Received 6 June 2020; | Revised xx June 2020; | Accepted xx June 2020

DOI: xxx/xxxx

ARTICLE TYPE

SVR Prediction Algorithm for Crack Propagation of Aviation Aluminum Alloy†

Zhihang Wang[1] | Jincai Chang*[2] | Qingyu Zhu[3]

[1]College of Sciences, North China University of Science and Technology, Zhihang Wang, Tangshan, China
[2]College of Sciences, North China University of Science and Technology, Jincai Chang, Tangshan, China
[3]Avic China Aero-Polytechnology Establishment, Qingyu Zhu, Beijing, China

Correspondence
*Jincai Chang, College of Sciences, North China University of Science and Technology, Tangshan, 063210, China.
Email: jincai@ncst.edu.cn

Present Address
College of Sciences, North China University of Science and Technology, Tangshan, 063210, China

Summary

Aluminum alloy materials is an important component material in the safe flight of aircraft. It is very important and necessary to predict the fatigue crack growth between holes of aviation aluminum alloy materials. At present, the investigation on the prediction of the cracks between two holes and multi-holes is a key problem to be solved. Due to the fatigue crack growth test of aluminum alloy plate with two or three holes was carried out by MTS fatigue testing machine, the crack length growth data under different test conditions were obtained. In this paper, support vector regression (SVR) was used to fit the crack data, and the parameters of SVR are optimized by grid search algorithm at the same time. And then the model of SVR to predict the crack length was established. Discussion on the results show that the prediction model is effective. Furthermore, the crack growth between three holes were predicted accurately through the model of the crack law between two holes under the same load form.

KEYWORDS:
Crack growth prediction, Data analysis, Support Vector Regression (SVR), Scikit-learn machine learning

1 | INTRODUCTION

As we all know, the aluminum alloy materials studied in this paper, as an important component material in the aviation field, is widely used in the aviation industry[1], and plays an important role in ensuring safety of the aviation aircraft flight. Therefore, we should attach great importance to carry out relevant research work on it.

The investigation on fatigue crack growth prediction has a history of several decades: as early as the middle of the 19th century, Wohler, a German railway engineer, put forward the concept of stress-life ($S-N$) curve and fatigue limit, and pointed out the influence degree of factors affecting materials fatigue. After that, some researchers developed Wohler's investigation from 1870 to 1890[2,3,4,5]. After decades of development, there are many researches on fatigue crack propagation, which are micro investigation on materials and mechanism analysis by finite element method.Paris et al.[6] put forward Paris formula to express crack growth law on the basis of fracture mechanics method, which is the most widely used method in engineering. Besides, Fathi A. Alshma et al.[7] confirmed all parameters of Paris Law in the experiment, and found the crack growth speed through the experiment and analysis.On the basis of previous studies, the investigation on crack growth and fracture is in the ascendant in recent years[8,9,10,11,12], and the main progress is described as follows: Beibei LEI[13] takes 2024-T4 aluminum alloy as the investigate object, and uses the finite element method to investigate the influence rule and mechanism of overload condition on the change trend of subsequent fatigue crack growth rate (da/dN).For the problem of crack propagation in a plate with a single

†The supports from the National Natural Science Foundation of China (Nos. 51674121 and Nos. 61702184), the Tangshan Team Project (No. 18130209B) and Aviation Fund Project (No. 2017ZD41006) are gratefully acknowledged.

