Applied Soft Computing 117 (2022) 108357

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

A least square support vector machine approach based on bvRNA-GA for modeling photovoltaic systems

Xiu Liu [a], Ning Wang [a,*], Daniel Molina [b], Francisco Herrera [b,c]

[a] State Key Laboratory of Industrial Control Technology, Institute of Cyber-systems and control, Zhejiang University, Hangzhou 310027, PR China
[b] Andalusian Research Institute in Data Science and Computational Intelligence (DaSCI), Department of Computer Science and Artificial Intelligence, University of Granada, 18071-Granada, Spain
[c] Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia

ARTICLE INFO

Article history:
Received 29 March 2020
Received in revised form 18 November 2021
Accepted 16 December 2021
Available online 24 December 2021

Keywords:
RNA computing
RNA genetic algorithm
Least square support vector machine (LSSVM)
Modeling
Photovoltaic systems

ABSTRACT

Accurate model plays an important role in designing, assessing, and controlling photovoltaic (PV) systems. In this work, the least-squares support vector machine (LSSVM) is adopted to model the current–voltage (V–I) characteristic curves of different PV systems. A novel RNA genetic algorithm (bvRNA-GA) is proposed to determine the parameters of LSSVM. The bvRNA-GA is featured by designing the bulge loop crossover operator and the virus-induced mutation operator, they are employed to balance the exploration and exploitation capacities. Different experiments with 10 benchmark functions are conducted to show that the search efficiency of bvRNA-GA is better than the other four state-of-art algorithms. The outputs of bvRNA-GA optimized LSSVM models can better agree with the real outputs of different PV systems, the modeling results demonstrate the effectiveness of bvRNA-GA in solving real-world problems.

1. Introduction

The photovoltaic (PV) systems as one of the greatest technologies for humans can directly convert the photons into electrical energy and do not produce harmful substances [1]. However, the PV systems are sensitive to the working temperatures and irradiation which lower the conversion efficiency of PV systems and limit the use of solar energy [2]. In order to better design the structure and enhance the performance, accurate models must be established for PV systems [3]. However, it is a tricky problem to model PV systems because they are featured with complex nonlinear characteristics and are easily influenced by outside conditions [4].

Researchers usually adopt two ways to model complex nonlinear systems, they are the parametric modeling method and the non-parametric modeling method. In recent years, the parametric modeling method has been applied to model the mathematical models for PV systems. The mathematical models use the physical processes and the related variables to represent the characteristics of PV systems, and researchers have applied many different algorithms to extract the unknown parameters in the mathematical models. For example, the parallel swarm algorithm (PSA) and the ant lion optimizer are respectively used in [5] and [6] to estimate the parameters of a simple PV system which is called the single diode model (SDM). Then, the double diode model (DDM) is studied, and some optimization algorithms are applied to detect the parameters for both SDM and DDM [4,7–9]. To detect the performance of this kind of modeling method, more and more researchers take the modeling problems of PV solar modules into consideration [10–13]. These researches demonstrate the better performance of the parametric modeling method. However, there is little research about non-parametric modeling in modeling PV systems.

Non-parametric is a kind of black-box modeling method, it only takes the mapping relationship of the system input–output into account and neglects the internal characteristics of PV systems. Among the existing non-parametric modeling methods, the support vector machine (SVM) attracts wide concern from researchers [14,15]. SVM has highly competitive performance in solving many engineering problems, such as face recognition, data mining, image processing, nonlinear system modeling, and other domains [16]. However, the constrained quadric programming optimization problem in SVM can cause the datasets to be rough and to fluctuate, severely limiting the applications in solving regression problems [17]. The least-squares SVM (LSSVM) was proposed by Suykens and Vandewalle, and it improves the performance of SVM in two ways [18]. Firstly, LSSVM modifies the inequality constraints in the SVM to equality constraints. Secondly, LSSVM adopts the training error square in the cost function

* Corresponding author.
E-mail addresses: nwang126@126.com, nwang@iipc.zju.edu.cn (N. Wang), dmolina@decsai.ugr.es (D. Molina), herrera@decsai.ugr.es (F. Herrera), ZJU_liuxiu@outlook.com (X. Liu).

https://doi.org/10.1016/j.asoc.2021.108357
1568-4946/

instead of the slack variables [19]. Therefore, it transforms the original quadratic programming problem into solving a set of linear equations, which thus improves the computation speed and accuracy significantly [20]. Furthermore, there are fewer hyper-parameters to be determined, which means we can acquire better modeling performance by optimizing fewer parameters. LSSVM has been made some attempts in complex non-linear modeling problems.

Rostami *et al.* [21] applied LSSVM to model a correlative model for CO_2 solubility in both dead and live oil systems. Zendehboudi [22] developed the LSSVM modeling approach to estimate the parameters of solid desiccant wheels. Bemani *et al.* [23] used the LSSVM for estimation of the cetane number of biodiesel in terms of fatty acid methyl esters composition. These works show the effectiveness of LSSVM in dealing with engineering problems, therefore, the LSSVM is adopted to model PV systems in this work. But the vital thing is that the performance of LSSVM highly depends on the regularization parameter and the width of the kernel function. An effective method should be investigated to determine the LSSVM parameters.

Recently, intelligent optimization algorithms have been adopted to tune the parameter values for LSSVM. For example, Gao and Liu [24] optimized LSSVM parameters with hybrid fruit fly optimization (FFO) and Particle swarm optimization (PSO) for cancer diagnosis. Tian [25] proposed an improved firefly algorithm to set the LSSVM parameters, which achieves better modeling accuracy in predicting wind speed . Wen and Cao [26] utilized the enhanced butterfly optimization algorithm (BOA) to determine LSSVM parameters for predicting residential CO_2 emissions. Wu *et al.* [27] used the artificial bee colony (ABC) optimized LSSVM to diagnose the system faults, which obtain good pattern recognition performance. Wu *et al.* [28] utilized the LSSVM for predicting aero-optic imaging deviation based on the chaotic particle swarm optimization (CPSO). Liu *et al.* [29] came up with a robust reliability model with the chaos modified PSO (CMPSO) algorithm to estimate the weighted LSSVM parameters.

Genetic algorithm (GA) proposed by professor J. Holland in 1975 imitates the biological evolution process [30]. GA is featured with good random searching ability and extensibility, it has been widely used to determine system parameters and select features [31–36]. Although GA has outstanding advantages, the defects of premature and low precision hinder its applications. To overcome these disadvantages, inspired by DNA computing [37], Tao and Wang proposed the RNA genetic algorithm (RNA-GA) [38]. According to the experimental results in [38], RNA-GA owns superiority over the basic GA both in searching precision and population diversity. After that, variants of RNA-GAs are proposed by researchers [39–44]. Inspired by RNA molecular operations the new operators are designed to enhance the performance of other intelligent algorithms as well [45,46].

In this work, we propose a novel RNA genetic algorithm (bvRNA-GA) that incorporates the newly designed opertors: the bulge loop crossover operator and the virus-induced mutation operator. The bvRNA-GA is tested with some benchmark functions and the results are compared against four state-of-art algorithms. Subsequently, bvRNA-GA is used to optimize the LSSVM parameters for modeling different PV systems (include the SDM, the DDM, and three different PV solar modules) under various working conditions. The main contributions of this work are summarized as follows:

- We propose a novel RNA genetic algorithm (bvRNA-GA). In bvRNA-GA, the bulge loop crossover operator and the virus-induced mutation operator are designed to balance exploration and exploitation capacities.
- The performance superiority of bvRNA-GA against four kinds of state-of-art algorithms is performed with some benchmark functions.
- We model the PV systems with the bvRNA-GA based LSSVM. This modeling method obtains higher accuracy PV system characteristic curves.

The remainder of the paper is organized as follows: Brief descriptions about the PV systems and the LSSVM are shown in Section 2. In Section 3, we propose the novel RNA genetic algorithm (bvRNA-GA) and the genetic strategies. The numerical experiments on benchmark functions are performed in Section 4. In Section 5, the bvRNA-GA is adopted to optimize the LSSVM for modeling PV systems. Finally, the conclusion is given in Section 6.

2. Photovoltaic system modeling

In this section, a brief mathematical description of different photovoltaic (PV) cell systems and the least-squares support vector machine (LSSVM) modeling approach are given below. In this work, the researched PV systems include the single diode model (SDM), the double diode model (DDM), and the three kinds of photovoltaic (PV) cell solar modules.

2.1. The PV systems

2.1.1. SDM and DDM

SDM and DDM attract widely studied by researchers, their equivalent circuits are shown in Fig. 1 [47,48].

According to Kirchhoff's current law (KCL) the terminal current I_L of the SDM is calculated as follows [47]:

$$I_L = I_{ph} - I_{sd} \cdot \left[\exp\left(\frac{V_L + R_s \cdot I_L}{a \cdot V_t} \right) - 1 \right] - \frac{V_L + R_s \cdot I_L}{R_{sh}} \tag{1}$$

where I_{ph}, I_{sd}, R_s, R_{sh}, and a represent the photocurrent, the diode saturation current, the series resistance, the shunt resistance, and the diode ideality factor, respectively. V_L is the output voltage. $V_t = kT/q$ is the thermal voltage of the diode, where the Boltzmann constant $k = 1.3806503 \times 10^{-23}$ J/K, and the electron charge $q = 1.60217646 \times 10^{-10}$C, T is the Kelvin temperature (K).

Accordingly, by applying the KCL to DDM the terminal current I_L is obtained [48]:

$$I_L = I_{ph} - I_{sd1} \cdot \left[\exp\left(\frac{V_L + R_s \cdot I_L}{a_1 \cdot V_t} \right) - 1 \right] - I_{sd2} \cdot \left[\exp\left(\frac{V_L + R_s \cdot I_L}{a_2 \cdot V_t} \right) - 1 \right] - \frac{V_L + R_s \cdot I_L}{R_{sh}} \tag{2}$$

The same as SDM, I_{sd1} and I_{sd2} are the diffusion and saturation current for each diode of DDM. Where a_1, a_2 are the diode ideality factors of those two diodes.

In this work, we use the LSSVM to model the I-V and P-V characteristic curves for SDM and DDM under working temperatures 25 °C, 33 °C, 50 °C, 75 °C, 100 °C (with irradiation level 1000 W/m^2).

2.1.2. The PV solar modules

In this work, three kinds of PV solar modules include the S75, the SM55, and the ST40 are studied, their manufacturers' information sheets are respectively shown in [49–51]. Where the S75 module contains 36 series-connected multi-crystalline silicon PV cells. SM55 was established by 36 series mono-crystalline silicon PV cells. ST40 is composed of a monolithic structure of series-connected Copper Indium Diselenide (CIS) based PV cells.

We model the I-V curves of S75, SM55, ST40 both working at different temperatures and different irradiance. Specifically, five kinds of working temperatures include 20 °C, 30 °C, 40 °C, 50 °C, 60 °C (with irradiation level 1000 W/m^2), and different irradiations 200 W/m^2, 400 W/m^2, 600 W/m^2, 800 W/m^2, 1000 W/m^2 (with standard temperature 25 °C) are studied.

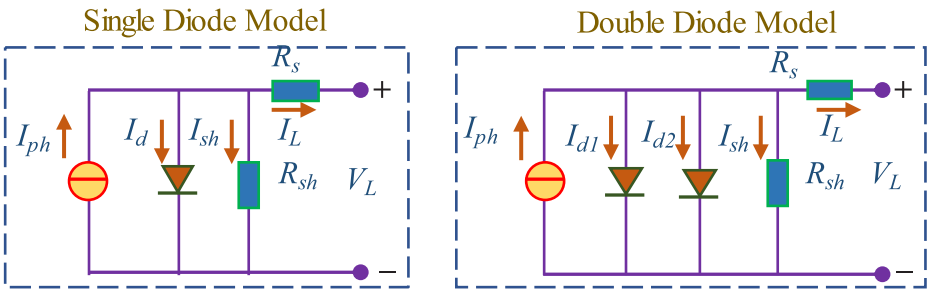


Fig. 1. The circuit diagram of the SDM [47] (the left) and the DDM (the right) [48].

2.1.3. Sampling datasets

We sample the datasets of SDM and DDM (includes the training and testing dataset) that can be used for the LSSVM modeling method according to the following steps, for the convenience of understanding, taking SDM working at temperatures 25 °C as an example:

Step 1: Use MATLAB Simulink's tool and the SDM's mathematical models described above to build the simulation model.

Step 2: Input the model parameters of SDM (25 °C, 1000 W/m^2) obtained in [3] to the simulation model.

Step 3: Input the voltage (V), where the range of voltage is from −0.2 V to 0.6 V; Sample the output current (I) and the sampling interval is 2.6E−03 V; Finally, the dataset with 300 pairs $\{(V_k, I_k)|k = 1,\ldots,300\}$ samples are obtained.

The datasets of SDM and DDM under different working conditions can be obtained use step 1 to step 2.

The way to sample the datasets of PV solar modules is different from that of SDM and DDM. The real characteristic curves of S75, SM55, ST40 under the researched outside conditions are respectively shown in their manufacturers' information sheets [49–51], and we directly extracted their modeling dataset from the manufacturers' information datasheet in [49–51]. For each working condition of these PV solar modules, the dataset with 300 pairs $\{(V_k, I_k)|k = 1,\ldots,300\}$ samples is obtained by directly measuring the characteristic curves in the manufacturers' information sheets.

2.2. The least-squares support vector machine

Usually, the parameters of the support vector machine (SVM) are determined by the quadratic programming method, while LSSVM achieves such purpose by a set of linear equations. Suppose the training dataset with M samples$\{(\boldsymbol{x}_i, \boldsymbol{y}_i)|i = 1,2,\ldots, M\}$, $\boldsymbol{x}_i \in R^{1\times I_n}$ and $\boldsymbol{y}_i \in R^{1\times O_n}$ are respectively the input variable and corresponding output, where I_n and O_n are dimensions (in this work $I_n = 6$, $O_n = 1$). The LSSVM model structure can be given as the following equation [24]:

$$\hat{\boldsymbol{y}}(\boldsymbol{x}) = \boldsymbol{w}^T \Phi(\boldsymbol{x}) + b \tag{3}$$

where $\boldsymbol{w}$ is the weight vector; b is the error variable; $\Phi(\bullet)$ represents the nonlinear mapping function that maps the input samples to a high dimensional feature space. Solving the following constrained optimization problem can be used to obtain the parameters of the linear regression function [24]:

$$\begin{cases} \min\limits_{\boldsymbol{w},b,\xi} J(\boldsymbol{w},\xi) = \dfrac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + \dfrac{1}{2}\gamma\sum\limits_{i=1}^{M}\xi_i^2 \\ s.t.\, \boldsymbol{y}_i = \boldsymbol{w}^T\Phi(\boldsymbol{x}_i) + b + \xi_i \end{cases} \tag{4}$$

where $J(\boldsymbol{w},\xi)$ is the lost function; γ is the regularization parameter; ξ_i is the deviation variable. The following Lagrangian function is constructed to solve this optimization problem [24]:

$$L = \frac{1}{2}\|\boldsymbol{w}\|^2 + \gamma\sum_{i=1}^{M}\xi_i^2 - \sum_{i=1}^{M}\alpha_i\{\boldsymbol{w}^T\varphi(\boldsymbol{x}_i) + b + \xi_i - \boldsymbol{y}_i\} \tag{5}$$

where L denotes the multiplier of Lagrangian. According to the Karush-Kuhn–Tucker (KKT) condition, the following equations are obtained [24]:

$$\begin{cases} \partial L/\partial \boldsymbol{w} = 0 \Rightarrow \boldsymbol{w} = \sum\limits_{i=1}^{M}\alpha_i\Phi(\boldsymbol{x}_i) \\ \partial L/\partial b = 0 \Rightarrow \sum\limits_{i=1}^{M}\alpha_i = 0 \\ \partial L/\partial \xi_i = 0 \Rightarrow \alpha_i = \gamma\xi_i \\ \partial L/\partial \alpha_i = 0 \Rightarrow \boldsymbol{w}^T\Phi(\boldsymbol{x}_i) + b + \xi_i - \boldsymbol{y}_i = 0 \end{cases} \tag{6}$$

After eliminating ω and ξ by the above equation, the optimization problem can be transformed into a linear equation problem shown as follows [24]:

$$\begin{bmatrix} 0 & \boldsymbol{I}^T \\ \boldsymbol{I} & \Omega + \dfrac{1}{\gamma}\boldsymbol{I} \end{bmatrix}\begin{bmatrix} b \\ \boldsymbol{\alpha} \end{bmatrix} = \begin{bmatrix} 0 \\ \boldsymbol{y} \end{bmatrix} \tag{7}$$

where $\boldsymbol{\alpha} = [\alpha_1,\ldots, \alpha_M]^T$; $\boldsymbol{I} = [1, \ldots, 1]^T$; $\Omega_{k,i} = K(x_k, x_i)$ (k, i = 1,2,…, M) is the kernel matrix. Therefore, the regression model of LSSVM is described as follows [22]:

$$\boldsymbol{y}(\boldsymbol{x}) = \sum_{i=1}^{M}\alpha_i K(\boldsymbol{x}, \boldsymbol{x}_i) + b \tag{8}$$

where $K(,)$ is the kernel function of LSSVM. In this work, the Radial Basis Function (RBF) is adopted in the LSSVM regression model, which is formulated as follows [24]:

$$K(\boldsymbol{x}, \boldsymbol{x}_i) = \exp\left(\frac{-\|\boldsymbol{x} - \boldsymbol{x}_i\|^2}{2\sigma^2}\right) \tag{9}$$

where σ is the width of RBF which determines the sensitivity of LSSVM.

Therefore, the parameters σ and γ of LSSVM need to be determined. In this work, we proposed the bvRNA-GA to adjust these two parameters, and the bvRNA-GA based LSSVM is applied to model different PV systems.

3. The bvRNA-GA

To enhance the performance, the bulge loop crossover operator and the virus-induced mutation operator are designed in this work. Therefore, the novel RNA-GA is named bvRNA-GA, the bvRNA-GA and the operators are described in the following subsections.

3.1. RNA encoding

According to the description of the RNA-GA [38], RNA-GA adopts four integers {0, 1, 2, 3} to respectively replace those four kinds of nucleotide bases (adenine(A), uracil(U), guanine(G),

cytosine(C)) to form the potential solutions of a global optimization problem. To the global optimization problem with D dimensional variables, the encoding length of a potential solution is $L = D \times l_s$, where l_s is the encoding length of the single variable. In this work, the complementary bases theory is also introduced to perform a special operation. Complementary bases theory is that complementary bases A $\leftrightarrow$ U and G $\leftrightarrow$ C can be connected by the hydrogen bond in biological, there are many different ring/loop structures in RNA molecules because of their phenomenon. Therefore, the integers 0 $\leftrightarrow$ 3 and 1 $\leftrightarrow$ 2 are regarded as complementary integers.

