arg \min_{\boldsymbol{\beta}^{(j)} \in R^p}$$

$$\cdot \left\{ \left\| \mathbf{y} - \mathbf{X}\boldsymbol{\beta}^{(j)} \right\|_2^2 + \lambda \left\| \mathbf{y} - \mathbf{X}\boldsymbol{\beta}^{(j)} \right\|_2 \left\| \boldsymbol{\beta}^{(j)} \right\|_1 \right.$$
$$\left. + \frac{\eta \left\| \mathbf{y} - \mathbf{X}\boldsymbol{\beta}^{(j)} \right\|_2}{2} \left\| \boldsymbol{\beta}^{(j)} \right\|_2^2 \right\}. \tag{10}$$

To solve (9) and (10), threshold functions so-called Θ-estimators [41] are applied in our work. The definitions of thresholding rules are given as below.

Definition 1. A thresholding function is a real valued function $\Theta(t; \lambda)$ defined for $-\infty < t < \infty$ and $0 \le \lambda < \infty$ such that

(1) $\Theta(-t; \lambda) = -\Theta(t; \lambda)$,

(2) $\Theta(t; \lambda) \le \Theta(t'; \lambda)$ for $t \le t'$,

(3) $\lim_{t \to \infty} \Theta(-t; \lambda) = \infty$,

(4) $0 \le \Theta(t; \lambda) \le t$ for $0 \le t < \infty$.

It can be told from Definition 1 that $\Theta(\cdot; \lambda)$ is an odd monotone unbounded shrinkage rule for t, at any λ. Θ can be used in a vector manner if either t or λ is given as a vector. The LASSO, SCAD, and EN thresholding functions are provided as follows:

$$\Theta_{\text{SOFT}}(t; \lambda) = \text{sgn}(t)(|t| - \lambda)_+,$$

$$\Theta_{\text{SCAD}}(t; \lambda) = \begin{cases} \lambda, & \text{if } t \le \lambda, \\ \dfrac{(a\lambda - t)}{(a - 1)}, & \text{if } \lambda \le t \le a\lambda, \\ 0, & \text{if } t \ge a\lambda. \end{cases} \tag{11}$$

$$\Theta_{\text{EN}}(t; \lambda, \eta) = \frac{\text{sgn}(t)(|t| - \lambda)_+}{1 + \eta},$$

where $\eta > 0$, $\lambda > 0$ are two regularization parameters.

2.2. Parameter Tuning. Parameter tuning work is of great importance in assuring the performances of forecasting methods. Notice that there are two tuning parameters λ and η used in the proposed method. Cross-validation (CV) is a famous data-driven method which has been widely applied in machine learning community. Given a fixed value for λ and η, the in-sample data will be randomly partitioned into K pieces of roughly equal size. The forecasting model
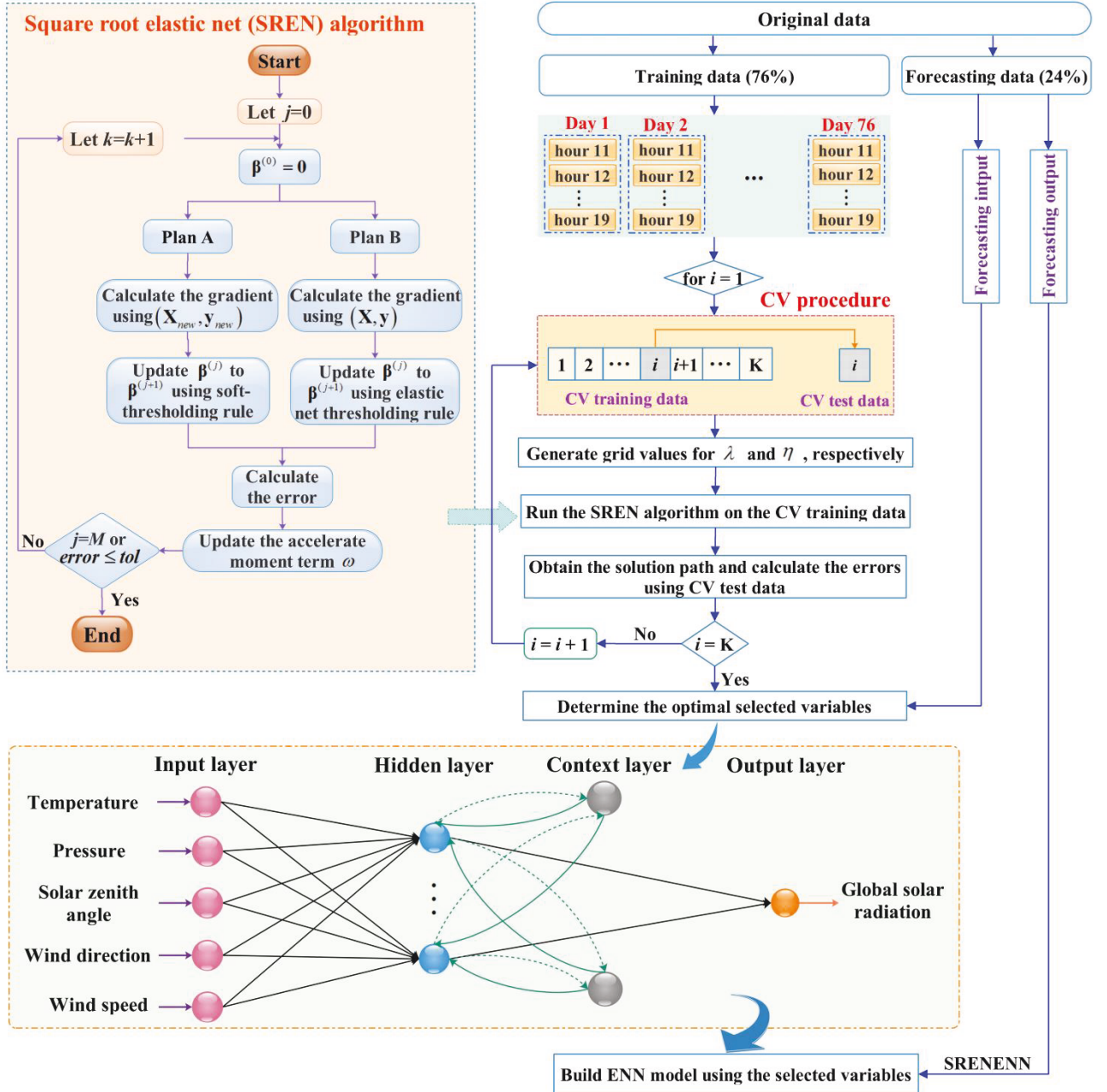
FIGURE 2: Flowchart of SRENENN model.

will be trained using $K - 1$ pieces of in-sample data, and the test error is computed using the Kth piece. CV will repeat this procedure for K times. The CV errors are obtained by adding the test errors, and the optimal regularization parameter is determined by the smallest CV error.

2.3. Elman Neural Network. Elman neural network (ENN) was first advocated by Elman in 1990 to solve speech recognition problem. It is a typically global feed forward local recurrent network. Its main network structure is consist of input layer, hidden layer, and output layer which are also the structure of three-layer feed-forward neural network [42] and backpropagation network [43].

The weights between different layers are going to be trained based on learning rule. The feedback connection has sets of neurons that record the output, and the weights are fixed. There is also a context layer in ENN which stored the output of hidden layer in the previous time point. Comparing with multilayer perceptron, ENN has a short memory and performs the task based on sequence prediction which adapts to time-varying characteristics. The schemes of ENN can be described in the following way:

$$
\begin{aligned}
\mathbf{h}_t &= T_h\left(\mathbf{W}_h \mathbf{x}_t + \mathbf{U}_t \mathbf{h}_{t-1} + \mathbf{b}_h\right), \\
y_t &= T_y\left(\mathbf{W}_y \mathbf{h}_t + \mathbf{b}_y\right),
\end{aligned}
\tag{12}
$$

Inputs: X (centered and scaled), **y** (centered), M: maximum number of iterations, λ, η: tuning parameters, *tol*: error tolerance.
Outputs: forecasting errors.
Step 1. Data Splitting
1. Split the original dataset into Training data $\mathbf{D}_{trn}$ (76% of original data) and test
2. data $\mathbf{D}_{tst}$ (24% of original data).
Scale of data
3. $\mathbf{X} \leftarrow \mathbf{X}/\tau, \mathbf{y} \leftarrow \mathbf{y}/\tau$ with $\tau \geq \|\mathbf{X}\|_2 / \sqrt{2}$;
Step 2. Cross validation
4. Divide the training data $\mathbf{D}_{trn}$ into K folds;
5. **for** $i = 1$ **to** K;
6. Use i-th fold as CV test data **F** and the remaining folds are regarded as CV training data **T**;
7. Generate grid values of $\lambda^G = \{\lambda_u\}_{u=1}^S$ and $\eta^G = \{\eta_v\}_{v=1}^m$.
Step 3. Run SREN algorithm with AGM
8.　**for** $u = 1$ **to** s;
9.　　**for** $v = 1$ **to** m;
10.　　　Initialization: $j \leftarrow 0, \boldsymbol{\beta}^{(0)} \leftarrow 0$;
11.　　　**while** $\|\boldsymbol{\beta}^{(j+1)} - \boldsymbol{\beta}^{(j)}\| < tol$ or $j > M$ do
12.　　　　Step 1. $\boldsymbol{\xi}^{(j)} \leftarrow \boldsymbol{\beta}^{(j)} + (\omega^{(j-1)} - 1/\omega^{(j)})(\boldsymbol{\beta}^{(j)} - \boldsymbol{\beta}^{(j-1)})$
13.　　　　Step 2. (Plan A). $\boldsymbol{\gamma}_{new}^{(j)} \leftarrow \boldsymbol{\xi}^{(j)} + \mathbf{X}_{new}^T(\mathbf{y}_{new} - \mathbf{X}_{new}\boldsymbol{\xi}^{(j)})$
14.　　　　　(Plan B). $\boldsymbol{\gamma}^{(j)} \leftarrow \boldsymbol{\xi}^{(j)} + \mathbf{X}^T(\mathbf{y} - \mathbf{X}\boldsymbol{\xi}^{(j)})$
15.　　　　Step 3. (Plan A). $\boldsymbol{\beta}^{(j+1)} \leftarrow \overleftarrow{\Theta}_{SOFT}(\boldsymbol{\gamma}_{new}^{(j)}; \lambda_u \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}^{(j)}\|_2)$
16.　　　　　(Plan B). $\boldsymbol{\beta}^{(j+1)} \leftarrow \overleftarrow{\Theta}_{EL}(\boldsymbol{\gamma}^{(j)}; \lambda_u \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}^{(j)}\|_2, \eta_v \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}^{(j)}\|_2)$
17.　　　　Step 4. $\omega^{(j+1)} = 1 + \sqrt{1 + 4\omega^{(j)^2}}/2$
18.　　　**end while**
19.　　**end for**
20.　**end for**
21.　Obtain the solution path $B = \{b_{uv}\}$ and corresponding sparsity pattern $G = \{g_{uv}\}$ using CV training data **T**.
Calculate CV errors
22. Calculate CV errors using **F**, B and G. Find the optimal tuning parameters
23. λ_{opt} and η_{opt} with respect to the smallest CV error.
Step 4. Establish elman neural network
24. Determine the optimal model parameters using Training data $\mathbf{D}_{trn}$ with selected variables considered as inputs
Step 5. Evaluate the forecasting performance
26. Calculate the test error using Test data $\mathbf{D}_{tst}$
27. **End for**

ALGORITHM 1: The SRENENN algorithm.

where $\mathbf{x}_t$ is the input vector, $\mathbf{h}_t$ is hidden layer vector, y_t is output layer, **W**, **U**, and **b** are weights and biases of the ENN, and T_h and T_y are activation functions.