[0]**Abbreviations:** SVR, Support vector regression

hole central crack, Shaoqin Zhang et al.[14] introduced a new Z fracture criterion which can well predict the crack propagating direction of mode I crack in carbon-fiber reinforced composite laminates, and proposed new concepts of in-plane average strain, in-plane dilatational strain energy density factor and reciprocal characteristic function.Furthermore, M Hajimohamadi et al.[15] investigated the analytical solution of stress field and stress intensity in infinite plane with elliptical holes with unequal length of prefabricated cracks. For the crack growth problem of two holes plate, R. R. Bhargava et al.[16] proposed a mathematical model of two unequal-collinear crack growth.Additionally, S Singh et al.[17] proposed an improved strip saturation model based on the combination of two internal electric saturation bands and studied two-dimensional (2D) arbitrarily polarized semi permeable dielectric analytically. For the crack growth problem of porous plate, Jinfang Zhao[18] et al. applied the basic principle of complex stress function method and its approximate superposition method to the solution of stress intensity factor of collinear multiple cracks in infinite plates, and made a preliminary exploration on the solution of porous cracks. Moreover, Zhenghong Li et al.[19] used Eshelby inclusion theory and weight function method to give the approximate analytical solution of stress intensity factor of typical porous multi-crack problem, and combined with Paris crack growth formula to predict the fatigue crack growth life. These investigations on the prediction of fatigue crack growth are based on empirical formula, analytical method and finite element method. Based on the existing mechanism investigation and crack growth performance, the analytical expression of crack growth law is obtained. In view of the fact that the dealing with the actual aviation crack data which is an urgent proposition to find as way to find a way to get the crack growth law by combining the crack growth mechanism and data-driven method, the intelligent algorithm model of aviation fatigue crack growth based on the data-driven method will be built in the focus of this paper.

On the other hand, support vector regression (SVR) algorithm is an extension of support vector machine algorithm in regression problem, which was first proposed by Drucker et al.[20] After years of development, SVR algorithm is widely used in all aspects of scientific investigation. The application of SVR algorithm in crack growth and life prediction are as follows. Based on grey theory and support vector regression method, Dalian Yang et al.[21] proposed GMSVR model and parameter optimization method of artificial bee colony, and applied it to FCG prediction of 7075 aluminum alloy. Furthermore, Weizhen Song et al.[22] used XFEM and SVR to predict fatigue life of plate cracks. Therefore, it is feasible to apply SVR algorithm to the crack growth prediction of aluminum alloy plate, which has the advantage of solving such problems.

Specially, support vector regression (SVR)[23,24] algorithm is a data-driven based method, which is used to fit the crack growth data obtained from fatigue load test in this paper. Under the same load form, a model which can predict the crack rule between three circular holes by the crack rule between two holes is established, and its fitting effect is tested. The aviation aluminum alloy plate with two or three holes is the test materials of fatigue load test. First of all, the data of crack length and cycle number of plate crack growth under constant amplitude load and variable amplitude load are collected in this experiment. After that, to get the general rule of crack growth, data exploration, data pre-processing and so on are carried out. Finally, the applicable support vector regression algorithm is used to fit the crack growth data to get the prediction results of the model and evaluate the effect of the model. This model can be used in the field of fatigue crack growth prediction of porous edge of aluminum alloy materials for aircraft. It can predict the crack growth length by the number of cycles, and can also predict the number of cycles by the crack growth length through the model. The establishment of crack prediction model for aviation aluminum alloy plate can predict three holes crack rule from two-hole crack and solve the prediction problem of unknown crack growth rule between holes. It is a breakthrough of crack prediction between holes based on data-driven method. This model not only provides the basis for the prediction of the crack law of aviation aluminum alloy materials, but also can be applied to the prediction of safe flight and life of aviation aircraft, and guide the prediction of the crack growth law between holes of actual aviation aluminum alloy plate.

2 | TEST CONTENT

2.1 | Test introduction

In this section, to investigate the crack growth law of aviation aluminum alloy plate with two holes and three holes under constant amplitude load and variable amplitude load respectively is one purpose of this experimental investigation. And on the other hand, to use the suitable support vector regression algorithm to fit and analyze the experimental data is another purpose of this experimental investigation. Besides, MTS fatigue testing machine system is used in this test, which is used in fatigue test and crack growth test of typical aviation connection structure with multiple cracks. Moreover, The crack growth behavior was measured by high power optical microscope. Figures 1-2 show the MTS fatigue testing machine system:

FIGURE 1 MTS fatigue testing machine system.