3.2. Procedure of bvRNA-GA

The procedures of bvRNA-GA are as follows, and the flowchart is shown in Fig. 2:

Step 1: Set the parameters (N, L, G, etc.); generate the initial population $\boldsymbol{P}_T$.

Step 2: Calculate the fitness values of $\boldsymbol{P}_T$, and sort $\boldsymbol{P}_T$; Select $N/2$ individuals with better fitness values to make up the population $\boldsymbol{P}_b$; select $N/2$ individuals with worse fitness values to make up the population $\boldsymbol{P}_w$.

Step 3: Copy $\boldsymbol{P}_b$, perform the bulge loop crossover operator which is defined in Section 3.4 to $\boldsymbol{P}_b$ according to probability p_{bl}, and perform the inner loop crossover operator to the copied $\boldsymbol{P}_b$ according to probability p_{il}; N new individuals are obtained after this step and reserving them to $\boldsymbol{P}_{nb}$.

Step 4: Perform the virus-induced mutation operator to populations $\boldsymbol{P}_{nb}$ and $\boldsymbol{P}_w$; One obtains $3\times N/2$ new individuals after this step, and they are reserved to the new population $\boldsymbol{P}_{mt}$.

Step 5: Calculate the fitness values of $\boldsymbol{P}_{mt}$, and sort $\boldsymbol{P}_{mt}$; Then select N individuals with better fitness values from $\boldsymbol{P}_{mt}$.

Step 6: Perform the proportional selection operator to the selected N individuals in Step 5.

Step 7: More than N individuals are obtained after Step 6, but only reserve N individuals with better fitness values, and replace the population $\boldsymbol{P}_T$ with the reserved N individuals.

Step 8: If achieves the maximum iteration G, output the current best individual (the optimal solution). Else, go to Step 2.

The following subsections are going to describe the proportional selection operator, the bulge loop crossover operator, the inner loop crossover operator, and the virus-induced mutation operator in detail.

3.3. The proportional selection

The proportional selection in Step 6 duplicates the individual in proportion. The duplicating proportion N_i ($N_i = 1, 2, 3, \ldots, N$) is calculated according to the following equation.

$$N_i = \left| N \cdot \left\lfloor \frac{F_i}{\sum_{j=1}^{N} F_j} \right\rfloor \right| \tag{10}$$

where $F_i = 1/f_i$, and $f_i > 0$ is the fitness value of ith individual; $\lfloor \bullet \rfloor$ denotes maximal integer less than '$\bullet$'.

The number $\hat{N} = \sum_{i=1}^{N} N_i$ tells the reason why the population size is bigger than N after the proportional selection, and only N individuals with better fitness values are reserved.

3.4. The crossover operators

During the searching processes of GA, the crossover operator is the most important approach for information communication among individuals. The basic GA has the drawbacks of slow convergence and low precision. To enhance the performance, the bulge loop crossover operator and the inner loop crossover operator are designed in this work. The bulge loop crossover operator is inspired by the bulge loop structure of RNA molecule double-strand structure [52] (shown in Fig. 3). The procedures of the bulge loop crossover operator and the inner loop operator are summarized as follows.

3.4.1. The bulge loop crossover operator

For the convenience of understanding, one takes $L = 18$ as an example, and the schematic of the bulge loop crossover operator is shown in Fig. 3. As is shown in Fig. 3(2), from left to right, the nucleotide bases in positions $[1, L/2]$ are defined as high bits, and the nucleotide bases in positions $[L/2 + 1, L]$ are defined as low bits. The procedures of the bulge loop crossover operator can be divided into 2 steps:

Step (a): The formation of the bulge loop structure (BLS), in Fig. 3 from (1) to (4);

Step (b): Performing the crossover operator, in Fig. 3 from (5) to (8).

Firstly, performing the procedures of steps (a):

Step (a.1): Select an individual from $\boldsymbol{P}_b$ ($\boldsymbol{P}_b$ has been defined in Section 3.2; Respectively select two positions from the high bits (loc_1) and the low bits (loc_2) (shown in Fig. 3(2)). loc_1 and loc_2 satisfy the following condition:

$$|loc_2 - loc_1| \geq 2 \tag{11}$$

Therefore, there are $(L/2)^2 - 1$ combinations of loc_1, loc_2 for a single individual. For the convenience of descripting, one defines $t_t = 0$.

Step (a.2): If $t_t < (L/2)^2 - 1$ and the nucleotide bases in loc_1 and loc_2 are the complementary base (defined in Section 3.1), join them together to form the BLS, this complementary base is the root base of the formed BLS (shown in Fig. 3(3)). This individual is marked as the bulge loop strand, then go to Step (a.4). If the bases in loc_1, loc_2 is not the complementary base, $t_t += 1$ and go to Step (a.3).

Step (a.3): If $t_t \geq (L/2)^2 - 1$, this individual cannot find complementary base and is marked as normal strand, go to Step (a.1); Otherwise, reselect loc_1, loc_2 (current loc_1, loc_2 combination has not been selected) and go to Step (a.2);.

Step (a.4): Replace the original bulge loop strand without forming the BLS with the corresponding bulge loop strand with BLS, shown in Fig. 3(4). Go to Step (a.1) until all the individuals in P_b are operated.

Secondly, performing the procedures of steps (b):

Step (b.1): Randomly select two individuals (marked as I_1 and I_2) from $\boldsymbol{P}_b$ (once I_1 and I_2 are selected, remove them from $\boldsymbol{P}_b$), matching I_1, I_2 and align at left (shown in Fig. 3(5)).

Step (b.2): If both I_1, I_2 individuals are bulge loop strands (shown in Fig. 3(5)), and the root base of each bulge loop structure can form the allelic bases, then these two bulge loop structures are dissociated from the root bases and exchange with each other (shown in Fig. 3(6)), after that, switch to Step (b.5); Else switch to Step (b.3).

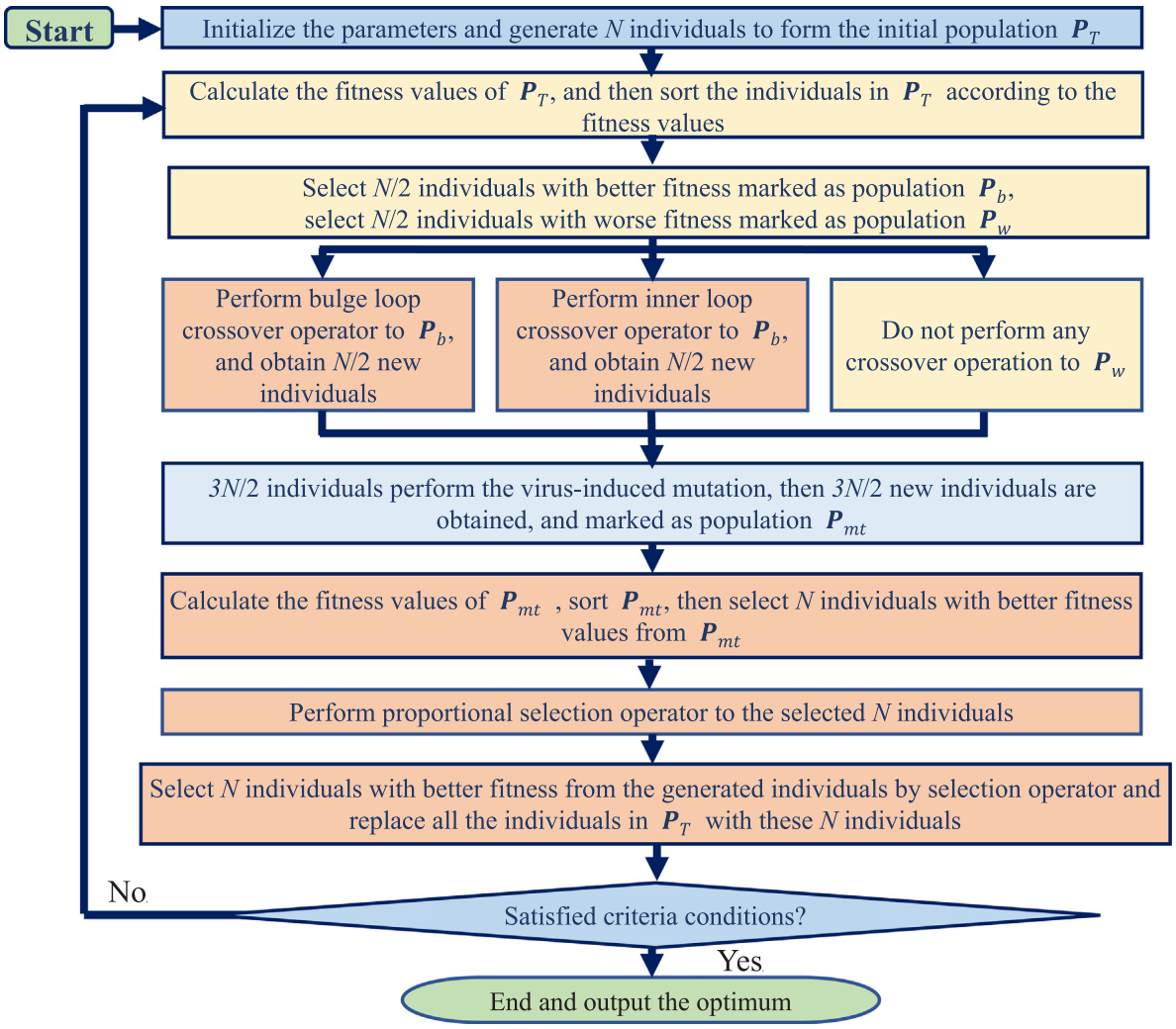


Fig. 2. The flowchart of the bvRNA-GA.

Step (b.3): If one individual is a bulge loop strand, and the other is a normal strand, the bulge loop will dissociate from the root bases and move to the allelic bases, after that, switch to Step (b.5); Else switch to Step (b.4).

Step (b.4): If both of the individuals are normal strands, they are directly reserved to $\boldsymbol{P}_{nb}$ without any operation, switch to Step (b.1); If $\boldsymbol{P}_b$ is none, stop performing the crossover operator.

Step (b.5): Restore the encoding length of I_1 and I_2 by moving the redundant bases of the longer one to the right of the shorter one (like Fig. 3(6)); Then two new individuals (nI_1 and nI_2) are obtained as Fig. 3(7).

Step (b.6): Put nI_1, nI_2 to $\boldsymbol{P}_{nb}$, and switch to Step (b.1); If $\boldsymbol{P}_b$ is none, stop performing the crossover operator.

3.4.2. The inner loop crossover operator

The procedures of the inner loop crossover operator:

Step 1: Randomly select two individuals from the copied $\boldsymbol{P}_b$ (once they are selected, remove them from the copied $\boldsymbol{P}_b$), matching these two individuals and align at the left.

Step 2: Traversal the allelic bases in the double-strands from high bits to low bits, the first complementary base position is defined as d_1. Then, searching from the position (d_1+1) to position L to find the second complementary base position d_2.

Step 3: If both positions d_1, d_2 are found, the inner loop structure is formed, rotate the inner loop structure for 180 degrees, the selected two individuals exchange their allelic bases between the positions d_1 and d_2, then two new individuals are obtained; Reserving them to $\boldsymbol{P}_{nb}$; Else if, directly reserving these two individuals to $\boldsymbol{P}_{nb}$.

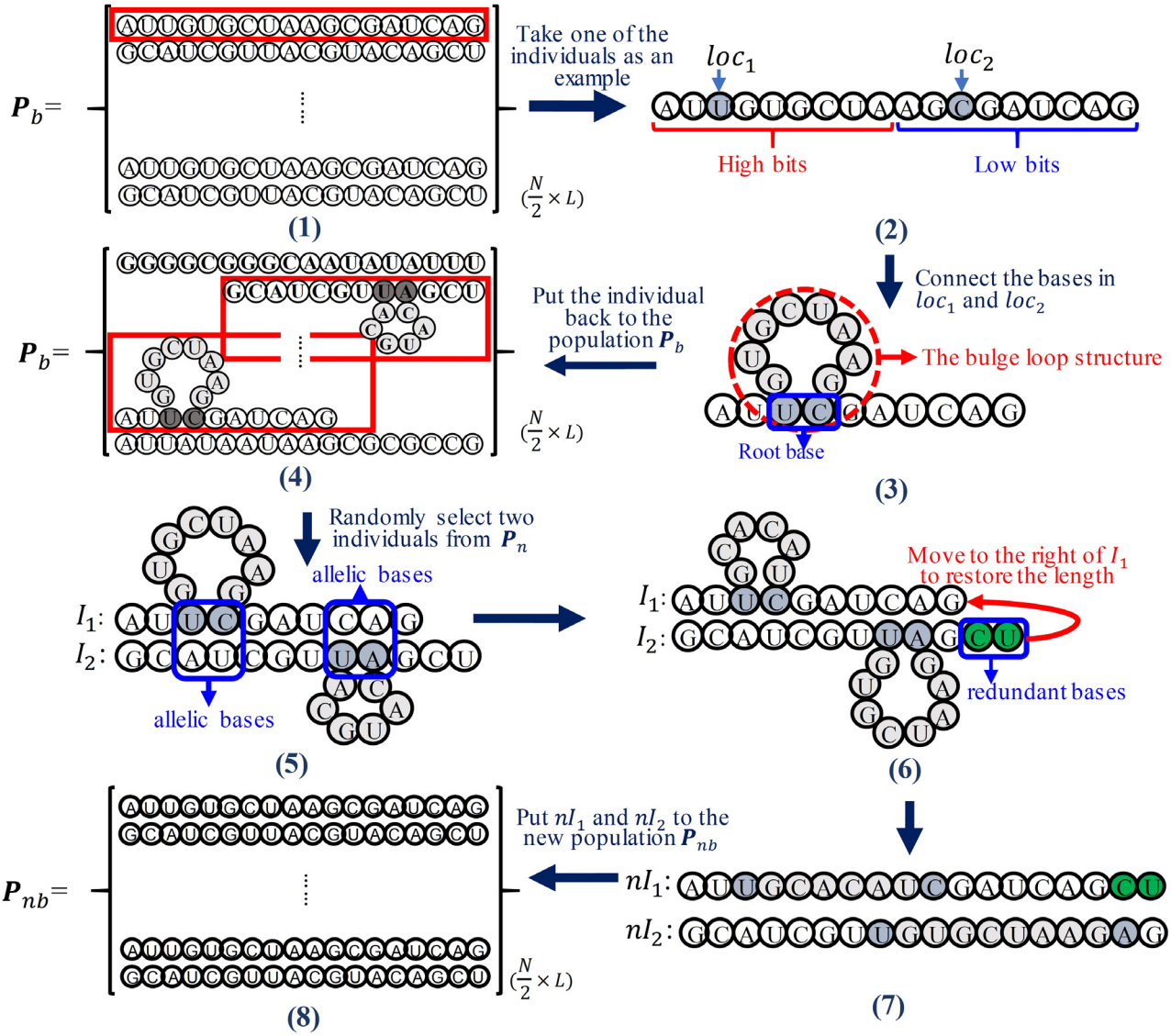
Step 4: If the copied $\boldsymbol{P}_b$ is none, stop performing the crossover operator; Else go to Step 1.

At last, after performing the bulge loop crossover operator and the inner loop crossover operator, there are N individuals in $\boldsymbol{P}_{nb}$.

3.5. The virus-induced mutation operator

The mutation operator is an important way to enhance the population diversity of GA. In this section, the virus-induced mutation operator is designed inspired by the genetic characteristic of virus individuals (the infection ability and the evolution ability) [53]. The infection ability is the capacity of the virus to transmit self-genetic information to the host individual. The evolution ability is the ability to change its self-genetic information with the genetic information of the host individual.

The genetic material in the virus usually contains many invalid genes, we represent these invalid bases with the symbol '*', an example of a virus individual is shown in Fig. 4(a)(1) and (b)(1).



(1) Population P_b before forming the bulge loop structure; (2) Randomly choose two positions; (3) Form the bulge loop structure; (4) Population P_b after forming the bulge loop structure; (5) Match I_1 and I_2, and align at left; (6) Perform the crossover operation and restore the encoding length of I_1 and I_2 ; (7) nI_1 and nI_2 are obtained after opening the structures; (8) The new population P_{nb} after bulge loop crossover operation is obtained.

Fig. 3. The schematic of the bulge loop crossover operator.

The virus population (P_v) is generated in each iteration. P_{nb} and P_w are called the host population in this section. It is nothing that, the virus load in P_v is determined by the total individual numbers of P_{nb} and P_w.

The procedures of virus-induced mutation operator are described as follows:

Step 1: Perform Step (a.1) to Step (a.3) (described in Section 3.4.1) to all the individuals in P_{nb} and P_w, and later, the individuals are marked as bulge loop strand and normal strand.

Step 2: Randomly generate $\lfloor N_t/10 \rfloor$ virus individuals to form the virus population P_v, where $N_t = 3N/2$. The length of virus individuals is L as well. When generating the bases of virus individuals, the base in current position L_c is the normal base or the invalid base is determined by probability p_m which is calculated by the following equation.

$$p_m = 0.02 + 0.08 \cdot \left(\frac{\varepsilon\left(\frac{L}{2} - L_c\right)}{1 + \exp(\delta \cdot (g - \frac{G}{2}))} + \frac{\varepsilon\left(L_c - \frac{L}{2} - 1\right)}{1 + \exp(-\delta \cdot (g - \frac{G}{2}))} \right) \tag{12}$$

where $\varepsilon(\tau)$ is jump function with two values $\{\varepsilon(\tau) = 1|\tau \geq 0\}$ and $\{\varepsilon(\tau) = 1|\ \tau < 0\}$; g is the current iteration; $\delta = 60/G$ which decides the changing rate of p_m.

Step 3: Randomly select a virus individual from P_v to infect the host individual select from the host population (shown in Fig. 4(a)(1) and Fig. 4(b)(1)). If the host individual is a normal strand (like Fig. 4(a)(2)), go to Step 4; If host individual is the bulge loop strand (shown in Fig. 4(b)(2)), go to Step 5.

