Visual Abstract

Optimization of S-N curve fitting based on neighborhood rough set reduction with improved firefly algorithm

Yangjinyu Li
Software Technology Institute, Dalian Jiaotong University, Dalian, China
djtulyjy@163.com

Li Zou
Software Technology Institute, Dalian Jiaotong University, Dalian, China
Liaoning Key Laboratory of Welding and Reliability of Rail Transportation Equipment, Dalian Jiaotong University, Dalian, China
Dalian Key Laboratory of Welded Structures and Its Intelligent Manufacturing Technology (IMT) of Rail Transportation Equipment, Dalian Jiaotong University, China
Lizou@djtu.edu.cn

Zhengjie Zhu
Software Technology Institute, Dalian Jiaotong University, Dalian, China
zzj1105031296@163.com

ABSTRACT. In order to reduce the S-N curve dispersion of titanium alloy welded joints and improve the prediction accuracy of fatigue life, a novel optimization method of S-N curve fitting based on neighborhood rough set reduction with improved firefly algorithm (IFANRSR) is proposed. Firstly, we propose an improved firefly algorithm (IFA) by updating the position and step size, combining IFA algorithm and neighborhood rough set into an IFANRSR algorithm for attribute reduction. Then, according to the fatigue data of titanium alloy welded joints, the fatigue decision system of welded joints is established, and the key factors affecting the fatigue life of welded joints are determined. Next, according to the set of key influencing factors obtained based on IFANRSR algorithm, the fatigue characteristics domains are divided, and the S-N curves are fitted on each fatigue characteristics domain, to obtain a group of S-N curves. To demonstrate the effectiveness of IFA algorithm, six benchmark functions are used, then the availability of IFANRSR algorithm is evaluated in comparison with other algorithms on four UCI datasets. Finally, the results of the goodness-of-fit show that the dispersion of fatigue data is reduced, which can effectively improve the prediction accuracy of fatigue life.

KEYWORDS. S-N curve, Fatigue characteristics domain, Neighborhood rough set, Firefly algorithm.

OPEN ACCESS

Citation: Li, Y., Zou, L., Zhu, Z., Optimization of S-N curve fitting based on neighborhood rough set reduction with improved firefly algorithm, Frattura ed Integrità Strutturale, 64 (2023) 250-263.

Received: 02.02.2023
Accepted: 16.03.2023
Online first: 19.03.2023
Published: 01.04.2023

INTRODUCTION

Vast amounts of studies have been done discussing methods for the estimation of the fatigue life of welded joints by domestic and foreign scholars, then a variety of life prediction theories, models, and analysis methods have been put forward [1]. It mainly consists of the local stress-strain method, the life prediction method based on fracture mechanics and the stress-life approach. Among them, the stress-life approach is the most popular life prediction way at present.

Three different mathematical expressions describe the S-N curve in the stress-life approach, often known as the S-N curve method, containing Basquin, Langer, and three parameters life-stress models [2]. According to different basic stress parameters adopted, the S-N curve method can be divided into the nominal stress method [3], the hot spot stress method [4], the notch stress method [5,6], and the nodal force based structural stress method (also known as mesh-insensitive structural stress method) [7,8]. The expression of stress intensity factors including joint type, thickness, loading type, and stress concentration factors are established with fracture mechanics theory. The parameters in the expression are selected to generate an S-N curve that can demonstrate the fatigue life based on substantial fatigue data of welded joints. The fatigue life prediction process of welded joints is based on equivalent structural stress, as shown in Fig. 1.

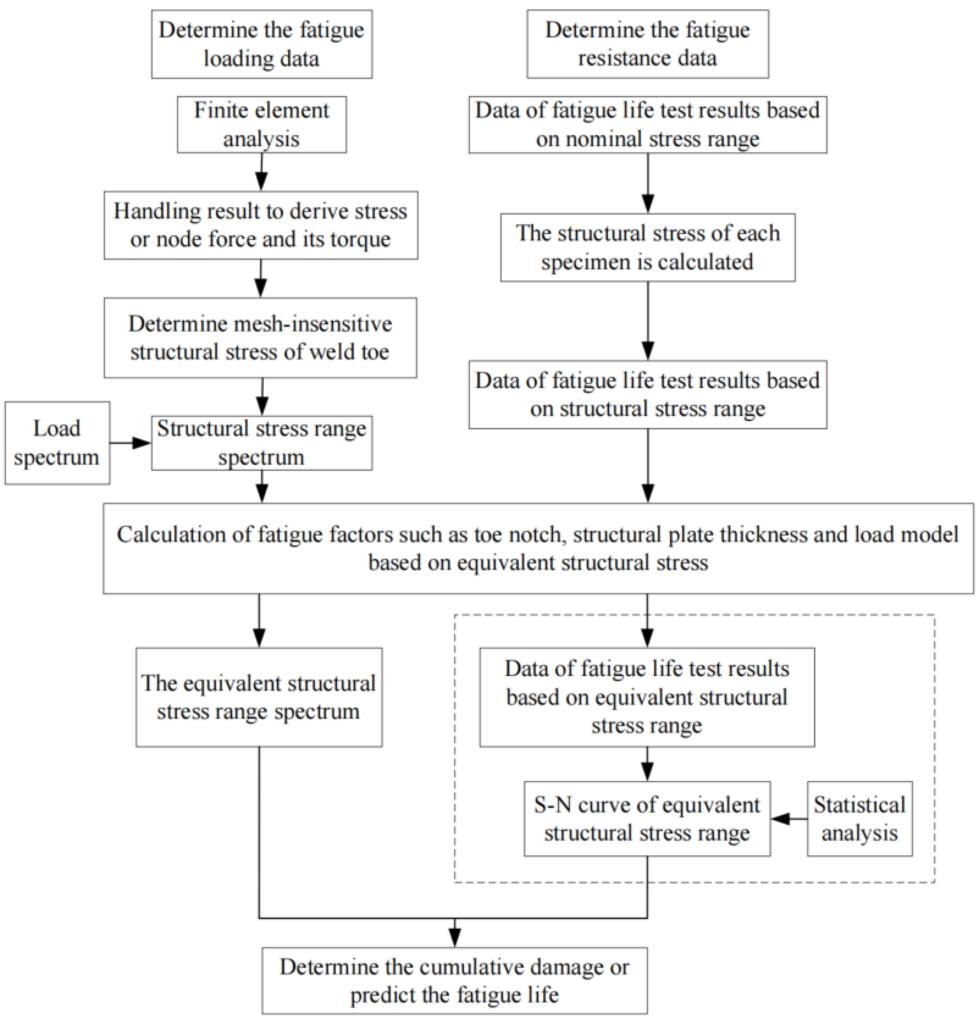

Figure 1: The prediction of fatigue life based on equivalent structural stress.