The weights in the network are trained by gradient-based backpropagation through time (BPTT). To reduce the model complexity of the neural network, L_2 regularization is often applied and the following optimization problem is considered:

$$\min_{\mathbf{w}} l(\mathbf{y}, \hat{\mathbf{y}}; \mathbf{X}, \mathbf{w}) + \frac{\eta^o}{2}\|\mathbf{w}\|_2^2, \tag{13}$$

where $l(\mathbf{y}, \hat{\mathbf{y}}; \mathbf{X}, \mathbf{w}) = (1/n)\|\mathbf{y} - \hat{\mathbf{y}}(\mathbf{Xw})\|$ denotes the mean square error of the forecasting model. The network weights needed to be estimated are given by **w**. $\hat{\mathbf{y}}$ represents the estimated forecasting value, and η^o is the tuning parameter. Notice that the complexity of neural network depends on the number of input layers and hidden layers. If variable selection method is applied appropriately, the number of
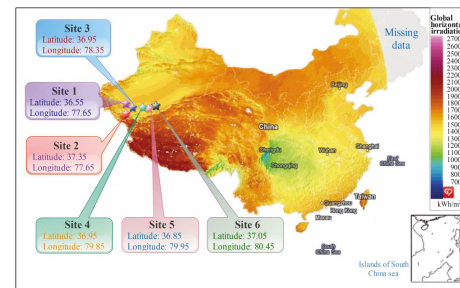

FIGURE 3: Location description about global solar radiation in the Xinjiang area.

inputs will be reduced so that a simple neural network can be constructed.

2.4. Square Root Elastic Net Elman Neural Network Model.
This paper combines the advantages of SREN and ENN

TABLE 1: Detailed description of the training and forecasting sets considered.

Training sets	Forecasting sets	The proportion of forecasting sets	Remark
Mar. 9, Mar. 12, Mar. 19, Mar. 21, Mar. 26, Mar. 29, Apr. 5, Apr. 6, Apr. 13, Apr. 15, Apr. 16, Apr. 21, Apr. 22, Apr. 23, Apr. 24, Apr. 25, Apr. 28, May 1, May 8	May 9, May 16, May 20, May 25, May 28, May 29	24%	Spring
Jun. 3, Jun. 6, Jun. 8, Jun. 11, Jun. 14, Jun. 15, Jun. 21, Jun. 24, Jun. 30, Jul. 4, Jul. 6, Jul. 13, Jul. 18, Jul. 31, Aug. 2, Aug. 3, Aug. 4, Aug. 8, Aug. 12	Aug. 14, Aug. 18, Aug. 23, Aug. 24, Aug. 27, Aug. 29	24%	Summer
Sep. 2, Sep. 4, Sep. 6, Sep. 7, Sep. 10, Sep 20, Sep. 21, Sep. 22, Oct. 4, Oct. 10, Oct. 11, Oct. 13, Oct. 18, Oct. 20, Oct. 21, Oct. 23, Oct. 26, Oct. 29, Nov. 1	Nov. 10, Nov. 12, Nov. 17, Nov. 26, Nov. 27, Nov. 29	24%	Autumn
Dec. 2, Dec. 3, Dec. 4, Dec. 9, Dec. 14, Dec. 16, Dec. 22, Dec. 31, Jan. 3, Jan. 9, Jan. 13, Jan. 22, Jan. 24, Jan. 29, Feb. 1, Feb. 2, Feb. 4, Feb. 7, Feb. 12	Feb. 16, Feb. 18, Feb. 19, Feb. 21, Feb. 22, Feb. 25	24%	Winter

and proposes a novel forecasting model called SRENENN model. The flowchart of SRENENN model is shown in Figure 2, which is designed in the following 5 main steps:

Step 1: Split the original global solar radiation dataset into training dataset and test dataset (cf. Section 3 for more details).

Step 2: CV procedure is applied to training data for selecting the optimal regularization parameters.

Step 3: SREN is used to select the important variables with regularization parameters.

Step 4: Elman neural network is established with variables selected by SREN.

Step 5: The forecasting performance is evaluated using the test dataset.

Algorithm 1 shows SRENENN algorithm with $\omega^{(-1)} = 0$, $\omega^{(0)} = 1$, and $\boldsymbol{\beta}^{(0)} = \boldsymbol{\beta}^{(-1)}$ defined. When $\tau \geq \|\mathbf{X}\|_2/\sqrt{2}$, SRENEENN algorithm converges. However, there is no need to let the algorithm run until convergence to reduce the computation time. The stop criterion of SRENNE algorithm is determined based on trial and error. The convergence error tol is set as $1e - 4$, and maximum number of iteration M is given as 100. SRENENN algorithm uses Accelerated Gradient Method (AGM) [44] to reduce the number of iterations so that the convergence can be achieved using less computation time. AGM has three advantages: (i) it does not involve any computation of inverse of matrix; (ii) paralleling the selection of unknown parameter and computation of gradient; (iii) making use of momentum to increase the convergence speed.

Define $F(\boldsymbol{\beta}) = \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2 + \lambda\|\boldsymbol{\beta}\|_1 + \eta\|\boldsymbol{\beta}\|_2^2/2$, and the convergence of SRENENN algorithm is guaranteed theoretically in Theorem 1 whose proof is shown in Appendix.

Theorem 1. *Let τ be the step size of SRENENN algorithm and $\tau \geq \|\mathbf{X}\|_2/\sqrt{2}$, and assume the following regularity condition*

TABLE 2: Parameter values of seven forecasting models at six sites.

Models	h	λ	η	N	Func1	Func2
ENN	5	—	—	2000	Tansig	Tansig
LASSOENN	10	4	—	2000	Tansig	Tansig
PCAENN	5	—	—	2000	Tansig	Tansig
SCADENN	15	1	—	2000	Tansig	Tansig
SRLENN	10	148.8	—	2000	Tansig	Tansig
SRENENN-A	5	0.0625	$5e - 5$	2000	Tansig	Tansig
SRENENN-B	15	0.0625	$5e - 5$	2000	Tansig	Tansig

hold $\inf_{\boldsymbol{\xi} \in A} \|\mathbf{X}\boldsymbol{\xi} - \mathbf{y}\|_2 > 0$, where A collects all the linear combination of $\boldsymbol{\beta}^{(j)}$ and $\boldsymbol{\beta}^{(j+1)}$. Then the following inequality holds

$$F\left(\boldsymbol{\beta}^{(j)}\right) - F\left(\boldsymbol{\beta}^{(j+1)}\right) \geq c\left\|\boldsymbol{\beta}^{(j+1)} - \boldsymbol{\beta}^{(j)}\right\|_2^2, \quad (14)$$

for some $c > 0$.

3. Case Studies

For the real data application, six sites from Xinjiang area in China are considered to demonstrate the advantages of the proposed SRENENN model via comparisons with traditional methods.

3.1. Data Description. Qinghai, Tibet, Xinjiang, and Inner Mongolia are suitable locations to install photovoltaic power station because sunshine is quite rich in these areas. That is why six sites (Site 1, Site 2, Site 3, Site 4, Site 5, and Site 6) are selected from these regions. The including latitudes and longitudes of six sites are provided in Figure 3. The dataset applied in this work is collected from National Renewable Energy Laboratory (NREL) which is available at http://www.nrel.gov/gis/solar.html. In addition to global solar radiation which is going to be studied, seven meteorological variables including solar zenith angle, precipitation, temperature, wind direction, wind speed, relative humidity, and pressure are

Table 3: Mean forecasting errors and computation time (in seconds) of forecasting models at six sites.

|	Models	MAE (W/m^2)	RMSE (W/m^2)	MAPE (%)	TIC (%)	Time (seconds)
	SVM	16.04	19.89	2.67	1.31	9.21
	ENN	20.98	27.38	5.59	1.8	7.12
	LASSOENN	14.5	18.98	3.77	1.25	2.15
	PCAENN	18.45	24.24	4.79	1.59	2.13
Site 1	SCADENN	16.44	21.36	4.23	1.4	5.82
	SRLENN	14.87	18.88	2.83	1.24	6.13
	SRENENN-A	14.43	18.26	2.73	1.2	3.14
	SRENENN-B	13.17	16.51	2.57	1.08	2.58
	SVM	15.88	22.87	3.69	1.75	15.12
	ENN	21.39	27.56	7.62	2.11	12.12
	LASSOENN	17.3	21.48	5.71	1.63	3.13
	PCAENN	19.06	24.2	6.31	1.85	2.95
Site 2	SCADENN	19.87	24.75	7.08	1.89	2.73
	SRLENN	16.77	20.72	3.7	1.57	2.82
	SRENENN-A	18.8	23.37	3.82	1.78	3.21
	SRENENN-B	13.63	18.41	3.19	1.41	3.54
	SVM	18.04	33.73	3.48	2.47	16.12
	ENN	19.44	26.29	7.01	1.91	14.12
	LASSOENN	16.32	21.04	5.45	1.53	3.51
	PCAENN	19.01	25.85	7.33	1.88	5.12
Site 3	SCADENN	20.19	26.16	6.32	1.9	4.78
	SRLENN	19.67	24.97	3.92	1.8	4.55
	SRENENN-A	10.61	13.80	2.99	1.15	4.82
	SRENENN-B	13.36	17.29	3.21	1.25	3.03
	SVM	17.31	28.89	4.36	2.29	18.21
	ENN	24.87	33.61	11.63	2.66	17.88
	LASSOENN	14.66	19.11	6.2	1.51	23.12
	PCAENN	17.97	24.08	8.85	1.9	3.54
Site 4	SCADENN	19.69	25.47	9.44	2.01	3.63
	SRLENN	18.63	24.04	4.87	1.89	4.15
	SRENENN-A	18.22	25.83	4.83	2.04	5.12
	SRENENN-B	13.50	17.52	3.81	1.39	3.09
	SVM	17.66	30.32	4.33	2.38	13.14
	ENN	24.85	33.74	11.32	2.64	10.25
	LASSOENN	14.62	19.04	6.06	1.49	4.53
	PCAENN	19.23	26.69	11.29	2.08	4.34
Site 5	SCADENN	19.84	25.61	9.23	2.00	3.12
	SRLENN	18.30	23.61	4.69	1.84	4.13
	SRENENN-A	18.26	25.79	4.76	2.01	3.58
	SRENENN-B	14.35	18.85	4.04	1.47	4.12
	SVM	15.87	22.75	4.91	1.83	12.27
	ENN	24.19	31.31	13.79	2.51	12.34
	LASSOENN	14.43	18.76	7.58	1.50	1.77
	PCAENN	21.27	28.38	13.63	2.26	1.93
Site 6	SCADENN	21.43	28.35	13.13	2.25	1.88
	SRLENN	18.11	23.52	5.27	1.87	2.12
	SRENENN-A	16.26	23.10	5.29	1.84	1.36
	SRENENN-B	13.10	16.91	4.23	1.35	1.22
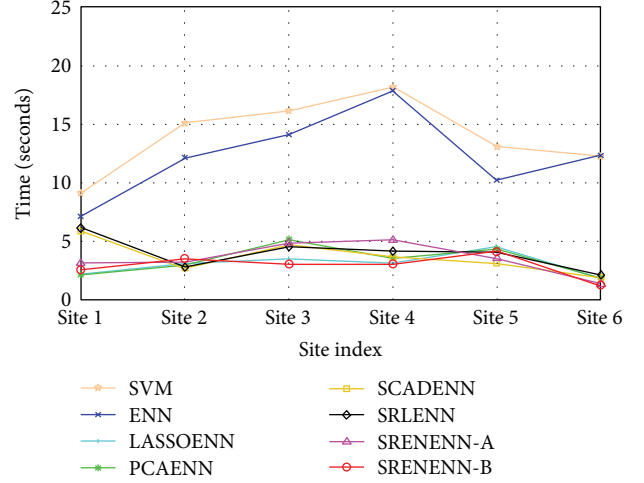
Figure 4: The computation time (in seconds) of eight models.