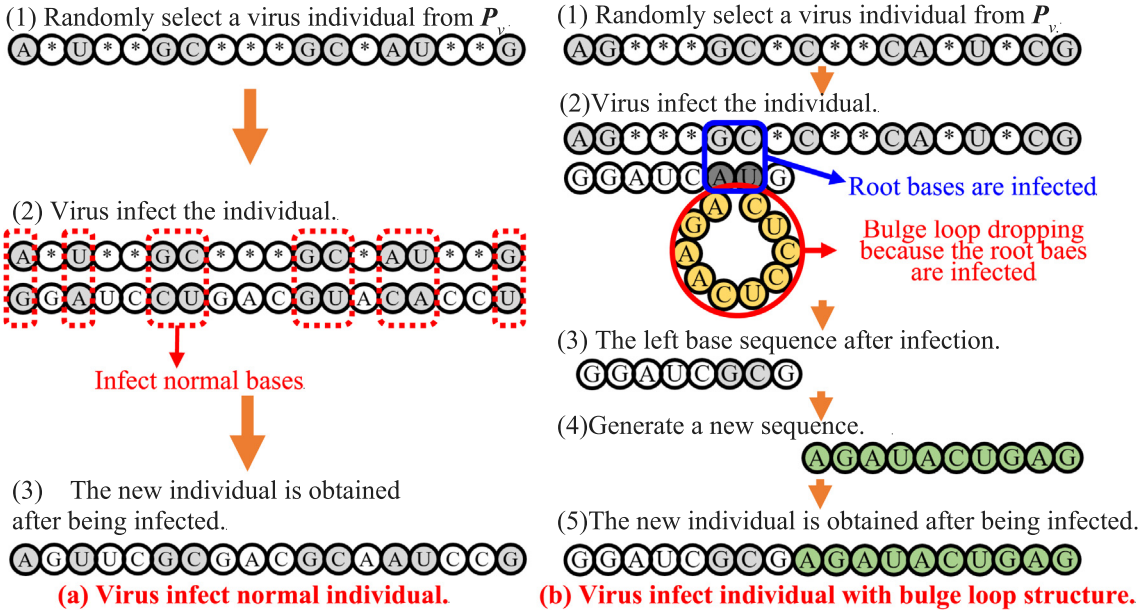


Fig. 4. The schematic operation of the virus-induced mutation operator.

Step 4: All the allelic bases (only the valid bases of virus individual can form allelic bases with the host individual) in the host individual are replaced by the valid bases of the virus, the infected process is shown in Fig. 4(a)(2), and the new individual is obtained after being infected and is shown in Fig. 4(a)(3), then reserving the new individual to population $\boldsymbol{P}_{mt}$, go to Step 8.

Step 5: There are two kinds of operation cases for bulge loop strand host individuals. Case 1: If the root bases of the bulge loop structure are infected by the valid bases of virus individual (shown in Fig. 4(b)(2)), go to Step 6; Case 2: If the host individual is the bulge loop strand, but the root bases are not infected or not completely infected by the valid bases of the virus, go to Step 7.

Step 6: Firstly, The bulge loop structure dropping, the left base sequence is shown in Fig. 4(b)(3) (other bases in host individual are not influenced by valid bases of virus individual); Secondly, to restore the length of the host individual, a random sequence is generated (shown in Fig. 4(b)(4)); Finally, the new individual is obtained after connecting the left base sequence with the random sequence (shown in Fig. 4(b)(5)), then reserving the new individual to population $\boldsymbol{P}_{mt}$, go to Step 8.

Step 7: All the allelic bases in the host individual are replaced by the valid bases of the virus; The new host individual is obtained by opening the bulge loop structure, reserving the new individual to population $\boldsymbol{P}_{mt}$, go to Step 8.

Step 8: When infecting the host individual, the valid bases and invalid bases of the virus individual is also changed according to probability. In other word, the valid bases may be changed into invalid bases or other different bases, and the invalid bases may be changed into a kind of valid bases; The changing probability for virus bases at the current location L_c is $p_r = 0.5 \cdot p_m$.

Step 9: Put the changed virus individual to $\boldsymbol{P}_v$, go to Step 3; If all the host individuals have been operated, stop the mutation operator.

4. Experimental data results and discussions

To assess the performance of the bvRNA-GA, benchmark function tests are carried out in this section. The experimental results of bvRNA-GA are compared with RNA-GA [38], Evolution Strategy with Covariance Matrix Adaptation (CMA-ES) [54], Success-History based Adaptive Differential Evolution incorporates Linear Population Size Reduction (L-SHADE) [55], Coyote Optimization Algorithm (COA) [56] to verify its searching capacity.

4.1. Benchmark functions and parameter setting

10 benchmark functions are used in this work, the detailed of these functions is shown in Table 1, where there are 5 typical benchmark functions (F1–F5) and 5 IEEE Congress on Evolutionary Computation (CEC)' 2014 benchmark functions (F6–F10) [57]. Among these functions, F1, F3 to F5, F6, F7 are unimodal problems, that is, there is only one global optimal value in the searching boundaries, they can better test the exploration capacity of the optimization algorithms. Especially, F6 and F7 are obtained by rotating typical benchmark functions which increase the searching difficulty. F2, F4, F8–F10 are multimodal functions, they have a global optimal value and a lot of local minimum. To obtain more credible experimental results, all the five algorithms are working in the same hardware and software, and the parameters of the RNA-GA [38], CMA-ES [54], L-SHADE [55], COA [56] are listed in Table 2, they are directly extracted from their original references. 50 times independently running are done on dimensions (D) 2, 10, and 30 for all the benchmark functions. The only termination criterion is to achieve the maximum iteration $G = 500 \cdot D$. The hardware and software environment used in this work is described as follows:

- **Hardware** Intel Core i5-7th CPU with frequency 2.5 GHz, 8 GB memory.
- **Software:** MATLAB 2017b working on the Windows 10 64-bit Professional operating system.

4.2. Discussion on the impacts of the improvement strategies

In this section, the performance of the crossover operators and the mutation operator is investigated. CRNA-GA is bvRNA-GA without using the virus-induced mutation operator. MRNA-GA is bvRNA-GA without using the bulge-loop crossover operator and

Table 1
A detailed description of the benchmark functions.

Typical benchmark functions				
ID	Function detail		Bounds	Optimum
F1	The ***Rosenbrock's***	$f_1(\boldsymbol{x}) = \sum_{i=1}^{D-1}\left[(x_i-1)^2 + 100\left(x_{i+1}-x_i^2\right)^2\right]$	$[-5, 5]$	0
F2	The ***Levy's***	$f_2(\boldsymbol{x}) = \sin^2(\pi w_1) + \sum_{i=1}^{D}(w_i-1)^2\left[1+10\sin^2(\pi w_i+1)\right] + (w_d-1)^2\left[1+\sin^2(2\pi w_d)\right]$, *where* $w_i = 1+\frac{x_i-1}{4}$	$[-10, 10]$	0
F3	The ***Rotated Hyber-Ellipsoid***	$f_3(\boldsymbol{x}) = \sum_{i=1}^{D}\sum_{j=1}^{i}x_j^2$	$[-65.36, 65.36]$	0
F4	The ***Sum of different powers***	$f_4(\boldsymbol{x}) = \sum_{i=1}^{D}\lvert x_i\rvert^{i+1}$	$[-1, 1]$	0
F5	The ***Sum Squares***	$f_5(\boldsymbol{x}) = \sum_{i=1}^{D} ix_i^2$	$[-5.12, 5.12]$	0
IEEE CEC' 2014 benchmark functions [57]				
ID	Function detail		Bounds	Optimum
F6	The ***Rotated High Conditioned Elliptic Function*** in IEEE CEC'2014 dataset		$[-100, 100]$	100
F7	The ***Rotated Discus Function*** in IEEE CEC'2014 dataset		$[-100, 100]$	300
F8	The ***Shifted and Rotated Ackley's Function*** in IEEE CEC'2014 dataset		$[-100, 100]$	500
F9	The ***Shifted and Rotated Katsuura Function*** in IEEE CEC'2014 dataset		$[-100, 100]$	1200
F10	The ***Shifted and Rotated Expanded Scaffer's Function*** in IEEE CEC'2014 dataset		$[-100, 100]$	1600

Table 2
Parameter setting of the referred algorithms.

Algorithm	N	G	Encoding length	Other parameters
bvRNA-GA	100		$L = 20{\cdot}D$	$p_{bl} = 0.5$, $p_{il} = 0.5$
RNA-GA	100		$L = 20{\cdot}D$	$p_c = 0.5$, $p_t = 1$, $a_1 = 0.02$, $b_1 = 0.2$, $g_0 = G/2$, $aa = 20/G$
CMA-ES	100	$500{\cdot}D$	*None*	$\sigma_F = 0.5$, $\mu = 5$, $\lambda = 10$
L-SHADE	100		*None*	$r^{N^{init}} = 18$, $r^{arc} = 2.6$, $p = 0.11$, $H = 6$
COA	100		*None*	$N_p = 20$, $N_c = 5$

Table 3
Comparison of the optimal result of bvRNA-GA, CRNA-GA, MRNA-GA, and RNA-GA on F1–F10 ($D = 10$).

No.	bvRNA-GA		CRNA-GA		MRNA-GA		RNA-GA		Rank
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	
F1	**2.8600E−01**	**9.6589E−02**	6.9464E+00	1.4092E+01	1.1127E+01	5.9792E+01	2.3775E+01	6.1851E+01	1/2/3/4
F2	**4.3190E−09**	**2.0672E−17**	6.0809E−03	8.4847E−05	2.2609E−02	8.9706E−04	3.8819E−03	1.0686E−05	1/3/4/2
F3	**6.5115E−08**	**3.2246E−14**	2.4033E−04	6.4765E−07	3.8627E+00	2.0723E+00	1.6899E+02	2.5887E+03	1/2/3/4
F4	**9.0953E−13**	**1.1737E−32**	1.5191E−08	3.0011E−15	9.8622E−10	4.1131E−18	6.7789E−06	7.7015E−12	1/2/3/4
F5	**3.8489E−09**	**1.2555E−16**	2.1922E−05	7.1773E−09	4.6460E−02	3.6463E−04	4.6454E−01	1.3483E−02	1/2/3/4
F6	3.0436E+05	7.4994E+10	7.0755E+06	3.3064E+13	**2.4597E+05**	**4.4127E+10**	6.1935E+05	2.4699E+11	2/4/1/3
F7	**1.5614E+03**	1.1430E+06	8.3391E+03	2.0623E+07	1.6543E+03	**9.4070E+05**	6.2080E+03	9.0555E+06	1/4/2/3
F8	5.1371E+02	4.9723E+01	5.2013E+02	3.0737E−03	**5.1191E+02**	**3.1298E+01**	5.1505E+02	1.9424E+01	2/4/1/3
F9	**1.2000E+03**	**6.1955E−06**	1.2005E+03	4.2425E−02	1.2002E+03	2.3709E−03	1.2002E+03	2.1394E−03	1/4/2/2
F10	1.6022E+03	9.1261E−02	1.6028E+03	1.3250E−01	**1.6018E+03**	1.0919E−01	1.6027E+03	**1.4909E−02**	2/4/1/3

the inner loop crossover operator. The working conditions and parameter setting of bvRNA-GA, CRNA-GA, MRNA-GA, RNA-GA stay the same with Section 4.1. The mean values (Mean) and standard deviation values (SD) on 10-dimensional benchmark functions are provided in Table 3.

First, according to Table 3, bvRNA-GA wins all the other variations in 7 of the 10 functions, so each component of the proposal (bulge-loop crossover operator, inner loop crossover operator, virus-induced mutation operator) is useful to improve the results. The CRNA-GA wins both MRNA-GA and RNA-GA on F1 to F5, however, MRNA-GA owns superior performance than CRNA-GA and RNA-GA on F6 to F10. From the above results, one can conclude that CRNA-GA with better local searching (exploration) ability, because it has good performance in solving easy problems like F1 to F5. But MRNA-GA achieves better results on much more complex problems like F6 to F10 for its good global searching (exploitation) capacity.

4.3. Discussion and comparison of the numerical results

Based on the parameters set and the conclusion in Sections 4.1 and 4.2, this section compared three evaluation indexes of bvRNA-GA (include the mean (Mean) value, the median (Median) value, and the standard deviation (SD) value) for the 10 benchmark functions with the results of RNA-GA, CMA-ES, L-SHADE, COA. The statistical results on 2-dimensional, 10-dimensional, 30-dimensional are respectively shown in Tables 4–6. The rank results are according to the mean values.

It can be seen from the experimental results in Table 4 that bvRNA-GA obtains the best results in 9 of 10 benchmark functions except for F6. In particular, the results for Mean, Median, SD of bvRNA-GA on F1–F5 are significantly superior to RNA-GA, CMA-ES, COA, L-SHADE. bvRNA-GA and RNA-GA find the best solution for F8, F9, F10, CMA-ES in the same searching performance with bvRNA-GA on F9, and L-SHADE also find the best solution for F10.

Table 4
Comparison of the numerical results by bvRNA-GA, RNA-GA, CMA-ES, COA, and L-SHADE for the 10 benchmark functions on 2-dimensional.

Function	Index	bvRNA-GA	RNA-GA	CMA-ES	COA	L-SHADE	Rank
F1	Mean	**1.2781E−11**	1.8811E−03	2.9621E−03	1.7408E−01	6.3857E−03	1/2/3/5/4
	Median	**1.7077E−15**	5.4786E−07	4.2356E−04	1.4247E−01	1.6911E−03	
	SD	**2.7105E−21**	6.6965E−05	4.7948E−05	2.6258E−02	6.3873E−05	
F2	Mean	**6.5528E−14**	1.3111E−06	3.9488E−01	3.9613E−01	4.2710E−01	1/2/3/4/5
	Median	**3.1739E−15**	1.5034E−08	4.2677E−01	4.0419E−01	4.2711E−01	
	SD	**2.1937E−26**	2.9911E−11	5.9597E−03	7.2037E−04	1.2639E−09	
F3	Mean	**4.6849E−11**	7.3972E−08	4.1930E−08	1.6649E+00	4.1388E−04	1/3/2/5/4
	Median	**1.0601E−20**	4.3326E−08	7.8633E−09	1.3776E+00	3.0274E−04	
	SD	**4.3895E−20**	1.2416E−14	5.4476E−15	2.0780E+00	2.3616E−07	
F4	Mean	**8.9849E−16**	5.1300E−09	1.7028E−05	5.2704E−03	1.6251E−05	1/2/4/5/3
	Median	**9.0949E−16**	3.4152E−09	1.1752E−05	4.2163E−03	9.3181E−06	
	SD	**2.6723E−33**	4.4486E−17	3.4300E−10	1.5124E−05	2.7654E−10	
F5	Mean	**4.2849E−23**	5.3000E−10	2.2624E−10	5.9996E−06	4.7773E−10	1/4/2/5/3
	Median	**2.1685E−23**	2.2256E−10	7.7633E−12	9.5133E−07	8.4933E−12	
	SD	**3.6971E−45**	1.1424E−18	5.9799E−19	2.1750E−10	2.2417E−18	
F6	Mean	3.5178E+02	3.5668E+02	1.1081E+02	5.3952E+03	**1.0540E+02**	3/4/2/5/1
	Median	3.2225E+02	2.8550E+02	1.0354E+02	4.4796E+03	**1.0175E+02**	
	SD	5.0784E+04	7.7155E+04	2.2270E+02	4.1183E+07	**1.2371E+02**	
F7	Mean	**3.2567E+02**	7.7280E+02	1.2170E+03	1.0407E+04	3.5085E+02	1/3/4/5/2
	Median	3.1770E+02	5.7058E+02	1.1829E+03	5.7336E+03	**3.0523E+02**	
	SD	**6.7547E+02**	2.4418E+05	3.7617E+05	1.6976E+08	1.5981E+04	
F8	Mean	**5.0019E+02**	5.0051E+02	5.2000E+02	5.0953E+02	5.0027E+02	1/3/4/5/2
	Median	**5.0000E+02**	**5.0000E+02**	5.2000E+02	5.1005E+02	5.0015E+02	
	SD	8.9968E−01	3.3356E−01	**3.1913E−04**	1.2665E+01	8.2912E−02	
F9	Mean	**1.2000E+03**	**1.2000E+03**	**1.2000E+03**	1.2007E+03	1.2008E+03	1/1/1/4/5
	Median	**1.2000E+03**	**1.2000E+03**	**1.2000E+03**	1.2007E+03	1.2008E+03	
	SD	1.3448E−03	**3.5506E−06**	2.5781E−05	1.8863E−01	1.1058E−01	
F10	Mean	**1.6000E+03**	**1.6000E+03**	1.6010E+03	1.6001E+03	**1.6000E+03**	1/1/5/4/1
	Median	**1.6000E+03**	**1.6000E+03**	1.6010E+03	1.6001E+03	**1.6000E+03**	
	SD	3.5595E−05	7.5531E−05	**1.6009E−21**	8.2462E−03	3.4260E−06	

Table 5
Comparison of the numerical results by bvRNA-GA, RNA-GA, CMA-ES, COA and L-SHADE for the 10 benchmark functions on 10-dimensional.

Function	Index	bvRNA-GA	RNA-GA	CMA-ES	COA	L-SHADE	Rank
F1	Mean	**2.8600E−01**	2.3775E+01	5.8776E+00	6.6564E+02	3.7415E+00	1/4/4/5/2
	Median	**9.0280E−02**	2.3980E+01	5.9033E+00	5.9250E+02	3.8361E+00	
	SD	**9.6589E−02**	6.1851E+01	2.7933E−01	1.6803E+05	1.6319E+00	
F2	Mean	**4.3190E−09**	3.8819E−03	2.5801E−01	3.1078E+00	4.1207E−01	1/2/3/5/4
	Median	**2.4359E−09**	3.8581E−03	2.1577E−01	3.1608E+00	4.2711E−01	
	SD	**2.0672E−17**	1.0686E−05	7.5132E−03	8.6125E−01	2.5344E−03	
F3	Mean	**6.5115E−08**	1.6899E+02	1.1848E−05	2.2807E+03	2.5858E−06	1/4/3/5/2
	Median	**1.2732E−09**	1.7559E+02	8.2695E−06	2.3685E+03	3.7032E−07	
	SD	**3.2246E−14**	2.5887E+03	1.0400E−10	1.0263E+06	1.9924E−11	
F4	Mean	**9.0952E−13**	6.7789E−06	1.1296E−05	1.6093E−02	1.2569E−12	1/3/4/5/2
	Median	9.0949E−13	6.5299E−06	9.1952E−06	1.5196E−02	**1.6779E−13**	
	SD	**1.1737E−32**	7.7015E−12	5.1728E−11	9.8316E−05	8.4103E−24	
F5	Mean	**3.8489E−09**	4.6454E−01	6.7018E−06	8.5483E+00	4.2107E−09	1/4/3/5/2
	Median	**1.4416E−10**	4.3248E−01	5.3768E−06	8.5069E+00	1.1569E−09	
	SD	1.2555E−16	1.3483E−02	2.3101E−11	1.1259E+01	**3.5871E−17**	
F6	Mean	3.0436E+05	6.1935E+05	4.5431E+04	8.1980E+06	**3.9784E+03**	3/4/2/5/1
	Median	1.6899E+05	4.8417E+05	3.2032E+04	7.3301E+06	**5.9173E+02**	
	SD	7.4994E+10	2.4699E+11	2.5602E+09	2.7981E+13	**6.3631E+07**	
F7	Mean	1.5614E+03	6.2080E+03	3.8890E+03	1.6654E+04	**3.0008E+02**	2/4/3/5/1
	Median	1.1960E+03	5.6136E+03	3.5827E+03	1.6851E+04	**3.0004E+02**	
	SD	1.1430E+06	9.0555E+06	3.2945E+06	2.9980E+07	**1.4714E−02**	
F8	Mean	**5.1371E+02**	5.1505E+02	5.2015E+02	5.2042E+02	5.2026E+02	1/2/3/5/4
	Median	5.1877E+02	**5.1377E+02**	5.2010E+02	5.2042E+02	5.2020E+02	
	SD	4.9723E+01	1.9424E+01	3.9240E−02	**7.1961E−03**	2.9823E−02	
F9	Mean	**1.2000E+03**	1.2002E+03	**1.2000E+03**	1.2010E+03	1.2007E+03	1/3/1/5/4
	Median	**1.2000E+03**	1.2002E+03	**1.2000E+03**	1.2011E+03	1.2005E+03	
	SD	**6.1955E−06**	2.1394E−03	6.7833E−03	3.8244E−02	1.7777E−01	
F10	Mean	**1.6022E+03**	1.6027E+03	1.6044E+03	1.6038E+03	1.6029E+03	1/2/5/4/3
	Median	**1.6023E+03**	1.6027E+03	1.6044E+03	1.6038E+03	1.6029E+03	
	SD	9.1261E−02	1.4909E−02	**4.9194E−04**	2.7474E−02	1.1862E−01	

According to Table 5, the search precision of bvRNA-GA in solving 10-dimensional problems is lower than before. But remain good performance than the other 4 algorithms, except losing the competition with L-SHADE on F6 and F7. From Table 6,

Table 6
Comparison of the numerical results by bvRNA-GA, RNA-GA, CMA-ES, COA and L-SHADE on the 10 benchmark functions when 30-dimensional.