Since 2007, the equivalent stress method has been adopted by many institutions, and the master S-N curve method based on equivalent structural stress is recommended for weld fatigue analysis in both ASME BPVC VIII-2-2015 standard [9] and API 579-1 standard [10]. At present, the master S-N curve method based on equivalent structural stress has been widely used. Zhou Shaoze et al. [11] proposed method under ultrasonic harmonic resonance, a better structural stress approach to assessing the fatigue life of welded joints, more accurately fitting the master S-N curve. Wang Ping et al. [12] by describing the braking load stress at the welded joints' position and combining the relationship between the S-N curve and the rate of crack growth, to fit a master S-N curve that could uniformly describe the analysis methods of cycle fatigue. Cheng Haigen et al. [13] conducted a comparative analysis between the S-N curve obtained by fatigue test fitting and fe-Safe software simulation technology. The results showed that the software simulation technique is more accurate at predicting the fatigue life of joints. Baoya Cao et al. [14] analyzed two parameters and grid sensitivity based on stress integral, the master S-N curve's applicability is confirmed, and further discussed the effect of the curve's slope on fatigue life under high cycle loading.

Literature analysis shows that, by the joint type, thickness, loading type, and stress concentration factors to carry out equivalent structural stress transformation, because all fatigue data points are packed into a narrow band, the dispersion of S-N data samples of welded joints is significantly reduced. Compared with other methods, for the calculation results and the accuracy of fatigue life, the nodal force based structural stress method can be used to analyze more precisely. We have studied the fatigue life prediction method for aluminum alloy welded joints in the early stage and achieved preliminary research results [15,16]. How to further reduce the scatter of fatigue data and improve the accuracy for fatigue life is a significant issue that needs to be solved, which is very important for the application of the method in engineering.

But in the present study, it is discovered that the dispersion of the fatigue data of titanium alloy welded joints is still relatively high, which result in unsatisfactory fatigue life prediction. At the same time, based on the prediction method for titanium alloy in the early stage, the classification accuracy of the attribute reduction of the neighborhood rough set is not well. Through reading the literature, we find the heuristic algorithm can improved it to make the reduction better. Many scholars have successfully used the heuristic algorithms to solve various optimization problems and applied them in various fields. Thanh Sang-To et al. [17,18] put forward a new Shrimp and Goby Association Search algorithm (SGA), which proved to be effective in avoiding local optimality, and applied it to solve large-scale global optimization problems. Furthermore, they propose The Planet Optimization Algorithm (POA), not only computation time is reduced, but also the accuracy of solving the optimal value is improved. Hoang-Le Minh et al. [19,20] aim at a damage assessment for a high-rise concrete structure, introduce a termite life cycle optimizer (TLCO) algorithm, and proved that can improve the convergence speed and accuracy, that can make a significant improvement in the damage identification of large-scale structures. Van-Thien Tran et al. [21] developed a BCMO-ANN algorithm that combines artificial neural network and balancing composite motion optimization, it can effectively solve the vibration and buckling behaviors optimization problems caused by the uncertainty of material properties.

In my research, a stress-life curve fitting optimization way based on neighborhood rough set reduction with improved firefly algorithm (IFANRSR) is proposed. By IFANRSR algorithm to determine critical influencing factors, the fatigue characteristics domain is delimited based on the critical factors set, and then every domain is matched with S-N curves. Furthermore, the dispersion of fatigue samples can be decreased, and the precision of the prediction of fatigue life is increased.

METHODOLOGY

Neighborhood rough set reduction

Hu Qinghua, by exploring mixed-data sets, in some discovered rough computer modelling and algorithms, the more systematic model of the neighborhood rough set is constructed. These models take the relationship between the rough set theory into the neighborhood space, and support processing discrete and continuous data. Attribute reduction is the key technology of neighborhood rough set. The method is to remove the redundant and irrelevant condition attributes without affecting the classification capability, and the decision system will be made simpler and the data processing efficiency can be increased.

Consider the following definition of a neighborhood decision system: $NDS = \langle U, C, D, V, f, \Delta, \delta \rangle$,where $U = \{\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_\mathrm{n}\}$ consists of a nonempty finite set of items; $C = \{\mathrm{a}_1, \mathrm{a}_2, \ldots, \mathrm{a}_\mathrm{c}\}$ is a set of condition features, D is the decision feature ; in $V = \bigcup_{\mathrm{a} \in \mathrm{C} \cup \mathrm{D}} V_\mathrm{a}$, V_a is a set of property *a*; $f : \mathrm{U} \times \{C \bigcup D\} \to V$ is a mapping function; $\Delta \to [0, \infty]$ is a distance function; δ is a Neighborhood radius parameter, and $(0 \le \delta \le 1)$. The shorthand for the neighborhood decision system is $NDS = \langle U, C, D, \delta \rangle$, in Reference [22] gives the following concepts:

Given a $NDS = \langle U, C, D, \delta \rangle$, for any samples $x, y \in U$, the condition feature subset $B \subseteq C$, Δ_B represents the

Euclidean distance function, then indicates a $NR_\delta(B)$ similarity relation that B decides as follows:

$$NR_\delta(B)=\{(\mathrm{x,y})\in U\times U \mid \Delta_B(\mathrm{x,y})\le\delta\} \tag{1}$$

$$\mathrm{n}_B^\delta(\mathrm{x})=\{\mathrm{y}\in U \mid \Delta_B(\mathrm{x,y})\le\delta\} \tag{2}$$

Given $NDS=\langle U,C,D,\delta\rangle$, for any $B\subseteq C$ and $X\subseteq U$, $x_\mathrm{k}\in U$, Hence, X on B's neighborhood upper and lower approximations are defined as:

$$\overline{B}(X)_\delta=\{\mathrm{x_k}\in U \mid \mathrm{n}_B^\delta(\mathrm{x_k})\cap X\ne\varnothing\} \tag{3}$$

$$\underline{B}(X)_\delta=\{\mathrm{x_k}\in U \mid \mathrm{n}_B^\delta(\mathrm{x_k})\subseteq X\} \tag{4}$$

Given $NDS=\langle U,C,D,\delta\rangle$, for any $B\subseteq C$, $U/D=\{D_1,D_2,\dots D_l\}$, then the neighborhood positive region of decision and neighborhood dependency of D on B are defined as:

$$POS_B^\delta(D)=\sum_{j=1}^{l}\underline{B}(\mathrm{d_j})_\delta \tag{5}$$

$$\gamma_B^\delta(D)=\frac{|POS_B^\delta(D)|}{|U|} \tag{6}$$

where, $D_j\in U/D$ and $j=1,2,\cdots,l$

Firefly algorithm

Firefly algorithm is a meta-heuristic algorithm constructed to imitate the glow behavior of fireflies in nature. In this algorithm, fireflies are attracted to each other by both brightness and attractiveness. In general, the brighter firefly is more attractive, and their ability to attract each other decreases with distance [23].

To put it simply, the firefly algorithm is to locate the brightest firefly in the solution space by its luminous properties, then approach the brightest firefly, further update the position information, and repeat the process until it reaches the optimum value.

The algorithm is described as: suppose the number of fireflies is *n*, and the search dimension is *d*.

1) The position of the individual i is defined as: $x_i=(x_{i1},x_{i2},\cdots,x_{id})$. The objective function value $f(x_i)$ is its brightness.