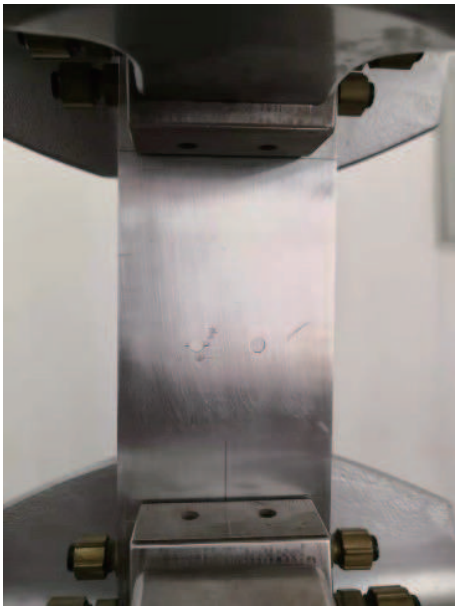

FIGURE 2 Test diagram of two holes plate specimen.

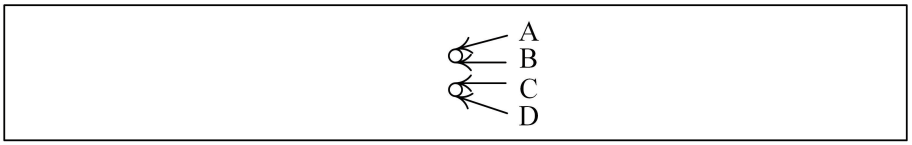

FIGURE 3 Schematic diagram of two holes plate specimen.

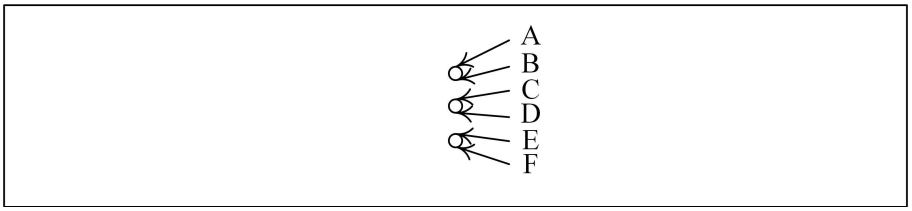

FIGURE 4 Schematic diagram of three holes plate specimen.

In addition, the most commonly used aluminum lithium alloy on the active advanced aircraft is the experimental materials used in this test. In order to explore the propagation law of cracks between holes in aviation aluminum alloy plate, the aviation aluminum alloy plate is divided into two holes and three holes. Besides, the schematic diagram of the two types of aviation aluminum alloy plates used is shown in Figures 3-4:

Besides, the diameter of the small circular holes in the aluminum alloy plate specimens are 4mm, while the distance between points B and C in the figure and the distance between points D and E are 12mm, which is shown in Figure 3 and Figure 4. The parameters of fatigue load test are shown in Table 1:

For points A, B, C, D, E and F in Figures 3-4, preset the initial crack length respectively, so as to obtain the crack growth in the test.

Constant amplitude load and variable amplitude load are the two categories that the style of the pre-set cracks accorded to the test purpose. Among them, there are 11 constant amplitude plate specimens and 9 variable amplitude plate specimens. As

TABLE 1 Table of fatigue load test parameters.

	Load range (Rad/s)[a]	Median load (Rad/s)[b]	Load amplitude	Variable load median	Amplitude of load variation
Constant amplitude load	1.3KN–13KN	7.15KN	5.85KN	–	–
Variable amplitude load	1.3KN–13KN	7.15KN	5.85KN	14.3KN	11.7KN

TABLE 2 Initial pre-set cracks of two holes specimen with constant amplitude.

Serial number	A	B	C	D
1	1	1	0	0
2	1	1.5	0	0
3	1	1	1	1
4	1	1.5	1	1
5	1	1.5	1.5	1
6	0.5	1.5	0.5	0.5
7	0.5	1.5	1.5	0.5
8	0.5	0.5	0.5	0.5
9	0.5	1.5	0	0
10	0.5	1.5	1.5	0.5
11	0.5	0.5	0	0

TABLE 3 Initial pre-set cracks of three holes specimen with constant amplitude.