Function	Index	bvRNA-GA	RNA-GA	CMA-ES	COA	L-SHADE	Rank
F1	Mean	**2.2115E+01**	3.4739E+03	2.5557E+01	4.3025E+03	2.6163E+01	1/4/2/5/3
	Median	2.8413E+01	3.4854E+03	2.5516E+01	4.4054E+03	**2.2733E+01**	
	SD	1.5846E+02	6.8767E+04	**1.7447E−01**	2.3636E+06	1.5274E+02	
F2	Mean	**4.6512E−06**	1.1893E+01	2.2634E−01	2.0917E+01	5.1303E−01	1/5/2/4/3
	Median	**3.1099E−06**	1.1824E+01	2.1578E−01	2.1048E+01	5.0713E−01	
	SD	**2.9186E−11**	1.6033E+00	2.2332E−03	1.8935E+01	5.9458E−02	
F3	Mean	4.6512E−06	1.9197E+04	**2.4761E−06**	2.7107E+04	8.6349E−06	2/4/1/5/3
	Median	3.1099E−06	1.9321E+04	**1.5095E−06**	2.7077E+04	5.4539E−06	
	SD	2.9186E−11	3.8211E+06	**8.1656E−12**	5.6916E+07	8.7813E−11	
F4	Mean	**1.2882E−10**	4.9137E−05	1.2037E−04	4.5965E−03	4.0816E−10	1/3/4/5/2
	Median	2.0539E−11	4.7263E−05	1.1421E−04	3.9359E−03	**7.0980E−12**	
	SD	**7.2327E−20**	1.3031E−10	5.4479E−09	1.1713E−05	2.4327E−18	
F5	Mean	**3.3114E−08**	9.1680E+01	3.1518E−06	1.6620E+02	4.3350E−08	1/4/3/5/2
	Median	3.0050E−08	9.2197E+01	1.4244E−06	1.6599E+02	**2.8919E−09**	
	SD	**3.6204E−16**	1.0825E+02	1.1060E−11	1.0735E+03	1.1114E−14	
F6	Mean	1.1381E+06	8.1698E+07	1.8407E+06	8.3476E+07	**8.6282E+05**	2/4/3/5/1
	Median	1.1701E+06	8.1548E+07	1.7653E+06	8.2667E+07	**7.7592E+05**	
	SD	**2.1700E+11**	1.3461E+14	3.0246E+11	5.8439E+14	3.1004E+11	
F7	Mean	9.5264E+03	2.4843E+04	2.1028E+04	5.0752E+04	**3.3813E+02**	2/3/4/5/1
	Median	8.9577E+03	2.5301E+04	2.1146E+04	4.9105E+04	**3.1481E+02**	
	SD	3.4700E+07	2.0981E+07	1.8915E+07	1.6825E+08	**2.1838E+03**	
F8	Mean	**5.2006E+02**	5.2085E+02	5.2059E+02	5.2072E+02	5.2044E+02	1/5/3/4/2
	Median	5.2010E+02	5.2084E+02	**5.2002E+02**	5.2074E+02	5.2040E+02	
	SD	1.3071E−01	**2.3545E−03**	4.4754E−01	1.2973E−02	2.3095E−02	
F9	Mean	**1.2000E+03**	1.2012E+03	**1.2000E+03**	1.2011E+03	1.2005E+03	1/5/1/4/3
	Median	**1.2000E+03**	1.2012E+03	**1.2000E+03**	1.2011E+03	1.2005E+03	
	SD	1.1999E−05	1.5349E−02	**6.5395E−06**	4.3896E−02	1.0794E−02	
F10	Mean	**1.6095E+03**	1.6117E+03	1.6130E+03	1.6128E+03	1.6108E+03	1/3/5/4/2
	Median	**1.6093E+03**	1.6117E+03	1.6130E+03	1.6128E+03	1.6108E+03	
	SD	5.5659E−01	4.3817E−02	**4.8814E−04**	6.4953E−02	1.2872E−01	

as the number of dimensions increases to 30, bvRNA-GA gains very competitive results on benchmark functions F1 to F5, F9, F10. Especially, the results for F2 to F5, F9, F10 are much closer to the optimum values of these benchmark functions. The results of bvRNA-GA on 2, 10, and 30-dimensional better verify the good searching capacity.

4.4. Analysis about the Wilconxon's Rank-Sum test and convergence

Statistical comparison is the crucial way to test the reliability of the intelligent algorithms [58,59], in this work, Wilconxon's Rank-Sum test [60] is performed to verify the superiority and significance of the bvRNA-GA. The experimental results obtained in Section 4.3 are adopted to compute the p-value at a 5% significance level of bvRNA-GA versus RNA-GA, CMA-ES, COA, and L-SHADE, respectively. The p-value results are collected in Table 7, p-value < 0.05 (marked with '+') means the bvRNA-GA has the statistically significant performance to the compared algorithms. If p-value > 0.05 (marked with '−') denotes bvRNA-GA shows worse performance against the compared algorithms. If p-value $= 0.05$ means bvRNA-GA has similar performance to the compared algorithms.

From Table 7, it can be seen that all p-value of bvRNA-GA *vs.* RNA-GA and bvRNA-GA *vs.* COA are <0.05. The p-value of bvRNA-GA *vs.* CMA-ES are <0.05 except F2 (on 2-dimensional), F2, F6 (on 10-dimensional), F1, F2, F3, F9 (on 30-dimensional). The p-value of bvRNA-GA *vs.* L-SHADE are <0.05 except F6, F7 (on 2-dimensional), F4, F6, F7 (on 10-dimensional), F1, F4, F5, F6, F7 (on 30-dimensional).

4.5. Analysis about the convergence

The average convergence curve of bvRNA-GA, RNA-GA, COA, L-SHADE for the 10 benchmark functions on 10-dimensional are plotted in Fig. 5 (CMA-ES is not included because it is not a population-based algorithm). Furthermore, some convergence curves are zoomed in to make their convergence trend easier to see. Fig. 5 demonstrates that bvRNA-GA converges faster to the best solution than other referred algorithms at the early stage with higher accuracy. Additionally, bvRNA-GA remains good searching performance at the middle and last stages.

The results as above indicate that the bvRNA-GA is very competitive to other algorithms, it also verifies the efficiency for the newly designed bulge loop crossover operator, the inner loop crossover operator, and the virus-induced mutation operator.

5. The modeling results of PV systems

PV systems are very important modules for converting light energy into electrical energy. However, the conversion efficiency of PV systems is easily influenced by the complex outside environment, such as ambient temperature, solar radiance. Therefore, it is important to research more accurate and efficient modeling solutions for PV systems. In this section, the LSSVM modeling approach is applied to model PV systems, and the bvRNA-GA, RNA-GA, CMA-ES, LSHADE, and COA are utilized to optimize the hyper-parameters of LSSVM. The modeling performance of the bvRNA-GA optimized LSSVM (bvRNA-GA-lSSVM) is compared with RNA-GA optimized LSSVM (RNA-GA-LSSVM), CMA-ES optimized LSSVM (CMA-ES-LSSVM), L-SHADE optimized LSSVM (LSHADE-LSSVM), and COA optimized LSSVM (COA-LSSVM). The parameters of the five algorithms are consistent with those in Table 2, except that the population size is $N = 40$, and the maximum iteration number $G = 50$. Furthermore, the modeling methods will independently run 20 times on each PV system.

Table 7
Test results of algorithms using Wilconxon's Rank-Sum test.

bvRNA-GA *vs.*								
D	No.	RNA-GA (*p*-value (+/ = /−))		CMA-ES (*p*-value (+/ = /−))		COA (*p*-value (+/ = /−))		L-SHADE (*p*-value (+/ = /−))	
2	F1	1.38E−10	+	7.25E−12	+	7.25E−12	+	7.25E−12	+
	F2	7.25E−12	+	1.00E+00	−	7.25E−12	+	7.25E−12	+
	F3	7.25E−12	+	5.08E−11	+	7.25E−12	+	7.25E−12	+
	F4	7.25E−12	+	7.25E−12	+	7.25E−12	+	7.25E−12	+
	F5	7.25E−12	+	7.25E−12	+	7.25E−12	+	7.25E−12	+
	F6	4.52E−01	+	1.00E+00	−	4.62E−07	+	1.00E+00	−
	F7	2.06E−06	+	1.45E−11	+	7.25E−12	+	9.09E−01	−
	F8	3.17E−02	+	7.25E−12	+	7.25E−12	+	1.52E−02	+
	F9	2.79E−02	+	4.74E−03	+	1.45E−11	+	7.25E−12	+
	F10	2.22E−01	+	7.25E−12	+	7.25E−12	+	7.25E−12	+
Number of (+/ = /−)		10/0/0		8/0/2		10/0/0		8/0/2	
10	F1	7.25E−12	+	7.25E−12	+	7.25E−12	+	7.25E−12	+
	F2	7.25E−12	+	1.00E+00	−	7.25E−12	+	7.25E−12	+
	F3	7.25E−12	+	7.25E−12	+	7.25E−12	+	5.63E−07	+
	F4	7.25E−12	+	7.25E−12	+	7.25E−12	+	9.97E−01	−
	F5	7.25E−12	+	7.25E−12	+	7.25E−12	+	9.05E−03	+
	F6	1.05E−02	+	1.00E+00	−	7.25E−12	+	1.00E+00	−
	F7	6.64E−09	+	1.07E−05	+	7.25E−12	+	1.00E+00	−
	F8	1.32E−01	−	6.84E−07	+	7.25E−12	+	7.25E−12	+
	F9	7.25E−12	+	7.04E−10	+	7.25E−12	+	7.25E−12	+
	F10	2.04E−07	+	7.25E−12	+	7.25E−12	+	2.54E−08	+
Number of (+/ = /−)		9/0/1		8/0/2		10/0/0		7/0/3	
30	F1	7.25E−12	+	9.97E−01	−	7.25E−12	+	9.74E−01	−
	F2	7.25E−12	+	1.00E+00	−	7.25E−12	+	7.25E−12	+
	F3	7.25E−12	+	9.97E−01	−	7.25E−12	+	2.56E−01	+
	F4	7.25E−12	+	7.25E−12	+	7.25E−12	+	9.64E−01	−
	F5	7.25E−12	+	7.25E−12	+	7.25E−12	+	9.99E−01	−
	F6	7.25E−12	+	8.83E−05	+	7.25E−12	+	9.75E−01	−
	F7	3.69E−09	+	2.51E−07	+	7.25E−12	+	1.00E+00	−
	F8	7.25E−12	+	1.63E−02	+	1.97E−09	+	6.80E−05	+
	F9	7.25E−12	+	1.00E+00	−	7.25E−12	+	7.25E−12	+
	F10	7.25E−12	+	7.25E−12	+	7.25E−12	+	6.71E−08	+
Number of (+/ = /−)		10/0/0		6/0/4		10/0/0		5/0/5	

5.1. The evaluation indicator

Researchers usually apply a certain evaluation indicator to evaluate the performance of the obtained models [61,62]. In this work, we adopt three different evaluation indicators to assess the performance of bvRNA-GA-LSSVM, RNA-GA-LSSVM, CMA-ES-LSSVM, L-SHADE-LSSVM, and COA-LSSVM for modeling PV systems. Three kinds of evaluation indicators include the root mean square error (RMSE), the mean absolute error (MAE), and the R-Square (R^2) are selected. Where RMSE was chosen as the evaluation criteria during the training process of the LSSVM model to help to optimize hyper-parameters during the training process. MAE and R^2 are applied to verify the performance of the obtained LSSVM model, because they can directly show the error deviation and fitting degree between the real output and the model output of LSSVM, respectively. They are respectively calculated according to the following equations:

$$RMSE(\boldsymbol{I},\boldsymbol{Y}) = \sqrt{\frac{1}{M}\bullet\sum_{i=1}^{M}(\boldsymbol{I}_i-\boldsymbol{Y}_i)} \tag{13}$$

$$MAE(\boldsymbol{I},\boldsymbol{Y}) = \frac{1}{M}\bullet\sum_{i=1}^{M}|\boldsymbol{I}_i-\boldsymbol{Y}_i| \tag{14}$$

$$\begin{cases} R^2(\boldsymbol{I},\boldsymbol{Y}) = \max\left(0, 1-\dfrac{MSE(\boldsymbol{I},\boldsymbol{Y})}{Var(\boldsymbol{I})}\right) \\ MSE(\boldsymbol{I},\boldsymbol{Y}) = RMSE(\boldsymbol{I},\boldsymbol{Y})^2 \end{cases} \tag{15}$$

where $\boldsymbol{I}$ is the real output of PV systems; $\boldsymbol{Y}$ represents the LSSVM model output; M is the dataset length; $Var(\bullet)$ is the standard deviation function.

5.2. Implementing the PV system modeling procedures

The modeling framework of bvRNA-GA optimized LSSVM for modeling PV systems is shown in Fig. 6. To facilitate the description of the LSSVM modeling procedure for the PV system, we take bvRNA-GA-LSSVM modeling of SDM (25 °C, 1000 W/m^2) as an example. The procedures about how to pre-process the dataset, initialize LSSVM and bvRNA-GA, modeling implementation and model validation are described as follows:

Step 1: Sample dataset. According to the described sampling method in Section 2.1.1, we sample the dataset of SDM (working at 25 °C, 1000 W/m^2) with 400 pairs $\{(V_k, I_k)|k = 1,\ldots,300\}$ samples. Where 300 pairs samples are randomly selected as the training dataset ($M = 200$), and the rest 100 pairs samples are taken as the testing dataset to verify the effectiveness of the obtained LSSVM model.

Step 2: Initialize the parameters of bvRNA-GA and set the LSSVM model structure, where the topological structure of LSSVM has six inputs and one output. The form of one of the row in the training dataset $\boldsymbol{\Theta}$ (with size $M\times 6$) before normalization and in output $\boldsymbol{Y}$ (with size $M\times 1$) after anti-normalization are respectively described as follows:

$$\begin{cases} Y_i = \left[\hat{I}\right]_i \\ \boldsymbol{\Theta}_i = [V(t-1), V(t-2), V(t-3), \\ \qquad I(t), I(t-1), I(t-2)] \\ i = 1, 2, \ldots, M \end{cases} \tag{16}$$

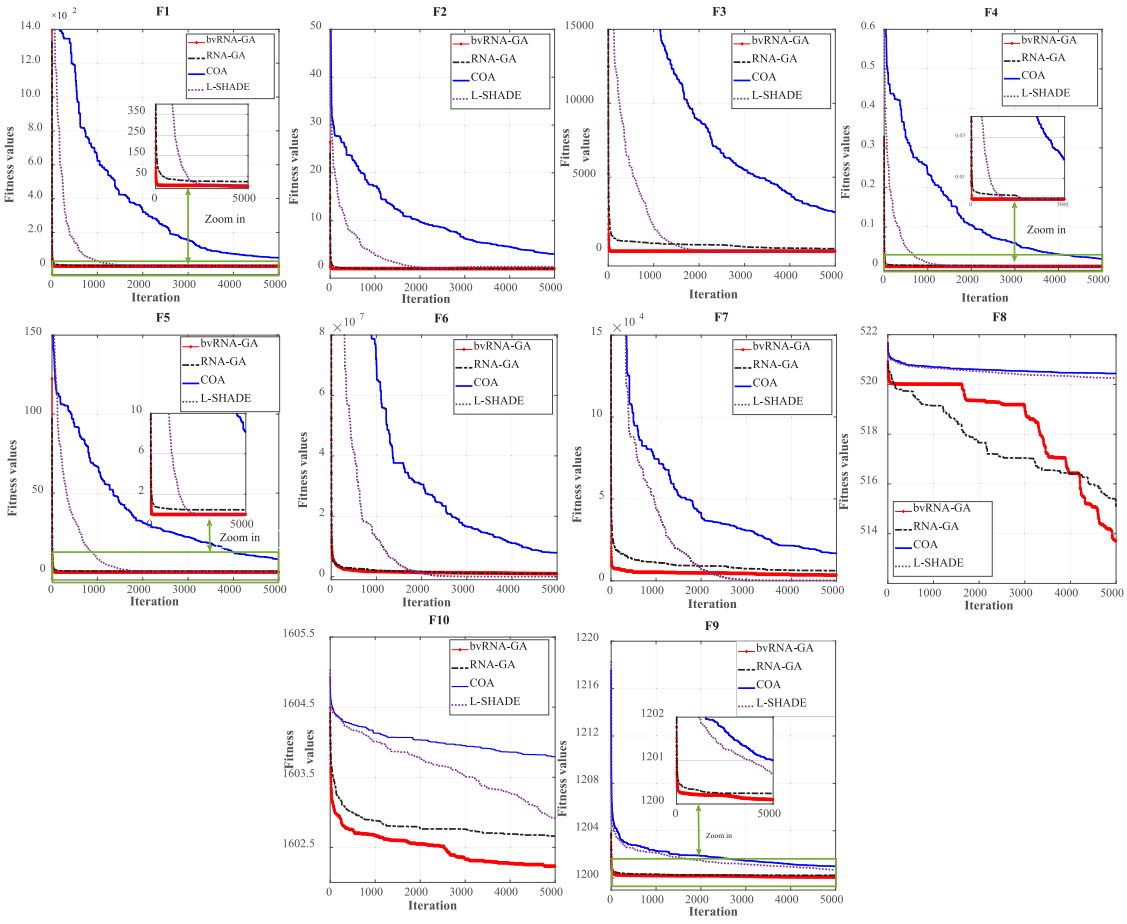


Fig. 5. The average evolution curves of bvRNA-GA, RNA-GA, COA and L-SHADE for the 10 benchmark functions on 10-dimensional.

