

## Yield Stress Prediction Model of RAFM Steel Based on the Improved GDM-SA-SVR Algorithm

Sifan Long<sup>1</sup>, Ming Zhao<sup>2,\*</sup> and Xinfu He<sup>3</sup>

**Abstract:** With the development of society and the exhaustion of fossil energy, researcher need to identify new alternative energy sources. Nuclear energy is a very good choice, but the key to the successful application of nuclear technology is determined primarily by the behavior of nuclear materials in reactors. Therefore, we studied the radiation performance of the fusion material reduced activation ferritic/martensitic (RAFM) steel. The main novelty of this paper are the statistical analysis of RAFM steel data sets through related statistical analysis and the formula derivation of the gradient descent method (GDM) which combines the gradient descent search strategy of the Convex Optimization Theory to get the best value. Use GDM algorithm to upgrade the annealing stabilization process of simulated annealing algorithm. The yield stress performance of RAFM steel is successfully predicted by the hybrid model which is combined by simulated annealing (SA) with support vector machine (SVM) as the first time. The effect on yield stress by the main physical quantities such as irradiation temperature, irradiation dose and test temperature is also analyzed. The related prediction process is: first, we used the improved annealing algorithm to optimize the SVR model after training the SVR model on a training data set. Next, we established the yield stress prediction model of RAFM steel. The model can predict up to 96% of the data points with the prediction in the test set and the original data point in the  $2\sigma$  range. The statistical test analysis shows that under the condition of confidence level  $\alpha=0.01$ , the calculation results of the regression effect significance analysis pass the  $T$ -test.

**Keywords:** Convex optimization theory, simulated annealing algorithm, reduced activation ferritic/martensitic steel, support vector regression.

### 1 Introduction

Reduced activation ferritic/martensitic (RAFM) has excellent thermal, physical and mechanical properties, such as high irradiation swelling and thermal expansion coefficient and high thermal conductivity. RAFM steel is often chosen as the preferred material for future fusion power stations and experimental reactors [Kemp, Cottrell and Bhadeshia (2006)].

<sup>1</sup> School of Software, Central South University, Changsha, 410075, China.

<sup>2</sup> School of Software, Central South University, Changsha, 410075, China.

<sup>3</sup> Laboratory of Irradiation Effect, China Institute of Atomic Energy, Beijing, 102413, China.

\* Corresponding Author: Ming Zhao. Email: meanzhao@csu.edu.cn.

In application, the yield stress is a highly important performance parameter of the material. Under certain conditions, the yield stress is a nonlinear function of the deformation velocity, deformation temperature and degree of deformation [Barnes and Walters (1985)]. There are many related variables and many uncontrollable factors in the measurement process. Therefore, it is difficult to describe the performance of RAFM steel in engineering applications. In this paper, the yield stress prediction model of RAFM steel is established by the support vector regression (SVR) model in the field of machine learning [Joachims (1999)]. The effects and joint effects of different parameters on the yield stress can be analyzed, obtaining further analysis of the correlation prediction curve change before and after irradiation. Prediction in the range of data based on the training set can be performed. This article uses SVR to establish the prediction model of yield stress for RAFM steel. SVM has been applied successfully in the field of text classification since it was put forward by Vapnik [Cortes and Vapnik (1995)]. Through the training of the efficient sequential minimal optimization (SMO) algorithm which was designed by Platt [Platt and John (1998)], SVM has been successfully promoted in industry. With the excellent performance, whether it is to address regression or classification tasks, SVM is always one of the best algorithms compared to other similar algorithms [Gain and Roy (2016); Jian, Shen and Li (2017)]. At present, the main research direction for the algorithm is the application combined with other algorithms and improvement [Keerthi, Shevade and Bhattacharyya (2014)]. For example, Ghamisi used it for feature selection [Ghamisi, Couceiro and Benediktsson (2015)], Ananthi and others have used it for voice recognition [Ananthi and Dhanalakshmi (2015)], and Darmatasia combined it with the CNN algorithm for handwriting recognition [Darmatasia, Fanany and Ivan (2017)]. Although the SVM algorithm shows excellent performance, there are still several problems. For example, the selection of the kernel function is a serious and outstanding problem. To date, there are many relevant studies. For example, Peng et al. used multicore SVM emotion recognition [Peng, Hu and Dang (2017)]. Tan et al. used SVM for hyperspectral image classification [Kun and Peijun (2010)]. The SVM algorithm has been used for unmanned aerial vehicle (UAV) fault diagnosis [Ye, Luo and Li (2014)]. For studies on the application of fusion material RAFM steel, Kemp and others used a neural network to study the performance of RAFM steel [Kemp, Cottrell and Bhadeshia (2006)]. Neelamegam et al. studied the optimization parameters of RAFM steel based on a hybrid intelligent model to obtain the desired weld bead shape parameters and heat affected zone (HAZ) width [Neelamegam, Sapineni and Muthukumaran (2013)]. These researchers combined the genetic algorithm to optimize parameters of the RAFM steel material on welding. In recent years, the research tends to be purely physical like the neutron irradiation study carried out by Gaganidze et al. [Gaganidze and Aktaa (2013)]. Babu et al. performed a fatigue crack propagation study [Babu, Mukhopadhyay and Sasikala (2016)]. Compared with this study, previous studies were too monotonous to establish a systematic model to describe the influence mechanism of related variables on yield stress. Although a number of traditional hybrid algorithms, combined heuristic algorithm with machine learning, like SVM, artificial neural network (ANN), naive bayes (NB) and other hybrid model are applied to other research fields.

For example, the hybrid model composed of genetic algorithm (GA) and SA algorithm combined with ANN is used to predict the groundwater level [Bahrami, Ardejani and Baafi (2016)]. DJ Armaghani use particle swarm optimization combined with neural network to predict the ultimate bearing capacity of rock-socketed pile [Zhang and Zhou (2016)]. Faced with the same task, there are also significant differences between different heuristic algorithms. For example, Jia F and others have compared the genetic algorithm and particle swarm optimization algorithm in the first-order design of TLS network [Jia and Lichti (2017)]. The global optimization algorithm has stronger search ability in the solution space but it cannot search accurately in a small range. Therefore, the heuristic algorithm combined with the fine search algorithm can effectively avoid the shortcomings of both and can effectively play their advantages.

In the prediction of RAFM steel, no one has used the hybrid model composed of simulated annealing algorithm and support vector machine to analyze and study the yield stress of RAFM steel before. This research has been proved by a lot of experiments, such as outlier detection, clustering analysis and feature selection. This study establishes the GDM-SA-SVR model based on the improved annealing algorithm. Within the  $2\sigma$  error range of the test set, the prediction under the conditions of a given test temperature and radiation dose can achieve a prediction accuracy of 96%. The  $T$ -test showed that there is 99% confidence that the regression effect is significant.

This article is composed by 5 sections. Section 2 makes some necessary analysis on the data set, this section explains that the characteristics of RAFM steel data set distribution a relatively complex distribution. Section 3 is to improve the simulated annealing algorithm and combine with SVR to form a hybrid prediction model for the yield stress of RAFM steel. Section 4 is to compare the model proposed in this research and other similar models on the test set, which is to illustrate the effectiveness of the hybrid model proposed in this paper. Section 5 used the hybrid model proposed in this paper to analyze the effect of the main parameters of RAFM steel on the yield stress, and finally gives the prediction results.

## **2 Dataset analysis**

The dataset is a part of the result in a radiation experiment. In the field of data mining, the quality of the data directly affects the result of the experiment. Therefore, in applications, data processing often takes up 80% of the work of the data analysts [Witten and Frank (2011)]. Using statistical methods and data visualization technology to analyze the original data set represents a valuable feature trend that usually is a necessary indicator of a successful selection for a suitable model.

### **2.1 RAFM steel dataset**

The data used in this paper are data points on RAFM steel irradiation experiments when the radiation dose is in the range of 0~90 (dpa). The data used in this paper consist of a total of 1811 experimental data points and 37 associated characteristic attribute parameters. In other words, this paper addresses a multielement nonlinear function with 37 elements, and

the ordinary linear model cannot reach the actual requirements. Therefore, it is necessary to find the appropriate model. The attribute distribution of the data is shown in the Appendix A [Kemp, Cottrell and Bhadeshia (2006)].

## 2.2 Symmetry, correlation analysis and outlier detection

Tab.6 in Appendix A shows that the distribution of data samples in a single dimension is asymmetrical and is accompanied by many outliers. That situation is not conducive to the establishment of an accurate model. The predictability between the data sample and the result label cannot be predicted. Therefore, a more comprehensive analysis is needed from the microscopic perspective. To determine the predictability of data samples and whether there is multicollinearity between attributes, it is necessary to use the statistical Pearson correlation coefficient [Taylor (1990)] (simple correlation coefficient) to measure the intimate degree among attributes. The calculation formula is as follows:

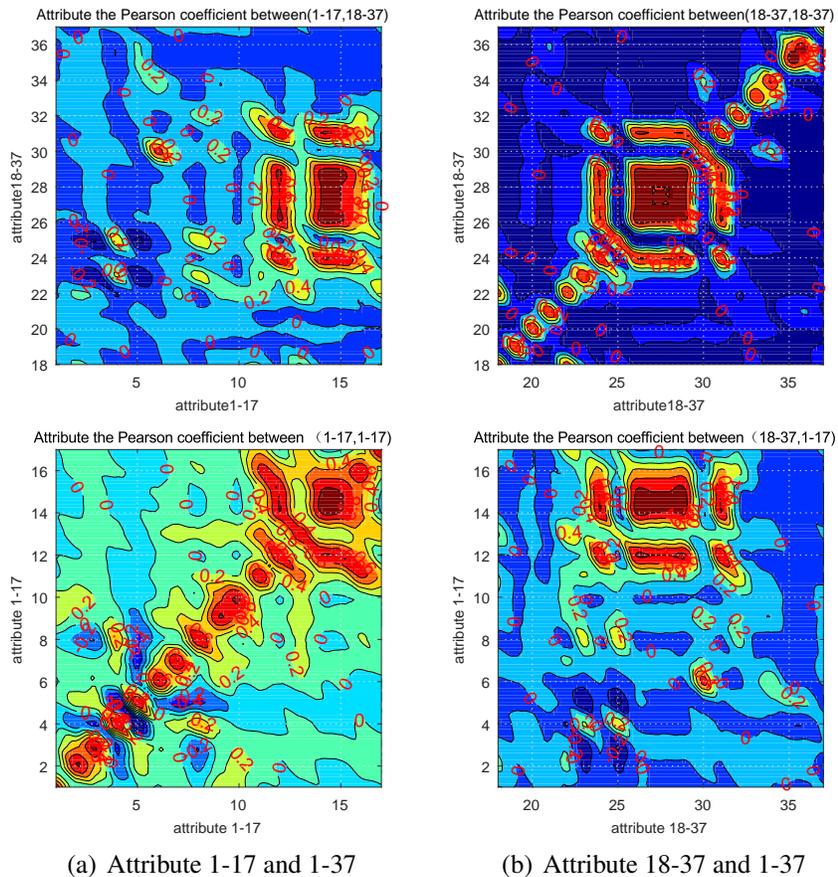
$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad (1)$$

The Pearson correlation coefficient can simply judge whether there is a linear relationship between two attributes, that is, whether the attributes are collinear, where X and Y are two one-dimensional data sets,  $\bar{X}$  is the mean of X, and  $\bar{Y}$  is the mean of Y. Because the Eq. (1) is symmetric, the Pearson coefficient has no effect on the order of computation, that is,  $r_{XY} = r_{YX}$ . Relevant criteria for evaluation are usually used to determine the degree of correlation [Nicewander (1988)]. Given the significant level of  $\alpha$  (generally  $\alpha=0.05$ ) and the degree of freedom (n-2), where  $\alpha$  is a confidence level, for example, when  $\alpha=0.01$ , we can get 99% probability to make decision inference. Then, we check the R distribution table for  $R\alpha$ , if  $|r| > R\alpha$ , the correlation coefficient is significant. To facilitate the discrimination, ensure that values of the  $|r|$  can meet the following relationship according to the simplified criteria.