Table 4: The total scores of compared methods for evaluation criteria.

Models	MAE	RMSE	MAPE	TIC	Total
SVM	20	34	13	34	101
ENN	46	47	47	47	187
LASSOENN	16	16	30	16	78
PCAENN	34	35	41	35	145
SCADENN	42	34	38	34	148
SRLENN	30	20	21	20	91
SRENENN-A	21	23	19	23	86
SRENENN-B	7	7	7	7	28

provided in the dataset. The samples of this dataset are collected based on the global solar radiation from 11:00 am to 19:00 pm in 2014 because the solar sources are very abundant during this time interval. The main purpose of this paper was to choose important variables from seven meteorological variables to perform the forecasting task. The strategy of splitting the data into training data and test data is given as follows: 19 days of months in each season are randomly selected as the training data to establish the forecasting model. The test data consists of 6 days which are also selected randomly from the remaining days in every season. Thus, the size of training data is 684 (19 days × 9 hours × 4 seasons = 684), and the size of test data is 216 (6 days × 9 hours × 4 seasons = 216). Furthermore, experiments on each season are also implemented based on the training samples (19 days × 9 hours = 171) and forecasting samples (6 days × 9 hours = 54) presented in Table 1; it is observed that the forecasting performance of different models is going to be tested using four seasons, and the forecasting samples in each season take up approximately 24%, which is a reasonable proportion.

3.2. Evaluation Criterion. To evaluate the forecasting performances of the proposed method and other comparing

Table 5: Mean forecasting errors of forecasting models for each season at Site 1, Site 2, and Site 3.

	MAE (W/m^2)	RMSE (W/m^2)	MAPE (%)	TIC (%)
Site 1				
Spring				
SVM	13.93	16.07	2.30	1.05
ENN	15.21	18.44	2.30	1.22
LASSOENN	12.27	14.68	1.77	0.97
PCAENN	13.12	17.06	1.84	1.12
SCADENN	17.8	20.58	2.92	1.37
SRLENN	14.88	17.19	2.31	1.14
SRENENN-A	14.19	16.88	2.38	1.11
SRENENN-B	11.59	13.83	1.71	0.91
Summer				
SVM	15.36	19.12	2.03	1.07
ENN	21.57	26.04	2.49	1.46
LASSOENN	16.02	18.82	1.88	1.06
PCAENN	23.81	26.81	2.82	1.50
SCADENN	19.40	22.26	2.30	1.25
SRLENN	17.41	20.38	2.22	1.14
SRENENN-A	16.70	19.91	2.12	1.11
SRENENN-B	13.04	15.15	1.6	0.85
Autumn				
SVM	21.58	27.91	2.67	1.69
ENN	19.50	24.46	2.54	1.47
LASSOENN	10.79	14.50	1.44	0.87
PCAENN	14.56	18.55	1.90	1.11
SCADENN	13.59	17.98	1.78	1.08
SRLENN	16.18	22.92	2.03	1.36
SRENENN-A	16.12	20.79	2.01	1.24
SRENENN-B	8.81	11.08	1.20	0.66
Winter				
SVM	15.63	17.78	4.10	1.80
ENN	27.64	37.19	15.03	3.73
LASSOENN	18.92	25.70	9.99	2.56
PCAENN	22.33	31.59	12.61	3.17
SCADENN	21.16	28.53	11.11	2.86
SRLENN	11.04	13.76	4.77	1.37
SRENENN-A	10.73	14.83	4.42	1.48
SRENENN-B	10.71	13.37	4.15	1.34
Site 2				
Spring				
SVM	19.5	22.91	3.61	1.78
ENN	14.83	18.85	2.60	1.45
LASSOENN	13.35	16.47	2.41	1.27
PCAENN	14.74	18.89	2.90	1.84
SCADENN	17.77	21.03	3.06	1.61
SRLENN	16.44	19.89	3.44	1.53
SRENENN-A	18.92	23.01	3.81	1.78
SRENENN-B	10.53	12.87	2.1	0.99

Table 5: Continued.

	MAE (W/m^2)	RMSE (W/m^2)	MAPE (%)	TIC (%)
Summer				
SVM	22.12	34.33	2.90	2.23
ENN	24.32	29.48	3.26	1.92
LASSOENN	20.93	24.54	2.87	1.58
PCAENN	21.88	25.04	2.72	1.34
SCADENN	22.59	26.53	3.15	1.71
SRLENN	20.98	24.88	3.01	1.60
SRENENN-A	23.03	28.38	3.34	1.82
SRENENN-B	13.94	20.55	2.12	1.33
Autumn				
SVM	9.63	12.15	1.53	0.84
ENN	17.89	21.97	2.59	1.52
LASSOENN	14.08	17.56	2.19	1.21
PCAENN	13.17	15.65	1.89	1.22
SCADENN	14.84	18.14	2.39	1.25
SRLENN	16.95	21.03	2.58	1.44
SRENENN-A	18.10	22.07	2.61	1.52
SRENENN-B	8.95	10.84	1.39	0.75
Winter				
SVM	21.27	24.49	8.23	2.99
ENN	28.51	36.50	22.03	4.42
LASSOENN	20.86	25.78	15.37	3.12
PCAENN	25.91	26.29	17.84	2.81
SCADENN	24.31	31.23	19.73	3.76
SRLENN	12.73	16.14	5.78	1.92
SRENENN-A	15.13	19.03	5.51	2.27
SRENENN-B	12.16	17.07	5.67	2.04
Site 3				
Spring				
SVM	16.97	20.55	3.16	1.5
ENN	15.62	19.66	2.62	1.43
LASSOENN	13.33	16.38	2.32	1.19
PCAENN	11.66	14.22	1.97	1.03
SCADENN	14.93	18.49	2.56	1.34
SRLENN	17.39	20.70	3.18	1.51
SRENENN-A	12.39	15.61	2.54	1.14
SRENENN-B	9.36	12.44	1.75	0.91
Summer				
SVM	17.34	34.33	2.26	2.13
ENN	19.72	23.29	2.57	1.45
LASSOENN	19.16	22.95	2.38	1.42
PCAENN	22.52	25.87	3.06	1.60
SCADENN	26.78	32.93	3.49	2.04
SRLENN	23.64	29.05	3.36	1.79
SRENENN-A	8.88	11.28	1.25	0.70
SRENENN-B	12.78	17.88	1.91	1.10

TABLE 5: Continued.

	MAE (W/m^2)	RMSE (W/m^2)	MAPE (%)	TIC (%)
Autumn				
SVM	26.54	51.25	3.69	3.45
ENN	16.42	21.89	2.38	1.45
LASSOENN	12.63	16.50	1.89	1.09
PCAENN	14.68	19.15	2.25	1.27
SCADENN	14.31	18.32	2.23	1.21
SRLENN	21.70	28.29	3.11	1.86
SRENENN-A	11.71	15.31	1.71	1.01
SRENENN-B	8.46	11.1	1.21	0.74
Winter				
SVM	18.94	24.32	6.23	2.79
ENN	25.98	36.83	20.67	4.17
LASSOENN	20.18	26.52	15.20	3.00
PCAENN	27.22	37.90	22.05	4.27
SCADENN	24.75	31.25	17.00	3.53
SRLENN	15.99	20.51	6.05	2.29
SRENENN-A	12.69	16.37	6.97	1.84
SRENENN-B	11.97	14.83	6.07	1.66

approaches, several criteria including mean absolute percent error (MAPE), mean absolute error (MAE), root mean square error (RMSE), and Theil inequality coefficient (TIC) are applied as evaluation criteria [45]. Let y_i be true value, $\widehat{y}_i$ represents the estimated value, and N denotes to be the sample size of test data. The best forecasting model provides the lowest MAPE, MAE, RMSE, and TIC. The evaluation criteria are provided as below.

$$\text{MAPE} = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{y_i - \widehat{y}_i}{y_i} \right| \times 100\%,$$

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^{N} |y_i - \widehat{y}_i|,$$

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_i - \widehat{y}_i)^2},$$

$$\text{TIC} = \frac{\sqrt{(1/N) \sum_{i=1}^{N} (y_i - \widehat{y}_i)^2}}{\sqrt{(1/N) \sum_{i=1}^{N} \widehat{y}_i^2} + \sqrt{(1/N) \sum_{i=1}^{N} y_i^2}} \times 100\%.$$

(15)

3.3. Wilcoxon Signed-Rank Test. The Wilcoxon signed-rank test was applied to determine if the proposed SRENENN model was superior to the SVM, ENN, LASSOENN, PCAENN, SCADENN, and SRLENN models for global solar radiation. The Wilcoxon signed-rank test is a nonparametric statistical hypothesis test used when comparing two matched or related samples to assess whether their population median ranks differ (i.e., it is a paired difference test). It can be used as an alternative to the parametric t-test.