2) The updating formula of attractiveness between individuals is defined as:

$$\beta_{ij}=\beta_0 e^{-\gamma r_{ij}^2} \tag{7}$$

where, β_0 is the initial attraction, generally, the value is 1, γ is the absorption coefficient of the light intensity. Generally, the value is 1, r_{ij} is the Cartesian distance between individuals:

$$r_{ij}=\sqrt{\sum_{k=1}^{d}(x_{ik}-x_{jk})^2} \tag{8}$$

3) The update formula of the position to which the individual is attracted and moved is:

$$x_i^{k+1} = x_i^k + \beta_{ij}\left(x_j^k - x_i^k\right) + a\varepsilon_i \tag{9}$$

where, k is the iteration times of the algorithm, α is the step, ε_i is random number, which subject to uniform distribution or gaussian distribution.

CONSTRUCTION OF S-N CURVE FITTING OPTIMIZATION MODEL

Compare with the other methods, the mesh-insensitive structural stress approach benefits from accurate calculation outcomes. However, the scattering of fatigue test specimen data expressed by this method is still relatively high, which results in poor precision in predicting fatigue life. It is very important to decrease the dispersity of the fatigue test specimen data and to increase the accuracy of fatigue life prediction.

Processes and steps of the model

To improve the accuracy of fatigue life prediction obtained from the test, we propose an S-N fitting optimization method. Firstly, the fatigue decision system is built using the welded joint fatigue data. Then through the analysis, using IFANRSR algorithm, the evaluation model of key influencing factors is established. Next, the key factors set of obtained by the IFANRSR algorithm are taken as the basis of fatigue characteristics domain division, so that some fatigue data have similar features are dispersed in an area that is largely independent. In this way, the fatigue characteristics domain is obtained and the curve is fitted, and S-N curve fitting based on the fatigue characteristics domain is realized.
An S-N fitting optimization method based on the fatigue characteristics domain is clearly described in FIG. 2:

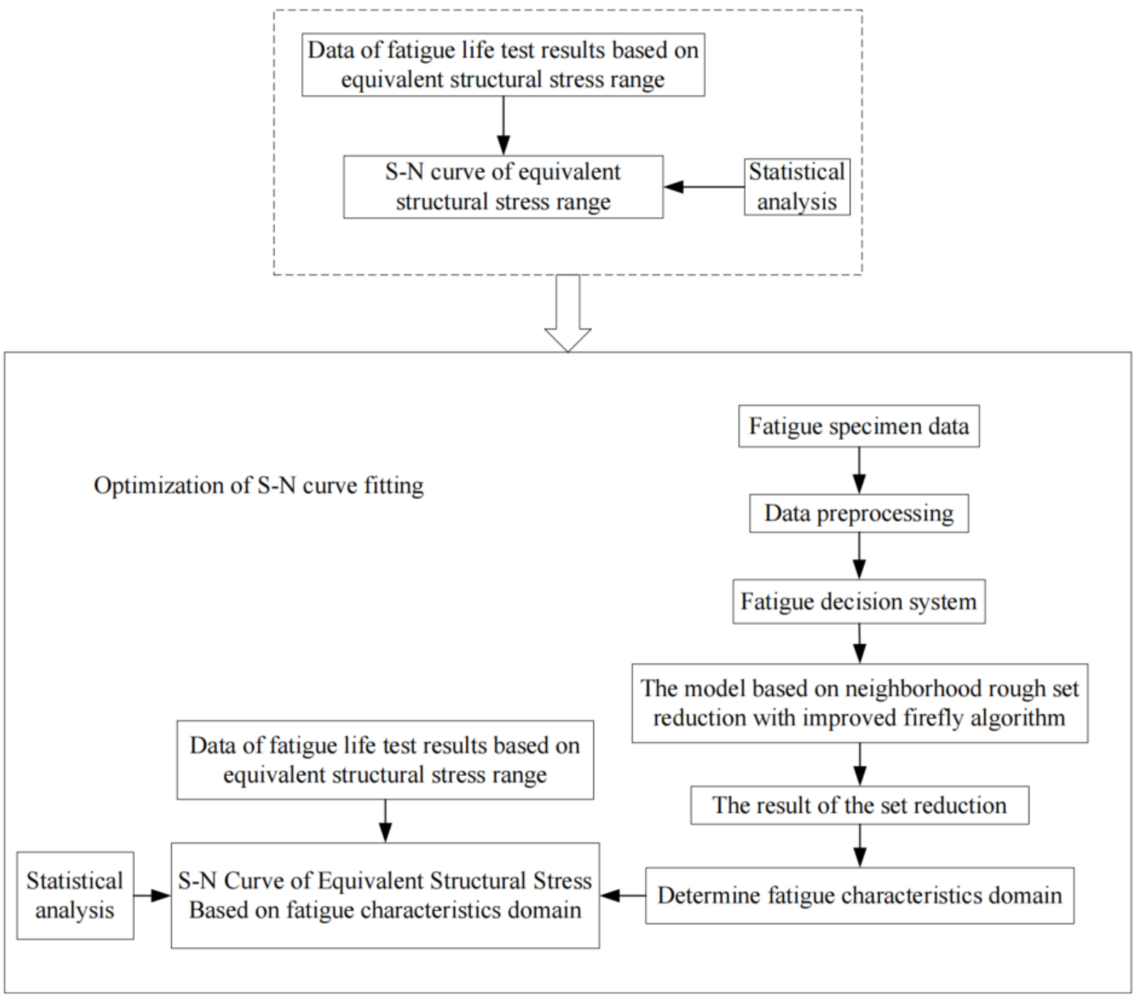

Figure 2: optimization of S-N curve fitting.

The improvement of firefly algorithm

Aiming at the shortcomings of traditional firefly algorithm in global optimization search, such as delayed convergence and fall down to local optimal trapping, an improved firefly algorithm is proposed.

Firstly, since the change of the third random term in Formula (9) is independent and does not use the information of individuals and populations, it is difficult for the changes to match the updates of individuals, and whenever the value is too small or too large, it is not good for the iterative renewal of the populations. Based on the above considerations, a firefly individual representing a higher fitness position within the search attraction range is introduced into the original position updating formula. It depicts the interaction between the population's members and the best-performing members in the current iteration of the population and is used to control the extent of influence of the best members in the current population on other members. The equation for the new position update is expressed as:

$$x_i^{k+1} = x_i^k + \beta_{ij}\left(x_j^k - x_i^k\right) + a\varepsilon_i + rand \times \left(x_n^k - x_i^k\right) \tag{10}$$

Where X_n represents the firefly individual at the highest fitness position in the current iteration.

Secondly, step plays a very important role in the firefly algorithm, setting the appropriate value of step will directly affect the algorithm's search functionality. The standard firefly algorithm adopts a fixed step value, which is not conducive to adjust according to the actual search situation. When the step is too large, the optimal solution cannot be obtained, and when the step is too small, it often gets into the local optimum. Nevertheless, the decreasing step can dynamically adjust the movement amplitude of the individual, so that the individual can conduct the global search with a significant step in the initial iteration and perform local optimization with a non-significant step in the later period. This work calculates the step by decreasing the number of iterations. As shown in Formula (11) :

$$\alpha_{t+1} = \alpha_t \times m \tag{11}$$

$$m = \left(\frac{1}{10}\right)^{\frac{5}{t \times \max imum}} \tag{12}$$

where m is the dynamic attenuation coefficient of the step, and t is the number of iterations.