**Table 1:** Criteria for determining the correlation coefficient

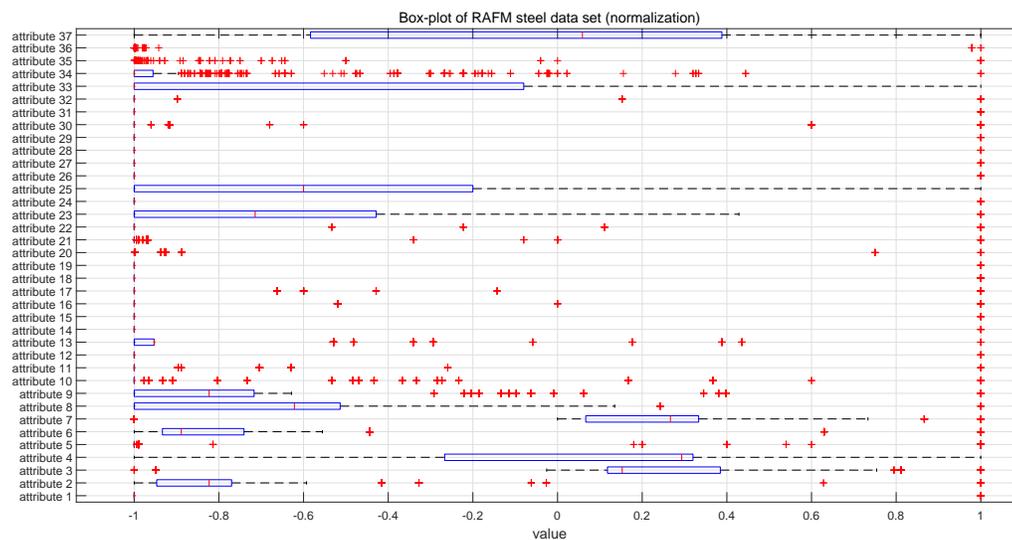
Index	$ r $	Degree of correlation
a	0.8-1	Extremely strong correlation
b	0.6-0.8	Strong correlation
c	0.4-0.6	Moderate degree correlation
d	0.2-0.4	Weak correlation
e	0-0.2	Extremely weak correlation or unrelated

In regression analysis, we need to understand the pure effect of each independent variable on the dependent variable. Multilinear means that there is a certain functional relationship between the independent variables. If there is a functional relationship between the two independent variables ( $x_1$  and  $x_2$ ), then  $x_2$  will change accordingly when  $x_1$  changes a



**Figure 1:** Correlation between attribute variables. The Pearson coefficient is calculated to detect the collinear relationship between attributes to infer the impact of the final model. Because of the properties between the Pearson coefficient and sequence independence, the image map is symmetrical

unit. At this time, you can not fix the other variables. Conditions, if we examine the effect of  $x_1$  on dependent variable  $y$  alone, the effect of  $x_1$  you observe is always mixed up with the effect of  $x_2$ , which leads to analysis errors, making the analysis of the effects of independent variables inaccurate, so we need to exclude the effect of multicollinearity in regression analysis. Tab. 1 shows that some attributes do have multiple collinear relationships, the relation coefficient between attributes is constructed into a matrix, similar to the adjacency matrix used to describe the adjacent relations of edges in graph theory. The  $(i, j)$  th element of the matrix represents the correlation coefficient between the  $i$  th attribute and the  $j$  th attribute, and then draws these matrix elements onto the thermograph. The diagonal correlation is the highest, and the overall image is axisymmetric because the diagonal is more self-related to the attribute. Secondly, the Pearson correlation coefficient is not in the order of calculation. In other words,  $R_{(i,j)}$  is equal to  $R_{(j,i)}$ . the degree of relevance is divided into 5 levels. When the correlation coefficient is above 0.8, we must process the data set, such as removing duplicate data, normalizing data and data cleaning. When the correlation coefficient is low, such as less than 0.4, we can not deal with it. Fig. 1 shows that there is no strong correlation between attributes except attribute 15 and attribute 28, which ensures that our model can be influenced by the collinearity of attributes.



**Figure 2:** The box-plot of 37 attribute values of RAFM steel shows that most of the attribute distributions are asymmetrical. In addition, there are many outliers

The distribution of data can affect the training process of the prediction model. Compared with the skewed distribution, the symmetry of the data contains a lot of information, and the distribution of information is more balanced, which will ease the learning task of the prediction model. In order to comprehensively analyze the data sets of RAFM steel, we use the box-plot in statistics to analyze it. It contains six data nodes, from which a set

of data is arranged from large to small, respectively, to calculate the upper edge of the data, the upper four digits Q3, the median, the lower four digits Q1, the lower edge, and an exception value(outliers). The specific content of the box diagram can be referred to the work of Robert McGill et al. [Mcgill, Tukey and Larsen (1978)]. Fig. 2 shows the data symmetry and outliers by using box diagrams to display the original data. The graph shows that the set of attributes presents asymmetrical distribution, and there are many outliers ('+' representing outliers), which has a great challenge to the prediction results. Therefore, we must be able to find a model with strong anti noise ability, strong robustness and high generalization performance. It is difficult to use the traditional method, which will affect the final stability of the model.

### **3 SVR yield stress model of RAFM steel**

Among a large number of heuristic search algorithms, different algorithms have different application scenarios. For example, the traditional genetic algorithm process is too complex for prediction of yield stress of RAFM steel [Zhang and Zhou (2016)], particle swarm optimization (PSO) algorithm is not good for solving discrete processing problem, it is easy to fall into the local optimum [Du (2016); Zhu and Cai (2016)]. Simulated annealing algorithm has many advantages, such as fast convergence speed and efficient algorithm. So, it is often applied to various optimization methods [Chen, Zou and Wang (2016)]. Based on these advantages, this study uses simulated annealing algorithm as the main optimization tool. This section will introduce the improved annealing method and the model of yield stress based on the SVR yield stress model of RAFM steel. Most search algorithms adopt the heuristic search method, but the ability to use information is deficient. Therefore, it is essential to improve the traditional annealing search algorithm by combining search information.

#### **3.1 GDM-SA search algorithm**

To accurately adjust the parameters of the SVR model, the simulated annealing algorithm must be improved to obtain the simulated annealing algorithm based on a smooth convex optimization and the gradient descent search algorithm, known as the GDM-SA algorithm. Compared with the traditional search algorithm, the GDM-SA algorithm can be used as a global search, and the algorithm can address the problem that the exact algorithm cannot be solved when the scale of the data is large. In the traditional simulated annealing algorithm, the general process is as follows.

1. Initialization: Set a high enough initial temperature  $T_0$ , and make  $T=T_0$ , random initialization solution  $S_1$ , markov chain length  $L$ .
2. A new solution  $S_2$  is generated from the current  $S_1$  random perturbation.
3. Calculating the increment of  $S_2$   $\Delta f=f(S_1) -f(S_2)$ .

4. If  $\Delta f < 0$ , accept the current solution, otherwise the new solution can only be accepted by the probability of  $\exp(-\Delta f/T)$ .
5. If the iterative condition is satisfied, save the optimal solution as is and end the program. Otherwise, perform do cooling according to the decay function and return to Step (2).

In the process of cooling, it is difficult to determine when the cooling achieves stable status. The traditional method is to fix the length of a markov chain directly. The problem is highly complex in that the objective functions in reality are always complex functions with multiple stagnation points, multiple saddle points, non-smoothness and other characteristics, and there are multiple local minima. Therefore, ordinary optimization methods will face more severe challenges. This paper uses a gradient descent algorithm to improve the range of the search region in the process of cooling. At the same time, this paper uses gradient data to give the algorithm local directivity.

In the process of annealing, if the number of iterations times is  $k$ , try to find the solution that meets the accuracy requirement in the  $K$  iterations. The algorithm tends to be stable at current temperature, and then proceed to the next iteration.

Given the following sets of training samples

$$(X_1, Y_1), (X_2, Y_2), (X_3, Y_3), \dots, (X_k, Y_k)$$

The error calculation formula used in the  $K$  iteration:

$$E_k = \frac{1}{2} \sum_{i=1}^l (\hat{y}_i^k - y_i^k)^2 \quad (2)$$

After  $K$  times iteration, define the measure of overall cumulative error

$$\mathfrak{R} \stackrel{def}{=} E_1 + E_2 + E_3 + E_4 + \dots + E_k \quad (3)$$

Satisfying

$$\frac{\partial E_i}{\partial E_j} = \xi_{ij} = \begin{cases} 1, & i = j \\ 0, & other \end{cases} \quad (4)$$

In fact,

$$\mathfrak{R} = \frac{1}{2} \sum_{i=1}^k \sum_{j=1}^m (\hat{y}_j^i - y_j^i)^2 \quad (5)$$

The above statements shall satisfy the independence assumption. Certainly, it is impossible to be completely independent in practice because there are overlapped parts of data samples

on the training set, and there will be correlations between the trained learners depending on the actual situation.

The error in the  $K$  iteration is  $E_k, \forall E_k \in [\tau_k, \eta_k], \forall \epsilon \rightarrow 0$ , according to the Hoeffding inequality, it has

$$P \left( |\mathfrak{R} - E[\mathfrak{R}]| \leq \frac{\sqrt{2}}{2} \epsilon \right) \geq 1 - 2 \exp \left[ - \frac{1}{\sum_{i=1}^k \left( \frac{\tau_i - \eta_i}{\epsilon} \right)^2} \right] \tag{6}$$

The above shows that the precision of system itself has an exponential increase after  $K$  times [Ruxton (2010)], where  $K$  is a given number.

Considering that the mathematical properties of the cost function  $E_k$  are not good, use the exponential loss function as an alternative function. Some researchers already proved the consistency of the exponential loss function and the square loss function [Zhang (2004)]. Thus, use the exponential loss function as the optimization objective, and the original formula becomes

$$\iota(k) = E_{k \sim \chi} [\exp(-g(k))] \tag{7}$$

Where  $g(k) = E_k$  and  $E_{k \sim \chi}$  is the mathematical expectation of the characteristic space  $\chi$ .

Then, use the method of smoothing convex optimization approach to optimize the new target function [Bubeck and bastien (2015)], that is to minimize the new objective function  $\iota(k)$

$$\text{min} \iota(k) \tag{8}$$

$$\text{subject to } \chi \rightarrow \Psi, k \in \chi \subset \Psi^n \tag{9}$$

where the  $\chi$  is the characteristic space of the training sample set, and the  $\psi$  is the characteristic space of the overall sample set.

There are many ways such as the Lagrangian multiplier method to optimize the objective function, coordinate the descent method, the Newton method, the gradient descent method, and the variant methods derived from them. Using different optimization methods will have different effects on search results. The cost function has been converted to the exponential loss function [Singh, Singh, Singh et al. (2008)], therefore, use the gradient descent method to do optimization since the exponential loss function has good mathematical properties.

The gradient descent method requires that the following sequence can be constructed:

$$K_{i+1} = K_i + \mu \Delta K_i, i = 1, 2, 3, \dots \tag{10}$$

And requiring

$$\iota(k_{i+1}) < \iota(k_i) \tag{11}$$

In this instance, we used the idea of relaxation and approximation, and constructed a series of relaxation variables to approach the real functions. With the increase in the iteration

times, and the contracted sequences, we can finally guarantee algorithm convergence. In practice,  $\mu$  is a step factor. To ensure convergence quickly, the direction of the descent should be in the negative direction of the gradient. The next section will deduce the optimization problem for this study.