Serial number	A	B	C	D	E	F
18	0.5	0.5	1.5	1.5	0.5	0.5

shown in the Table 2 and Table 3, the statistics of crack pre-set of plate specimen under constant amplitude load (the crack length refers to the length of crack propagation with the initial crack tip as the origin, unit: *mm*):

It can be seen that the Table 4 and Table 5 (unit: *mm*) is shown the statistics of the pre-set cracks of the plate specimen under variable amplitude load:

The vertical line of point B and C is taken as the axis of symmetry for two holes aluminum alloy plate in this paper. It can be seen that point A corresponds to point D and point B corresponds to point C. If the initial crack length of four points is the same, it is regarded as the case of initial crack symmetry. Similarly, the vertical line of point C and point D is taken as the axis of symmetry for three holes aluminum alloy plate. It can be seen that point A corresponds to point F, point B corresponds to point E, and point C corresponds to point D. In the case of the above initial crack configuration, the type of the initial crack length at each hole edge of the two-hole aluminum alloy plate is involved in this paper. In this way, the random configuration of the initial crack can be considered. It can provide the basis for the follow-up investigation work through the above analysis.

TABLE 4 Initial pre-set cracks of two holes specimen with variable amplitude.

Serial number	A	B	C	D
1	0.5	0.5	0	0
2	0.5	1.5	0	0
3	0.5	0.5	0.5	0.5
4	0.5	1.5	0.5	0.5
5	0.5	1.5	1.5	0.5

TABLE 5 Initial pre-set cracks of three holes specimen with variable amplitude.

Serial number	A	B	C	D	E	F
9	0.5	0.5	1.5	0.5	0.5	0.5

TABLE 6 Partial test data of constant amplitude two holes test specimen 03.

	A	B	C	D
13000	0.6	0.9	0.9	0.9
13200	0.6	0.9	0.9	1
13400	0.7	1	1.1	1
13600	0.8	1	1.1	1.05
13800	0.8	1.2	1.2	1.05
		...		
18900	5	4.7	5	5.6
18925	5.2	4.7	5	5.9
18950	5.5	4.7	5	6.2
18975	5.7	4.7	5	6.6
19000	6	4.7	5	7

2.2 | Test process

Firstly, the above-mentioned constant amplitude fatigue load spectrum and variable amplitude fatigue load spectrum (see Table 1) which are suitable for laboratory use are adopted. Secondly, the fatigue crack growth test of porous plate specimen under constant amplitude load spectrum is carried out. The fatigue crack growth data of multiple cracks are collected, and the growth rule is analyzed. Finally, the fatigue crack growth test of plate specimen with porous edge cracks under variable amplitude load is carried out. Therefore, the data of crack length and number of cycles are collected to provide test data for the prediction of crack length. According to the previous solution, it also can even establish the life of fatigue multi crack propagation law.

2.3 | Experimental data

In this section, the test data of crack growth length and cycle times of 20 groups of aluminum alloy specimens are obtained, which are saved in 20 Excel files respectively through the above fatigue loading test. Some original data of crack growth for the above-mentioned aluminum alloy specimens are presented in this paper. The initial cracks at points A, B, C and D in the constant amplitude two hole test specimen 03 are all 1.0 mm test data, some of which are shown in Table 6 (the data in the first column of the table is the number of cycles, and the rest are in *mm*):

TABLE 7 Partial test data of constant amplitude three holes test specimen 18.

	A	B	C	D	E	F
5000	0	0	0	0	0	0
28000	0	0	0	0	0	0
28500	0	0	0	0.1	0	0
29000	0	0	0	0.2	0	0
29500	0	0	0	0.3	0	0
			...			
64250	3.4	2	8	7.2	2.8	6.6
64500	3.7	2	8	7.2	2.8	7.5
64750	4	2	8	7.2	2.8	7.6
64973	6.2	2	8	7.2	2.8	10.2
64976	17.2	2	8	7.2	2.8	16.8

TABLE 8 Partial test data of variable amplitude two holes test specimen 03 (Complement).