The neighborhood rough set reduction with improved firefly algorithm

Because different feature reduction algorithms usually get different key factors set, which leads to the different division of fatigue characteristics domain. To obtain the fatigue characteristics domain with a better effect on the curve fitting, IFANRSR algorithm is selected for the feature reduction in this work.

On one side, the neighborhood rough set is suitable for the continuous attribute data, so it can swiftly and easily analyze numerical data in practical problems. Attribute reduction is the key technology of neighborhood rough set. It is to delete redundant and irrelevant condition attributes without affecting their classification ability. It can effectively make the decision system simple, thus improving the speed of data analysis and processing. On the other side, because it needs a lot of computation to get all the feature combinations, the traditional attribute reduction algorithm often cannot obtain a smaller reduction set, we use a meta-heuristic algorithm as the search method to find the best feature subset. Very few parameters, simplicity of implementation, and great global optimization capabilities are all benefits of the firefly algorithm, so it is better to choose the firefly algorithm as the search strategy. So the best feature reduction set based on IFANRSR algorithm can be used for the key factors set.

The flow chart of IFANRSR algorithm is shown in Fig. 3:

The steps of IFANRSR algorithm is as follows:

Step 1: Input the Fatigue decision system of welded joints $NDS = \langle U, C, D, \delta \rangle$. Initialize the parameters of firefly algorithm $n, \beta_0, \gamma, \alpha, k, \varepsilon_i$. Where, n is the count of firefly, β_0 is the initial attraction, γ is the absorption coefficient light intensity, α is the step, k is the iteration times of the firefly algorithm, and ε_i is the random number.

Step 2: Set up the bulletin board. The location and fitness value of the firefly, which stands in for all conditional attributes, serves as the bulletin board's initial value.

Step 3: Random initialized individual position of the firefly x_i, and it is an n-bit binary number. The location of each firefly represents a subset of conditional attributes, so represent the conditional attribute in the reduction set as "1" , and

not the decrease set as "0". Create an $n \times |C|$ dimension 0 matrix representing the positions of all fireflies in the current firefly population. Formula (6) to figure up the attribute dependency. Take the attribute dependency as the objective function of the firefly algorithm. Calculate the value of the firefly target function $f(x_i)$ and sort it, to get the maximum value, that is the location of the firefly have the most brightness.
Step 4: Formula (7) is used to calculate the attractiveness of fireflies in the group.
Step 5: Formula (10) is used to update firefly positions. According to formula (11), the update of iterative steps is carried out. Fireflies tend to have high fitness through searching.
Step 6: Revise the iteration count so that iteration = iteration +1. The count of conditional attributes represented by each firefly increases with the algorithm iteration.
Step 7: Whether the maximum number of iterations has been reached. If not, it carries out the next iteration, performs Step 3; or else, performs Step 8.
Step 8: When the fitness is larger than or equal to the bulletin board records, or the conditional attribute set is less than the bulletin board record. Update bulletin board.
Step 9: Decode the message on the bulletin board, and get the reduction set as the final output.

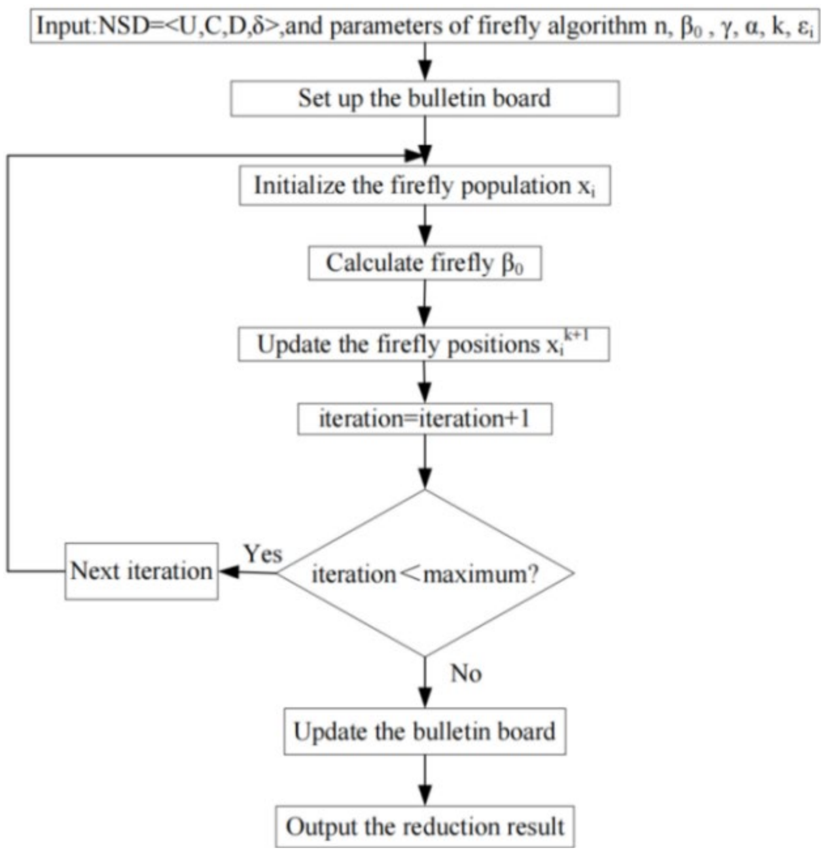

Figure 3: the flow chart of IFANRSR algorithm.

Experiment and discussion

Performance analysis of the improved firefly algorithm

In order to verify the optimization performance of IFA algorithm, six typical benchmark functions in CEC2005 (as shown in Tab. 1) are selected for testing to compare with FA, PSO, GWO, and FPA. Among them, F1~F4 are unimodal functions, which mainly the test local search ability of the algorithm. F5 is the multimodal function, which mainly tests the global search ability of the algorithms. F6 is a fixed dimension function, which mainly tests the balance between the local search ability and the global search ability of the algorithm.
The experiment is designed by MATLAB R2020b and tested on a computer running windows 10 with an Intel Core i5 2.5GHz processor and 4GB memory. In order to ensure the fairness of the experiment, all algorithms are set to

$n = 20, k_{\max} = 300$ and they are run independently 30 times respectively. And the algorithm parameters of both IFA and FA are set to $\beta_0 = 2, \gamma = 1, \alpha = 0.2, \varepsilon_i = 0.98$. The algorithm parameters of PSO [24], GWO [25], and FPA [26] are set to $\omega = 1, c_1 = 1.5, c_2 = 2$; $\alpha_{\max} = 2, a_{\min} = 0, r_1, r_2 \in [0,1]$; $p = 0.8, lamda = 1.5, alpha = 0.01$. The test results of optimal value, worst value, average value, and standard deviation of the algorithm in 30 dimensions on six benchmark functions are shown in Tab. 2.