Let N be the sample size or the number of pairs. The prediction sample size of each method was 216 (6 days × 9 hours × 4 seasons = 216), and thus, $N = 216$. For $i = 1, 2, \ldots, N$, let $y_{1,i}$ and $y_{2,i}$ be the forecasting values of two different approaches and consider the following hypothesis:

$$\text{Null hypothesis: } H_0 : \mu_1 \le \mu_2,$$

$$\text{Alternative hypothesis: } H_1 : \mu_1 > \mu_2,$$

(16)

where μ_1 and μ_2 are medians of sequences $Y_1 = \{y_{1,i} \mid i = 1, 2, \ldots, N\}$ and $Y_2 = \{y_{2,i} \mid i = 1, 2, \ldots, N\}$. The details of Wilcoxon signed-rank test process can be listed as follows [46–48]:

Step 1: For $i = 1, 2, \ldots, N$, calculate $|y_{2,i} - y_{1,i}|$ and sgn $(y_{2,i} - y_{1,i})$, where sgn $(\cdot)$ is the sign function.

Step 2: Exclude pairs with $|y_{2,i} - y_{1,i}| = 0$.

Step 3: Let N_r be the reduced sample size. Order the remaining N_r pairs from smallest absolute difference to largest absolute difference $|y_{2,i} - y_{1,i}|$. Rank the pairs, starting with the smallest as rank 1.

Step 4: Calculate the sum $\mathbf{W}^+$ of the positive ranks and the sum $\mathbf{W}^-$ of the negative ranks.

Step 5: As N_r increases, the sampling distribution of $\mathbf{W}$ converges to a normal distribution. Thus, for larger samples, $\mathbf{Z}$ statistic can be calculated as $\mathbf{Z} = (\mathbf{W} - N_r(N_r + 1))/4/\sqrt{(N_r(N_r + 1)(2N_r + 1))/6}$. If $\mathbf{Z} > \mathbf{Z}_{\text{critical}}$, then reject H_0. For small samples, $\mathbf{W}$ can be calculated as $\mathbf{W} = \min (\mathbf{W}^+, \mathbf{W}^-)$. If $\mathbf{W} \ge \mathbf{W}_{\text{critical},N_r}$, then reject H_0. Alternatively, a p value can be calculated from enumeration of all possible combinations of $\mathbf{Z}$ or $\mathbf{W}$ given N_r.

3.4. Friedman Test. Friedman test is a nonparametric statistical test which can be applied to evaluate the performances of forecasting methods based on different criteria on multiple datasets [49]. The Friedman test considers the following hypothesis:

$$\text{Null hypothesis: } H_0 : R_1 = R_2 = \cdots = R_k,$$

$$\text{Alternative hypothesis: } H_1 : \text{not } H_0,$$

(17)

where $R_j = (1/B)\sum_i r_i^j$ with r_i^j representing the rank of jth of k algorithms of B datasets. Based on the Friedman statistics [50]

$$\chi_F^2 = \frac{12B}{k(k+1)} \left[\sum_j R_j^2 - \frac{k(k+1)^2}{4} \right],$$

(18)

which follows chi square distribution $\chi_F^2(k-1)$, an F-distribution statistics $F(k-1, (k-1)(B-1))$ is calculated

$$F = \frac{(B-1)\chi_F^2}{B(k-1) - \chi_F^2}.$$

(19)

TABLE 6: Mean forecasting errors of forecasting models for each season at Site 4, Site 5, and Site 6.

	MAE (W/m²)	RMSE (W/m²)	MAPE (%)	TIC (%)
Site 4				
Spring				
SVM	11.99	14.45	2.33	1.15
ENN	17.06	20.67	3.28	1.64
LASSOENN	11.38	14.50	2.18	1.16
PCAENN	16.49	19.51	2.93	1.57
SCADENN	14.18	17.57	2.38	1.40
SRLENN	16.86	19.77	3.63	1.57
SRENENN-A	13.43	18.52	2.92	1.47
SRENENN-B	11.2	13.47	2.02	1.07
Summer				
SVM	27.37	49.61	3.71	3.37
ENN	30.07	41.78	4.46	2.81
LASSOENN	16.15	22.13	2.31	1.49
PCAENN	22.87	28.99	3.39	1.94
SCADENN	28.87	34.23	4.24	2.30
SRLENN	25.00	30.95	3.92	2.06
SRENENN-A	17.75	24.39	2.82	1.64
SRENENN-B	15.45	20.99	2.32	1.41
Autumn				
SVM	14.13	16.85	2.11	1.21
ENN	27.18	33.28	4.44	2.38
LASSOENN	15.23	18.13	2.48	1.29
PCAENN	19.43	25.01	3.07	1.78
SCADENN	14.96	17.97	2.61	1.28
SRLENN	19.23	26.72	2.91	1.90
SRENENN-A	21.42	28.89	3.31	2.05
SRENENN-B	8.15	10.00	1.30	0.71
Winter				
SVM	15.77	19.62	9.32	2.47
ENN	25.19	35.20	34.32	4.42
LASSOENN	15.89	20.81	17.83	2.59
PCAENN	18.37	25.93	26.93	3.22
SCADENN	20.77	28.14	28.53	3.51
SRLENN	13.9	17.51	8.71	2.17
SRENENN-A	20.29	29.96	10.28	3.71
SRENENN-B	13.44	15.8	9.01	1.96
Site 5				
Spring				
SVM	11.85	14.32	2.29	1.12
ENN	16.93	20.54	3.22	1.61
LASSOENN	12.53	15.59	2.16	1.24
PCAENN	19.09	23.18	3.04	1.85
SCADENN	13.65	17.28	2.16	1.37
SRLENN	16.01	18.61	3.39	1.46
SRENENN-A	13.55	18.24	2.90	1.43
SRENENN-B	7.84	9.61	1.36	0.76

TABLE 6: Continued.

	MAE (W/m²)	RMSE (W/m²)	MAPE (%)	TIC (%)
Summer				
SVM	28.51	52.75	3.81	3.56
ENN	30.20	42.22	4.45	2.81
LASSOENN	17.22	22.83	2.47	1.53
PCAENN	22.79	28.43	3.25	1.90
SCADENN	28.76	34.71	4.09	2.32
SRLENN	24.72	30.74	3.84	2.03
SRENENN-A	17.96	24.54	2.84	1.63
SRENENN-B	16.2	22.16	2.19	1.48
Autumn				
SVM	14.80	17.95	2.18	1.27
ENN	27.06	33.35	4.36	2.36
LASSOENN	18.35	20.48	2.89	1.45
PCAENN	21.32	26.12	3.31	1.85
SCADENN	15.63	19.97	2.69	1.41
SRLENN	18.73	26.20	2.81	1.84
SRENENN-A	21.16	28.65	3.26	2.01
SRENENN-B	11.28	13.70	1.73	0.97
Winter				
SVM	15.50	19.19	9.05	2.38
ENN	25.22	35.26	33.26	4.37
LASSOENN	15.61	20.49	16.91	2.53
PCAENN	18.37	25.71	25.61	3.18
SCADENN	21.96	28.54	27.46	3.54
SRLENN	13.77	15.86	8.71	1.94
SRENENN-A	20.37	30.09	10.11	3.68
SRENENN-B	13.44	16.45	8.01	2.03
Site 6				
Spring				
SVM	13.38	14.96	2.59	1.21
ENN	18.14	21.81	3.45	1.76
LASSOENN	19.96	23.4	4.37	1.9
PCAENN	18.29	20.70	3.26	1.66
SCADENN	16.02	19.22	2.97	1.55
SRLENN	18.05	21.84	4.23	1.76
SRENENN-A	12.47	18.35	2.68	1.49
SRENENN-B	10.68	13.4	2.09	1.09
Summer				
SVM	22.51	35.64	3.20	2.44
ENN	26.88	33.72	3.95	2.30
LASSOENN	15.37	20.50	2.17	1.39
PCAENN	22.66	28.81	3.41	1.95
SCADENN	29.16	34.75	4.32	2.36
SRLENN	20.60	26.34	3.19	1.78
SRENENN-A	15.07	21.66	2.34	1.47
SRENENN-B	11.35	14.38	1.71	0.98

TABLE 6: Continued.

	MAE (W/m^2)	RMSE (W/m^2)	MAPE (%)	TIC (%)
Autumn				
SVM	13.14	15.96	1.95	1.15
ENN	28.09	34.57	4.59	2.49
LASSOENN	16.30	18.74	2.71	1.35
PCAENN	22.33	29.98	3.75	2.14
SCADENN	21.03	28.11	3.50	2.01
SRLENN	18.86	25.69	2.99	1.84
SRENENN-A	23.31	29.66	3.60	2.13
SRENENN-B	9.52	12.37	1.48	0.89
Winter				
SVM	14.46	17.97	11.90	2.31
ENN	23.66	33.36	43.17	4.28
LASSOENN	15.37	21.36	23.33	2.70
PCAENN	21.83	32.64	44.10	4.14
SCADENN	19.54	29.11	41.72	3.70
SRLENN	14.91	19.57	10.66	2.46
SRENENN-A	14.20	21.17	12.54	2.67
SRENENN-B	11.57	15.39	9.36	1.94

If the null hypothesis H_0 is rejected which means there exist significant differences between the comparing algorithms, a post hoc test will be given based on critical difference (CD).

3.5. Statistical Analysis of Selected Variables. To test whether the selected variables are significant or not, the following F test statistic is considered.

$$F = \frac{(\mathrm{RSS}_1 - \mathrm{RSS}_0)/(p_1 - p_0)}{\mathrm{RSS}_0/(n - p_1)}, \qquad (20)$$

where RSS_0 is the residual sum of square for the least square fit of full model with p_1 variables. And the same for the smaller reduced model with p_0 variables. Under the Gaussian assumption and the null hypothesis that the smaller model is correct, the F test statistic will have a $F(\alpha, p_1 - p_0, n - p_1)$ distribution. If the F value is larger than the critical value, then the selected variables are determined to be significant.

4. Results and Discussion

In this paper, SREN is combined with the Elman neural network (ENN) to select the important variables and forecast the global solar radiation. SREN is a penalized variable selection method using convex penalty function which is computational efficient. Comparing with subset variable selection which considers all the possible combinations of the variables, SREN selected all of the important variables directly.

Lots of approaches are considered for global solar radiation such as SVM, ENN, LASSOENN, SCADENN, SRLENN, and PCAENN [50]. Comparisons between these methods and SRENENN are presented in this part.