During the iterative process, given that the error of the  $i$  time is  $E_i$ , perform the second order Taylor expansion here because when the error is very small, the higher term of the Taylor expansion is the high order infinitesimal about the step factor  $\mu$  which is also called the learning rate in a neural network. The objective function can be approximately replaced by the following quadratic polynomial.

$$\iota(k_{i+1}) \approx \iota(k_i) + \left[ \frac{\partial \iota(k_i)}{\partial k_i} \right]^\top (k_{i+1} - k_i) + \frac{1}{2} (k_{i+1} - k_i)^\top \frac{\partial \left( \frac{\partial \iota(k_i)}{\partial k_i} \right)}{\partial k_i} (k_{i+1} - k_i) \quad (12)$$

The upper form is the approximation function that is formed by the quadratic expansion structure. For the optimization function  $\iota(k)$ , here further are:

$$\iota(k_{i+1}) \approx \exp(-g(k_i)) \cdot (1 - \nabla g(k_i) \cdot (k_{i+1} - k_i)) + \frac{1}{2} H[\exp(-g(k_i))] \cdot (k_{i+1} - k_i)^2 \quad (13)$$

Mark  $\nabla$  is a gradient operator, and  $H[f(x)]$  is the second order Hessian matrix of the  $f(x)$  function [Powell (1979)]. For the definition of the Hessian matrix, please refer to convex optimization theory [Bubeck (2014)], and the corresponding Hessian matrix here is:

$$H[\exp(-g(k_i))] = \exp(-g(k_i)) [\nabla g(k_i)^2 - H[g(k_i)]] \quad (14)$$

In actual operation, because the calculation of the Hessian matrix absorbs extensive computing resources [Mcgill, Tukey and Larsen (1978)], the complexity grows exponentially with the increase of scale of the problem. Therefore, the relatively simple matrix  $(1/\mu)I$  that is like the Hessian matrix is used to approximate the substitution and  $I$  is the  $k$  identity matrix. From that, the new simplified formula is as follows:

$$\iota(k_{i+1}) \approx \exp(-g(k_i)) \cdot (1 - \nabla g(k_i) \cdot (k_{i+1} - k_i)) + \frac{1}{2\mu} \|k_{i+1} - k_i\|_2^2 \quad (15)$$

The gradient vector operator near point  $k_{i+1}$  is obtained by the quadratic approximation method.

$$\nabla \iota(k_{i+1}) = \frac{\partial \iota(k_{i+1})}{\partial k_{i+1}} = -\exp(-g(k_i)) \cdot \nabla g(k_i) + \frac{1}{\mu} \|k_{i+1} - k_i\|_1 \quad (16)$$

Refer to convex optimization theory for the second order gradient analysis of the real variable function [Mcgill, Tukey and Larsen (1978)]. The necessary condition for the existence of a first-order extreme point is to meet the requirement that the gradient is 0, using the following formula:

$$k_{i+1} = k_i + \mu \cdot \nabla g(k_i) \cdot \exp(-g(k_i)) \quad (17)$$

The upper form is the updated iteration formula of this algorithm. If the search ability cannot meet the specific data requirements, we can also use the second order gradient that is the information of the Hessian matrix to find a faster descent direction. The optimized objective function is as follows:

$$\min_{\Delta k}(k + \Delta k) = \iota(k) + (\nabla \iota(k))^\top \cdot \Delta k + \frac{1}{2}(\Delta k)^\top \cdot \nabla^2 \iota(k) \cdot \Delta k \quad (18)$$

With reference to convex optimization theory, the necessary and sufficient conditions for the second order gradient analysis of a real variable function shows that in order to obtain the local optimal point of the original function (it is the minimum here), the first order gradient of the iterative increment is 0 and the second order gradient is positive. The Hessian matrix is required to be a positive definite matrix. Therefore, it is necessary to find the constraint conditions.

The concept of a convex set, that is, for a given set  $S$  that is a convex set, two arbitrary points  $A$  and  $B$  belong to  $S$  and the line segments connecting them are also in the set  $S$ . It is as follows:

$$a, b \in S, \forall \tau \in [0, 1] \Rightarrow \tau a + (1 - \tau)b \in S \quad (19)$$

For the objective function, there is:

$$H[\iota(g(k))] = \nabla(-\nabla g(k) \cdot \exp(-g(k))) = \nabla g(k)^2 \cdot \exp(-g(k)) \geq 0 \quad (20)$$

The objective function  $\iota$  is a lower convex function. Therefore, any convex function satisfies the Jensen inequality

$$\forall k_i, k_{i+1} \in \varnothing, \forall \lambda \in [0, 1] \quad (21)$$

$$\iota(\lambda k_i + (1 - \lambda)k_{i+1}) \leq \lambda \iota(k_i) + (1 - \lambda)\iota(k_{i+1}) - \frac{\mu}{2}\lambda(1 - \lambda)(k_i - k_{i+1})^2 \quad (22)$$

when  $\mu > 0$ ,  $\iota$  is a strong convex function. Therefore, the sufficient and necessary condition for the strong convex function  $\iota$  to be an objective function that is defined on  $R$  is that if and only if  $\iota$  can be a quadratic differential and the Hessian matrix is a positive definite matrix.

$$H_k \iota(k) = \frac{\partial^2 \iota(k)}{\partial k \partial k^\top} \succ 0, \forall k \in \varnothing \quad (23)$$

From the above analysis, the objective function  $\iota$  satisfies the constraints of the second order convex optimization. After the above constraints are satisfied, the second order algorithm of the smooth convex optimization can be used to analyze the second order gradient of the objective function  $\iota$ . Therefore, it has the following:

$$\frac{\partial \iota(k + \Delta k)}{\partial \Delta k} = \nabla \iota(k) + \nabla^2 \iota(k) \Delta k = 0 \quad (24)$$

Thus, the second order iterative formula for this study is obtained.

$$\Delta k_{nt} = \exp(-g(k)) \cdot (\nabla^2 \iota(k))^{-1} \cdot \nabla g(k) \quad (25)$$

The upper form is called the Newton method, which uses the information of the second order gradient to optimize. The updated direction is more accurate, but the drawback is the inverse calculation of the Hessian matrix and the iterative process will also take up extensive of calculation resources. It is necessary to consider the actual situation for screening and this study gives two optimization methods above.

Applied to the cooling process during annealing, searching the optimal value in the process of cooling under the current temperature by the above algorithm, the algorithm will converge to an optical point and end the search. The algorithm will reach the stable status of the cooling process and start the next round of temperature cooling. The pseudo code of the whole process is represented in reference Tab. 2.

Tab. 2 is the use of pseudo code to explain the main program of the combinative annealing algorithm. The traditional annealing algorithm will set a fixed metropolis inner loop chain, except the Steps 5~16, as a short-term equilibrium constraint condition at the current temperature. This balance is called a steady state under the current temperature. However, this approach is relatively fixed, and it is difficult to know whether it reached the steady state or not after the iteration. Therefore, the GDM algorithm is designed to automatically adjust and optimize the algorithm to reach balance by the given precision value. Compared with traditional SA algorithms, the GDM algorithm is more flexible and the scope of the search space is also expended.

**Table 2:** Improved simulated annealing GDM-SA algorithm

<b>Input</b>	<i>Training data set <math>D=\{(x_1,y_1),(x_2,y_2),\dots,(x_n,y_n)\}</math>;</i>
<b>Main parameters</b>	<i>The annealing temperature <math>T</math>, the attenuation parameter <math>R</math>, Tolerance <math>T_{min}</math>, the maximum iteration number <math>I_{max}</math>, learning rate step factor <math>\mu</math>, and the stable precision <math>\epsilon</math> are set.</i>
1:	<i>Random initialization parameters</i>
2:	<b>While</b> $T > T_{min}$ <b>do</b>
3:	$L(x) \leftarrow newL(x)$ ;
4:	$\Delta L(x) \leftarrow L(x) - newL(x)$
5:	<b>If</b> $\Delta L \leq 0$
6:	<i>Accept new solutions and execute the GDM algorithm into the steady status</i>
7:	<b>For</b> $\epsilon' > \epsilon$ <b>do</b> //Execute GDM algorithm
8:	$K' \leftarrow K - \mu \Delta K$ ;
9:	<b>End for</b>
10:	<b>Elseif</b> $rand() > exp(-\Delta L/T)$
11:	<i>The new solution is accepted with a certain probability, and the GDM algorithm is executed into steady status</i>
12:	<b>For</b> $\epsilon' > \epsilon$ <b>do</b> //Execute GDM algorithm
13:	$K' \leftarrow K - \mu \Delta K$ ;
14:	<b>End for</b>
15:	<b>Else</b>
16:	<i>Do not accept new solutions</i>
17:	$T \leftarrow r * T$ ; //Annealing and cooling
18:	<b>End while</b>
19:	<i>output the best values;</i>

### 3.2 Comparison with unimproved algorithms

Compared with unimproved annealing algorithms, the simulated annealing algorithm of GDM has the advantages of a wider search range and easing of the premature local optimization. To verify the above advantages, chose the following complex functions as the test target to search the minimum of function  $f(x_1, x_2)$ . The function formula is

$$f(x_1, x_2) = (x_1^2 + x_2^2)^{0.4} \cdot \sin^2(40(x_1^2 + x_2^2)^{0.1}) + 1 \quad (26)$$

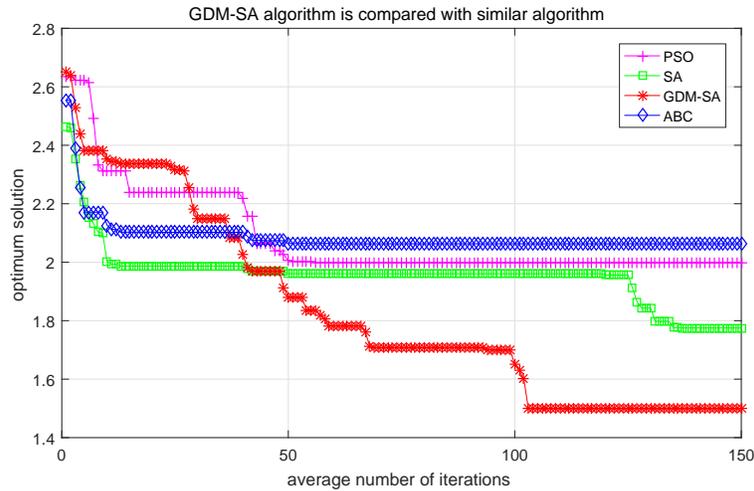
This function has numerous local minimum points, and it is easy to find that the global minimum of this function is  $x_1=0, x_2=0$ . At this point,  $f(x)$  takes the global minimum value of 1, that is

$$f(x_1, x_2)_{min} = 1 \quad (27)$$

$$x_1 = 0, x_2 = 0 \quad (28)$$

After selecting certain tasks, perform the comparison and analysis of the search capability between the simulated annealing algorithm improved by the GDM algorithm and the previous one. Set the same annealing parameters. The attenuation parameter is 0.95. The step factor is 10 (force to control the disturbance). The initial temperature is 100, and the tolerance is 0.0001. The learning rate of GDM is 0.001. The maximum number of iterations is 4000. The minimum change curve of function  $f(x)$  obtained in the iterative process is calculated as shown in Fig. 3 below.

Fig. 3 shows the results of calculating the minimum value of the given complex function on the examples given in this study. Under the same conditions, we compare the traditional simulated annealing algorithm, GDM-SA algorithm, particle swarm optimization algorithm and artificial bee colony (ABC) algorithm [Gao, Suganthan and Pan (2016)]. The average results of each algorithm are obtained by running several times, and then the graph of search results changing with the number of iterations is drawn. It can be seen that the traditional optimization algorithm falls into the local optimum after the search, such as the traditional simulated annealing algorithm with an optimum value of 1.775, particle swarm optimization with an optimum value of 1.996, artificial bee colony with an optimum value of 2.063. The optimal value of GDM-SA algorithm is 1.498, and the search space is wide. Our explanation is that in the process of searching, we can make full use of the advantages of accurate search to find the best value after falling into the local optimal value, and at the same time, we can jump out of the local optimal value by combining the idea of annealing algorithm. Finally, our improved algorithm obtains better results than a single heuristic algorithm. However, the improved annealing algorithm of GDM can almost achieve a global optimal solution. Therefore, Fig. 3 shows that the improved simulated annealing algorithm has a stronger searching ability, faster convergence speed and wider search scope. The above experiment is only a method to compare under the same conditions. When the temperature rises appropriately, the traditional simulated annealing algorithm can also approach the global minimum.