No.	Function	Formula	Range	Optimal value
F1	Sphere	$f_1(\mathrm{x}) = \sum_{i=1}^{d} x_i^2$	$x_i \in [-100,100]$	0
F2	Rosenbrock	$f_2(x) = \sum_{i=1}^{d} \left[100\left(x_{i+1} - x_i^2\right) + \left(x_i - 1\right)^2\right]$	$x_i \in [-30,30]$	0
F3	Step	$f_3(x) = \sum_{i=1}^{d} \lvert x_i + 0.5 \rvert^2$	$x_i \in [-100,100]$	0
F4	Quartic	$f_4(x) = \sum_{i=1}^{d} i x_i^4 + rand(0,1)$	$x_i \in [-1.28,1.28]$	0
F5	Penalized	$f_5(x) = 0.1\left\{\sin^2(3\pi x_1) + \sum_{i=1}^{d} (x_i - 1)^2 \left[1 + \sin^2(3\pi x_i + 1)\right] + (x_n - 1)^2 \left[1 + \sin^2(2\pi x_n)\right]\right\} + \sum_{i=1}^{n} u(x_i, 5, 100, 4)$	$x_i \in [-50,50]$	0
F6	Kowalik	$f_6(x) = \sum_{i=1}^{d} \left[a_i - \frac{x_1\left(b_i^2 + b_i x_2\right)}{b_i^2 + b_i x_3 + x_4}\right]^2$	$x_i \in [-5,5]$	0.0003

Table 1: Information of the benchmark function.

As can be seen from the statistical results in Tab. 2, for F2, F3, and F6, IFA algorithm is superior to the other four algorithms in terms of optimal value, worst value, average value, and variance. Especially on F3, the optimal value of IFA is improved by more than 6 exponential levels compared with other algorithms. For F1, the overall performance of IFA is a little bit worse than GWO, but the optimal value is 7.65E-10 higher than that of GWO. At the same time, compared with other algorithms, it improves several index levels. For F4, IFA is optimal except that the worst value of F4 is slightly worse than GWO 0.07. For multimodal function F5, although the performance of IFA is not ideal, it can be seen that compared with FA algorithm, it is still greatly improved, in which the average value is reduced by more than 20.
According to the optimal value and average value of the algorithm, it can be seen that the optimization ability of IFA algorithm is a great improvement over FA algorithm on both unimodal functions and multimodal functions, and is mostly better than the other four algorithms. Therefore, IFA has high optimization accuracy. According to the data analysis of standard deviation, IFA is the best among the five algorithms except for F1. So IFA has a stable iterative process. To sum up, IFA can overcome the precocious convergence problem effectively, enhance the global search and local search ability, and improve the optimization accuracy, which verifies the effectiveness of the proposed improved algorithm.

Performance analysis of the neighborhood rough set reduction with improved firefly algorithm

In this work, the parameter of the three algorithms are set to $n = 20, \beta_0 = 2, \gamma = 1, \alpha = 0.2, \varepsilon_i = 0.98, k_{\max} = 20, \delta = 0.1$. Using LIBSVM as a tool, the 10-Ford method is introduced to obtain the classification accuracy of the reduction set. Open source datasets from UCI: Wine, Lymphography, Heart, and Zoo (as shown in Tab. 3) are used as test datasets. Compared with the attribute reduction algorithm for forward greedy search (FARNeMF) and neighborhood rough set reduction with the firefly algorithm (FANRSR) on four datasets. And the reduction results, classification accuracy results of the algorithm, and standard deviation are obtained by independently repeating 20 experiments, as shown in Tab. 4.

No.	Algorithm	Best	Worst	Mean	STD
F1	FA	1.82E-05	0.23	3.51E-02	6.16E-02
	PSO	1.24E-02	0.36	6.98E-02	7.09E-02
	GWO	9.98E-10	2.97E-06	2.21E-07	6.47E-07
	FPA	3.97E+03	9.79E+03	6.77E+03	1.41E+03
	IFA	2.33E-10	4.10E-04	2.54E-05	8.94E-05
F2	FA	26.43	5.67E+02	1.30E+02	1.55E+02
	PSO	32.52	1.68E+03	2.76E+02	3.71E+02
	GWO	26.33	2.60E+02	39.41	50.72
	FPA	2.18E+06	6.49E+06	3.67E+06	1.24E+06
	IFA	10.40	2.04E+02	38.21	41.23
F3	FA	3.75E-04	0.17	2.57E-02	4.22E-02
	PSO	1.17E-02	0.54	0.10	50.7
	GWO	0.75	2.23	1.45	0.47
	FPA	4.06E+03	1.20E+04	7.55E+03	1.77E+03
	IFA	2.99E-10	2.69E-02	1.82E-03	3.71E+02
F4	FA	2.48E-02	0.98	0.13	0.21
	PSO	0.19	1.16	0.62	0.29
	GWO	1.92E-03	0.12	1.68E-02	3.14E-02
	FPA	0.54	4.32	2.10	1.09
	IFA	2.54E-03	0.19	6.65E-02	4.68E-02
F5	FA	2.22E-04	73.8	52.13	21.28
	PSO	1.81E-02	3.05	0.77	0.29
	GWO	0.68	1.66	1.13	0.25
	FPA	2.71E+05	2.44E+07	7.21E+06	7.03E+06
	IFA	1.10E-02	58.42	29.50	4.68E-02
F6	FA	1.67E-03	2.06E-02	4.11E-03	6.93E-03
	PSO	7.86E-04	1.78E-02	2.79E-03	4.62E-03
	GWO	3.98E-04	2.04E-02	5.55E-03	8.56E-03
	FPA	7.61E-04	5.79E-03	2.45E-03	1.26E-03
	IFA	3.70E-04	1.09E-03	6.87E-04	1.62E-04

Table 2: Comparison of calculation results.

Data set	Sample	Feature	Classes
Zoo	101	16	7
Wine	178	13	3
Heart	270	13	2
Lymphography	148	18	4

Table 3: Information of the data set

Data set	Algorithm	Reducts	Accuracy		
			Best	Mean	STD
Zoo	IFANRSR	4,6,8,12,13,14	90.18	83.72	2.72
	FANRSR	4,6,8,12,13,14	87.09	80.30	2.42
	FARNeMF	4,6,8,12,13	79.36	77.45	0.42
Wine	IFANRSR	1,7,11,13	98.33	94.72	1.10
	FANRSR	1,2,11,13	96.57	89.21	2.23
	FARNeMF	1,2,13	81.99	80.06	1.45
Heart	IFANRSR	1,3,5,7,8,13	88.89	81.74	1.34
	FANRSR	1,3,5,7,13	86.59	79.70	2.02
	FARNeMF	1,4,5	71.48	69.07	1.74
Lymphography	IFANRSR	2,3,10,13,14,15,18	93.33	81.21	4.57
	FANRSR	2,3,13,14,15,18	86.67	77.46	3.68
	FARNeMF	2,13,14,15,16,18	73.05	69.20	2.87

Table 4: Experimental results of three algorithms.