Table 2 shows the parameters applied in establishing the comparing forecasting models. The regularization parameters λ and η are selected as 0.0625 and $5e-5$ in SRENENN methods using 10-fold cross-validation. In LASSOENN and SCADENN, the regularization parameters are chosen as 4 and 1. N represents the maximum number of iterations to establish neural network and is set as 2000. The activation function from input layer to hidden layer (Func1) is given as Tansig transfer function based on trial and error. Similarly, denote Func2 to be the activation function from hidden layer to output layer and it is set as Tansig transfer function. The selection of the number of hidden neurons N_h which determines the model complexity is important in constructing an ENN. The best value for N_h is selected from a generated grid values {5, 10, 15, 20}. The back propagation through time (BPTT) is applied to construct an ENN with the weights initialized using random values from the uniform distribution $U(0, 1)$. Based on trial and error, the gradient descent with momentum and adaptive learning rate are set as 0.9 and 0.01. SVM is implemented using R package "e1071" with rbf kernel function with two unknown parameters (γ, C) selected from two grids $\{2^{-5}, 2^{-4}, \ldots, 2^{-1}\}$ and $\{2^2, 2^3, 2^4\}$ using 10-fold cv. All the parameters are selected in all the models by proper tuning work.

The results presented in Table 3 reveal that SRENENN-B achieves the best results in terms of forecasting accuracy on average in all the sites except Site 3 where SRENENN-A has the best performance. The significant differences are observed among SRENENN-B, SVM, and ENN methods which do not involve any dimension reduction. For instance, MAE obtained by SRENENN-B is much lower than ENN in Site 2. The error has been reduced by about 37.23% using fewer variables. Comparing with SVM which also has a good performance, SRENENN-B improves the forecasting accuracy by 17.89%. In Site 4, SRENENN-B has boosted the RMSE of the ENN and SVM by 47.87% and 39.36%, respectively. Comparing with PCAENN, LASSOENN, and SCADENN which performs better than ENN, SRENENN-B is still the winner in terms of MAE, RMSE, MAPE, and TIC. It is easy to observe that the PCAENN, LASSOENN, and SCADENN provide similar performances but LASSOENN delivers better results than PCAENN, SCADENN, and SRLENN in almost all the sites. From the aspect of MAPE, LASSOENN provides better results in all sites except Site 1 and Site 2. Further, it was noticed that the performance of SVM is better than SRLENN in terms of MAE in Sites 2–6 and SRENENN-A outperforms SRLENN except Site 2. On the other hand, the computation time of different forecasting methods is shown in the last column of Table 3. Obviously, it takes SRENENN-B less computation time than other approaches. Both SVM and ENN which take all the variables as inputs use more computation time than penalized ENN and PCAENN. The computation time of other forecasting methods is comparable. The corresponding plot is shown in

TABLE 7: The results of Wilcoxon signed-rank tests between SRENENN-B and other competitors (p values are in the parenthesis).

Sites	SVM	ENN	LASSOENN	PCAENN	SCADENN	SRLENN	SRENENN-A
Site 1	5.58 (0.00)	3.91 (0.00)	4.67 (0.00)	4.48 (0.00)	5.47 (0.00)	4.92 (0.00)	3.22 (0.00)
Site 2	6.30 (0.00)	4.33 (0.00)	4.31 (0.00)	5.69 (0.00)	6.37 (0.00)	4.68 (0.00)	2.68 (0.00)
Site 3	3.51 (0.00)	2.33 (0.00)	5.85 (0.00)	4.30 (0.00)	5.95 (0.00)	4.90 (0.00)	5.71 (0.00)
Site 4	4.95 (0.00)	4.41 (0.00)	6.24 (0.00)	6.18 (0.00)	6.06 (0.00)	4.53 (0.00)	3.69 (0.00)
Site 5	2.78 (0.00)	3.91 (0.00)	5.20 (0.00)	6.18 (0.00)	6.06 (0.00)	4.53 (0.00)	4.36 (0.00)
Site 6	5.69 (0.00)	3.02 (0.00)	4.81 (0.00)	5.14 (0.00)	5.99 (0.00)	5.34 (0.00)	3.69 (0.00)

TABLE 8: The results of Wilcoxon signed-rank tests between SRENENN-B and other competitors for each season at Sites 1–6 (p values are in the parenthesis).

	Models	ENN	LASSOENN	PCAENN	SCADENN	SRENENN-A	SRENENN-B
	Spring	0.38 (0.00)	1.53 (0.00)	2.47 (0.00)	0.45 (0.00)	2.04 (0.00)	3.07 (0.00)
	Summer	1.16 (0.00)	5.12 (0.00)	0.52 (0.00)	2.88 (0.00)	0.05 (0.00)	1.15 (0.00)
Site 1	Autumn	1.60 (0.00)	2.42 (0.00)	2.99 (0.00)	3.18 (0.00)	2.44 (0.00)	4.50 (0.00)
	Winter	3.28 (0.00)	2.69 (0.00)	2.09 (0.00)	1.78 (0.00)	2.31 (0.00)	3.18 (0.00)
	Spring	1.63 (0.00)	1.65 (0.00)	2.70 (0.00)	3.20 (0.00)	3.53 (0.00)	2.71 (0.00)
	Summer	2.10 (0.00)	2.18 (0.00)	2.74 (0.00)	3.20 (0.00)	2.38 (0.00)	1.72 (0.00)
Site 2	Autumn	2.32 (0.00)	2.10 (0.00)	1.57 (0.00)	3.20 (0.00)	2.90 (0.00)	3.49 (0.00)
	Winter	1.27 (0.00)	0.56 (0.00)	0.62 (0.00)	3.20 (0.00)	1.76 (0.00)	2.31 (0.00)
	Spring	3.07 (0.00)	1.56 (0.00)	2.25 (0.00)	1.38 (0.00)	2.70 (0.00)	5.94 (0.00)
	Summer	2.26 (0.00)	2.17 (0.00)	1.58 (0.00)	2.08 (0.00)	2.18 (0.00)	2.31 (0.00)
Site 3	Autumn	1.21 (0.00)	2.62 (0.00)	2.87 (0.00)	3.06 (0.00)	1.81 (0.00)	2.99 (0.00)
	Winter	1.22 (0.00)	2.02 (0.00)	0.97 (0.00)	1.87 (0.00)	2.11 (0.00)	0.75 (0.00)
	Spring	0.19 (0.00)	1.24 (0.00)	3.45 (0.00)	2.22 (0.00)	5.02 (0.00)	2.88 (0.00)
	Summer	2.95 (0.00)	2.38 (0.00)	2.28 (0.00)	2.73 (0.00)	3.13 (0.00)	1.19 (0.00)
Site 4	Autumn	0.08 (0.00)	2.39 (0.00)	3.18 (0.00)	2.83 (0.00)	2.19 (0.00)	2.86 (0.00)
	Winter	0.95 (0.00)	2.79 (0.00)	2.25 (0.00)	2.12 (0.00)	2.98 (0.00)	2.39 (0.00)
	Spring	1.97 (0.00)	0.84 (0.00)	0.45 (0.00)	0.22 (0.00)	4.02 (0.00)	3.02 (0.00)
	Summer	2.99 (0.00)	3.17 (0.00)	2.28 (0.00)	2.73 (0.00)	3.13 (0.00)	1.41 (0.00)
Site 5	Autumn	3.08 (0.00)	2.83 (0.00)	3.18 (0.00)	2.83 (0.00)	2.19 (0.00)	2.03 (0.00)
	Winter	1.87 (0.00)	3.02 (0.00)	2.25 (0.00)	2.12 (0.00)	2.98 (0.00)	2.88 (0.00)
	Spring	2.78 (0.00)	1.39 (0.00)	1.74 (0.00)	3.56 (0.00)	3.56 (0.00)	2.01 (0.00)
	Summer	3.11 (0.00)	3.09 (0.00)	1.18 (0.00)	2.11 (0.00)	1.87 (0.00)	1.53 (0.00)
Site 6	Autumn	1.91 (0.00)	2.58 (0.00)	1.81 (0.00)	2.11 (0.00)	1.04 (0.00)	3.14 (0.00)
	Winter	2.37 (0.00)	1.54 (0.00)	1.43 (0.00)	2.55 (0.00)	1.73 (0.00)	2.47 (0.00)

Figure 4. Therefore, SRENENN-B delivers better forecasting results with less computation time.

Table 4 depicts the scores of the compared models. The best model will give the lowest total score. Obviously, SRENENN-B provides the lowest score among all the compared methods (see the last column), followed by LASSOENN, SRENENN, SRLENN, SVM, PCAENN, SCADENN, and ENN. Tables 5 and 6 show the performances of compared forecasting approaches including four seasons. It is not difficult to find SRENENN-B that gives the highest forecasting accuracy. SVM provides better results than other methods in spring, autumn, and winter while ENN gives the worst result in four seasons. The results are quite similar as what we observe in Table 3.

The results using Wilcoxon signed-rank test between SRENENN-B and other forecasting approaches are summarized in Tables 7 and 8, which show Z statistic values and p values. In this study, the significant level is set as 0.05 so that the Z critical value is 1.96. From the tables, it is easy to observe that all of the Z statistic values are larger than 1.96 and p values are much smaller than 0.05. Thus, the null hypothesis is rejected and we decide that the proposed SRENENN-B model is significantly different from the other models. Since SRENENN-B has provided the smallest errors at all sites, it is concluded that SRENENN-B is superior over other models in terms of forecasting accuracy.

Table 9 reveals MAE, RMSE, MAPE, and TIC values of the SRENENN and other forecasting approaches. The results

TABLE 9: The results based on Friedman test for compared methods.

	Site 1	Site 2	Site 3	Site 4	Site 5	Site 6	Average	χ^2_F	F_F
MAE									
SVM	5	2	4	3	3	3	3.33		
ENN	8	8	6	8	8	8	7.67		
LASSOENN	3	4	3	2	2	2	2.67		
PCAENN	7	6	5	4	6	6	5.67	34.70	23.86
SCADENN	6	7	8	7	7	7	7.00		
SRLENN	4	3	7	6	5	5	5.00		
SRENENN-A	2	5	1	5	4	4	3.50		
SRENENN-B	1	1	2	1	1	1	1.17		
MAPE									
SVM	2	2	3	2	2	2	2.17		
ENN	8	8	7	8	8	8	7.83		
LASSOENN	5	5	5	5	5	5	5.00		
PCAENN	7	6	8	6	7	7	6.83	39.50	79.00
SCADENN	6	7	6	7	6	6	6.33		
SRLENN	4	3	4	4	3	3	3.50		
SRENENN-A	3	4	1	3	4	4	3.17		
SRENENN-B	1	1	2	1	1	1	1.17		
RMSE									
SVM	5	4	8	7	7	3	5.67		
ENN	8	8	7	8	8	8	7.83		
LASSOENN	4	3	3	2	2	2	2.67		
PCAENN	7	6	5	4	6	7	5.83	31.90	15.77
SCADENN	6	7	6	5	4	6	5.67		
SRLENN	3	2	4	3	3	5	3.33		
SRENENN-A	2	5	1	6	5	4	3.83		
SRENENN-B	1	1	2	1	1	1	1.17		
TIC									
SVM	5	4	8	7	7	3	5.67		
ENN	8	8	7	8	8	8	7.83		
LASSOENN	4	3	3	2	2	2	2.67		
PCAENN	7	6	5	4	6	7	5.83	31.89	15.77
SCADENN	6	7	6	5	4	6	5.67		
SRLENN	3	2	4	3	3	5	3.33		
SRENENN-A	2	5	1	6	5	4	3.83		
SRENENN-B	1	1	2	1	1	1	1.17		

of Friedman test show that the F distribution statistics follows $F(7, 35)$ distribution and the critical value of it is 0.39. Thus, the null hypothesis that the ranks of compared methods are equal with each other is rejected. This means that a post hoc test based on Bonferroni-Dunn test is needed to make more comparisons. The CD value is calculated as 3.59 based on [49]. Therefore, SRENENN performs significantly better than ENN, PCAENN, SCADENN, LASSOENN, and SVM for RMSE and TIC. This is because the average ranks between SRENENN-B and these competitors are larger than 3.59. On the other hand, SRENENN-B does not show great improvement over SRENENN-A and SRLENN in terms of evaluation criteria.