**Figure 3:** The search results of GDM-SA algorithm and traditional simulated annealing algorithm, particle swarm optimization algorithm and artificial bee colony algorithm on function extremum optimization problem. Each algorithm performs at least three experiments, and then the search results are averaged to plot the optimal value and the number of iterations.

The results above show that the SA algorithm combined with the GDM algorithm, known as the GDM-SA algorithm, has a strong search ability. Next, the parameters of SVR are optimized by the GDM-SA algorithm to get the spatial grid research results within the interval range of optimal value. This paper introduces the SVR model of GDM-SA below and later establishes the yield stress prediction model of SVR.

### 3.3 GDM-SA-SVR yield stress model

This study use the SVR model to establish the yield stress model [Huang and Tsai (2009)]. The principle of the SVR model is referred to the work of Vapnik [Cortes and Vapnik (1995)].

Referring to the data distribution of RAFM steel in Appendix A, the magnitude difference between different attributes is large. For example, carbon element content and the test temperature have 10000 orders of magnitude difference. In scientific calculation, it is necessary to prevent two very large numbers from calculating directly, which will lead to the loss of calculation precision and data information. Therefore, the normalization process is needed [Wang and Tang (2015)]. The minimum and maximum normalized normalization formula is used here.

$$v'_i = \frac{v_i - \min_A}{\max_A - \min_A} (\text{newmax}_A - \text{newmin}_A) + \text{newmin}_A \quad (29)$$

The strategy used in this paper is using the improved simulated annealing algorithm GDM-

SA to optimize the penalty factor  $C$  of SVR and the kernel function parameter  $G$  (gamma) to get an approximate interval of extreme value. Meanwhile, for the regression error threshold parameter  $\epsilon$  (general experience value is 0.1), when the error is less than the modified value, it will not be punished, and the value will be punished when the error is bigger than the modified value. That is equivalent to the expected regression precision, so it has a significant effect on the result. In order to obtain more accurate results, we set a more subtle adjustment interval. Then, use the space grid search algorithm for accurate search [Liu, Liu and Yang (2006)]. The combination of rough and precise search can effectively approximate the global optimal solution.

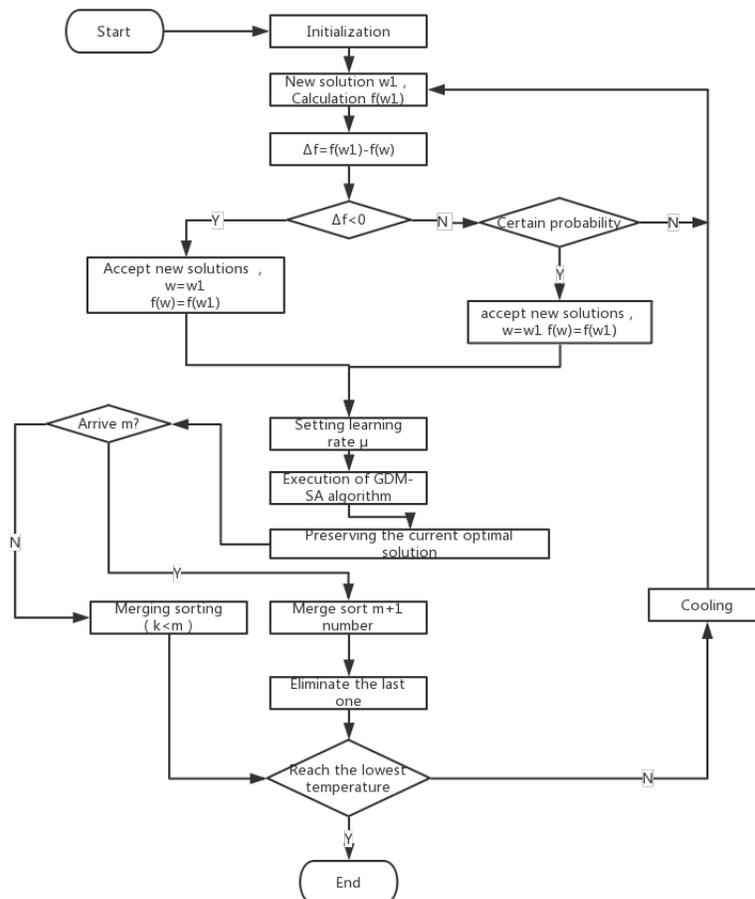
Train the model on a training set by the annealing algorithm [Ginneken and Van (2016)]. Calculate the mean square error. The formula used in the  $K$  iteration is the mean square error formula  $E_k$ , above. (Refer to the definition in the last chapter).

Combine with the above design model to design a preliminary algorithm to iterate and optimize parameters  $C$  and  $G$  of SVR. To obtain more solutions, a more robust method is used to save the number of  $m$  optimal solutions after each iteration. Next, the number of  $M$  optimal solutions is sorted by the merging sort and stored in the matrix structure and the final product is output. To date, the annealing algorithm SA was improved by the GDM algorithm to optimize the SVR model, and saved the number of  $m$  best optimal solutions of the SVR model. The whole flow chart is shown as the Fig. 4. The related attenuation parameters are set to 0.95, the step factor is 10 (force to control disturbance), the initial temperature is 80, the tolerance is 0.0001, the GDM learning rate is 0.001, the maximum iteration number is 2000. Tab. 3 shows that 10 optimal values were selected after an iterated annealing algorithm.

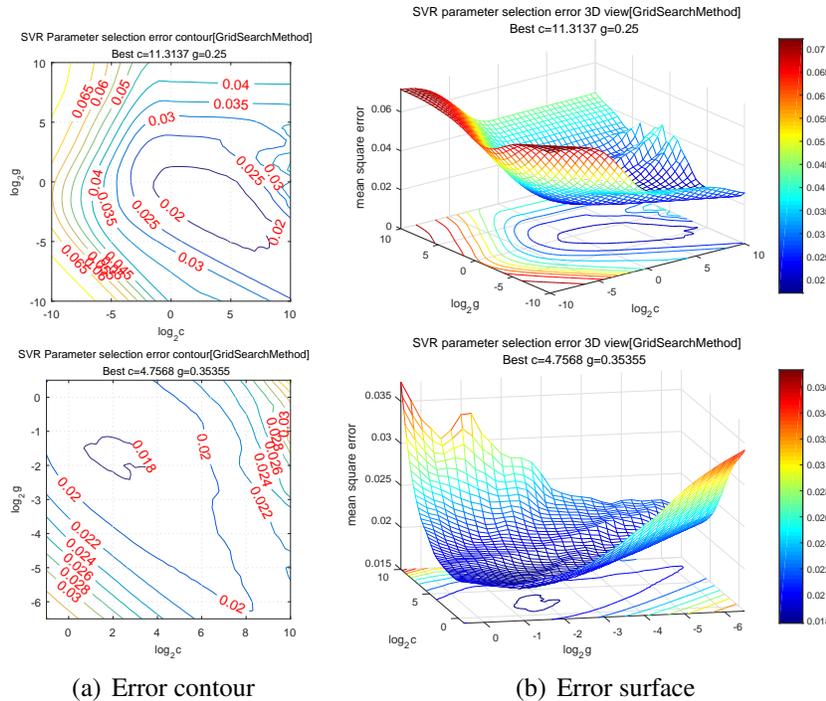
**Table 3:** The best parameter values from different groups in a certain iteration (here  $\epsilon = 0.1$ )

Index	Best C	Best G	Error
1	25.29710369	17.08638799	0.008325685
2	23.93748805	15.58713842	0.008328292
3	25.31997575	14.78919021	0.008322496
4	25.29052213	14.68531137	0.008321027
5	23.89358021	14.84428247	0.008326067
6	25.44601581	17.68439861	0.008327954
7	24.66109982	15.72346235	0.008327622
8	25.36246968	15.13926097	0.008320294
9	25.10177947	15.10707552	0.008323473
10	24.96402485	17.34268876	0.008327509

Excessive penalty parameters will increase the accuracy of the training set but will lead to overfitting. In practice, we are inclined to choose relatively small parameter values. After many times of simulated annealing algorithm optimization, the final optimal value is  $C=16$ ,  $G=8$  and  $\epsilon=0.0065$ . After initial determination of the optimal range, start the space grid search [Liu, Liu and Yang (2006)].



**Figure 4:** Flowchart of improved simulated annealing algorithm. The improved algorithm process is more complex, and search ability is enhanced



**Figure 5:** Error contour and surface map after multiple iterations. Every iteration, the search area will be further reduced, and we can then determine the global optimal solution through repeated iterations

Take the sampling points of two dimensions in the optimal value interval calculated by the annealing algorithm, which is the test point of the parameters  $C$  and  $G$ . Taking the mean square error as the evaluation standard, make the three-dimensional coordinate diagram of the error of parameters  $C$  and  $G$  and draw the corresponding isogram. By analyzing the contour lines, the search range is further reduced and is repeatedly iterative to obtain the parameters that meet the requirements of the error precision and finally end the algorithm.

After repeated iteration through the above methods, the parameters  $C$  and  $G$  with small errors are finally approximated, and the value is assigned to the SVR model to retrain the model on the training set and finally obtain a relatively ideal model. Through the preceding analysis, for the extremum optimization problem of complex functions, find the optimal value by a heuristic algorithm (most are locally optimal) and then, after repeatedly iterative analysis, it can approach the global optimum infinitely [Zien, Kramer and Sonnenburg (2009)].

To date, an SVR model of 37 attribute variables on yield strength YS has been established. The following is the test and prediction analysis of the mode.

## 4 Test and analysis of the model

### 4.1 Model learning

When evaluating the performance of a model, a very important index is the learning situation of the training samples. Lack of learning will lead to underfitting, and the prediction results will reflect the high deviation characteristics. Overlearning will learn some features of the sample itself (for example, sample noise), and there will be overfitting at this time. The prediction results have a high variance characteristic. The ideal situation is to make a compromise between underfitting and overfitting. The generalization performance of the model is the best at this time. Overfitting and underfitting always accompanied the whole process of machine learning. Many factors are needed to be considered in the training model, and another important reason is that there is not a complete theory to guide a model for meeting the main problems in the process of establishing the model, such as training guidance and evaluating sample information. Even so, there still are some computational learning theories that can be used to analyze the learning model.

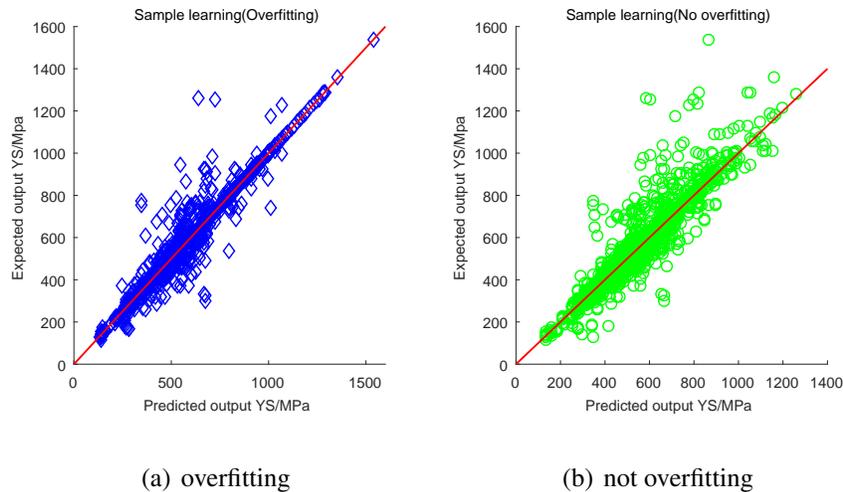
For a training set, we assume that learning satisfies the requirement by reaching a certain precision,  $\epsilon$  [Zeugmann (2016)], and it is recorded as time  $h$ , the number of samples is  $t$ , and in the iterative training process, define the hypothesis space  $\mathbb{R}$ . Next comes

$$P(\exists h \in \mathbb{R} \rightarrow |E(h) - \hat{E}(h)| > \epsilon) = P((|E_{h_1} - \hat{E}_{h_1}| > \epsilon) \vee \dots \vee (|E_{h_j} - \hat{E}_{h_j}| > \epsilon)) \quad (30)$$

$$\leq \sum_{h \in \mathbb{R}} P(E(h) - \hat{E}_h > \epsilon) \leq 2\mathbb{R} \exp(-2t\epsilon^2) \quad (31)$$

when the number of samples  $t$  is big, the empirical error of  $H$  can be approximated to replace the generalization error. Given the hypothetical space  $\mathbb{R}$ , when  $t \rightarrow \infty$ , the right side of the inequality is 0. There must be a hypothesis that the generalization error is minimal to guarantee the learning process. The following Fig. 6 shows the learning of the sample after the final model training.