In the experiment of attribute reduction, two evaluation criteria are very important, namely reduction results and classification accuracy. The classification ability improves with increased classification accuracy. Standard deviation is an important standard to evaluate the stability of a set of data while ensuring the accuracy of classification. The smaller the standard deviation is, the higher the stability will be. Consequently, it is possible to effectively assess the reduction set's classification performance. In Tab. 4, the optimum and mean value of the reduction accuracy obtained from the IFANRSR algorithm are more significant than the FARNeMF and FANRSR algorithm. The standard deviation of IFANRSR is the smallest in the Wine and Heart datasets, and it is smaller than FANRSR in the Lymphography datasets. So the classification accuracy of IFANRSR algorithm is better, and it's more stable and more suitable for attribute reduction.

Establishment of fatigue database

According to the fatigue test of reference [27], five different kinds of welded joints are used in the tensile and fatigue experiment. The stress ratio applied to the specimen is R=0 and the Frequency of the cycle is 5*Hz*. The welded joints are made by manual TIG welding, and a JIS Z3331 YTB35 welding rod with a 1.6 or 2.4*mm* diameter is used. A fatigue life calculation method based on equivalent structural stress range is proposed according to the reference [28]. A fatigue analysis database is established using fatigue data gathered from the titanium alloy welded joints obtained in the above experiments. A total of 44 specimens are collected in this database and the S-N curve is fitted according to this. Continuous data includes thickness, nominal stress range, Eq. structural stress range, and SFC; discrete data includes a joint type. There are two plate thicknesses, including 2 and 10mm; SFC is the stress concentration factor. SFC has SCF Membrane, SCF Bend, and SFC Total; five joint types are LT, CT, CB, CT, and LL. Due to space constraints, only a portion of experimental data is displayed in Tab. 5.

Thickness (*mm*)	SCF Membrane	SCF Bend	SFC Total	Joint type	Nominal Stress Range (*MPa*)	Eq. Structural Stress Range (*MPa*)	Life Cycles
2	1	0.4	1.4	LT	167	258	675006
2	1	0.4	1.4	CT	150	231	4213944
2	1.9	0.7	2.8	LL	114	354	447552
10	1	0.4	1.4	LT	102	206	2233310
10	1	0	1	CB	204	299	7946640
10	1	0.4	1.4	CT	143	383	833690
10	1.5	0.7	2.2	LL	87	279	833690
				......			

Table 5: Partial fatigue data

Features extraction based on the neighborhood rough set reduction with improved firefly algorithm

The IFANRSR is utilized in this study for attribute reduction in order to identify the key factors affecting the fatigue life of titanium alloy welded joints. Following that, fatigue characteristics domains are separated based on these major factors and fit the S-N curve in their respective fields.

Based on fatigue test data of titanium alloy, a decision system of influencing factors of fatigue life is established. a is used to represent each attribute, then the set of conditional attributes be expressed as: C={{ a1 - Thickness }, { a2 - SFC Membrane }, { a3 - SFC bend }, { a4 - SFC }, { a5 - Joint type }, { a6 - Eq. Structural Stress Range }}, D ={ lg(N) }, the specific is shown in Tab. 6. Where, N is the fatigue life.

The reduction result set is: { a1{Thickness}, a5 {Joint type}, a6 {Eq. Structural Stress Range}}. It shows that compared with other condition attributes, the effect of thickness, joint type, and Eq. Structural Stress Range on the fatigue life is larger. For this reason, the reduction results are regarded as the critical influence nature for the fatigue life of titanium alloy welded joints.

In this work, Twenty independent replicates are performed. Among them, the optimal accuracy is 100%, and the average accuracy is 86.5%, indicating that the IFANRSR algorithm has a good classification effect on this data. The variance is 11.5, which is relatively small. It shows that IFANRSR algorithm has high stability.

Fitting the S-N Curves and statistical analysis

The S-N curve is described by three arithmetic expressions, comprising the Basquin, Langer, and three parameter stress-life equations. In this study, the S-N curve of equivalent structural stress is fitted using the least squares method based on the Basquin equation.

U	Condition attributes						Decision attribution
	C1	*C2*	*C3*	*C4*	*C5*	*C6*	*D*
1	2	1	0.4	1.4	LT	258	675006
2	2	1	0.4	1.4	CT	231	4213944
3	2	1.9	0.7	2.8	LL	354	447552
4	10	1	0.4	1.4	LT	206	2233310
5	10	1	0	1	CB	299	7946640
6	10	1	0.4	1.4	CT	383	833690
7	10	1.5	0.7	2.2	LL	279	833690
			......				

Table 6: Part of the decision table

Basquin equation is as follows:

$$S^m N = C \tag{13}$$

where m and C are the coefficients determined by different materials. Then by taking the logarithm of both sides of this equation and get:

$$\lg N + m \lg S = \lg C \tag{14}$$

The S-N curve based on Eq. SS Range as shown below.

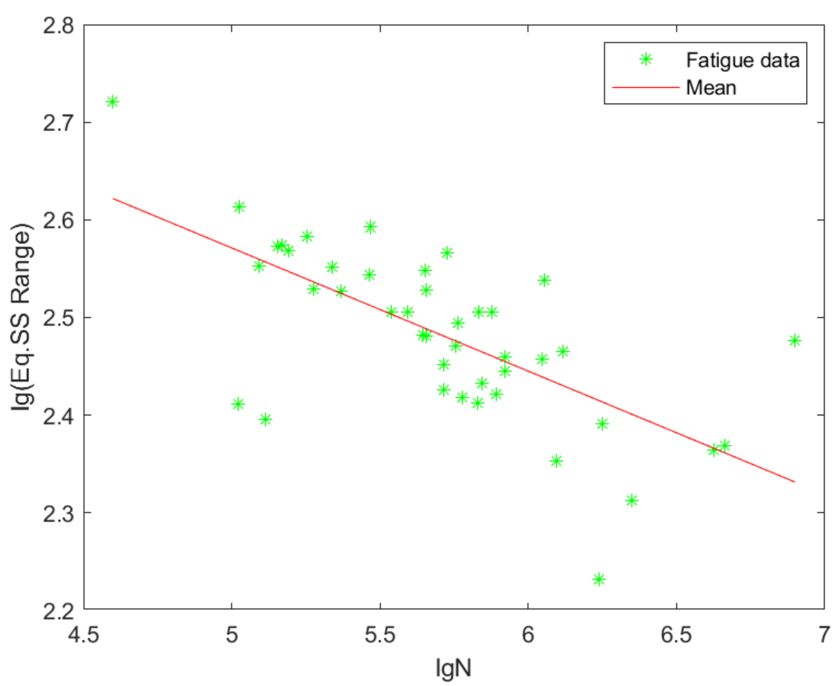

Figure 4: S-N curve based on Eq. SS Range.