Figure 5 summarized the results of estimated values against the true value. It is not difficult to tell that the estimated values of SRENENN-B are closer to the true value than other compared approaches. ENN provides the worst results because all the variables are employed as inputs. Thus, there must be some redundant features contained in the ENN. SRENENN-B gives good forecasting results in all the sites which demonstrates that the selected variables *temperature, pressure, solar zenith angle, wind direction,* and *wind speed* are considered to be important for inputs of ENN. Table 10 reveals statistical analysis of selected variables. It is observed that all the variables selected are significant because F values are much greater than critical
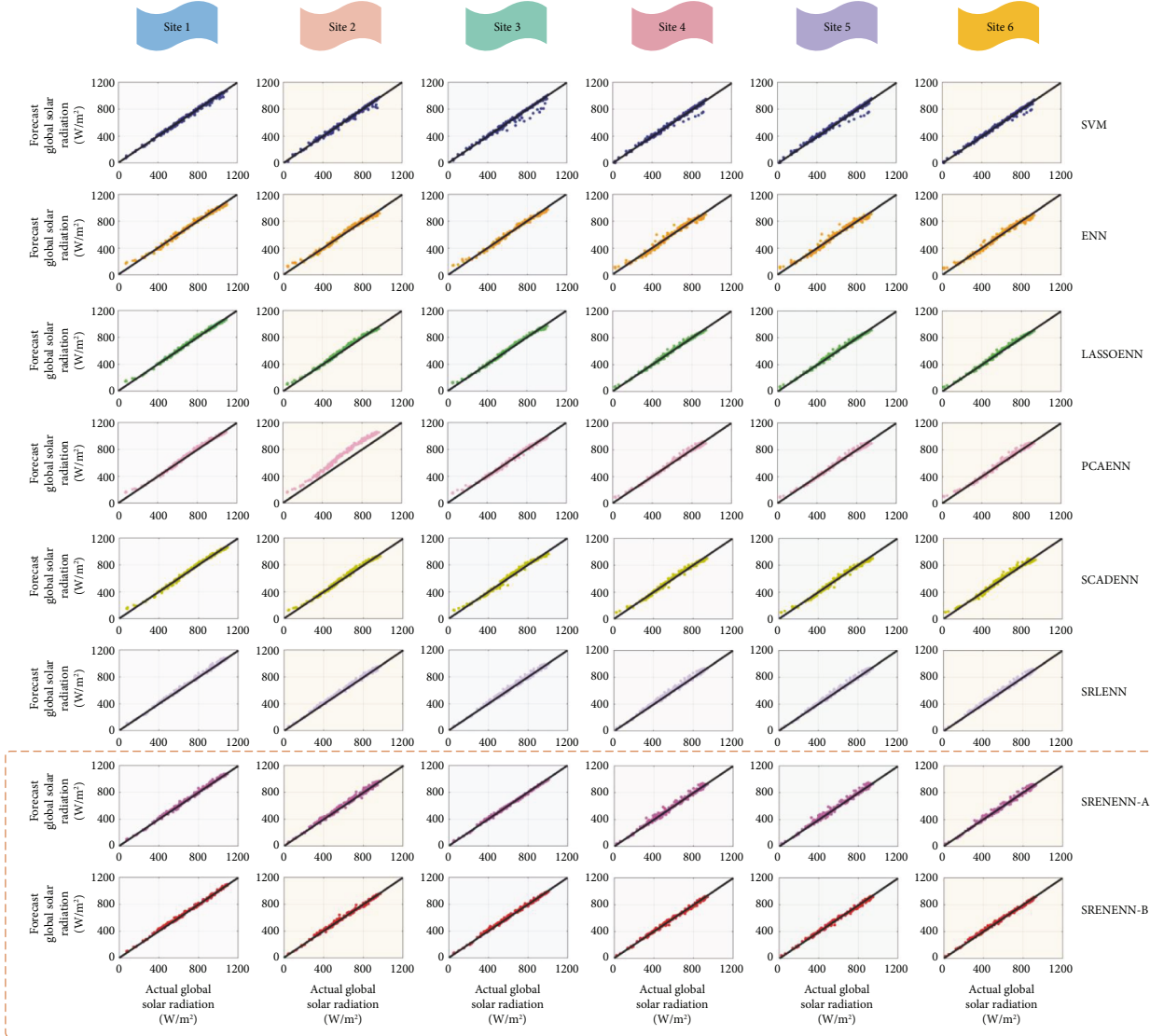
FIGURE 5: The scatter plots of actual and forecast global solar radiation by seven models: bold black dash line represents the perfect fit.

TABLE 10: The results of statistical analysis of selected variables by SREN at Sites 1–6.

Sites	R	F_{crit}	F value	Significant level α
Site 1	0.98	0.06	318.23	0.05
Site 2	0.99	0.05	250.12	0.05
Site 3	0.99	0.15	212.34	0.05
Site 4	0.99	0.06	10.12	0.05
Site 5	0.99	0.08	15.13	0.05
Site 6	0.99	0.09	28.74	0.05

values F_{crit}. Furthermore, the coefficients of determination R^2 are approximately one which indicates that the established model is trustworthy.

Figure 6 reveals the boxplots in terms of RMSEs of the compared forecasting models everyday in order to reveal the benefits of SRENENN. Figures 6(a)–6(f), (A), (C), (E), (G), (I), and (K), show the RMSEs of each model with limitation including all the outliers. Median is applied here to make the comparisons because it is less sensitive to outliers. Here, RMSEs of ENN are far larger than other forecasting approaches at all the sites. From Figures 6(a)–6(f), (B), (D), (F), (H), (J), and (L), obviously, SRENENN-B gives the lowest RMSE values. Therefore, based on boxplots (A)–(L) in Figures 6(a)–6(f), SRENENN-B delivers better forecasting performances.

5. Conclusions

Global solar radiation is a vital and hot research topic. Looking for a way to predict the global solar radiation accurately is crucial. There are a number of methods derived to achieve this goal. Our work investigated and studied the

1: SVM   2: ENN   3: LASSOENN   4: PCAENN   5: SCADENN
6: SRLENN   7: SRENENN-A   8: SRENENN-B

It is obvious that the model SRENENN-A and SRENENN-B are better than other comprtitors beacause their abnormal forecast values are lower than others.
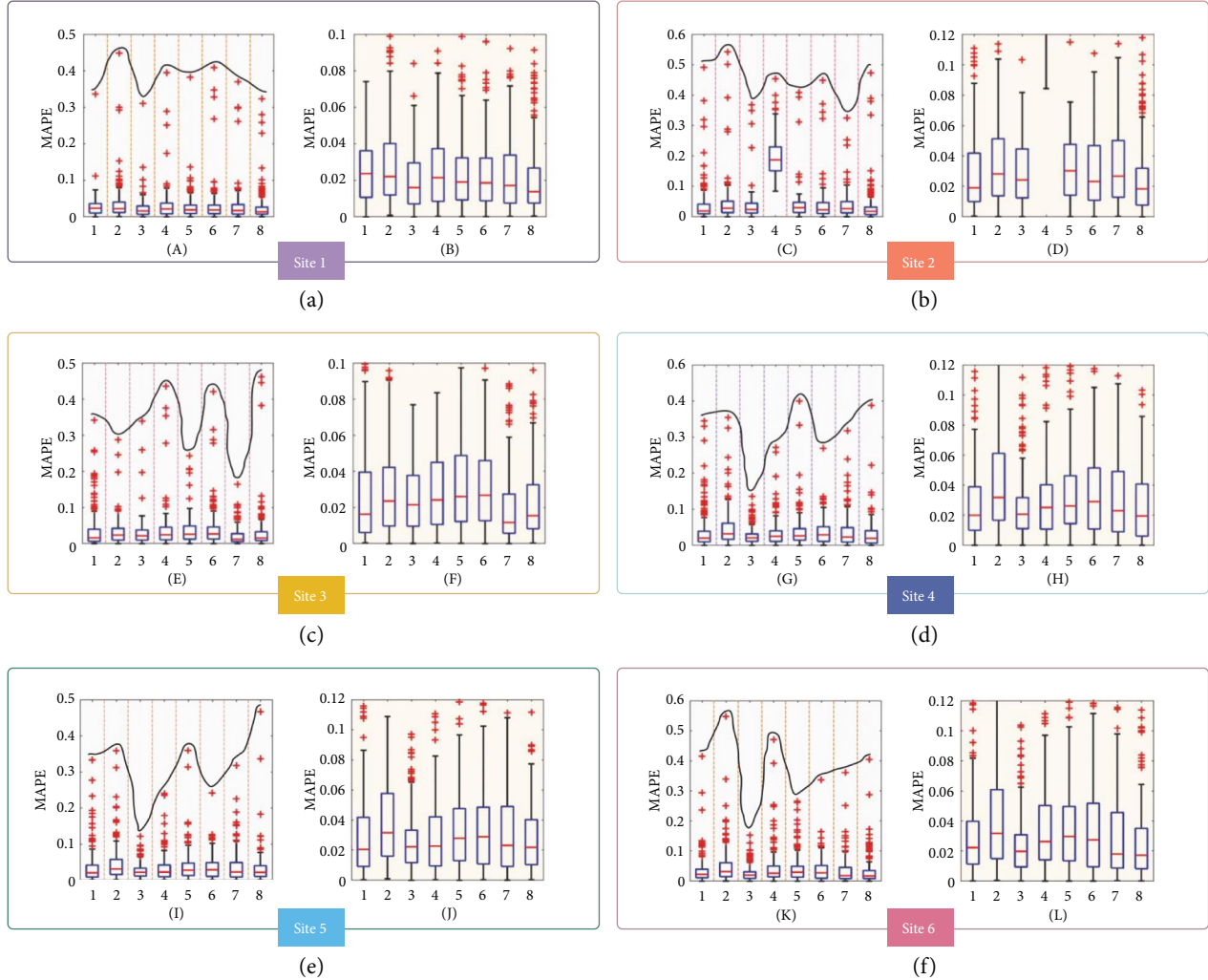

FIGURE 6: Boxplot of RMSEs of ENN, (a) LASSOENN, (b) PCAENN, (c) SCADENN, (d) SRLENN, (e) SRENENN-A, and (f) SRENENN-B for six sites in one year. (A, C, E, G, I, K) presents the RMSEs of each model without any limitation, and (B, D, F, H, J, L) presents the RMSEs of each model with the limitation.