The training set used in this paper selects 1711 data training models from a random disorder sequence, and the remaining 100 data training models are used to verify the model. In the process of training the model, Fig. 6(a) is over fitting learning the situation. Fig. 6(b) is the learning situation between under fitting and over fitting, that is, learning the characteristics of samples, and noisy data not learned here, so there are some large deviations. This phenomenon is a normal. The data itself not only contains the noise data points but also the contradictory data points.

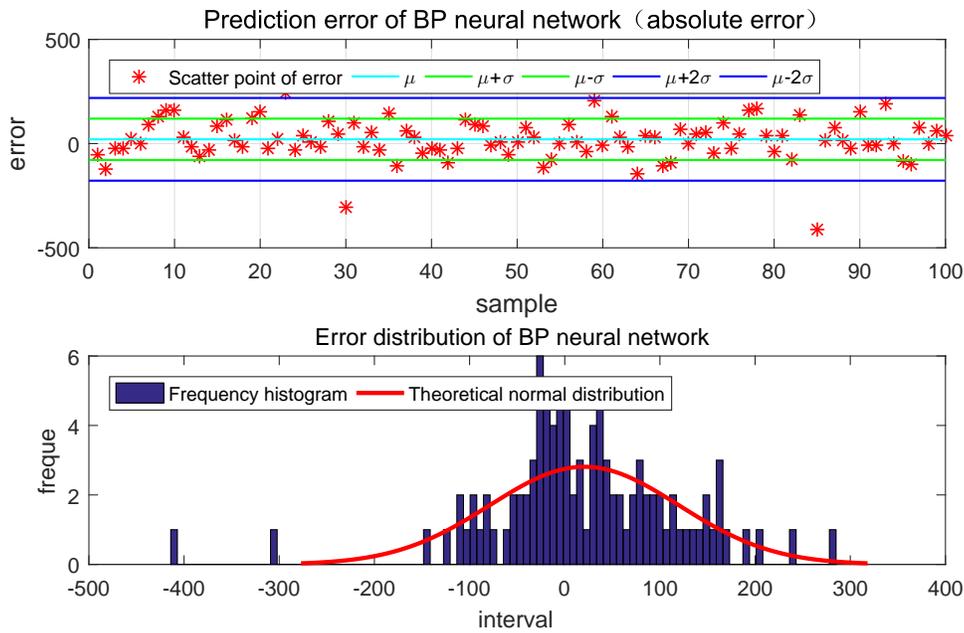


**Figure 6:** Overfitting and not overfitting sample learning

#### 4.2 Model comparison

To verify the performance of the model, compare the model with other similar models including ANN, random forest, linear regression and general regression neural network (GRNN). Train these models on the same set of data and then test these models on the same test set. Get the absolute error (residual) and error distribution of the predicted results. Perform test results statistics with the similar models. Fig. 7 to Fig. 11 is the result of error analysis.

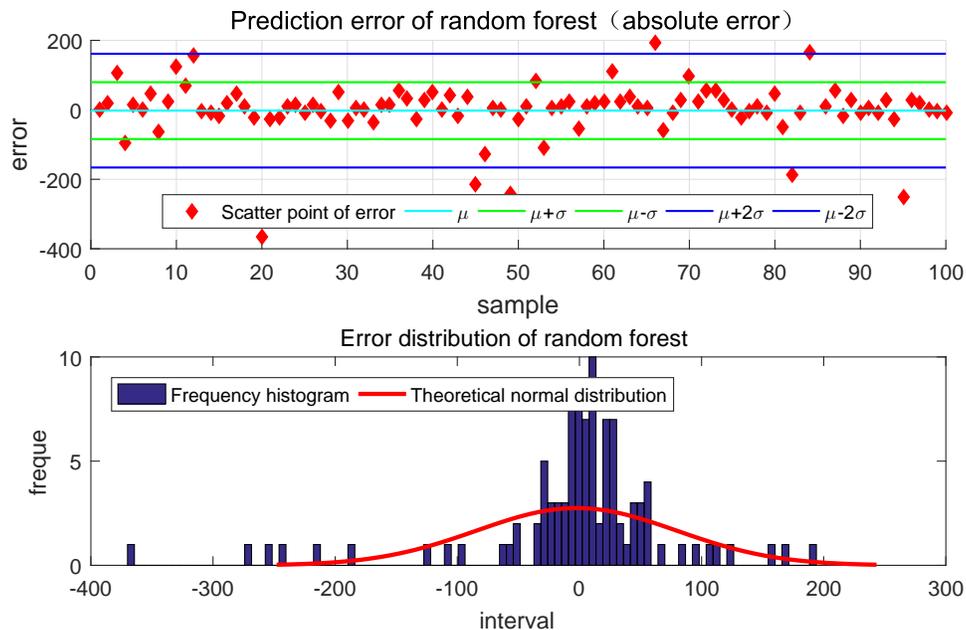
The Tab. 4 is a specific statistical analysis table. In the field of machine learning, use statistical methods to test regression problems. and then, we obtain the residuals of prediction results and expected results on test sets. Perform statistical analysis of residual data by the commonly statistics including mean, maximum deviation, variance, standard deviation and mean square error (MSE) and goodness of fit (also called coefficient of determination). In general, the mean value of error obeys a normal distribution, and the closer to 0, the better the mean value of error is. The smaller the value of the maximum deviation, variance, standard deviation and mean square error are, the better the parameters themselves are. When the goodness of fit, known as the coefficient of determination  $R^2$ , is the closer to 1, the better the effect of the regression is. Tab. 4 is the BP neural network, random forest, linear regression, generalized regression neural network and the GDM-SA-SVR algorithm used in this paper. It is obvious that GDM-SA-SVR has much better performance than other similar algorithm models in addition to the relevant smaller disadvantages. Therefore, the GDM-SA-SVR algorithm model is effective.



**Figure 7:** Absolute error and frequency distribution histogram of BP neural network

**Table 4:** Performance comparison of BP neural network, random forest, linear regression, GRNN (generalized regression neural network) and GDM-SA-SVR algorithm on test set

Error S-tatistics	BP Neural Network	Random Forest	Linear Regression	GRNN	GDM-SA-SVR
mean value	20.51900694	-2.350357661	-18.70689077	106.6010736	<b>-1.5253643</b>
maximum deviation	-410.4	-367.6	-433.2	498.6	<b>-317.2</b>
variance	9.84E+03	6.70E+03	1.59E+04	1.38E+04	<b>4.26E+03</b>
standard deviation	99.21	81.85	126.01	117.43	<b>65.27</b>
mean square error	1.02E+04	6.64E+03	1.61E+04	2.50E+04	<b>4.35E+03</b>
goodness of fit	0.732619977	0.804896505	0.60397564	0.635439089	<b>0.880247139</b>



**Figure 8:** Absolute error and frequency distribution histogram of random forest

#### 4.3 Significance test of the regression effect

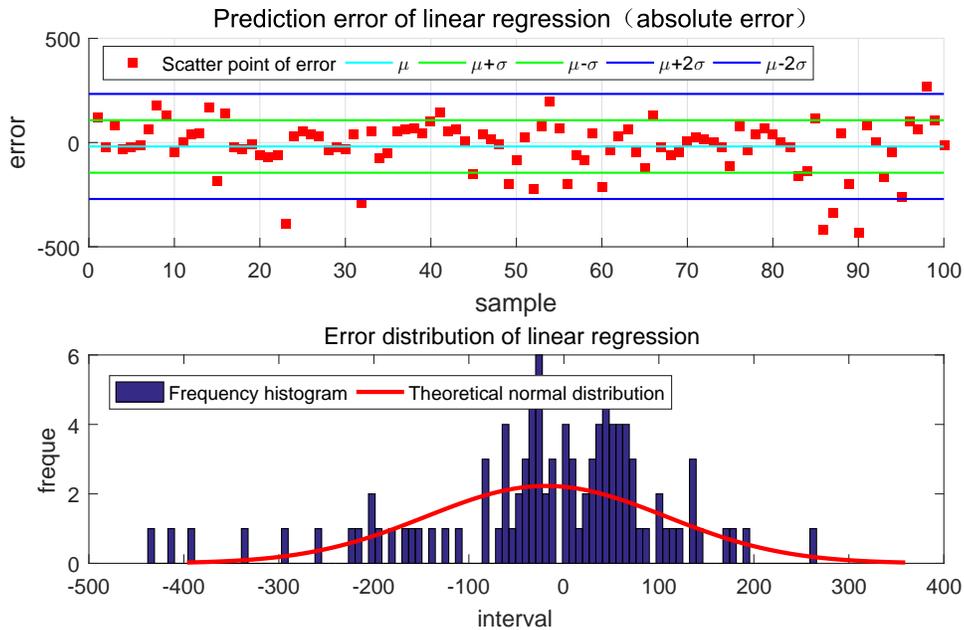
The content of the previous section has explored the learning situation of the model for sample data and only assesses the ability of a model to store characteristic information of a data sample. Since the partition of a training set is completely independent, the performance of a test set is truly to show the predictability of the model when the model is used to predict. Given a test set that is not trained, the prediction effect of the model is different from the reality.

Fig. 12 shows that 96 of the 100 data points on the test set fall within the two times sigma range of the prediction curve, and only four of the number are out of the range. The regression effect is very obvious from preliminary judgment, but for further analysis, it is necessary to test the results and the actual results by the *T*-test [Ruxton (2010)], it uses the *t* distribution theory to deduce the probability of occurrence of differences, so as to compare whether the difference between the two averages is significant. Therefore, it is possible to test whether there is a significant difference between the predicted result and the actual result.

For the original problem, it is assumed that there is a linear relationship between the predicted results and the actual results, and the regression coefficient set as *b*.

1. Putting forward the original hypothesis of regression coefficient

$$H_0 : b = 0 \quad (32)$$



**Figure 9:** Absolute error and frequency distribution histogram of linear regression

2. Structural statistics  $T$ , and satisfy

$$T = \frac{\hat{b}}{\hat{\sigma}} \sqrt{L_{xx}} \overset{H_0}{\underset{true}{\sim}} t(n-2) \quad (33)$$

3. Given the significance level  $\alpha (0 < \alpha < 1)$  making

$$P(|T| > t_{\alpha/2}(n-2)) = \alpha \quad (34)$$

4. Get a rejection region

$$|T| > t_{\alpha/2}(n-2) \quad (35)$$

Calculate  $T$  statistics through the experimental results and the test results. Accept the original hypothesis if the rejection region is not satisfied, that means the predicted effect does not have a linear relationship. Otherwise, the prediction effect has a linear relationship. The effect of regression is remarkable.