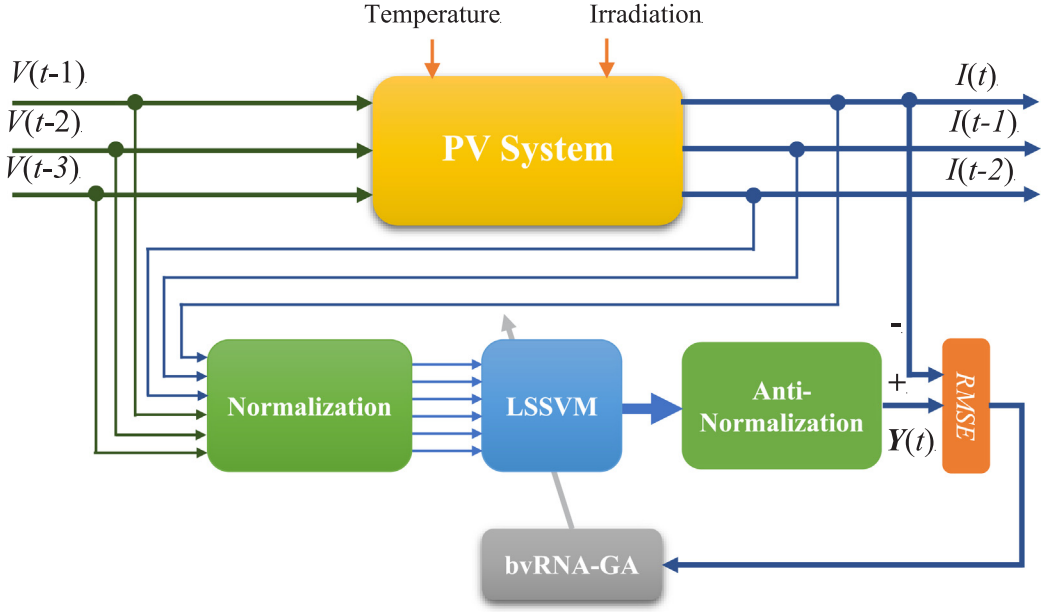


Fig. 6. The framework of bvRNA-GA optimized LSSVM for modeling PV systems.

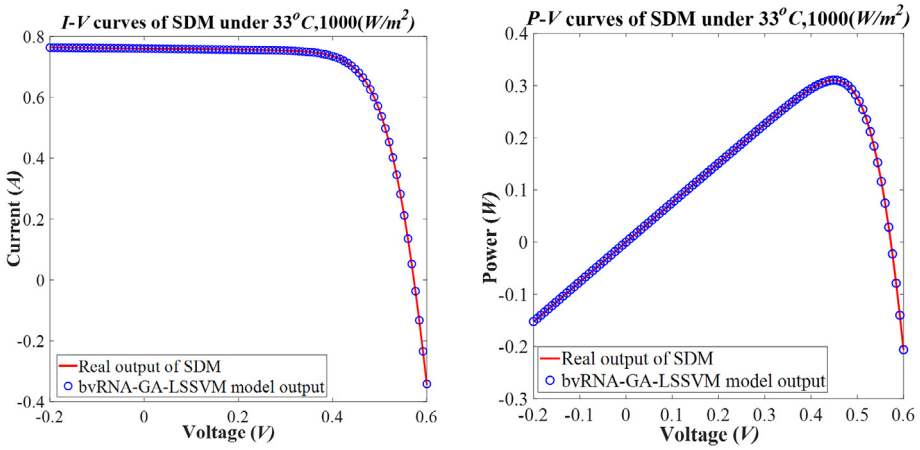


Fig. 7. *I-V* and *P-V* curves comparison between the real output of SDM and bvRNA-GA-LSSVM model output when the temperature is 33 °*C*.

where $\hat{I}$ represents the LSSVM model output, t is the sampling interval.

Step 3: Normalize the input training dataset $\boldsymbol{\Theta}$ according to the following equation:

$$\widehat{\boldsymbol{\Theta}} = \frac{\boldsymbol{\Theta} - \boldsymbol{\varpi}}{\boldsymbol{\delta}} \tag{17}$$

where $\widehat{\boldsymbol{\Theta}}$ is the normalized training dataset in same size with $\boldsymbol{\Theta}$. $\boldsymbol{\varpi}$ and $\boldsymbol{\delta}$ are respectively the mean value vector and standard deviation vector of $\boldsymbol{\Theta}$, and they all have length 6.

Step 4: Implement bvRNA-GA to optimize the hyper-parameters (σ and γ) of the LSSVM, according to the modeling framework in Fig. 6.

Step 4.1: Randomly generate N individuals which form the initial population of bvRNA-GA. Each individual in the population represents the potential solution of hyper-parameters σ and γ.

Step 4.2: In order to obtain the corresponding RMSE value for each individual, one successively inputs each individual to LSSVM and uses the normalized training dataset ($\widehat{\boldsymbol{\Theta}}_{M\times 6}$) to train LSSVM. After that, one obtains the LSSVM output and easily gets the RMSE value by using Eq. (13).

Step 4.3: Implement the optimization procedure of bvRNA-GA to perform the bulge loop crossover operator, the inner loop crossover operator, the virus-induced mutation operator, and the proportional selection orderly in the population.

Step 4.4: If achieve the maximum iteration, output the best individual which is the obtained optimal solution of hyper-parameters σ and γ; Otherwise, switch to Step 4.1.

Step 5: Validation. The model output is acquired after inputting the testing dataset to LSSVM (with the optimal hyper-parameters). Using RMSE, MAE to calculate the error between model output and the real output of SDM, and applying R^2 to evaluate the fitting performance of the obtained LSSVM model.

One can easily obtain the modeling results for different PV systems by changing the dataset in step 1 to different PV systems' datasets and performing step 1 to step 5.

The modeling results of RNA-GA-LSSVM, CMA-ES-LSSVM, L-SHADE-LSSVM, COA-LSSVM are acquired through the above modeling procedure except that replace bvRNA-GA with the corresponding optimization algorithm.

Based on the above parameter setting and the described modeling procedures, the modeling results of different modeling methods for the PV systems (SDM, DDM, S75, SM55, S40) are obtained. Therefore, we make a comparison about the performance of bvRNA-GA-LSSVM against the RNA-GA-LSSVM, CMA-ES-LSSVM, L-SHADE-LSSVM, COA-LSSVM in the following sections.

5.3. Modeling results of SDM and DDM

The modeling results of SDM and DDM at different working temperatures are collected in Tables 8 and 9, where the Mean is the mean values, SD is the standard deviation values of RMSE, MAE, R^2, the best results are bolded.

In terms of the RMSE results in Table 8, the results reveal the superiority of bvRNA-GA-LSSVM in modeling SDM with different working temperatures compared with the other four algorithms. Moreover, bvRNA-GA-LSSVM gains the first rank on 33 °*C*, 50 °*C*, 75 °*C*. LSHADE-LSSVM gets the best results on 25 °*C*, and COA-LSSVM wins bvRNA-GA-LSSVM on 100 °*C*. According to Table 9, bvRNA-GA-LSSVM achieves higher precision on the RMSE and MAE, which means that bvRNA-GA-LSSVM has more successful modeling performance than the other methods. Although it is concluded that better modeling results are respectively obtained by COA-LSSVM for DDM when the temperature is 75 °*C*. The R^2 results in Table 8, Table 9 for both SDM and DDM are close to 1.0 which proves the good fitting performance of the LSSVM obtained by bvRNA-GA.

To directly show the difference between the bvRNA-GA-LSSVM and the real output of SDM, DDM, the best-obtained model outputs are plotted in Figs. 7 to 10. Because the 33 °*C* is the standard working temperature of SDM and DDM, therefore, we draw the *I-V* curves and *P-V* curves in Figs. 7 and 8, separately. The power (W) values are obtained by multiplying the voltage with the current of the bvRNA-GA-LSSVM model output. As can be seen from Figs. 7 and 8, the *I-V* and *P-V* characteristic curves obtained by bvRNA-GA-LSSVM display little fluctuation than that of the real output of SDM and DDM. The bvRNA-GA-LSSVM model output almost matches all the points with the real output demonstrates the good mapping capacity of the bvRNA-GA optimized LSSVM. Figs. 9 and 10 further prove the good performance of the proposed modeling method, the bvRNA-GA-LSSVM model output can also better agree with the real output of SDM and DDM for other temperatures (25 °*C*, 50 °*C*, 75 °*C*, 100 °C).

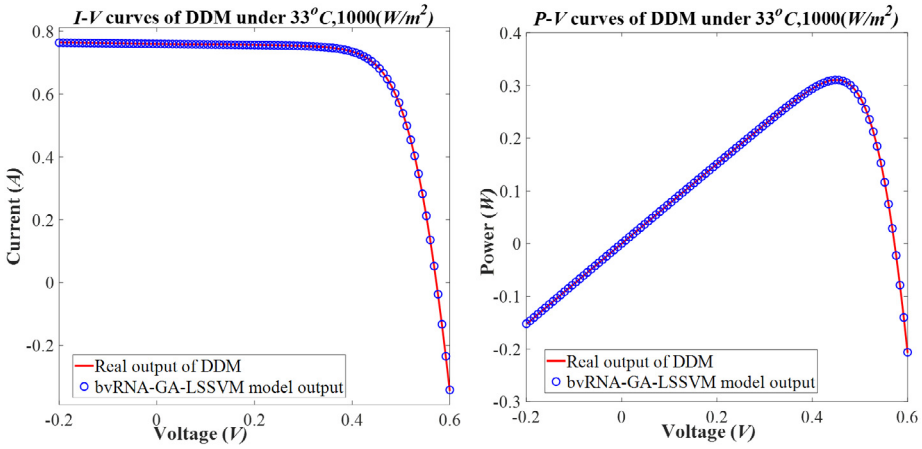


Fig. 8. *I-V* and *P-V* curves comparison between the real output of DDM and bvRNA-GA-LSSVM model output when the temperature is 33 °*C*.

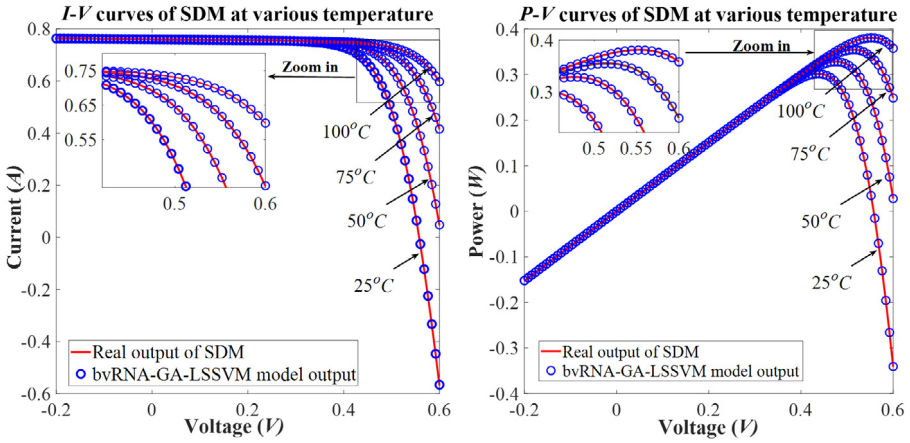


Fig. 9. *I-V* curves comparison between the real output of SDM and bvRNA-GA-LSSVM model output when the temperature is 25 °C, 50 °*C*, 75 °*C*, 100 °*C*, respectively.

Table 8
The modeling results comparison of different modeling approaches for SDM at various temperatures.

Algorithm		bvRNA-GA-LSSVM			RNA-GA-LSSVM			CMA-ES-LSSVM			LSHADE-LSSVM			COA-LSSVM			Rank
Indicator		RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	
25 °C	Mean	4.26E−04	1.35E−04	**1.00E+00**	4.26E−04	1.35E−04	9.97E−01	7.08E−02	5.75E−02	8.69E−01	**4.23E−04**	**1.34E−04**	**1.00E+00**	4.25E−04	1.35E−04	**1.00E+00**	3/3/5/1/2
	SD	**8.46E−29**	3.01E−28	**4.61E−32**	3.97E−17	3.55E−18	9.55E−21	4.74E−04	3.18E−04	6.45E−03	1.64E−10	3.60E−12	3.77E−14	1.89E−11	1.24E−12	4.40E−15	
33 °C	Mean	**4.07E−04**	**1.31E−04**	**1.00E+00**	8.07E−04	2.31E−04	9.99E−01	6.35E−02	5.20E−02	8.95E−01	4.20E−04	1.34E−04	9.90E−01	4.08E−04	1.32E−04	9.91E−01	1/4/5/3/2
	SD	**6.58E−29**	5.87E−29	**1.49E−32**	3.51E−17	7.62E−18	7.74E−21	3.94E−04	2.69E−04	7.08E−03	8.30E−11	1.29E−11	1.94E−14	1.92E−12	2.81E−13	4.28E−16	
50 °C	Mean	**3.87E−04**	1.44E−04	**1.00E+00**	3.84E−03	1.48E−03	9.98E−01	**3.53E−02**	2.92E−02	9.68E−01	3.94E−04	1.32E−04	**1.00E+00**	**3.87E−04**	1.43E−04	**1.00E+00**	1/4/5/3/1
	SD	**7.12E−14**	2.37E−11	**1.47E−17**	3.77E−14	2.41E−11	8.11E−18	2.77E−04	1.97E−04	1.18E−03	6.22E−11	9.15E−11	1.34E−14	7.26E−13	2.82E−11	1.43E−16	
75 °C	Mean	**2.88E−04**	1.77E−04	**1.00E+00**	8.77E−03	3.73E−03	1.00E+00	1.02E−02	7.80E−03	9.98E−01	3.07E−04	1.72E−04	1.00E+00	2.89E−04	1.76E−04	1.00E+00	1/4/5/3/2
	SD	**9.52E−30**	2.55E−29	**5.84E−33**	9.29E−17	2.53E−17	1.02E−20	1.54E−05	1.29E−05	3.53E−06	3.29E−10	7.12E−11	4.60E−14	1.97E−11	1.36E−11	2.28E−15	
100 °C	Mean	2.82E−04	2.11E−04	**1.00E+00**	2.84E−04	2.16E−04	**1.00E+00**	7.54E−03	4.97E−03	9.99E−01	2.87E−04	1.97E−04	9.99E−01	**2.81E−04**	2.02E−04	**1.00E+00**	2/3/5/4/1
	SD	1.26E−11	1.22E−10	1.32E−15	**8.90E−18**	4.69E−15	**1.00E−21**	4.37E−06	3.10E−06	5.24E−07	6.65E−11	5.66E−10	7.46E−15	1.39E−11	2.31E−10	1.46E−15	

Table 9
The modeling results comparison of different modeling approaches for DDM at various temperatures.

Algorithm		bvRNA-GA-LSSVM			RNA-GA-LSSVM			CMA-ES-LSSVM			LSHADE-LSSVM			COA-LSSVM			Rank
Indicator		RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	
25 °C	Mean	**4.20E−04**	**1.34E−04**	**1.00E+00**	1.99E−03	3.36E−03	**1.00E+00**	6.50E−02	5.30E−02	8.98E−01	4.26E−04	1.34E−04	**1.00E+00**	4.24E−04	1.34E−04	**1.00E+00**	1/4/5/3/2
	SD	**2.88E−26**	**3.43E−28**	**6.68E−30**	1.02E−16	1.30E−17	2.39E−20	1.93E−04	1.31E−04	2.39E−03	4.79E−11	4.24E−12	1.15E−14	8.79E−12	1.46E−14	2.09E−15	
33 °C	Mean	**3.95E−04**	**1.27E−04**	**1.00E+00**	9.56E−04	2.71E−04	9.97E−01	5.22E−02	4.29E−02	9.36E−01	4.17E−04	1.32E−04	**1.00E+00**	3.97E−04	1.27E−04	**1.00E+00**	1/4/5/3/2
	SD	**2.30E−26**	**5.89E−27**	4.76E−30	4.13E−17	9.66E−18	8.59E−21	2.05E−04	1.41E−04	1.22E−03	1.26E−10	5.83E−12	2.88E−14	4.88E−11	9.14E−13	1.09E−14	
50 °C	Mean	**3.69E−04**	**1.45E−04**	**1.00E+00**	3.87E−04	1.97E−04	9.98E−01	2.92E−02	2.42E−02	9.76E−01	3.91E−04	1.36E−04	9.99E−01	3.80E−04	1.44E−04	**1.00E+00**	1/3/5/4/2
	SD	**6.46E−25**	**5.48E−26**	1.23E−28	3.19E−17	4.03E−18	6.08E−21	2.64E−04	1.88E−04	1.51E−03	3.81E−11	1.18E−10	7.88E−15	8.62E−12	2.54E−11	1.68E−15	
75 °C	Mean	2.78E−04	**1.81E−04**	**1.00E+00**	2.88E−04	1.86E−04	9.98E−01	1.04E−02	7.85E−03	9.98E−01	2.96E−04	1.76E−04	**1.00E+00**	**2.80E−04**	1.79E−04	9.99E−01	2/3/5/4/1
	SD	**1.27E−27**	**3.70E−28**	**1.62E−31**	2.08E−17	2.17E−18	2.12E−21	1.73E−05	1.42E−05	4.40E−06	1.72E−10	7.47E−11	2.14E−14	7.76E−11	4.14E−11	8.98E−15	
100 °C	Mean	**2.81E−04**	**2.06E−04**	**1.00E+00**	2.83E−04	2.14E−04	9.99E−01	6.82E−03	4.26E−03	9.99E−01	2.86E−04	2.05E−04	**1.00E+00**	**2.81E−04**	2.02E−04	**1.00E+00**	1/3/5/4/1
	SD	1.83E−11	1.75E−10	1.91E−15	**3.82E−12**	**3.65E−11**	4.00E−16	1.83E−06	8.52E−07	1.60E−07	5.63E−11	2.40E−10	6.28E−15	1.44E−11	2.10E−10	1.50E−15	

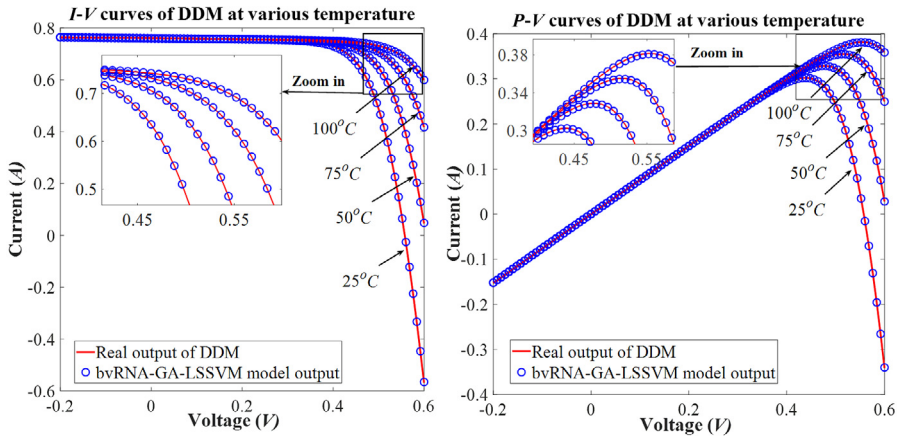


Fig. 10. *I-V* and *P-V* curves comparison between the real output of DDM and bvRNA-GA-LSSVM model output when the temperature is 25 °C, 50 °C, 75 °C, respectively.