SRENENN method. The key findings of this paper are demonstrated as follows:

(1) To build an interpretable model and overcome the selection inconsistency of existing variable selection methods, this work studies the square root loss function and elastic net and proposes the square root elastic net (SREN) which is a novel variable selection method.

(2) To boost the computational capacity of ENN and other penalize ENN models, this paper derives a simple-to-implement and fast algorithm to implement SREN. The experiments of computation time demonstrate the computation efficiency of the proposed algorithm.

(3) To improve the forecasting accuracy of ENN and other penalize ENN models, this paper establishes SRENENN model using the inputs selected by SREN.

To sum up, the proposed SRENENN model provides better forecasting performances than the traditional methods based on the real data application of six locations in Xinjiang area of China. For the future research, the following research directions will be focused on: (i) explore the performance of SRENENN model on a solar radiation problem under complex weather condition; (ii) investigate the application of SREN on other time series forecasting models; and (iii) study the application of SREN on high-dimensional data.

Appendix

Proof 1. A proof is provided based on [51, 52]. Define the object function as $F(\boldsymbol{\beta}) = \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2 + \lambda\|\boldsymbol{\beta}\|_1 + \eta\|\boldsymbol{\beta}\|_2^2/2$. A surrogate function is defined as

$$G(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2 + \frac{1}{\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2} (\boldsymbol{\gamma} - \boldsymbol{\beta})^{\mathrm{T}} \mathbf{X}^{\mathrm{T}} (\mathbf{X}\boldsymbol{\beta} - \mathbf{y})$$
$$+ \frac{\|\boldsymbol{\beta} - \boldsymbol{\gamma}\|_2^2}{2\|\mathbf{X}\boldsymbol{\beta} - \mathbf{y}\|_2} + \lambda \|\boldsymbol{\beta}\|_1 + \frac{\eta}{2} \|\boldsymbol{\beta}\|_2^2. \tag{A.1}$$

After some simple algebra, $\min_{\boldsymbol{\gamma}} G(\boldsymbol{\beta}, \boldsymbol{\gamma})$ is the same as

$$\min_{\boldsymbol{\gamma} \in R^p} \frac{1}{\|\mathbf{X}\boldsymbol{\beta} - \mathbf{y}\|_2} (\boldsymbol{\gamma} - \boldsymbol{\beta})^{\mathrm{T}} \mathbf{X}^{\mathrm{T}} (\mathbf{X}\boldsymbol{\beta} - \mathbf{y}) + \frac{\|\boldsymbol{\beta} - \boldsymbol{\gamma}\|_2^2}{2\|\mathbf{X}\boldsymbol{\beta} - \mathbf{y}\|_2}$$
$$+ \lambda \|\boldsymbol{\beta}\|_1 + \frac{\eta}{2} \|\boldsymbol{\beta}\|_2^2, \tag{A.2}$$

which can be reformulated as

$$\min_{\boldsymbol{\gamma} \in R^p} \frac{1}{\|\mathbf{X}\boldsymbol{\beta} - \mathbf{y}\|_2} \left(\frac{1}{2} \left\| \boldsymbol{\gamma} - \left(\boldsymbol{\beta} + \mathbf{X}^T \mathbf{y} - \mathbf{X}^T \mathbf{X} \boldsymbol{\beta} \right) \right\|_2^2 \right.$$
$$\left. + \lambda \|\mathbf{X}\boldsymbol{\beta} - \mathbf{y}\|_2 \|\boldsymbol{\beta}\|_1 + \frac{\eta}{2} \|\mathbf{X}\boldsymbol{\beta} - \mathbf{y}\|_2 \|\boldsymbol{\beta}\|_2^2 \right). \tag{A.3}$$

Applying Lemma 7 in reference [53], we obtain that

$$\boldsymbol{\gamma}_{opt}^{(j)} = \overrightarrow{\Theta}_{EL} \left(\boldsymbol{\beta}^{(j)} + \mathbf{X}^T \mathbf{y} - \mathbf{X}^T \mathbf{X} \boldsymbol{\beta}^{(j)}; \lambda \left\| \mathbf{X}\boldsymbol{\beta}^{(j)} - \mathbf{y} \right\|_2, \eta \left\| \mathbf{X}\boldsymbol{\beta}^{(j)} - \mathbf{y} \right\|_2 \right), \tag{A.4}$$

in Plan A. For Plan B we also have

$$\boldsymbol{\gamma}_{opt}^{(j)} = \overrightarrow{\Theta}_{SOFT} \left(\boldsymbol{\beta}^{(j)} + \mathbf{X}_{new}^T \mathbf{y}_{new} - \mathbf{X}_{new}^T \mathbf{X}_{new} \boldsymbol{\beta}^{(j)}; \lambda \left\| \mathbf{X}\boldsymbol{\beta}^{(j)} - \mathbf{y} \right\|_2 \right) \tag{A.5}$$

and notice that

$$G\left(\boldsymbol{\beta}, \boldsymbol{\gamma}_{opt} + \delta \right) - G\left(\boldsymbol{\beta}, \boldsymbol{\gamma}_{opt} \right) \geq \frac{\|\delta\|_2^2}{2\|\mathbf{X}\boldsymbol{\beta} - \mathbf{y}\|_2}. \tag{A.6}$$

Combining with the iterates defined by (9) or Eq. (10) and using Taylor expansion, we can get

$$F\left(\boldsymbol{\beta}^{(j+1)} \right) + \frac{1}{2} \left(\boldsymbol{\beta}^{(j+1)} - \boldsymbol{\beta}^{(j)} \right)^{\mathrm{T}} \left(\frac{\mathbf{I}}{\left\| \mathbf{X}\boldsymbol{\beta}^{(j)} - \mathbf{y} \right\|_2} - \frac{\mathbf{X}^{\mathrm{T}}\mathbf{X}}{\left\| \mathbf{X}\boldsymbol{\xi}^{(j)} - \mathbf{y} \right\|_2} \right) \left(\boldsymbol{\beta}^{(j+1)} - \boldsymbol{\beta}^{(j)} \right)$$
$$\leq G\left(\boldsymbol{\beta}^{(j)}, \boldsymbol{\beta}^{(j+1)} \right) \leq G\left(\boldsymbol{\beta}^{(j)}, \boldsymbol{\beta}^{(j)} \right) - \frac{1}{2\left\| \mathbf{X}\boldsymbol{\beta}^{(j)} - \mathbf{y} \right\|_2} \left(\boldsymbol{\beta}^{(j+1)} - \boldsymbol{\beta}^{(j)} \right)^{\mathrm{T}} \left(\boldsymbol{\beta}^{(j+1)} - \boldsymbol{\beta}^{(j)} \right) \tag{A.7}$$
$$= F\left(\boldsymbol{\beta}^{(j)} \right) - \frac{1}{2\left\| \mathbf{X}\boldsymbol{\beta}^{(j)} - \mathbf{y} \right\|_2} \left(\boldsymbol{\beta}^{(j+1)} - \boldsymbol{\beta}^{(j)} \right)^{T} \left(\boldsymbol{\beta}^{(j+1)} - \boldsymbol{\beta}^{(j)} \right),$$

for some $\boldsymbol{\xi}^{(j)} = \nu^{(j)} \boldsymbol{\beta}^{(j)} + (1 - \nu^{(j)}) \boldsymbol{\beta}^{(j+1)}$ with $\nu^{(j)} \in (0, 1)$. A simple reformulation of (A.7) yields that

$$F\left(\boldsymbol{\beta}^{(j)} \right) - F\left(\boldsymbol{\beta}^{(j+1)} \right) \geq \frac{1}{2} \left(\frac{2}{\left\| \mathbf{X}\boldsymbol{\beta}^{(j)} - \mathbf{y} \right\|_2} - \frac{\|\mathbf{X}\|_2^2}{\left\| \mathbf{X}\boldsymbol{\xi}^{(j)} - \mathbf{y} \right\|_2} \right)$$
$$\cdot \left\| \boldsymbol{\beta}^{(j+1)} - \boldsymbol{\beta}^{(j)} \right\|_2^2 \tag{A.8}$$

Under the regularity condition $\inf_{\boldsymbol{\xi} \in A} \|\mathbf{X}\boldsymbol{\xi} - \mathbf{y}\|_2 > \varepsilon$ and τ large enough, $F(\boldsymbol{\beta}^{(j)})$ is monotone decreasing. Define $M := F(\boldsymbol{\beta}^{(0)})$ and let $\|\mathbf{X}\|_2^2 < 2\varepsilon/M$, we have $C = 2/\|\mathbf{X}\boldsymbol{\beta}^{(j)} - \mathbf{y}\|_2 - \|\mathbf{X}\|_2^2/\|\mathbf{X}\boldsymbol{\xi}^{(j)} - \mathbf{y}\|_2$. Thus, using the optimal conditions, $\boldsymbol{\beta}^{(j)}$ has a unique limit point $\boldsymbol{\beta}^*$. Furthermore, $\boldsymbol{\beta}^*$ satisfies the KKT condition which means it is a global minimum. This completes our proof.

Nomenclature

Abbreviation

ENN: Elman neural network
EL: Elastic net
MAE: Mean absolute error
MAPE: Mean absolute percentage error
RMSE: Root mean square error
SRL: Square root LASSO
SREN: Square root elastic net
TIC: Theil inequality coefficient
CD: Critical difference.

English Symbols

h: Number of nodes in the hidden layer
n: Sample size
N: Maximum number of iterations in ENN
p: Number of variables

X: Data matrix
y: Response variable.

Greek Symbols

β^*: True variables
$\beta^{(j)}$: The estimate in the jth iteration
η: Regularization parameter for l_2 part
Θ: Thresholding rules
λ: Regularization parameter for l_1 part
σ: Noise level.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this article.

Acknowledgments

References

[1] K. P. Lin and P. F. Pai, "Solar power output forecasting using evolutionary seasonal decomposition least-square support vector regression," *Journal of Cleaner Production*, vol. 134, pp. 456–462, 2016.

[2] C. Voyant, T. Soubdhan, P. Lauret, M. David, and M. Muselli, "Statistical parameters as a means to a priori assess the accuracy of solar forecasting models," *Energy*, vol. 90, pp. 671–679, 2015.

[3] S. A. Kalogirou, "Solar thermal collectors and applications," *Progress in Energy and Combustion Science*, vol. 30, no. 3, pp. 231–295, 2004.

[4] J. Wu, C. K. Chan, Y. Zhang, B. Y. Xiong, and Q. H. Zhang, "Prediction of solar radiation with genetic approach combing multi-model framework," *Renewable Energy*, vol. 66, no. 3, pp. 132–139, 2014.