For a given confidence level  $\alpha=0.01$ , calculate the results  $T_1=35.7328$ . There are 100 data points in the test set. When the degree of freedom is greater than 45, the  $T$  statistic is approximates the standard normal distribution, where  $T \approx Z$ . Therefore, in the  $\alpha=0.01$  confidence level, check the table to get  $Z=2.57$ , that is,  $T_1 > T$ . Therefore, reject the original hypothesis, and there is 99% confidence that the predictive value and the actual value of the regression effect is obvious.

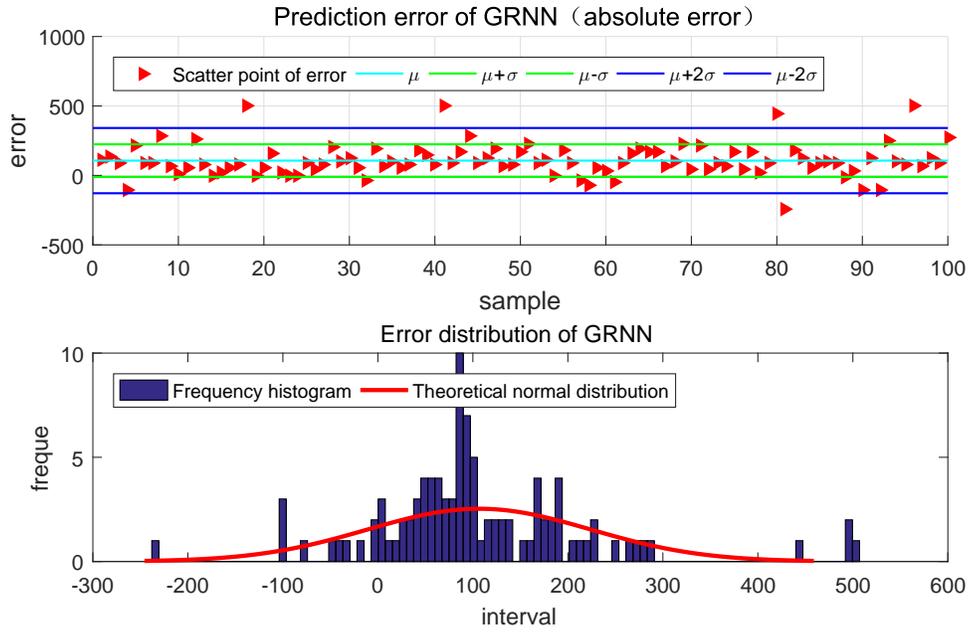


Figure 10: Absolute error and frequency distribution histogram of GRNN

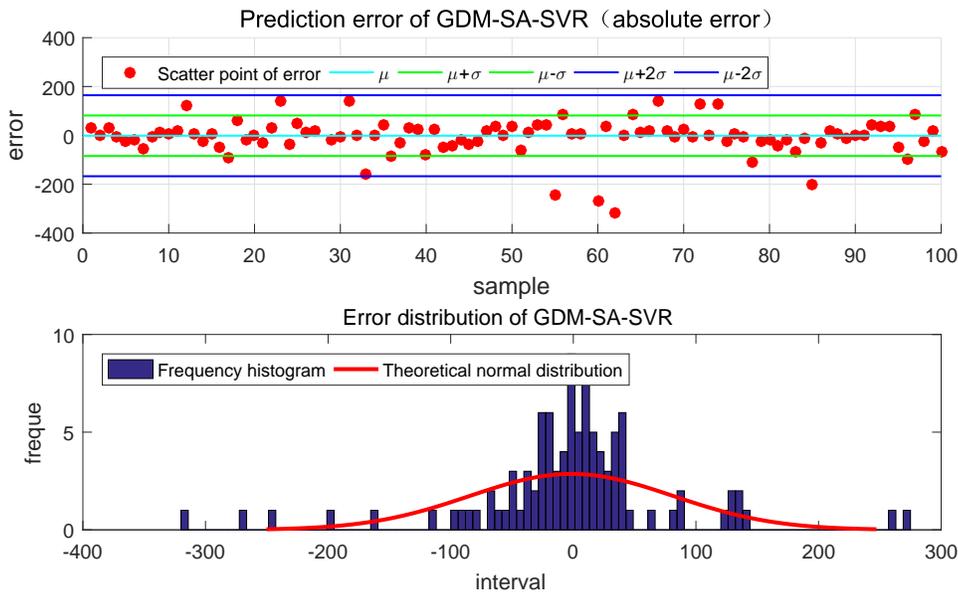
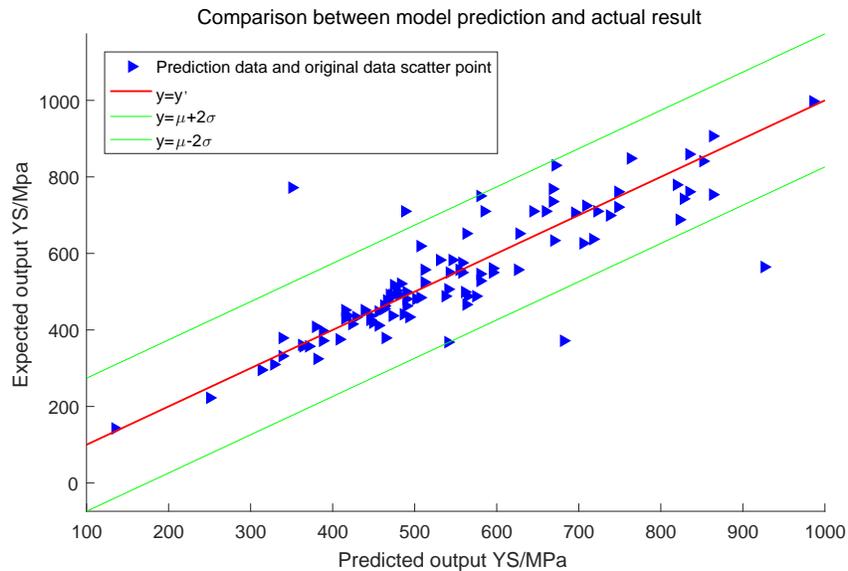


Figure 11: Absolute error and frequency distribution histogram of GDM-SA-SVR



**Figure 12:** The predictive effect of the GDM-SA-SVR model on the test set. From the graph, we can see that in the 100 training sets, (assuming that the error obeys normal distribution), only 4 points fall outside the prediction interval of  $2\sigma$

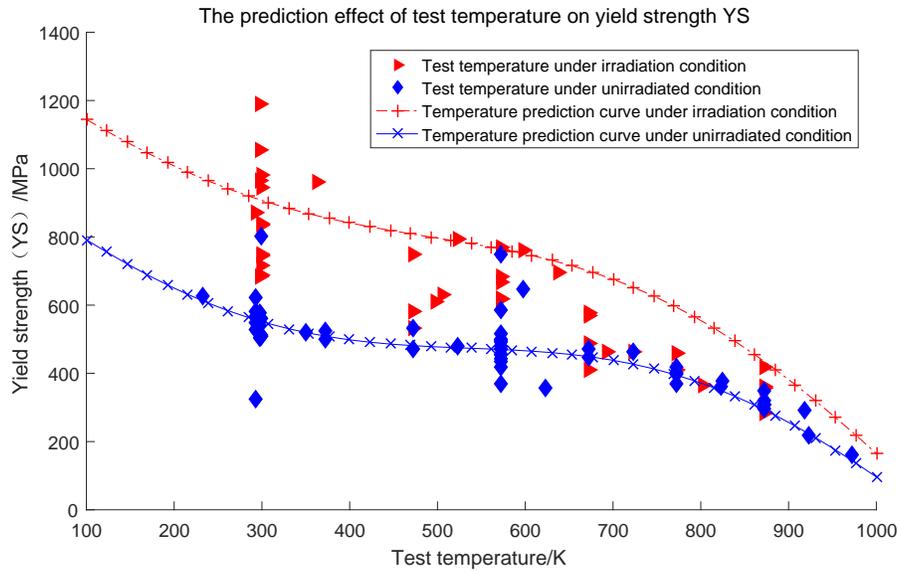
In addition, there is another method (the correlation coefficient method used previously) that can test the effect of regression. This article does not make a detailed introduction about the correlation coefficient method and only gives the final result. Through calculation, the correlation coefficient fluctuates within the range of more than 0.8. At that point, the quantitative model set by statistical analysis evaluates that the SVR model has a significant regression effect on the test set.

## 5 Results and discussion

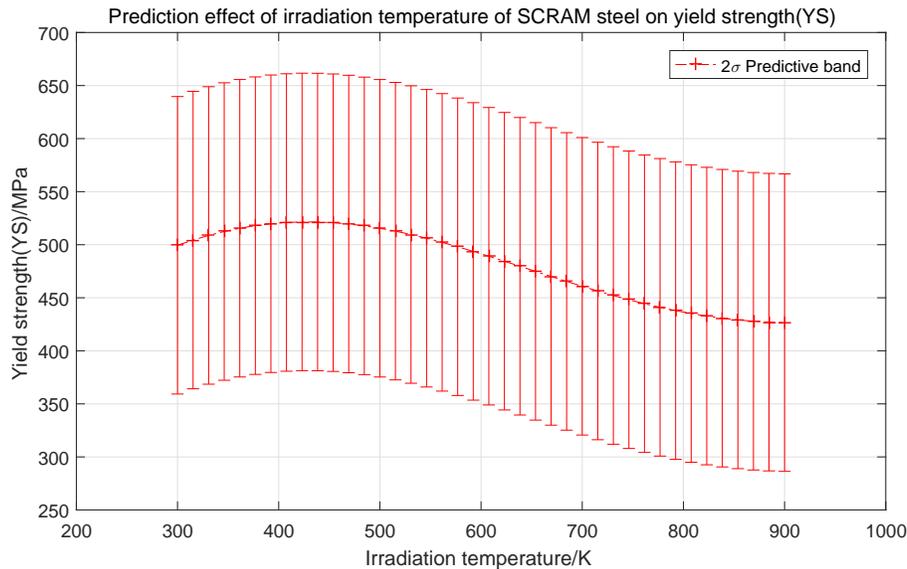
### 5.1 Prediction of test temperature

The test set data (100 pieces) is remarked before and after irradiation, and a scatter plot is drawn. The data are influenced by the mutual influence of different at tributes. Therefore, the data have a certain dispersion, but the overall trend is more obvious. Extrapolation prediction is carried out within the extrapolated range of [100,1000] and a test temperature (K) [Aktaa and Schmitt (2006)]. Use the trained SVR model to make the prediction curves of Fig. 13 with the constraints before and after irradiation. Compared with the true scatter point data of the test set, the prediction curve has a better prediction effect.

The model prediction shows that in the temperature range of 100 K to 1000 K, when the temperature increases, the yield stress decreases as the curve in the Fig. 13. The contrast of before irradiating and after irradiating indicates that doing the experiment in the test temperature range, within a given dose (the different effects from different doses, the size



**Figure 13:** The predictive effect of the GDM-SA-SVR model on the test set. The trained model is given 0 or a number of irradiation parameters, and then the prediction curve is compared with the original data points. The original data points are distributed near the prediction curve. The GDM-SA-SVR algorithm has a good prediction effect



**Figure 14:** Prediction interval of SCRAM steel in  $2\sigma$ , where  $\sigma$  is the variance of the normal distribution of error

of irradiation dose in the curve diagram is approximately 0.024 dpa), the yield strength difference will decrease before and after irradiation [Wang, Zhang and Zhao (2017)]. Since the sample of the training set includes a variety of alloy materials, the prediction of the RAFM steel alloy material is in line with the test set.