The three forms of the S-N curve are fitted using statistical analysis, and the statistical parameters of goodness-of-fit are mostly made up of SSE, R-Square, adjusted R-Square, and RMSE.
Where SSE is the sum of error squares. The smaller the SSE, the smaller the error, and the better the model effect; its optimal value is 0. It is defined as:

$$SSE=\sum_{i=1}^{n}\omega_i\left(y_i-\hat{y}_i\right)^2 \tag{15}$$

SSR is the sum of squares of the regression, it is defined as:

$$SSR=\sum_{i=1}^{n}\omega_i\left(\hat{y}_i-\bar{y}\right)^2 \tag{16}$$

SST is the total sum of squares, it is defined as:

$$SST=\sum_{i=1}^{n}\omega_i\left(y_i-\bar{y}\right)^2 \tag{17}$$

R-square is the coefficient of determination, and it varies from 0 to 1. It mostly depicts the accuracy of the fitting curve through the alteration of data. When it is closer to 1, indicates that the data with a better fitting effect on the model. It is defined as:

$$\mathrm{R}-square=\frac{SSR}{SST}=1-\frac{SSE}{SST} \tag{18}$$

The *adjusted R-square* is the adjusted coefficient of determination. The *adjusted R-square* can obtain anything with a value of 1 or smaller. Generally, when the value of the *adjusted R-square* is close to 1, indicates that the data with a better fitting effect on the model, and is defined as:

$$adjusted\ \mathrm{R}-square=1-\frac{SSE\left(n-1\right)}{SST\left(v\right)} \tag{19}$$

RMSE is the fit standard deviation of the regression coefficient, when the model fitting effect is better, its value is 0. It is defined as:

$$RMES=\sqrt{MSE}-\sqrt{\frac{SSE}{v}} \tag{20}$$

S-N Curve Based on Fatigue Characteristics Domain

Compared with other methods, the nodal force based structural stress method can shrink the dispersion of the fatigue specimen data. However, the numerical goodness-of-fit value is still not very good in terms of design, which will result in poor fatigue life forecast accuracy. In order to improve the problem, the notion of fatigue characteristics domain is developed, to decrease the dispersion level and standard deviation of fatigue samples, and to provide a more precise test foundation for the later assessment of fatigue life. Then, S-N curves are fitted according to the different domains obtained.
In this work, the data in Tab. 4. is used as the experimental data set. IFANRSR algorithm is adopted to reduce the attribution of the fatigue decision system. Based on the reduction result, seven fatigue characteristics domains can be defined: $S1$:$\{X\in U|X_{C1=2}$ and $X_{C5=LT}\}$; $S2$:$\{X\in U|X_{C1=2}$ and $X_{C5=CT}\}$; $S3$:$\{X\in U|X_{C1=2}$ and $X_{C5=LL}\}$; $S4$:$\{X\in U|X_{C1=10}$ and $X_{C5=LT}\}$; $S5$:$\{X\in U|X_{C1=10}$ and $X_{C5=CB}\}$; $S6$:$\{X\in U|X_{C1=10}$ and $X_{C5=CT}\}$; $S7$:$\{X\in U|X_{C1=10}$ and $X_{C5=LL}\}$.

Fig. 5 shows S-N curves based on each fatigue characteristic domain, and their respective goodness-of-fit statistics are shown in Tab. 7.

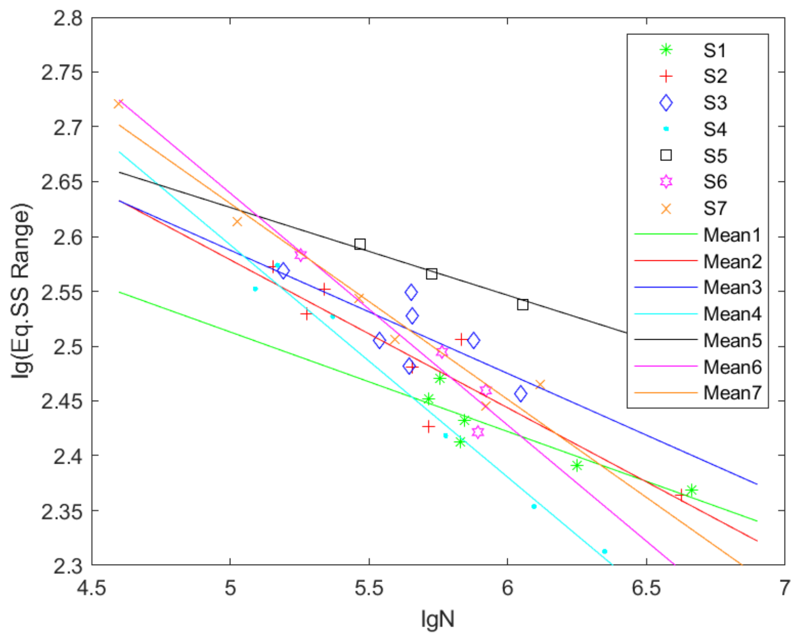

Figure 5: S-N curves based on fatigue characteristics domain.

	Mean	*Mean1*	*Mean2*	*Mean3*	*Mean4*	*Mean5*	*Mean6*	*Mean7*
SSE	0.1842	0.0015	0.0054	0.0034	0.0011	4.324e-0.5	0.0014	0.0025
R-square	0.4522	0.7978	0.8319	0.6176	0.9819	0.9943	0.9053	0.9540
Adjusted R-square	0.4385	0.7472	0.7983	0.5412	0.9774	0.9914	0.8580	0.9425
RMSE	0.0669	0.0191	0.0329	0.0260	0.0165	0.0046	0.0260	0.0248

Table 7: Goodness-of-fit statistics of Mean1-Mean7.

From Tab. 7 that according to the statistical analysis results of goodness-of-fit: the *SSE* form Mean in the whole domain is 0.1842, and the SSE depending on the fatigue characteristics domain from Mean1 to Mean7 is all lower than in the entire domain. From Mean1 to Mean7, each R-square and adjusted R-square approaches 1 rather than the Mean. The *RMSE* from Mean1 to Mean7 are all reduced by at least half than that of Mean, and they are closer to 0 than the Mean.
As little *SSE* and *RMSE* as possible, as larger *R-square* and *adjusted R-square* as possible show that separating fatigue characteristics domain improves the accuracy of the fatigue life prediction, and further reduced the scattering rate of fatigue data. Therefore, attribute reduction is carried out by the IFANRSR algorithm, and then fatigue characteristics domain is divided, and the S-N curve fitting of the fatigue design is realized.

CONCLUSION

In the whole fatigue life analysis of titanium alloy welded joints, the design and fitting of the S-N curve have become a key technical link. To reduce the dispersion and standard deviation of fatigue data, the prediction accuracy of fatigue life can be improved. This work mainly proposes an S-N curve fitting optimization method based on IFANRSR algorithm.

In this work, firstly, the fatigue decision system of welded joints is constructed, then the fatigue data set of titanium alloy welded joints is processed by IFANRSR algorithm to obtain the key influencing factors. IFANRSR algorithm combines IFA with neighborhood rough set for attribute reduction. Because the classification accuracy of neighborhood rough set attribute reduction is not high enough, a heuristic algorithm is used to improve it. FA algorithm has the advantages of fewer parameters, easy implementation, and strong global optimization ability, so it is selected as the search strategy. However, FA algorithm also has some disadvantages such as slow convergence speed and easy falling into local optimization. Given these shortcomings, IFA algorithm is proposed by improving position update and step size and then used in the IFANRSR algorithm. Next, the fatigue characteristics domain is divided according to the key influencing factors. Finally, seven S-N curves are fitted according to the fatigue characteristics domain divided. According to the statistical analysis results, the fatigue calculation results will be more accurate by adopting the S-N curve fitting optimization method based on IFANRSR algorithm. It can effectively reduce the dispersion of fatigue data, to improve the accuracy of fatigue life prediction.