[5] C. Renno, F. Petito, and A. Gatto, "ANN model for predicting the direct normal irradiance and the global radiation for a solar application to a residential building," *Journal of Cleaner Production*, vol. 135, pp. 1298–1316, 2016.

[6] J. Boland, M. David, and P. Lauret, "Short term solar radiation forecasting: island versus continental sites," *Energy*, vol. 113, pp. 186–192, 2016.

[7] E. Akarslan, F. O. Hocaoglu, and R. Edizkan, "A novel M-D (multi-dimensional) linear prediction filter approach for hourly solar radiation forecasting," *Energy*, vol. 73, pp. 978–986, 2014.

[8] P. G. Kosmopoulos, S. Kazadzis, K. Lagouvardos, V. Kotroni, and A. Bais, "Solar energy prediction and verification using operational model forecasts and ground-based solar measurements," *Energy*, vol. 93, pp. 1918–1930, 2015.

[9] A. Mellit, "Artificial intelligence technique for modelling and forecasting of solar radiation data: a review," *International Journal of Artificial Intelligence and Soft Computing*, vol. 1, no. 1, pp. 52–76, 2008.

[10] H. W. Sun, N. Zhao, X. F. Zeng, and D. Yan, "Study of solar radiation prediction and modeling of relationships between solar radiation and meteorological variables," *Energy Conversion and Management*, vol. 105, pp. 880–890, 2015.

[11] M. David, F. Ramahatana, P. J. Trombe, and P. Lauret, "Probabilistic forecasting of the solar irradiance with recursive ARMA and GARCH models," *Solar Energy*, vol. 133, pp. 55–72, 2016.

[12] V. H. Quej, J. Almorox, M. Ibrakhimov, and L. Saito, "Empirical models for estimating daily global solar radiation in Yucatán Peninsula, Mexico," *Energy Conversion and Management*, vol. 110, pp. 448–456, 2016.

[13] A. R. el Ouderni, T. Maatallah, S. el Alimi, and S. Ben Nassrallah, "Experimental assessment of the solar energy potential in the gulf of Tunis, Tunisia," *Renewable & Sustainable Energy Reviews*, vol. 20, pp. 155–168, 2013.

[14] C. Paoli, C. Voyant, M. Muselli, and M. L. Nivet, "Forecasting of preprocessed daily solar radiation time series using neural networks," *Solar Energy*, vol. 84, no. 12, pp. 2146–2160, 2010.

[15] Y. Kashyap, A. Bansal, and A. K. Sao, "Solar radiation forecasting with multiple parameters neural networks," *Renewable & Sustainable Energy Reviews*, vol. 49, pp. 825–835, 2015.

[16] J. Z. Wang, H. Jiang, Y. J. Wu, and Y. Dong, "Forecasting solar radiation using an optimized hybrid model by cuckoo search algorithm," *Energy*, vol. 81, pp. 627–644, 2015.

[17] H. Jiang, "Model forecasting based on two-stage feature selection procedure using orthogonal greedy algorithm," *Applied Soft Computing*, vol. 63, pp. 110–123, 2018.

[18] K. Benmouiza and A. Cheknane, "Forecasting hourly global solar radiation using hybrid k-means and nonlinear autoregressive neural network models," *Energy Conversion and Management*, vol. 75, pp. 561–569, 2013.

[19] S. X. Chen, H. B. Gooi, and M. Q. Wang, "Solar radiation forecast based on fuzzy logic and neural networks," *Renewable Energy*, vol. 60, pp. 195–201, 2013.

[20] S. Salcedo-Sanz, C. Casanova-Mateo, A. Pastor-Sánchez, and M. Sánchez-Girón, "Daily global solar radiation prediction based on a hybrid coral reefs optimization-extreme learning machine approach," *Solar Energy*, vol. 105, pp. 91–98, 2014.

[21] K. Gairaa, A. Khellaf, Y. Messlem, and F. Chellali, "Estimation of the daily global solar radiation based on box-Jenkins and ANN models: a combined approach," *Renewable & Sustainable Energy Reviews*, vol. 57, pp. 238–249, 2016.

[22] C. Zhang, J. Z. Zhou, C. S. Li, W. L. Fu, and T. Peng, "A compound structure of ELM based on feature selection and parameter optimization using hybrid backtracking search algorithm for wind speed forecasting," *Energy Conversion and Management*, vol. 143, pp. 360–376, 2017.

[23] M. Benghanem, A. Mellit, and S. N. Alamri, "ANN-based modelling and estimation of daily global solar radiation data: a case study," *Energy Conversion and Management*, vol. 50, no. 7, pp. 1644–1655, 2009.

[24] A. Rahimikhoob, "Estimating global solar radiation using artificial neural network and air temperature data in a semi-arid environment," *Renewable Energy*, vol. 35, no. 9, pp. 2131–2135, 2010.

[25] X. Y. Qing and Y. G. Niu, "Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM," *Energy*, vol. 148, pp. 461–468, 2018.

[26] M. Vakili, S. R. Sabbagh-Yazdi, S. Khosrojerdi, and K. Kalhor, "Evaluating the effect of particulate matter pollution on estimation of daily global solar radiation using artificial neural network modeling based on meteorological data," *Journal of Cleaner Production*, vol. 141, pp. 1275–1285, 2017.

[27] A. Rohani, M. Taki, and M. Abdollahpour, "A novel soft computing model (Gaussian process regression with K-fold cross validation) for daily and monthly solar radiation forecasting (part: I)," *Renewable Energy*, vol. 115, pp. 411–422, 2018.

[28] R. R. Hocking, "A biometrics invited paper. The analysis and selection of variables in linear regression," *Biometrics*, vol. 32, no. 1, pp. 1–49, 1976.

[29] H. Wang, "Forward regression for ultra-high dimensional variable screening," *Journal of the American Statistical Association*, vol. 104, no. 488, pp. 1512–1524, 2009.

[30] S. Jović, O. Aničić, M. Marsenić, and B. Nedić, "Solar radiation analyzing by neuro-fuzzy approach," *Energy and Buildings*, vol. 129, pp. 261–263, 2016.

[31] M. Almaraashi, "Investigating the impact of feature selection on the prediction of solar radiation in different locations in Saudi Arabia," *Applied Soft Computing*, vol. 66, pp. 250–263, 2018.

[32] A. Aybar-Ruiz, S. Jiménez-Fernández, L. Cornejo-Bueno et al., "A novel grouping genetic algorithm-extreme learning machine approach for global solar radiation prediction from numerical weather models inputs," *Solar Energy*, vol. 132, pp. 129–142, 2016.

[33] H. Mori, "Application of graphical modelling to selecting input variables for solar radiation forecasting," *IFAC-PapersOnLine*, vol. 48, no. 30, pp. 137–142, 2015.

[34] H. Jiang and Y. Dong, "Global horizontal radiation forecast using forward regression on a quadratic kernel support vector machine: case study of the Tibet autonomous region in China," *Energy*, vol. 133, pp. 270–283, 2017.

[35] A. Moghaddamnia, R. Remesan, M. H. Kashani, M. Mohammadi, D. Han, and J. Piri, "Comparison of LLR, MLP, Elman, NNARX and ANFIS models—with a case study in solar radiation estimation," *Journal of Atmospheric and Solar-Terrestrial Physics*, vol. 71, no. 8-9, pp. 975–982, 2009.

[36] M. Luo, C. Li, X. Zhang, R. Li, and X. An, "Compound feature selection and parameter optimization of ELM for fault diagnosis of rolling element bearings," *ISA Transactions*, vol. 65, pp. 556–566, 2016.

[37] R. Tibshirani, "Regression shrinkage and selection via the lasso: a retrospective," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 73, no. 3, pp. 273–282, 2011.

[38] J. Fan and R. Li, "Variable selection via nonconcave penalized likelihood and its Oracle properties," *Journal of the American Statistical Association*, vol. 96, no. 456, pp. 1348–1360, 2001.

[39] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 67, no. 2, pp. 301–320, 2005.

[40] A. Belloni, V. Chernozhukov, and L. Wang, "Square-root Lasso: pivotal recovery of sparse signals via conic programming," *Biometrika*, vol. 98, no. 4, pp. 791–806, 2011.

[41] Y. Y. She, "Thresholding-based iterative selection procedures for model selection and shrinkage," *Electronic Journal of Statistics*, vol. 3, pp. 384–415, 2009.

[42] P. Auer, H. Burgsteiner, and W. Maass, "A learning rule for very simple universal approximators consisting of a single layer of perceptrons," *Neural Networks*, vol. 21, no. 5, pp. 786–795, 2008.

[43] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representations by error propagation," in *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, vol. 1*, pp. 318–362, MIT Press, Cambridge, MA, USA, 1988.

[44] Y. Nesterov, "A method of solving a convex programming problem with convergence rate O(1/k^2)," *Soviet Mathematics Doklady*, vol. 27, no. 2, pp. 372–376, 1983.

[45] Y. Dong, J. Z. Wang, H. Jiang, and X. M. Shi, "Intelligent optimized wind resource assessment and wind turbines selection in Huitengxile of Inner Mongolia, China," *Applied Energy*, vol. 109, pp. 239–253, 2013.

[46] C. J. Lu, T. S. Lee, and C. C. Chiu, "Financial time series forecasting using independent component analysis and support vector regression," *Decision Support Systems*, vol. 47, no. 2, pp. 115–125, 2009.

[47] A. C. Pollock, A. Macaulay, M. E. Thomson, and D. Önkal, "Performance evaluation of judgemental directional exchange rate predictions," *International Journal of Forecasting*, vol. 21, no. 3, pp. 473–489, 2005.

[48] J. J. Wang, W. Y. Zhang, J. Z. Wang, T. T. Han, and L. B. Kong, "A novel hybrid approach for wind speed prediction," *Information Sciences*, vol. 273, no. 18, pp. 304–318, 2014.

[49] J. Demšar, "Statistical comparisons of classifiers over multiple data sets," *Journal of Machine Learning Research*, vol. 7, pp. 1–30, 2006.

[50] R. L. Iman and J. M. Davenport, "Approximations of the critical region of the fbietkan statistic," *Communications in Statistics - Theory and Methods*, vol. 9, no. 6, pp. 571–595, 1980.

[51] Y. Y. She, J. P. Wang, H. H. Li, and D. P. Wu, "Group iterative spectrum thresholding for super-resolution sparse spectral selection," *IEEE Transactions on Signal Processing*, vol. 61, no. 24, pp. 6371–6386, 2013.

[52] F. Bunea, J. Lederer, and Y. Y. She, "The group square-root Lasso: theoretical properties and fast algorithms," *IEEE Transactions on Information Theory*, vol. 60, no. 2, pp. 1313–1325, 2014.

[53] Y. Y. She, "Selective factor extraction in high dimensions," *Biometrika*, vol. 104, no. 1, pp. asw059–asw110, 2017.