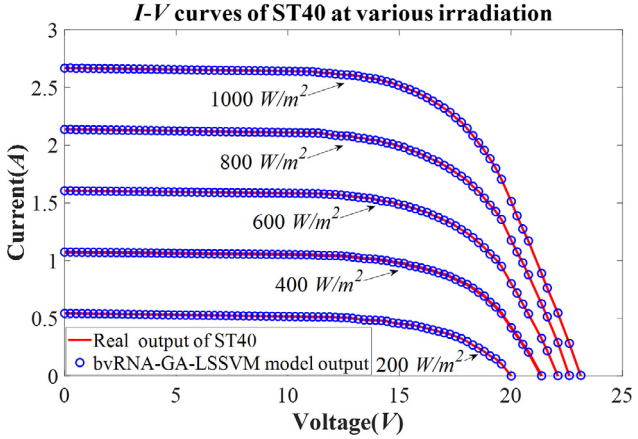


Fig. 11. *I-V* curves comparison between the real output and bvRNA-GA-LSSVM model output under different irradiance for ST40 module.

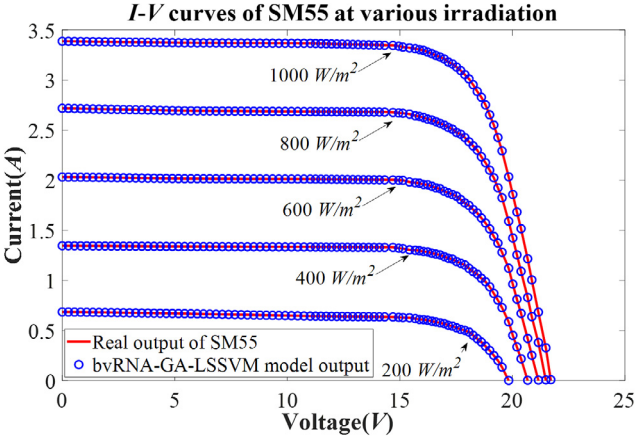


Fig. 12. *I-V* curves comparison between the real output and bvRNA-GA-LSSVM model output under different irradiance for SM55 module.

5.4. Modeling results of PV solar modules under different working irradiation

To further test the practicability of the proposed modeling method, bvRNA-GA-LSSVM, RNA-GA-LSSVM, CMA-ES-LSSVM, L-SHADE-LSSVM, COA-LSSVM are employed to model the characteristic curves for three different PV solar modules (ST40, SM55, S75). In this work, they work under five kinds of irradiations (1000 W/m^2, 800 W/m^2, 600 W/m^2, 400 W/m^2, 200 W/m^2) with temperature is 25 °C are studied. In the same way, all the modeling methods run 20 times for every working condition of each PV solar module, and the Mean, SD values of RMSE, MAE, R^2 are shown in Tables 10 to 12. The *I-V* curves of ST40, SM55, S75 are respectively obtained by inputting the test dataset to LSSVM with corresponding optimal hyper-parameters, and the characteristic curves under different working irradiations are shown in Figs. 11–13. In terms of the performance indicators comparison results in Table 10, the RMSE, MAE values of the bvRNA-GA-LSSVM are smaller than the other modeling methods, except that the RNA-GA-LSSVM and COA-LSSVM are respectively achieving the same precision on 200 W/m^2 and 1000 W/m^2 for ST40. The obtained characteristic curves of the ST40 by the bvRNA-GA-LSSVM model are plotted in Fig. 11, it is observable that bvRNA-GA-LSSVM produces accurate curves in all irradiation conditions.

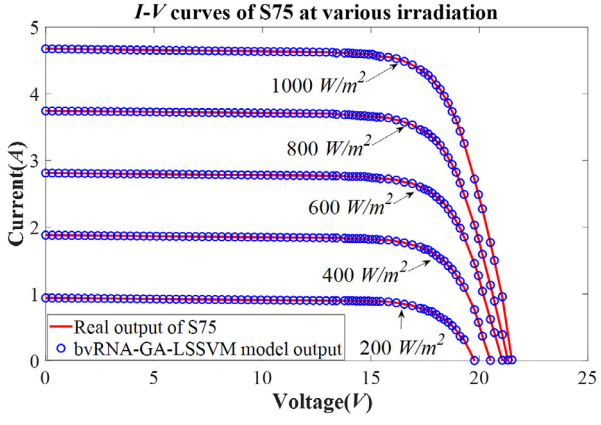


Fig. 13. *I-V* curves comparison between the real output and bvRNA-GA-LSSVM model output under different irradiance for S75 module.

Table 11 shows the comparison results for SM55 under different working irradiations. bvRNA-GA-LSSVM ranks first on 1000 W/m^2, 800 W/m^2, 600 W/m^2, 400 W/m^2, 200 W/m^2. The SD values for RMSE, MAE, R^2 proves the stability of the bvRNA-GA in tuning the hyper-parameters of LSSVM. Otherwise, COA-LSSVM also gets the best performance when irradiation is 800 W/m^2

Table 10
The modeling results comparison of different modeling approaches for ST40 under various working irradiations.

Algorithm		bvRNA-GA-LSSVM			RNA-GA-LSSVM			CMA-ES-LSSVM			LSHADE-LSSVM			COA-LSSVM			Rank
Indicator		RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	
1000 W/m^2	Mean	**2.13E−02**	**9.72E−03**	**1.00E+00**	2.96E−02	9.78E−03	**1.00E+00**	1.50E+00	1.16E+00	9.30E−01	2.14E−02	9.97E−03	**1.00E+00**	**2.13E−02**	9.89E−03	1.00E+00	1/4/5/3/1
	SD	5.83E−14	2.66E−11	7.76E−20	**1.37E−14**	**1.27E−11**	**2.00E−20**	1.78E−01	1.15E−01	1.94E−03	1.46E−08	5.57E−08	1.52E−14	1.26E−09	1.96E−08	1.33E−15	
800 W/m^2	Mean	**2.13E−02**	**9.64E−03**	**1.00E+00**	2.15E−02	9.67E−03	**1.00E+00**	1.31E+00	1.01E+00	9.46E−01	2.15E−02	1.01E−02	**1.00E+00**	2.14E−02	9.84E−03	**1.00E+00**	1/3/5/3/2
	SD	1.96E−15	**5.18E−14**	2.18E−21	**1.02E−15**	**2.40E−12**	**7.66E−22**	1.58E−01	9.92E−02	1.44E−03	6.56E−09	6.03E−08	7.34E−15	7.68E−09	7.68E−08	8.53E−15	
600 W/m^2	Mean	**8.20E−03**	**6.61E−03**	**1.00E+00**	8.21E−03	6.61E−03	**1.00E+00**	1.18E+00	9.24E−01	9.53E−01	8.54E−03	6.92E−03	**1.00E+00**	8.37E−03	6.76E−03	9.76E−01	1/2/5/4/3
	SD	**3.05E−14**	**6.18E−12**	**5.30E−21**	3.91E−14	1.68E−11	7.32E−21	1.81E−01	1.11E−01	1.24E−03	5.97E−08	9.50E−08	1.13E−14	1.20E−07	1.22E−07	2.36E−14	
400 W/m^2	Mean	**2.15E−02**	**9.44E−03**	**1.00E+00**	2.16E−02	9.67E−03	9.76E−01	7.92E−01	6.18E−01	9.76E−01	2.17E−02	9.85E−03	9.99E−01	2.17E−02	9.87E−03	**1.00E+00**	1/2/5/3/3
	SD	2.75E−14	4.41E−12	3.85E−20	**2.54E−14**	**2.62E−13**	**3.35E−20**	1.26E−01	8.80E−02	5.96E−04	9.98E−09	7.37E−08	1.32E−14	2.53E−08	2.48E−07	3.43E−14	
200 W/m^2	Mean	**4.65E−02**	**1.58E−02**	**1.00E+00**	**4.65E−02**	1.58E−02	**1.00E+00**	4.13E−01	2.80E−01	9.94E−01	4.67E−02	1.66E−02	**1.00E+00**	4.67E−02	1.65E−02	**1.00E+00**	1/1/5/2/2
	SD	2.31E−11	**2.94E−09**	2.24E−16	**4.16E−14**	5.79E−11	**3.38E−19**	6.27E−03	3.53E−03	5.03E−06	6.42E−09	2.06E−07	5.09E−14	2.23E−08	6.88E−07	1.86E−13	

Table 11
The modeling results comparison of different modeling approaches for SM55 under various working irradiations.

Algorithm		bvRNA-GA-LSSVM			RNA-GA-LSSVM			CMA-ES-LSSVM			LSHADE-LSSVM			COA-LSSVM			Rank
Indicator		RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	
1000 W/m^2	Mean	**1.18E−02**	**8.60E−03**	**1.00E+00**	1.19E−02	8.68E−03	**1.00E+00**	2.23E+00	1.82E+00	8.22E−01	1.30E−02	9.43E−03	**1.00E+00**	1.19E−02	8.63E−03	**1.00E+00**	1/2/5/4/2
	SD	**1.97E−29**	**7.26E−28**	**0.00E+00**	1.21E−12	9.46E−13	4.11E−19	1.93E−01	1.21E−01	9.13E−03	2.49E−07	1.07E−07	1.03E−13	1.00E−08	1.06E−08	3.53E−15	
800 W/m^2	Mean	**1.24E−02**	**1.08E−02**	**1.00E+00**	1.29E−02	1.09E−02	**1.00E+00**	1.96E+00	1.59E+00	8.48E−01	1.26E−02	1.10E−02	**1.00E+00**	**1.24E−02**	**1.08E−02**	1.00E+00	1/4/5/3/1
	SD	7.60E−14	1.29E−11	2.30E−20	**4.96E−15**	**2.30E−12**	**2.07E−21**	4.10E−01	2.91E−01	1.31E−02	4.83E−08	6.33E−08	1.99E−14	9.26E−11	2.63E−09	3.75E−17	
600 W/m^2	Mean	**1.02E−02**	**8.30E−03**	1.00E+00	1.18E−02	8.34E−03	**1.00E+00**	1.83E+00	1.53E+00	8.78E−01	1.04E−02	8.60E−03	**1.00E+00**	**1.02E−02**	8.38E−03	**1.00E+00**	1/4/5/3/1
	SD	2.41E−14	6.22E−12	5.87E−21	**2.09E−14**	**3.44E−12**	**5.31E−21**	1.76E−01	1.12E−01	5.33E−03	4.15E−08	4.66E−08	1.20E−14	1.06E−08	1.93E−08	2.98E−15	
400 W/m^2	Mean	**2.93E−02**	**1.15E−02**	**1.00E+00**	2.96E−02	1.51E−02	**1.00E+00**	2.03E+00	1.66E+00	8.35E−01	2.94E−02	1.18E−02	**1.00E+00**	2.97E−02	1.16E−02	1.00E+00	1/3/5/2/4
	SD	**1.18E−16**	**9.05E−13**	**3.44E−22**	2.78E−16	9.02E−13	7.93E−22	2.88E−01	1.75E−01	8.55E−03	6.65E−09	6.65E−08	1.65E−14	1.00E−08	1.35E−07	2.45E−14	
200 W/m^2	Mean	**8.29E−03**	**7.37E−03**	**1.00E+00**	**8.29E−03**	**7.37E−03**	**1.00E+00**	4.48E−01	3.37E−01	9.93E−01	9.25E−03	8.10E−03	**1.00E+00**	8.88E−03	7.81E−03	**1.00E+00**	1/1/5/4/3
	SD	**9.71E−14**	**1.43E−11**	**1.94E−20**	7.62E−13	1.74E−11	1.65E−19	2.31E−02	1.66E−02	2.12E−05	3.09E−07	1.66E−07	8.58E−14	3.94E−07	2.48E−07	1.09E−13	

Table 12
The modeling results comparison of different modeling approaches for S75 under various working irradiations.

Algorithm		bvRNA-GA-LSSVM			RNA-GA-LSSVM			CMA-ES-LSSVM			LSHADE-LSSVM			COA-LSSVM			Rank
Indicator		RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	
1000 W/m^2	Mean	**4.96E−02**	**1.85E−02**	**1.00E+00**	8.96E−02	4.85E−02	9.98E−01	1.71E+00	1.39E+00	8.75E−01	**4.96E−02**	1.88E−02	**1.00E+00**	**4.96E−02**	1.86E−02	9.98E−01	1/4/5/1/1
	SD	3.82E−15	2.17E−11	2.51E−21	**1.13E−16**	**4.36E−13**	**6.23E−22**	4.28E−01	2.97E−01	2.01E−02	2.63E−09	4.13E−08	1.64E−14	8.45E−10	4.95E−08	5.97E−15	
800 W/m^2	Mean	**3.03E−02**	**1.54E−02**	**1.00E+00**	3.27E−02	1.54E−02	9.98E−01	1.70E+00	1.36E+00	8.93E−01	**3.03E−02**	1.55E−02	**1.00E+00**	**3.03E−02**	1.55E−02	**1.00E+00**	1/4/5/1/1
	SD	1.46E−14	1.68E−11	6.58E−20	**1.38E−16**	9.54E−13	**8.19E−22**	2.03E−01	1.41E−01	3.93E−03	3.93E−09	1.93E−08	9.30E−15	8.26E−10	2.24E−08	2.08E−15	
600 W/m^2	Mean	**6.94E−02**	**2.50E−02**	**1.00E+00**	6.98E−02	2.75E−02	**1.00E+00**	1.46E+00	1.20E+00	9.12E−01	6.99E−02	2.53E−02	9.98E−01	6.98E−02	2.53E−02	**1.00E+00**	1/2/5/4/2
	SD	**1.73E−16**	**9.01E−13**	1.13E−20	2.52E−16	1.41E−12	2.36E−20	3.29E−01	2.30E−01	6.74E−03	9.92E−10	6.93E−08	1.60E−14	1.03E−09	1.02E−07	1.75E−14	
400 W/m^2	Mean	**8.07E−02**	**3.24E−02**	**1.00E+00**	8.63E−02	3.84E−02	**1.00E+00**	9.79E−01	8.00E−01	9.60E−01	8.07E−02	3.28E−02	**1.00E+00**	8.07E−02	3.28E−02	9.98E−01	1/4/5/1/1
	SD	2.93E−14	2.21E−12	7.36E−19	**1.32E−16**	**6.92E−13**	**2.19E−20**	1.99E−01	1.49E−01	1.76E−03	1.51E−09	1.22E−07	3.84E−14	2.05E−09	1.83E−07	5.32E−14	
200 W/m^2	Mean	**1.06E−01**	**4.19E−02**	**1.00E+00**	1.61E−01	4.19E−02	9.98E−01	5.27E−01	3.77E−01	9.90E−01	1.68E−01	4.24E−02	1.00E+00	1.66E−01	4.23E−02	**1.00E+00**	1/2/5/4/3
	SD	5.27E−13	2.17E−13	9.41E−19	**1.95E−15**	**2.64E−13**	2.05E−19	3.31E−02	2.50E−02	7.18E−05	1.16E−09	1.07E−07	5.77E−14	1.42E−09	1.46E−07	7.94E−14	

Table 13
The modeling results comparison of different modeling approaches for ST40 under different temperatures.

Algorithm		bvRNA-GA-LSSVM			RNA-GA-LSSVM			CMA-ES-LSSVM			LSHADE-LSSVM			COA-LSSVM			Rank
Indicator		RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	
20 °C	Mean	**4.15E−01**	**1.73E−01**	**9.96E−01**	5.53E−01	1.79E−01	9.91E−01	2.10E+00	1.55E+00	8.76E−01	4.67E−01	1.73E−01	9.90E−01	4.56E−01	1.77E−01	9.95E−01	1/4/5/3/2
	SD	**4.04E−29**	**4.06E−25**	**1.18E−30**	1.70E−15	1.38E−13	1.34E−18	2.18E−02	2.01E−02	4.16E−04	5.49E−09	3.39E−07	4.09E−12	3.99E−10	3.17E−08	3.20E−13	
30 °C	Mean	**2.95E−01**	**1.14E−01**	**9.98E−01**	2.98E−01	1.14E−01	9.97E−01	1.77E+00	1.30E+00	9.08E−01	**2.95E−01**	1.15E−01	9.91E−01	2.98E−01	1.86E−01	9.92E−01	1/3/5/1/3
	SD	**4.85E−29**	**6.40E−25**	**1.40E−31**	6.50E−16	2.24E−13	4.33E−19	4.90E−02	3.76E−02	5.58E−04	1.44E−09	2.95E−07	5.78E−13	7.96E−11	2.50E−08	5.09E−14	
40 °C	Mean	**1.39E−01**	**6.36E−02**	**1.00E+00**	1.40E−01	6.36E−02	9.99E−01	1.59E+00	1.14E+00	9.22E−01	1.40E−01	6.46E−02	9.99E−01	1.40E−01	6.36E−02	9.99E−01	1/2/5/2/2
	SD	**1.00E−29**	**2.59E−25**	**8.43E−33**	1.77E−14	1.67E−13	9.59E−19	3.24E−02	2.73E−02	4.40E−04	6.14E−08	7.75E−07	3.57E−12	2.18E−12	4.95E−11	1.34E−16	
50 °C	Mean	**7.56E−02**	**3.14E−02**	**1.00E+00**	7.70E−02	3.49E−02	**1.00E+00**	1.39E+00	1.02E+00	9.34E−01	**7.56E−02**	3.16E−02	**1.00E+00**	**7.56E−02**	3.17E−02	**1.00E+00**	1/4/5/1/1
	SD	**1.44E−17**	**2.17E−14**	**7.07E−22**	1.90E−17	5.62E−13	1.53E−20	5.67E−02	3.62E−02	7.12E−04	2.66E−10	1.89E−08	5.69E−15	7.19E−10	3.56E−08	1.56E−14	
60 °C	Mean	**3.63E−02**	**1.61E−02**	**1.00E+00**	3.64E−02	1.66E−02	**1.00E+00**	1.28E+00	9.07E−01	9.42E−01	3.64E−02	1.63E−02	**1.00E+00**	**3.63E−02**	1.62E−02	**1.00E+00**	1//3/5/3/1
	SD	3.59E−14	6.58E−13	1.83E−19	**6.06E−17**	**8.99E−14**	**5.07E−22**	3.47E−02	2.13E−02	3.66E−04	4.27E−09	3.89E−08	2.09E−14	7.53E−10	1.05E−08	3.79E−15	

and 600 W/m^2, respectively. The comparison between real output of SM55 and bvRNA-GA-LSSVM model output in Fig. 12 further demonstrate the superiority of the proposed modeling approach.