ACKNOWLEDGMENTS

This research was supported by the National Science Foundation of China under Grant (52005071).

REFERENCE

[1] Zhao, W.Z., Li, X.W. and Dong, P.S. (2017). Fatigue Resistance Design Theory and Method of Welded Structure. Welding Technology, 46(08), pp. 64. DOI:10.13846/j.cnki.cn12-1070/tg.2017.08.036.
[2] Fang, H.Y. (2008). Welding Structure. Beijing, China Machine Press.
[3] Hobbacher, A. F. (2016). Recommendations for Fatigue Design of Welded Joints and Components. Springer, Cham. DOI: 10.1007/978-3-319-23757-2.
[4] Wang, W.J., Bai, J.Y. and Liu, W. (2017). Fatigue assessment of weld structure based on the hot spot stress method. Journal of Beijing Jiaotong University, 41(6), pp. 82-87. (in Chinese)
[5] Liu, X., Zhang, K.L., Yao, Y. and Wang, X.P. (2016). Fatigue analysis of welded joints based on notch stress method. Engineering Mechanics, 33(6), pp. 209-214+249. (in Chinese)
[6] Zhou, Z.Y., Wang, Y.Z. and Yang, X. (2018). Fatigue cumulative damage assessment of longitudinal fillet welded gusset in welded frame based on different stress approaches. Transactions of the China welding institution, 39(08), pp. 18-22+129-130. (in Chinese)
[7] Zhao, F., Li, A.Q., Li, W.R. and Sheng, S. (2017). Wind-Induced fatigue analysis of high-rise steel structures using equivalent structural stress method. Applied Sciences, 7(1). DOI: 10.3390/app7010071.
[8] Alencar, G., Hong, J.K., de Jesus, A., da Silva, J.G.S. and Calçada, R. (2021). The master S-N curve approach for fatigue assessment of welded bridge structural details. International Journal of Fatigue, pp. 106432. DOI: 10.1016/J.IJFATIGUE.2021.106432.
[9] ASME. (2015). BPVC VIII-2-2015 ASME Boiler and Pressure Vessel Code. The American Society of Mechanical Engineers, New York.
[10] API. (2007). API 579-1/ASME FFS-1 2007 Fitness-for-Service. The American Petroleum Institute, Washington.
[11] Zhou, S.Z., Guo, S., Chen, B.Z. and Zhao, W.Z. (2022). Master S-N curve fitting and life prediction method for very high cycle fatigue of welded structures. Transactions of the China Welding Institution, 43(05), pp. 76-82+118. (in Chinese)
[12] Wang, P., Pei, X.J., Qing, H.L. and Dong, P.S. (2021). Unique fatigue design method of welded structures and application. Journal of Mechanical Engineering, 57(16), pp. 349-360. (in Chinese)
[13] Cheng, H.G., Hu, C. and Jiang, Y. (2021). Numerical simulation and experimental verification of fatigue life of Q420D high-strength steel transverse-cruciform weld joint. Materials for Mechanical Engineering, 45(12), pp. 78-83+89. (in Chinese)
[14] Cao, B.Y., Ding, Y.L., Fang, Z. and Song, Y.S. (2019). Influence of weld parameters on the fatigue life of deck-rib welding details in orthotropic steel decks based on the improved stress integration approach. Applied Sciences, 9(18). DOI: 10.3390/app9183917.

[15] Zou, L., Ren, S.Y., Li, H.X., Yang, X.H. An optimization of master S-N curve fitting method based on improved neighborhood rough set, IEEE ACCESS, Volume 9, pp. 8404-8420. DOI: 10.1109/ACCESS.2021.3049403.
[16] Zou, L., Ren, S.Y., Yang, G. and Yang, X.H. (2021). Analysis of factors affecting fatigue life of welded joints based on improved conditional neighborhood entropy. Transactions of the China Welding Institution, 42(11), pp. 43-50+99-100. (in Chinese)
[17] Sang-To, T., Hoang-Le, M., Wahab, M.A. and Cuong-Le, T. (2023). A new metaheuristic algorithm: Shrimp and Goby association search algorithm and its application for damage identification in large-scale and complex structures. Advances in Engineering Software, 176, DOI: 10.1016/J.ADVENGSOFT.2022.103363.
[18] Sang-To, T., Hoang-Le, M., Wahab, M.A. and Cuong-Le, T. (2022). An efficient Planet Optimization Algorithm for solving engineering problems. Scientific Reports, 12:8362, DOI: https://doi.org/10.1038/s41598-022-12030-w.
[19] Hoang-Le, M., Sang-To, T., Theraulaz, G., Wahab, M.A., and Cuong-Le, T. (2023). Termite life cycle optimizer. Expert Systems With Applications, 213(PC), DOI: 10.1016/J.ESWA.2022.119211.
[20] Hoang-Le, M., Sang-To, T., Khatir, S., Wahab, M.A. and Cuong-Le, T. (2023). Damage identification in high-rise concrete structures using a bio-inspired meta-heuristic optimization algorithm. Advances in Engineering Software, 176, 103399, DOI: 10.1016/j.advengsoft.2022.103399.
[21] Van-Thien, T., Trung-Kien, N., Nguyen-Xuan H. and Wahab, M.A. (2023). Vibration and buckling optimization of functionally graded porous microplates using BCMO-ANN algorithm. Thin-Walled Structures, 182(PB), DOI: 10.1016/J.TWS.2022.110267.
[22] Hu, Q.H., Yu, D., Liu, J.F. and Wu, C.X. (2018). Neighborhood rough set based heterogeneous feature subset selection. Information Sciences, 178, pp. 3577-3594. DOI: 10.1016/j.ins.2008.05.024.
[23] Li, J., Wei, X.Y., Li, B. and Zeng, Z.G. (2022). A survey on firefly algorithms. Neurocomputing, 500, pp. 662-678. DOI: 10.1016/J.NEUCOM.2022.05.100.
[24] Kennedy, J. and Eberhart, R. (1995). Particle swarm optimization. IEEE International Conference on Neural Networks, 1942-1948.
[25] Joshi, H. and Arora, S. (2017). Enhanced Grey Wolf Optimization Algorithm for Global Optimization. Fundamenta Informaticae, 153(3), pp. 235-264. DOI: 10.3233/FI-2017-1539.
[26] Yang, X.S. (2013). Flower Pollination Algorithm for Global Optimization. International Conference on Unconventional Computing and Natural Computation, abs/1312.5673.
[27] Iwata, T. and Matsuoka, K. (2004). Fatigue strength of CP grade 2 titanium fillet welded joint for Ship Structure. Welding in the World, 48(7-8), pp. 40-47. DOI: 10.1007/BF03266442.
[28] Dong, P., Prager, M. and Osage, D. (2007). The design master S-N curve in ASME Div 2 rewrite and its validations. Welding in the World, 51(5-6), pp. 53-63. DOI: 10.1007/BF03266573.