### **5.2 Prediction of radiation temperature**

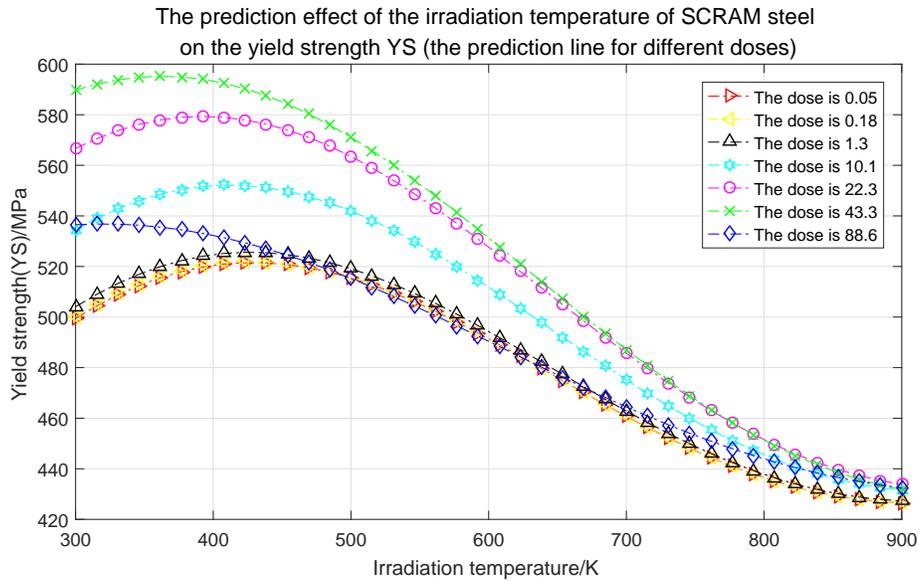
Below is the introduction of the reduced activation ferritic martensitic steel SCRAM steel, which was developed by a university with vacuum induction electric furnace smelting and argon protective atmosphere electroslag re-melting technology. The main chemical components are shown in Tab. 5.

For the above SCRAM steel, give the same test temperature and irradiation dose to investigate the effect of different irradiation temperatures on the yield strength (YS) of SCRAM steel.

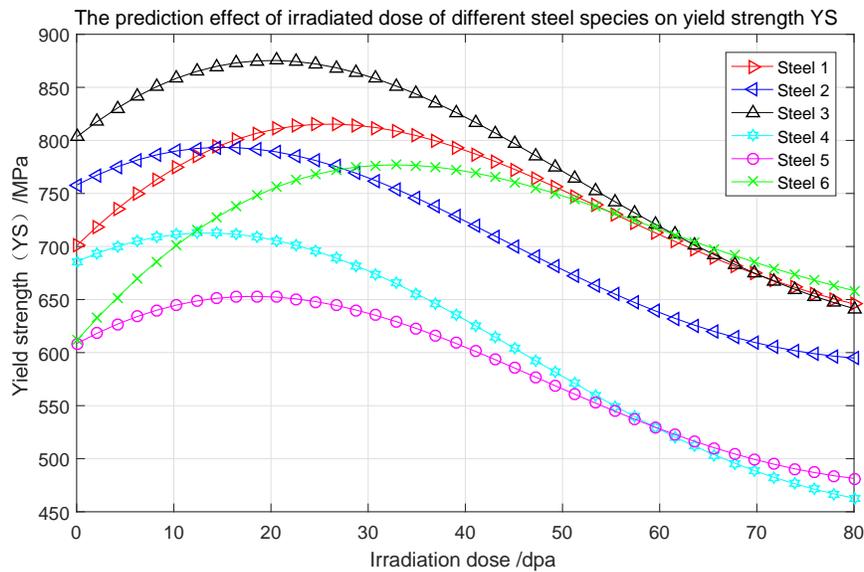
**Table 5:** Chemical composition of SCRAM steel (wt%)

<b>Element</b>	<b>Si</b>	<b>P</b>	<b>Ti</b>
<b>Content</b>	0.28	0.0056	0.011
<b>Element</b>	<b>V</b>	<b>Cr</b>	<b>S</b>
<b>Content</b>	0.094	9.3	0.002
<b>Element</b>	<b>Mn</b>	<b>W</b>	<b>C</b>
<b>Content</b>	0.5	2.28	0.088
<b>Element</b>	<b>N/ppm</b>	<b>O/ppm</b>	<b>Fe</b>
<b>Content</b>	150	38	Remnant

Fig. 14 is the prediction interval of SCRAM steel in  $2\sigma$ . In the irradiation temperature range [300,900] (unit K), the yield strength variation interval slowly increases first in the range of [200, 800], and then decreases (MPa). Most of the radiation temperature range of the training set is arranged in 273 K to 900 K, so the prediction of extrapolation to SCRAM steel has a certain reliability. More accurate analysis needs to be compared with real experimental data.



**Figure 15:** The prediction line of radiation temperature of SCRAM steel at different doses. The radiation dose was 0.05, 0.18, 1.3, 10.1, 22.3, 43.3 and 88.6, respectively, and the corresponding prediction curve was calculated by GDM-SA-SVR model



**Figure 16:** When the element composition of RAFM steel is changed, the prediction curves of different types of steels under different radiation doses are given

Fig. 15 is the change of the yield strength of radiation at several doses in 0.05, 0.18, 1.3, 10.1, 22.3, 43.3, and 88.6 (dpa) [Gaganidze and Aktaa (2013)]. The prediction curve of the irradiation temperature in the range of [300, 900] shows that the mutual influence of the physical mechanism of the two attributes is very complex, and in the studied dose range and a given range of irradiation temperature, the effect of the yield strength is not monotonic as there are many maximum or minimum values. If the physical background test is verified, the test points can be selected at the intersection point given by the forecast line.

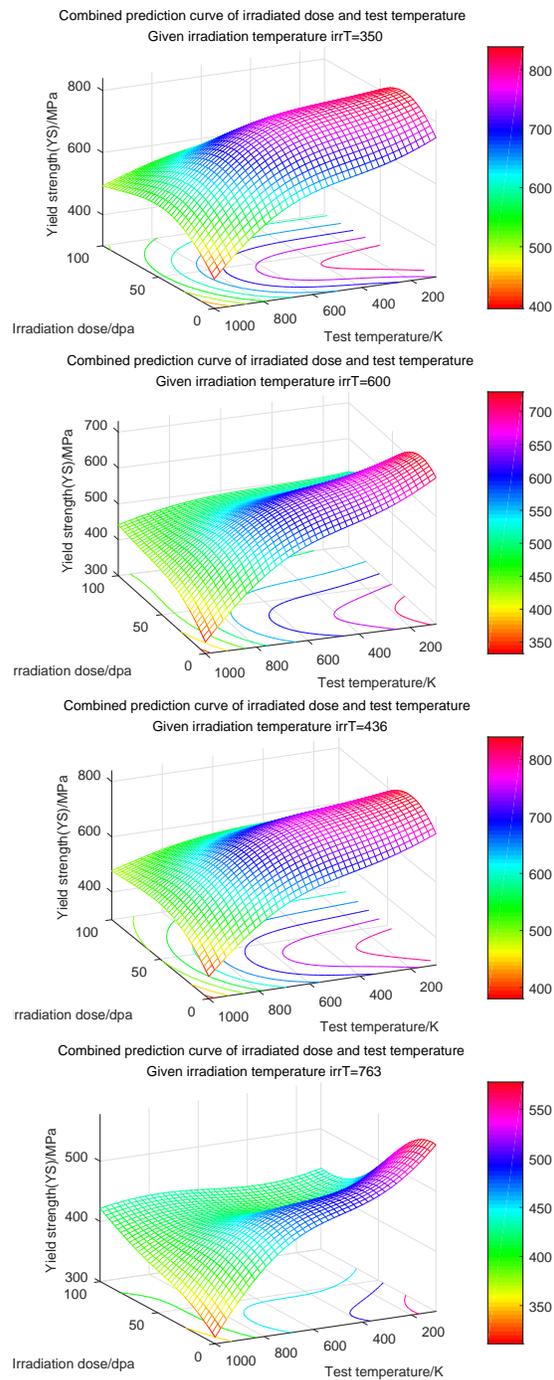
The different irradiation doses have different effects on the yield strength. For example, the yield strength of the zirconium alloy will increase under irradiation, but the ductility will decrease [Lucon (2002)]. This paper presents a research method to study the effects of different materials. First, the irradiation dose is evenly divided in a certain range, and then the SVR yield stress model is used to make the corresponding curve, so it can further study the effect of irradiation dose on the different results. This method can be evaluated and contrasted in reference to the study of yield stress [Virgil'Ev, Kalyagina and Makarchenko (1979)].

### **5.3 Prediction of temperature**

The effect of different steels on the irradiation dose is also different. Fig. 16 is the prediction curve made through the SVR prediction model under the conditions of extracting 6 types of RAFM steels with different chemical compositions from the test data set and a given dose in range of [0, 80] (dpa). The prediction curves all are first increased and then decreased, and different chemical compositions have different effects. Most of the data used in this dataset are in the range of [0, 50]. Therefore, a comparative analysis should be performed with the future experimental points if the over-dose prediction is done. In practical engineering applications, generally, there are no standard experimental data points as a contrast indication, so the method needs to be studied in combination with other methods [Lucon (2002)].

### **5.4 Combination prediction of test temperature and irradiation dose**

The previous analysis of the original data shows that there are multiple collinear dependencies among some attributes. To study the effect of this dependency on the model, a three-dimensional graph must be made to do visualization analysis of the SVR model. In Fig. 17, 350 K, 436 K, 600 K and 783 K are taken as irradiated temperature points corresponding to the combined influence of radiation dose and test temperature on the predicted surface. Clearly, the images of different irradiated surfaces have changed greatly. At the same time, it is worthy noticing that there are few data points in the boundary range, so the model will cause a deviation in the process of learning.



**Figure 17:** The combined effect of test temperature and irradiation dose. The irradiation temperature is 350 K, 436 K, 600 K and 763 K, and the combined effect of the irradiation dose and the test temperature on yield strength is calculated by the GDM-SA-SVR model

## 6 Conclusions

In this paper, we use data mining technology, through statistical analysis of data distribution symmetry to further prepare for our screening model. Correlation can describe the collinearity of attributes and analyze the stability of the model. Next, we use the technique of convex optimization theory to improve the traditional SA algorithm. Then, the GDM-SA algorithm is obtained with stronger search ability. After tests with complex optimal problems test, compared with the traditional SA algorithm under the same conditions, the GDM-SA algorithm has faster convergence speed, less falling into local optima and wider search ability. Therefore, use the improved simulated annealing algorithm (GDM-SA) to optimize the parameters of the SVR model. Finally, a GDM-SA-SVR yield stress prediction model is derived for RAFM steel yield strength.

By comparison, the RAFM steel yield strength prediction model is better than ANN, linear regression, random forest and GRNN. In summary, it is concluded that the hybrid model composed of simulated annealing algorithm and support vector machine can achieve superior performance for compared with ANN, linear regression, GRNN, and random forest. The possible reason is that, from the distribution of the dataset, the data distribution is sparsely presented, which fits the prediction mechanism of the support vector machine. Because of data dispersity and sparsity, SVM can classify data better when it is mapped to the feature space. The soft interval is small, and the error is also small. Therefore, it is also very important to do data mining for original data, which is helpful for subsequent research. These results may help to guide the development direction of the model in the future as follows.

The SVR model uses a single kernel model while the best kernel function in the experiment is the Gauss kernel or the sigmoid kernel. Therefore, the use of multi-core models will achieve better results in facing of different data sets, which is also a heavily studied research field in SVR. In addition to the nature of the model itself, the data sets shall also be considered. If there are new data in the future, adding the new data into the data set to retrain the model can get better prediction ability than the current dataset, and the model is more accurate.

**Acknowledgement:** The research is supported by “National Natural Science Foundation of China” under Grant No. 61572526, and thanks to Mr. He from the material radiation effect team of the China Institute of Atomic Energy. With the help and guidance of Mr. He and Mr. Deng, the experiment was successfully conducted, and the results were greatly improved, which enhanced the structure of this article. Thanks to the editor for giving detailed comments, the quality of the article can be improved.