From Table 12, the Mean values for RMSE and MAE of bvRNA-GA-LSSVM are smaller than the other algorithms, the R^2 evaluation indicator is close to 1 also indicate the good fitting capacity on modeling S75. The ranking order of bvRNA-GA-LSSVM is always the first, while LSHADE-LSSVM and COA-LSSVM have similar performance to bvRNA-GA-LSSVM when irradiation is 1000 W/m^2, 800 W/m^2, 400 W/m^2. It is shown in Fig. 13 that the *I-V* curves output by bvRNA-GA optimized LSSVM can better fit the real output of S75 under five different working irradiations.

In conclusion, we can observe that LSSVM tuned by bvRNA-GA obtains the best results for modeling PV solar modules at different working irradiations.

5.5. Modeling results of PV solar modules under different working temperatures

In this section, the performance of bvRNA-GA-LSSVM on modeling ST40, SM55, S75 under five kinds of working temperatures (20 °C, 30 °C, 40 °C, 50 °C, 60 °C when irradiation level is 1000 W/m^2) are discussed. The RMSE, MAE, R^2 values are shown in Tables 13 to 15. In the same way, the *I-V* curves output by the studied PV solar modules and by bvRNA-GA-LSSVM are respectively shown in Figs. 14–16.

According to the results in Table 13, the results of bvRNA-GA-LSSVM remain its superiority on modeling ST40 under different working temperatures. LSHADE-LSSVM and COA-LSSVM respectively reach the same accuracy with the proposed modeling method for ST40 at 30 °C, 50 °C, 60 °C, however, the SD values are worse than bvRNA-GA-LSSVM. Fig. 14 shows that the output

Table 14
The modeling results comparison of different modeling approaches for SM55 under different temperatures.

Algorithm		bvRNA-GA-LSSVM			RNA-GA-LSSVM			CMA-ES-LSSVM			LSHADE-LSSVM			COA-LSSVM			Rank
Indicator		RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	
20 °C	Mean	**4.22E−01**	**1.70E−01**	**9.97E−01**	4.25E−01	1.71E−01	9.95E−01	3.55E+00	2.76E+00	4.50E−01	4.25E−01	1.72E−01	9.95E−01	4.25E−01	1.71E−01	9.95E−01	1/2/5/2/2
	SD	**4.52E−28**	**3.73E−24**	**6.70E−30**	3.02E−14	1.38E−13	1.69E−17	1.20E−02	2.42E−02	4.68E−03	7.77E−08	3.21E−07	4.32E−11	3.41E−08	2.64E−07	2.25E−11	
30 °C	Mean	**1.81E−01**	**7.00E−02**	**9.99E−01**	1.83E−01	7.19E−02	9.79E−01	2.48E+00	1.94E+00	7.57E−01	1.81E−01	7.22E−02	9.89E−01	1.84E−01	7.19E−02	9.89E−01	1/3/5/1/4
	SD	**4.72E−29**	**6.93E−25**	**5.90E−32**	9.46E−16	2.05E−14	1.10E−19	1.21E−01	8.62E−02	7.53E−03	4.54E−09	1.51E−07	3.32E−13	7.44E−10	1.00E−08	8.77E−14	
40 °C	Mean	**1.58E−01**	**5.60E−02**	**9.99E−01**	1.59E−01	5.68E−02	**9.99E−01**	2.25E+00	1.74E+00	8.03E−01	**1.58E−01**	5.61E−02	**9.99E−01**	**1.58E−01**	**5.60E−02**	**9.99E−01**	1/4/5/1/1
	SD	**1.03E−28**	**4.86E−25**	**5.06E−32**	6.39E−16	2.80E−14	5.52E−20	8.06E−02	5.68E−02	3.73E−03	2.91E−10	1.53E−08	2.01E−14	5.16E−11	2.83E−10	1.96E−15	
50 °C	Mean	**5.06E−02**	**2.20E−02**	**1.00E+00**	5.62E−02	2.20E−02	**1.00E+00**	2.19E+00	1.63E+00	8.01E−01	5.09E−02	2.23E−02	**1.00E+00**	6.00E−02	2.29E−02	**1.00E+00**	1/3/5/2/4
	SD	**1.13E−27**	**2.63E−25**	**9.34E−32**	1.42E−14	2.30E−14	1.30E−19	1.08E−01	6.85E−02	4.75E−03	2.28E−08	5.76E−08	2.10E−13	2.54E−09	8.41E−09	2.38E−14	
60 °C	Mean	**1.97E−02**	**1.04E−02**	**1.00E+00**	2.00E−02	1.42E−02	9.99E−01	2.19E+00	1.60E+00	7.97E−01	2.00E−02	1.07E−02	**1.00E+00**	**1.97E−02**	**1.04E−02**	9.99E−01	1/3/5/3/1
	SD	**3.24E−27**	**1.57E−25**	**1.75E−32**	1.72E−14	1.19E−14	2.62E−20	2.58E−02	1.96E−02	1.57E−03	4.06E−08	2.08E−08	6.39E−14	2.43E−09	1.75E−09	3.74E−15	

Table 15
The modeling results comparison of different modeling approaches for S75 under different temperatures.

Algorithm		bvRNA-GA-LSSVM			RNA-GA-LSSVM			CMA-ES-LSSVM			LSHADE-LSSVM			COA-LSSVM			Rank
Indicator		RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	
20 °C	Mean	**1.12E+00**	**2.90E−01**	9.67E−01	1.24E+00	8.95E−01	**9.68E−01**	3.71E+00	2.87E+00	4.07E−01	1.13E+00	2.92E−01	9.67E−01	**1.12E+00**	**2.90E−01**	**9.68E−01**	1/4/5/3/1
	SD	**7.69E−27**	**4.33E−25**	**6.45E−28**	1.73E−12	1.87E−12	6.23E−15	1.43E−02	2.10E−02	6.44E−03	5.18E−06	1.90E−06	1.87E−08	2.64E−06	1.48E−06	9.49E−09	
30 °C	Mean	**8.89E−02**	**4.11E−02**	**1.00E+00**	1.09E−01	4.61E−02	**1.00E+00**	3.17E+00	2.47E+00	5.82E−01	8.98E−02	4.12E−02	**1.00E+00**	8.91E−02	4.12E−02	**1.00E+00**	1/4/5/3/2
	SD	**5.08E−24**	7.61E−24	**1.09E−28**	1.67E−14	1.92E−14	3.53E−19	5.99E−02	5.91E−02	9.96E−03	3.67E−08	2.64E−08	7.81E−13	7.35E−08	6.12E−08	1.57E−12	
40 °C	Mean	**1.07E−01**	**3.61E−02**	**1.00E+00**	1.68E−01	3.61E−02	9.76E−01	2.76E+00	2.09E+00	6.97E−01	**1.07E−01**	3.65E−02	**1.00E+00**	**1.07E−01**	3.61E−02	1.00E+00	1/4/5/1/1
	SD	**7.95E−29**	**8.69E−25**	**1.62E−32**	3.92E−15	3.36E−14	1.36E−19	7.80E−02	6.33E−02	6.41E−03	1.32E−08	4.39E−08	4.37E−13	**2.03E−34**	4.40E−13	3.69E−14	
50 °C	Mean	**7.69E−02**	**2.40E−02**	**1.00E+00**	**7.69E−02**	2.40E−02	**1.00E+00**	2.59E+00	1.94E+00	7.24E−01	7.71E−02	2.44E−02	**1.00E+00**	**7.69E−02**	2.41E−02	**1.00E+00**	1/1/5/4/1
	SD	**1.30E−27**	**2.22E−24**	**4.74E−32**	1.27E−14	7.42E−14	2.54E−19	9.29E−02	5.93E−02	6.03E−03	1.80E−08	8.26E−08	3.58E−13	2.25E−08	1.54E−08	4.24E−13	
60 °C	Mean	**1.12E−02**	**6.47E−03**	**1.00E+00**	1.64E−02	6.76E−03	9.76E−01	2.28E+00	1.72E+00	7.74E−01	1.16E−02	6.72E−03	**1.00E+00**	1.13E−02	6.59E−03	**1.00E+00**	1/4/5/3/2
	SD	**3.41E−26**	**1.81E−24**	**2.21E−32**	2.49E−14	6.82E−15	1.07E−20	1.92E−01	1.24E−01	1.11E−02	6.40E−08	2.58E−08	2.95E−14	1.68E−07	1.12E−07	8.07E−14	

of bvRNA-GA-LSSVM agrees well with the real out of ST40 in all studied working temperatures.

The results in Table 14 illustrate that bvRNA-GA-LSSVM wins other modeling algorithms in both accuracy and stability for modeling SM55 under different working temperatures. LSHADE-LSSVM gets the same precision with bvRNA-GA-LSSVM when the temperature is 30 °C, 40 °C. COA-LSSVM obtains the same performance as bvRNA-GA-LSSVM on 40 °C and 60 °C. Fig. 15 indicates that bvRNA-GA-LSSVM model can better reflect the actual characteristic of SM55 under various outside temperature conditions.

From Table 15, the RMSE, MAE values indicate the good modeling accuracy and stability of bvRNA-GA-LSSVM on modeling S75. The *I-V* curves comparison between the real output and bvRNA-GA-LSSVM model output Fig. 16 also demonstrates the validity of bvRNA-GA-LSSVM modeling method.

According to the description above, bvRNA-GA-LSSVM can provide reliable modeling results for ST40, SM55, S75 modules under various working temperatures.

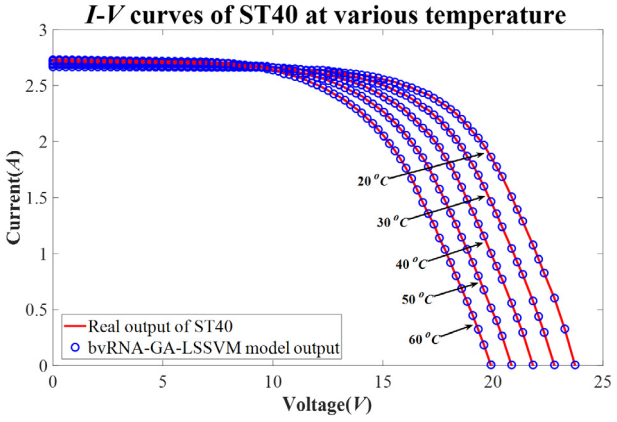


Fig. 14. *I-V* curves comparison between the real output and bvRNA-GA-LSSVM model output under different temperatures for ST40 module.

5.6. Convergence and performance analysis

To intuitively show the performance of bvRNA-GA, RNA-GA, L-SHADE, COA for optimizing the parameters of LSSVM, the computational time are collected in Tables 16 and 17, and the total average convergence curves (TTACs) are shown in Fig. 17 (in the training process). It worth mentioned here, the working environment (software and hardware) is the same as in Section 4.1. The computation time is adopted to evaluate how long does the optimization process take to run, the time in Tables 16 and 17 are the average time that bvRNA-GA, RNA-GA, CMA-ES, L-SHADE, COA spend to getting the optimal hyper-parameters for LSSVM (the cost CPU time by step 4.1 to step 4.4 in Section 5.2). Because CMA-ES is not a population-based algorithm, we cannot plot the convergence curves in Fig. 17. Otherwise, the TTAC means that we directly add up the evolution curves in different conditions (temperature or irradiation) of the same PV system. The TTAC can effectively reflect the convergence speed and the modeling accuracy of the modeling methods for each PV system.

From Table 16, CMA-ES-LSSVM and LSHADE-LSSVM are faster than the other three algorithms for modeling SDM and DDM. bvRNA-GA-LSSVM, RNA-GA-LSSVM take extra time to perform decoding operations and special searching operations, therefore, the running time is larger. However, comparing with RNA-GA-LSSVM on modeling SDM, DDM, and COA-LSSVM on modeling DDM, bvRNA-GA-LSSVM costs less time than them. In terms of Table 17, the efficiency of bvRNA-GA is increasing to some degrees, and the consumption of time is closer with CMA-ES-LSSVM and LSHADE-LSSVM.

According to Fig. 17, all the referred modeling approaches can converge to similar accuracy, but bvRNA-GA obtained better parameters for LSSVM at the early stage. We can also conclude that bvRNA-GA-LSSVM can obtain good modeling convergence for every working condition of each PV system because the TTACs of bvRNA-GA-LSSVM converge significantly faster than another modeling method. The results in Fig. 17 also indicate that the bvRNA-GA can obtain better parameters for LSSVM even if one reduces the iteration times.

Table 16
Average computational time (s) over by different algorithms for modeling SDM, DDM at various temperatures.

Model	SDM					DDM				
Temperature (°C)	25	33	50	75	100	25	33	50	75	100
bvRNA-GA-LSSVM	115.23	113.53	113.69	114.60	114.08	111.63	111.33	110.52	110.18	109.93
RNA-GA-LSSVM	166.45	163.50	164.23	165.73	169.53	167.83	168.15	166.31	165.92	168.01
CMA-ES-LSSVM	81.54	82.27	85.77	85.58	85.27	86.52	89.14	83.82	84.15	87.42
LSHADE-LSSVM	87.26	82.53	82.16	81.72	81.44	82.49	82.49	82.54	81.81	81.71
COA-LSSVM	92.35	90.83	91.08	95.23	120.36	121.97	122.81	122.31	122.66	132.75

Table 17
Average computational time (s) by different algorithms for modeling ST40, SM55, S75 modules at various temperatures and irradiations.

Irradiation (W/m^2)	Temperature (°C)	1000	800	600	400	200	20	30	40	50	60
bvRNA-GA-LSSVM		78.92	80.63	75.33	73.52	69.53	82.77	77.24	73.37	73.44	64.69
RNA-GA-LSSVM		112.02	110.34	112.89	107.94	100.33	116.78	117.47	111.30	99.50	98.78
CMA-ES-LSSVM	ST40	67.53	55.79	69.84	53.34	60.86	88.34	55.37	56.51	43.91	66.63
LSHADE-LSSVM		56.22	76.97	62.75	44.85	41.40	70.10	60.75	46.12	58.79	48.66
COA-LSSVM		108.51	111.52	77.19	82.32	79.68	99.14	76.16	93.28	62.00	96.36
bvRNA-GA-LSSVM		80.36	85.53	82.83	81.14	72.85	90.79	82.86	77.76	75.69	75.63
RNA-GA-LSSVM		121.06	118.05	126.20	121.22	113.03	126.98	128.70	113.65	118.29	111.94
CMA-ES-LSSVM	SM55	77.30	86.74	81.34	66.46	63.36	63.04	82.91	79.57	75.74	69.05
LSHADE-LSSVM		62.91	84.22	52.89	57.82	69.88	69.08	76.16	53.53	58.74	49.85
COA-LSSVM		137.68	126.35	98.68	82.77	80.82	84.44	95.55	76.27	81.80	77.56
bvRNA-GA-LSSVM		74.08	73.77	72.21	69.94	68.26	97.67	91.02	89.29	86.50	80.56
RNA-GA-LSSVM		100.37	101.88	101.66	96.94	97.43	149.74	137.74	131.58	119.71	115.68
CMA-ES-LSSVM	S75	72.01	64.14	48.77	68.97	50.66	98.12	67.09	59.46	76.75	77.78
LSHADE-LSSVM		58.51	47.28	46.27	50.86	44.82	77.17	60.33	74.52	63.99	68.85
COA-LSSVM		73.99	79.51	87.46	93.33	75.33	102.55	130.24	89.75	130.06	88.22

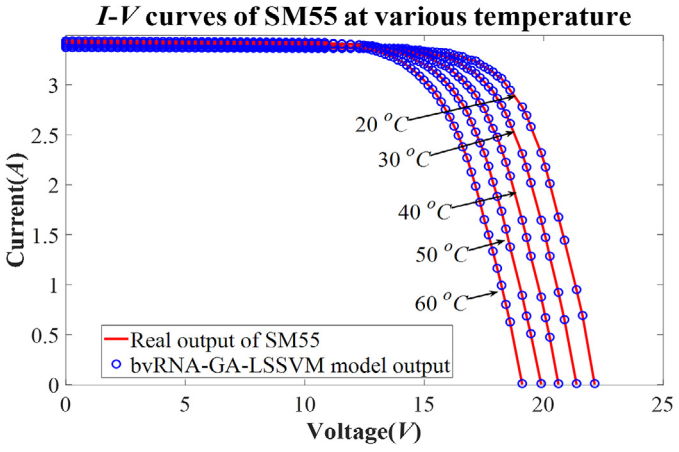


Fig. 15. I-V curves comparison between the real output and bvRNA-GA-LSSVM model output under different temperatures for SM55 module.

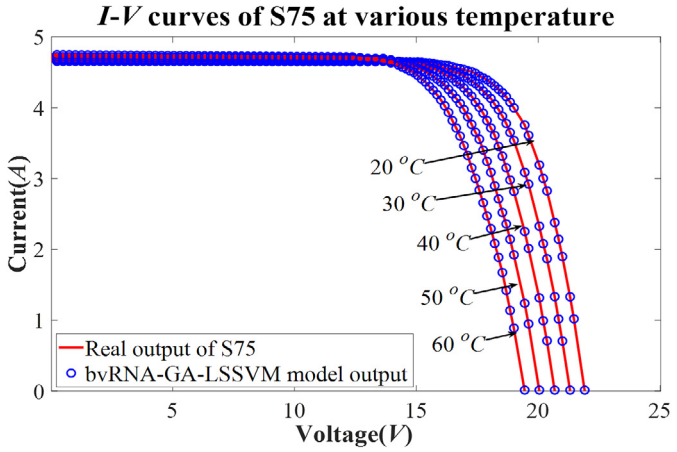


Fig. 16. I-V curves comparison between the real output and bvRNA-GA-LSSVM model output under different temperatures for S75 module.