**References**

- Aktaa, J.; Schmitt, R.** (2006): High temperature deformation and damage behavior of rafm steels under low cycle fatigue loading: experiments and modeling. *Fusion Engineering & Design*, vol. 81, no. 19, pp. 2221-2231.
- Ananthi, S.; Dhanalakshmi, P.** (2015): SVM and HMM modeling techniques for speech recognition using LPCC and MFCC features. *Proceedings of the 3rd International Conference on Frontiers of Intelligent Computing: Theory and Applications*, pp. 519-526.
- Babu, M. N.; Mukhopadhyay, C. K.; Sasikala, G.** (2016): Study of fatigue crack growth in rafm steel using acoustic emission technique. *Journal of Constructional Steel Research*, vol. 126, pp. 107-116.
- Bahrani, S.; Ardejani, F. D.; Baafi, E.** (2016): Application of artificial neural network coupled with genetic algorithm and simulated annealing to solve groundwater inflow problem to an advancing open pit mine. *Journal of Hydrology*, vol. 536, pp. 471-484.
- Barnes, H. A.; Walters, K.** (1985): The yield stress myth? *Rheologica Acta*, vol. 24, no. 4, pp. 323-326.
- Bubeck, S.** (2014): Theory of convex optimization for machine learning. *Foundations and Trends in Machine Learning*, vol. 8, no. 3-4, pp. 231-357.
- Bubeck, S.; Bastien** (2015): Convex optimization: algorithms and complexity. *Foundations and Trends® in Machine Learning*, vol. 8, no. 3-4, pp. 231-357.
- Chen, D.; Zou, F.; Wang, J.** (2016): Samcctbo: a multi-class cooperative teaching-learning-based optimization algorithm with simulated annealing. *Soft Computing*, vol. 20, no. 5, pp. 1-23.
- Cortes, C.; Vapnik, V.** (1995): Support-vector networks. *Machine Learning*, vol. 20, no. 3, pp. 273-297.
- Darmatasia; Fanany, M. I.** (2017): Handwriting recognition on form document using convolutional neural network and support vector machines (cnn-svm). *5th International Conference on Information and Communication Technology*, pp. 1-6.
- Du, C.** (2016): An improved particle swarm algorithm with immune mechanism for traffic matrix estimation. *International Conference on Natural Computation*, pp. 269-274.
- Gaganidze, E.; Aktaa, J.** (2013): Assessment of neutron irradiation effects on rafm steels. *Fusion Engineering & Design*, vol. 88, no. 3, pp. 118-128.
- Gain, A.; Roy, D.** (2016): A novel algorithm for object detection in low contrast image and recognition by svm. *International Journal of Scientific Research*, vol. 4, no. 10, pp. 506-510.
- Gao, K. Z.; Suganthan, P. N.; Pan, Q. K.** (2016): An improved artificial bee colony algorithm for flexible job-shop scheduling problem with fuzzy processing time. *Expert Systems with Applications An International Journal*, vol. 65, no. C, pp. 52-67.
- Ghamisi, P.; Couceiro, M. S.; Benediktsson, J. A.** (2015): A novel feature selection approach based on fodpso and svm. *IEEE Transactions on Geoscience & Remote Sensing*, vol. 53, no. 5, pp. 2935-2947.
- Ginneken; Van, L. P. P. P.** (2016): The annealing algorithm revisited. *Kluwer International*, vol. 72, pp. 109-111.

- Huang, C. L.; Tsai, C. Y.** (2009): A hybrid soft-margin SVM with a filter-based feature selection for stock market forecasting. *Expert Systems with Applications*, vol. 36, no. 2, pp. 1529-1539.
- Jia, F.; Lichti, D.** (2017): A comparison of simulated annealing, genetic algorithm and particle swarm optimization in optimal first-order design of indoor TLS networks. vol. IV-2/W4, pp. 75-82.
- Jian, L.; Shen, S.; Li, J.** (2017): Budget online learning algorithm for least squares SVM. *IEEE Transactions on Neural Networks & Learning Systems*, vol. 28, no. 9, pp. 2076-2087.
- Joachims, T.** (1999): Making large-scale support vector machine learning practical. *Advances in Kernel Methods*, pp. 169-184.
- Keerthi, S. S.; Shevade, S. K.; Bhattacharyya, C.** (2014): Improvements to Platt's SMO algorithm for SVM classifier design. *Neural Computation*, vol. 13, no. 3, pp. 637-649.
- Kemp, R.; Cottrell, G. A.; Bhadeshia, H. K. D. H.** (2006): Neural-network analysis of irradiation hardening in low-activation steels. *Journal of Nuclear Materials*, vol. 348, no. 3, pp. 311-328.
- Kun, T.; Peijun, D.** (2010): Classification of hyperspectral image based on morphological profiles and multi-kernel SVM. *Hyperspectral Image and Signal Processing: Evolution in Remote Sensing*, pp. 1-4.
- Liu, R.; Liu, E.; Yang, J.** (2006): *Optimizing the Hyper-parameters for SVM by Combining Evolution Strategies with a Grid Search*. Springer Berlin Heidelberg.
- Lucon, E.** (2002): Mechanical tests on two batches of oxide dispersion strengthened RAFM steel (Eurofer97). *Fusion Engineering & Design*, vol. 61, no. 2, pp. 683-689.
- McGill, R.; Tukey, J. W.; Larsen, W. A.** (1978): Variations of box plots. *American Statistician*, vol. 32, no. 1, pp. 12-16.
- Neelamegam, C.; Sapineni, V.; Muthukumar, V.** (2013): Hybrid intelligent modeling for optimizing welding process parameters for reduced activation ferritic-martensitic (RAFM) steel. *Journal of Intelligent Learning Systems & Applications*, vol. 5, no. 1, pp. 39-47.
- Nicewander, W.** (1988): Thirteen ways to look at the correlation coefficient. *American Statistician*, vol. 42, no. 1, pp. 59-66.
- Peng, Z.; Hu, Q.; Dang, J.** (2017): Multi-kernel SVM based depression recognition using social media data. *International Journal of Machine Learning & Cybernetics*, no. 4, pp. 1-15.
- Platt, J.; John, C.** (1998): Sequential minimal optimization: A fast algorithm for training support vector machines. *Advances in Kernel Methods-Support Vector Learning*, pp. 212-223.
- Powell, M. J. D.** (1979): On the estimation of sparse Hessian matrices. *Siam Journal on Numerical Analysis*, vol. 16, no. 6, pp. 1060-1074.
- Ruxton, G. D.** (2010): The unequal variance t-test is an underused alternative to Student's t-test and the Mann-Whitney U test. *Behavioral Ecology*, vol. 17, no. 4, pp. 688-690.
- Singh, R.; Singh, S. K.; Singh, U.; Singh, G. P.** (2008): Bayes estimator of generalized-exponential parameters under LINEX loss function using Lindley's approximation. *Data Science Journal*, vol. 7, pp. 65-75.

- Taylor, R.** (1990): Interpretation of the correlation coefficient: A basic review. *Journal of Diagnostic Medical Sonography*, vol. 6, no. 1, pp. 35-39.
- Virgil'Ev, Y. S.; Kalyagina, I. P.; Makarchenko, V. G.** (1979): Radiation alteration of the properties of graphite over a wide range of irradiation temperature and neutron flux. *Soviet Atomic Energy*, vol. 46, no. 3, pp. 210-213.
- Wang, C.; Zhang, C.; Zhao, J.** (2017): Microstructure evolution and yield strength of clam steel in low irradiation condition. *Materials Science & Engineering A*, vol. 682, pp. 563-568.
- Wang, W. L.; Tang, M. H.** (2015): A normalization process to standardize handwriting data collected from multiple resources for recognition. *Procedia Computer Science*, vol. 61, pp. 402-409.
- Witten, I. H.; Frank, E.** (2011): Data mining: practical machine learning tools and techniques with java implementations. *ACM Sigmod Record*, vol. 31, no. 1, pp. 76-77.
- Ye, H.; Luo, Q.; Li, Y.** (2014): Application of wavelet and multi-kernel svm in uav sensors fault diagnosis. *Electronic Measurement Technology*, vol. 1, pp. 22-30.
- Zeugmann, T.** (2016): *PAC Learning* . Springer.
- Zhang, J.; Zhou, Z.** (2016): An improved genetic algorithm and its applications to the optimisation design of an aspirated compressor profile. *International Journal for Numerical Methods in Fluids*, vol. 79, no. 12, pp. 640-653.
- Zhang, T.** (2004): Statistical behavior and consistency of classification methods based on convex risk minimization. *Annals of Statistics*, vol. 32, no. 1, pp. 56-84.
- Zhu, S.; Cai, J.** (2016): Application of improved affine particle swarm optimization algorithm in parameter estimation. *Journal of Electronic Measurement & Instrumentation*, vol. 12, pp. 1-24.
- Zien, A.; Kramer, N.; Sonnenburg, S.** (2009): The feature importance ranking measure. *Statistics*, vol. 5782, pp. 694-709.

## Appendix A. Statistic information of sample data

Reference Tab. 6.

**Table 6:** Basic information of the input parameters

Index	Input Variable	Minimum	Maximum	Average	Variance	Standard Deviation
1	Processing parameters(%)	0	10	0.0828	0.8836	0.94
2	C (Wt%)	0.087	0.2	0.0971	0.000169	0.013
3	Cr(Wt%)	2.25	12	8.3269	1.054729	1.027
4	W (Wt%)	0	3	1.4843	0.605284	0.778
5	Mo(Wt%)	0	1	0.1592	0.131769	0.363
6	Ta (Wt%)	0	0.54	0.0635	0.010404	0.102
7	V (Wt%)	0	0.3	0.1822	0.002916	0.054
8	Si(Wt%)	0	0.37	0.0546	0.002704	0.052
9	Mn (Wt%)	0	0.13	0.1445	0.041616	0.204
10	N (Wt%)	0	0.06	0.0025	0.006561	0.081
11	Al (Wt%)	0	0.054	0.0008	0.001369	0.037
12	As (Wt%)	0	0.005	0	0.00000009	0.0003
13	B (Wt%)	0	0.0085	0.0007	0.00000169	0.0013
14	Bi (Wt%)	0	0.005	0	0.00000009	0.0003
15	Ce (Wt%)	0	0.036	0.0001	0.00000484	0.0022
16	Co (Wt%)	0	0.01	0.0002	0.00000081	0.0009
17	Cu (Wt%)	0	0.035	0.0006	0.00001024	0.0032
18	Ge (Wt%)	0	1.2	0.0132	0.016384	0.128
19	Mg (Wt%)	0	0.01	0	0.00000036	0.0006
20	Nb (Wt%)	0	0.16	0.00164	0.000121	0.011
21	Ni (Wt%)	0	2	0.0566	0.0961	0.31
22	O (Wt%)	0	0.009	0.0002	0.00000121	0.0011
23	P (Wt%)	0	0.007	0.0013	0.00000196	0.0014
24	Pb (Wt%)	0	0.005	0	0.00000009	0.0003
25	S(Wt%)	0	0.005	0.0012	0.00000121	0.0011
26	Sb(Wt%)	0	0.003	0	0.00000004	0.0002
27	Se(Wt%)	0	0.003	0	0.00000004	0.0002
28	Sn(Wt%)	0	0.003	0	0.00000004	0.0002
29	Te(Wt%)	0	0.005	0	0.00000009	0.0003
30	Ti(Wt%)	0	0.25	0.01	0.002116	0.046
31	Zn(Wt%)	0	0.005	0	0.00000009	0.0003
32	Zr(Wt%)	0	0.059	0.003	0.00001296	0.0036
33	Irradiation temperature(Wt%)	273	925	401.0928	32148.49	179.3
34	Irradiation dose(dpa)	0	90	3.4287	100.8016	10.04
35	Irradiated He(dpa)	0	5000	35.697	129456.04	359.8
36	He_dpa(dpa)	0	6315.7895	56.1696	343119.8151	585.7643
37	Test temperature(k)	123	973	549.78	43848.36	209.4