The obtained modeling results and the discussion in the above sections demonstrate that the bvRNA-GA can better determine the LSSVM's parameters. The bvRNA-GA-LSSVM can modeling different PV systems with high accuracy.

6. Conclusion

In this work, LSSVM modeling method is utilized to model the characteristic curves for PV systems. To determine the values of kernel parameter and relative weight for LSSVM, a novel RNA-GA (bvRNA-GA) is proposed. In bvRNA-GA, the bulge loop crossover operator, the inner loop crossover operator, and the virus-induced mutation operator are designed to balance the exploration and exploitation capacities. To verify the searching performance of the bvRNA-GA, four state-of-art algorithms (RNA-GA, CMA-ES, L-SHADE, COA) are used to make the comparison. Different experiments with 10 benchmark functions have been conducted to compare the five mentioned optimization algorithms. According to the results, bvRNA-GA has the advantage in solving low dimensional problems and remains good convergence and global searching capacity when increasing the dimensions of the problems. In the modeling part, the results of bvRNA-GA optimized LSSVM obtain high-quality characteristic models for SDM, DDM, three different kinds of PV solar modules under various working conditions. The modeling results denote the effectiveness of the proposed modeling method in modeling complex non-linear systems. In the future, we will focus on reducing the influence of low searching efficiency caused by RNA encoding. Furthermore, we will attempt to develop multi-object bvRNA-GA to solve more real problems.

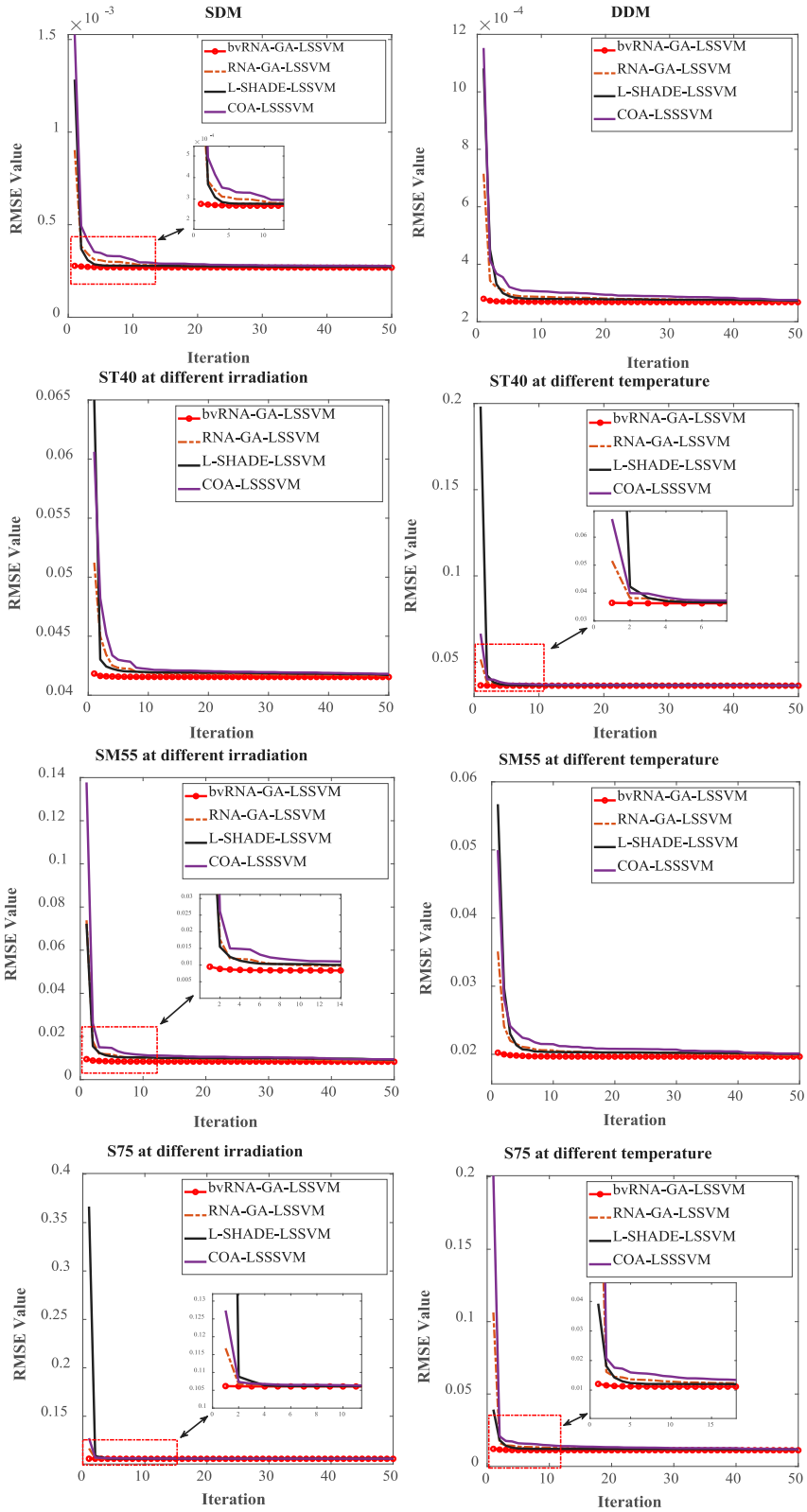


Fig. 17. Total average convergence curves of bvRNA-GA-LSSVM, RNA-GA-LSSVM, L-SHADE-LSSVM, COA-LSSVM for different PV systems.

CRediT authorship contribution statement

Xiu Liu: Software, Writing – review & editing, Investigation, Data curation, Writing – original draft. **Ning Wang:** Conceptualization, Methodology, Revising, Writing – original draft, Review and correction, Supervision. **Daniel Molina:** Revising, Supervision, Editing. **Francisco Herrera:** Revising, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported in part by National Natural Science Foundation (NNSF) of China under Grant 61573311, and it is also supported in part by the Spanish Ministry of Science grant PID2020-119478GB-I00.

The authors would like to thank Senior engineer Mr. Guoping Wang and Ms. Guangrong Yi for their help.

References

[1] B. Zhao, M. Hu, X. Ao, Q. Xuan, G. Pei, Spectrally selective approaches for passive cooling of solar cells: A review, Appl. Energy 262 (2020) 114548.
[2] J. Liu, X. Chen, S. Cao, H. Yang, Overview on hybrid solar photovoltaic-electrical energy storage technologies for power supply to buildings, Energy Conv. Manag. 187 (2019) 103–121.
[3] M. Abd Elaziz, D. Oliva, Parameter estimation of solar cells diode models by an improved opposition-based whale optimization algorithm, Energy Conv. Manag. 171 (2018) 1843–1859.
[4] B.S.S. Ganesh Pardhu, V. Reddy Kota, Radial movement optimization based parameter extraction of double diode model of solar photovoltaic cell, Sol. Energy 213 (2021) 312–327.
[5] T.O. Ting, J. Ma, K.S. Kim, K. Huang, Multicores and GPU utilization in parallel swarm algorithm for parameter estimation of photovoltaic cell model, Appl. Soft. Comput. 40 (2016) 58–63.
[6] G. Kanimozhi, H. Kumar, Modeling of solar cell under different conditions by ant lion optimizer with lambertw function, Appl. Soft. Comput. 71 (2018) 141–151.
[7] B. Lekouaghet, A. Boukabou, C. Boubakir, Estimation of the photovoltaic cells/modules parameters using an improved rao-based chaotic optimization technique, Energy Conv. Manag. 229 (2021) 113722.
[8] T.V. Luu, N.S. Nguyen, Parameters extraction of solar cells using modified JAYA algorithm, Optik 203 (2020) 164034.
[9] A.K. Abdulrazzaq, G. Bognár, B. Plesz, Evaluation of different methods for solar cells/modules parameters extraction, Sol. Energy 196 (2020) 183–195.
[10] D. Kler, Y. Goswami, K.P.S. Rana, V. Kumar, A novel approach to parameter estimation of photovoltaic systems using hybridized optimizer, Energy Conv. Manag. 187 (2019) 486–511.
[11] H.G.G. Nunes, J.A.N. Pombo, P.M.R. Bento, S.J.P.S. Mariano, M.R.A. Calado, Collaborative swarm intelligence to estimate PV parameters, Energy Conv. Manag. 185 (2019) 866–890.
[12] S. Li, W. Gong, X. Yan, C. Hu, D. Bai, L. Wang, L. Gao, Parameter extraction of photovoltaic models using an improved teaching-learning-based optimization, Energy Conv. Manag. 186 (2019) 293–305.
[13] X. Yang, W. Gong, Opposition-based JAYA with population reduction for parameter estimation of photovoltaic solar cells and modules, Appl. Soft. Comput. 104 (2021) 107218.
[14] V. Vapnik, The Nature of Statistic Learning Theory, New York, 1995.
[15] V. Vapnik, Statistical Learning Theory, New York, 1998.
[16] G. Acampora, F. Herrera, G. Tortora, A. Vitiello, A multi-objective evolutionary approach to training set selection for support vector machine, Knowledge-Based Syst. 147 (2018) 94–108.
[17] A. Rosales-Pérez, S. García, J.A. Gonzalez, C.A. Coello Coello, F. Herrera, An evolutionary multi objective model and instance selection for support vector machines with Pareto-based ensembles, IEEE Trans. Evol. Comput. 21 (6) (2017) 863–877.
[18] J.A.K. Suykens, J. Vandewalle, Least squares support vector machine classifiers, Neural Process. Lett. 9 (3) (1999) 293–300.
[19] X. Yang, L. Tan, L. He, A robust least squares support vector machine for regression and classification with noise, Neurocomputing 140 (2014) 41–52.
[20] A. Yang, W. Li, Short-term electricity load forecasting based on feature selection and Least Squares Support Vector Machines, Knowledge-Based Syst. 163 (2019) 159–173.
[21] A. Rostami, M. Arabloo, M. Lee, A. Bahadori, Applying SVM framework for modeling of CO2 solubility in oil during CO2 flooding, Fuel 214 (2018) 73–87.
[22] A. Zendehboudi, Implementation of GA-LSSVM modelling approach for estimating the performance of solid desiccant wheels, Energy Conv. Manag. 127 (2016) 245–255.
[23] A. Bemani, Q. Xiong, A. Baghban, S. Habibzadeh, A.H. Mohammadi, M.H. Doranehgard, Modeling of cetane number of biodiesel from fatty acid methyl ester (FAME) information using GA-, PSO-, and HGAPSO- LSSVM models, Renew. Energy 150 (2020) 924–934.
[24] X. Gao, X. Liu, A novel effective diagnosis model based on optimized least squares support machine for gene microarray, Appl. Soft. Comput. 66 (2018) 50–59.
[25] Z. Tian, Short-term wind speed prediction based on LMD and improved FA optimized combined kernel function LSSVM, Eng. Appl. Artif. Intell. 91 (2020) 103573.
[26] L. Wen, Y. Cao, A hybrid intelligent predicting model for exploring household CO2 emissions mitigation strategies derived from butterfly optimization algorithm, Sci. Total Environ. 727 (2020) 138572.
[27] J. Wu, S. Huang, Z. Kou, Research and optimization of intelligent diagnosis algorithm based on rope tension, Measurement 147 (2019) 106892.
[28] Y. Wu, W. Xue, L. Xu, X. Guo, D. Xue, Y. Yao, S. Zhao, N. Li, Optimized least-squares support vector machine for predicting aero-optic imaging deviation based on chaotic particle swarm optimization, Optik 206 (2020) 163215.
[29] X. Liu, Y. Gu, S. He, Z. Xu, Z. Zhang, A robust reliability prediction method using weighted least square support vector machine equipped with chaos modified particle swarm optimization and online correcting strategy, Appl. Soft. Comput. 85 (2019) 105873.
[30] J.H. Holland, Adaptation in Natural and Artificial Systems, MIT Press, 1975.
[31] X. Pei, Y. Zhou, N. Wang, A Gaussian process regression based on variable parameters fuzzy dominance genetic algorithm for B-TFPMM torque estimation, Neurocomputing 335 (2019) 153–169.
[32] Y. Villacampa, F.J. Navarro-González, P. Compañ Rosique, R. Satorre-Cuerda, A guided genetic algorithm for diagonalization of symmetric and hermitian matrices, Appl. Soft. Comput. 75 (2019) 180–189.
[33] J.M. Adanez, B.M. AI-Hadithi, A. Jimenez, Multidimensional membership functions in T–s fuzzy models for modelling and identification of nonlinear multivariable systems using genetic algorithms, Appl. Soft. Comput. 75 (2019) 607–615.
[34] T. Vuolio, V.V. Visuri, A. Sorsa, S. Ollila, T. Fabritius, Application of a genetic algorithm based model selection algorithm for identification of carbide-based hot metal desulfurization, Appl. Soft. Comput. 92 (2020) 106330.
[35] Z.I. Erzurum Cicek, Zehra Kamisli Ozturk, Optimizing the artificial neural network parameters using a biased random key genetic algorithm for time series forecasting, Appl. Soft. Comput. 102 (2021) 107091.
[36] J.D. Hemanth, J. Anitha, B.K. Ane, Fusion of artificial neural networks for learning capability enhancement: Application to medical image classification, Expert Syst. 34 (6) (2017) 1–20.
[37] L.M. Adleman, Molecular computation of solutions to combination problems, Science 266 (5187) (1994) 1021–1024.
[38] J.L. Tao, N. Wang, DNA computing based RNA genetic algorithm with applications in parameter estimation of chemical engineering processes, Comput. Chem. Eng. 31 (12) (2007) 1602–1618.
[39] L. Zhang, N. Wang, An adaptive RNA genetic algorithm for modeling of proton exchange membrane fuel cells, Int. J. Hydrog. Energy 38 (1) (2013) 219–228.
[40] Q.Q. Zhu, N. Wang, L. Zhang, Circular genetic operators based RNA genetic algorithm for modeling proton exchange membrane fuel cells, Int. J. Hydrog. Energy 39 (31) (2014) 17779–17790.
[41] K.T. Wang, N. Wang, A protein inspired RNA genetic algorithm for parameter estimation in hydrocracking of heavy oil, Chem. Eng. J. 167 (1) (2011) 228–239.
[42] N. Wang, D. Wang, Y. Xing, L. Shao, S. Afzal, Application of co-evolution RNA genetic algorithm for obtaining optimal parameters of SOFC model, Renew. Energy 150 (2020) 221–233.
[43] X.X. Zhu, N. Wang, Hairpin RNA genetic algorithm based ANFIS for modeling overhead cranes, Mech. Syst. Signal Proc. 165 (2022) 108326.
[44] L. Zhang, N. Wang, A modified DNA genetic algorithm for parameter estimation of the 2-chlorophenol oxidation in supercritical water, Appl. Math. Model. 37 (3) (2013) 1137–1146.
[45] X. Liu, N. Wang, A novel gray wolf optimizer with RNA crossover operation for tackling the non-parametric modeling problem of FCC process, Knowl. Based Syst. 216 (2021) 106751.
[46] X.X. Zhu, N. Wang, Splicing process inspired cuckoo search algorithm based ENNs for modeling FCCU reactor-regenerator system, Chem. Eng. J. 354 (2018) 1018–1031.

[47] J.C.H. Phang, D.S.H. Chan, J.R. Phillips, Accurate analytical method for the extraction of solar cell model parameters, Electron. Lett. 20 (1984) 406–408.
[48] D.S.H. Chan, J.C.H. Phang, Analytical methods for the extraction of solar-cell single-and double-diode model parameters from I-V characteristics, IEEE Trans. Electron Devices 34 (2) (1987) 286–293.
[49] Shell S75 Photovoltaic Solar Module. <http://www.aeet-service.com/pdf/shell/Shell-Solar_S75.pdf>.
[50] Shell ST40 Photovoltaic SolarModule. <http://www.aeet-service.com/pdf/shell/Shell-Solar_ST40.pdf>.
[51] Shell SM55 Photovoltaic Solar Module. <http://www.aeet-service.com/pdf/shell/Shell-Solar_SM55.pdf>.
[52] M.P. Robertson, G.F. Joyce, Highly efficient self-replicating RNA enzymes, Chem. Biol. 21 (2) (2014) 238–245.
[53] N. Kubota, K. Shimojima, T. Fukuda, Virus-evolutionary genetic algorithm-ecological model on planar grid, in: Proceedings of North American Fuzzy Information Processing, Berkeley, CA, USA, 1996, pp. 505-509.
[54] N. Hansen, S.D. Müller, P. Koumoutsakos, Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation (CMA-ES), Evol. Comput. 11 (1) (2003) 1–18.
[55] R. Tanabe, A.S. Fukunaga, Improving the search performance of SHADE using linear population size reduction, in: IEEE Congress on Evolutionary Computation (CEC), Beijing, China, 2014, pp. 1658-1665.
[56] J. Pierezan, L.D. Santos Coelho, Coyote Optimization Algorithm: A new metaheuristic for global optimization problems, IEEE Congress on Evolutionary Computation (CEC), Rio de Janeiro, Brazil, 2018, pp. 1-8.
[57] J. Liang, B. Qu, P.N. Suganthan, Problem definitions and evaluation criteria for the CEC 2014 special session and sompetition on single objective real-parameter numerical optimization, in: Computational Intelligence Laboratory and Nanyang Technological University, Tech. Rep., China and Singapore, 2013-2014.
[58] J. Derrac, S. Garcia, S. Hui, P.N. Suganthan, F. Herrera, Analyzing convergence performance of evolutionary algorithms: A statistical approach, Inform. Sci. 289 (2014) 41–58.
[59] J. Carrasco, S. García, M.M. Rueda, S. Das, F. Herrera, Recent trends in the use of statistical tests for comparing swarm and evolutionary computing algorithms: Practical guidelines and a critical review, Swarm Evol. Comput. 54 (2020) 100665.
[60] J. Perolat, I. Couso, K. Loquin, O. Strauss, Generalizing the wilcoxon rank-sum test for interval data, Int. J. Approx. Reason. 56 (Part A) (2015) 108–121.
[61] E.L. Hedrea, R.E. Precup, C.A. Bojan-Dragos, Results on tensor product-based model transformation of magnetic levitation systems, Acta Polytech. Hung. 16 (9) (2019) 93–111.
[62] Y. Zhai, X. Ding, X. Jin, L. Zhao, Adaptive LSSVM based iterative prediction method for NOx concentration prediction in coal-fired power plant considering system delay, Appl. Soft. Comput. 89 (2020) 106070.