	A	B	C	D	
11000	0	0.1	0.1	0.1	
11500	0	0.1	0.1	0.1	
12000	0.1	0.1	0.1	0.1	
12500	0.1	0.1	0.1	0.1	
13000	0.1	0.2	0.2	0.1	
		...			
15750	0.5	0.5	0.6	0.6	
16000	0.6	0.5	0.6	0.6	Added a variant
16500	0.6	0.5	0.6	0.6	
		...			
42100	6.3	4.8	6.2	7.2	
42200	6.5	4.8	6.2	7.5	
42300	7	4.8	6.2	7.8	
42400	8	4.8	6.2	9	
42448	17.1	4.8	6.2	18.2	

The initial crack configuration of the constant amplitude three-hole specimen 18 is as follows: the initial crack at point A, B, E and F is 0.5mm and the initial crack at point C and point D is 1.5mm. Some of the data are shown in Table 7 (the data in the first column of the table is the number of cycles, and the rest are in *mm*):

However, the initial cracks at points A, B, C and D in the two hole variable amplitude test specimen 03(Complement) are all 0.5 mm test data, some of which are shown in Table 8 (the data in the first column of the table is the number of cycles, and the rest are in *mm*):

The initial crack configuration of variable amplitude three-hole specimen 09 is as follows: the initial crack at point A, B, D, E, F is 0.5mm and the initial crack at point C is 1.5mm. Some of the data are shown in Table 9 (the data in the first column of the table is the number of cycles, and the rest are in *mm*):

TABLE 9 Partial test data of variable amplitude three holes test specimen 09.

	A	B	C	D	E	F	
5000	0	0	0	0	0	0	
10000	0	0	0	0	0	0	
15000	0	0	0	0	0	0	
20000	0	0	0	0	0	0	
25000	0	0	0	0	0	0	
			...				
46000	0	0.1	1.8	1.1	0.2	0	
47000	0	0.1	1.9	1.2	0.2	0	Added a variant
48000	0	0.1	1.9	1.2	0.2	0	
			...				
132500	6.3	2.8	7.2	8.7	2.3	5.3	
133000	8	2.8	7.2	8.7	2.3	5.9	
133250	9.1	2.8	7.2	8.7	2.3	6.7	
133500	10.5	2.8	7.2	8.7	2.3	7.7	
133532	17.3	2.8	7.2	8.7	2.3	18	

3 | THE THEORY OF SUPPORT VECTOR REGRESSION

As we know, support vector regression (SVR) is a development of support vector machine algorithm, which was first extended to support vector regression by Drucker et al.[20] Support vector regression is a small sample learning method with solid theoretical basis. And the traditional process from induction to deduction can be avoid by it. By this means, SVR simplified greatly the usual regression problem. Support vector regression algorithm avoids "dimension disaster" in a sense, and has good robustness. The main principles of SVR are as follows:

Given training samples $T = \{(x_1, y_1), (x_2, y_2), ..., (x_m, y_m)\}$, $y_i \in \mathbb{R}$. The regression model shaped as $f(x) = w^T x + b$ is the learning objective function. Moreover, the model parameters which make $f(x)$ as close as possible to y, w and b are to be determined.

Next, support vector regression[23] assumes that we can tolerate the maximum deviation of ε between $f(x)$ and y. That is to say, the loss is only calculated when the absolute value of the difference between $f(x)$ and y is greater than ε. As shown in Figure 5, this is equivalent to building a spacing band (i.e., the part sandwiched between the two dashed lines in the figure) with a width of ε centered on $f(x)$ (i.e., the solid line in the figure). Whereas if the training samples fall into this interval, they are considered to be correctly predicted.

Therefore, the SVR problem can be formalized as

$$\min_{w,b} \frac{1}{2}\|w\|^2 + b \tag{1}$$

Where, $b = C\sum_{i=1}^{m} l_\varepsilon(f(x_i) - y_i)$, C is the regularization constant. It can be seen from Fig.6 that, l_ε is the ε-insensitive loss function shown in Figure 6.

$$l_\varepsilon(z) = \begin{cases} 0, & \text{if } |z| \le \varepsilon; \\ |z| - \varepsilon, & \text{otherwise.} \end{cases} \tag{2}$$

By introducing relaxation variables ξ_i and $\hat{\xi}_i$, equation (1) can be rewritten as

$$\min_{w,b,\xi_i,\hat{\xi}_i} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m}(\xi_i + \hat{\xi}_i) \tag{3}$$

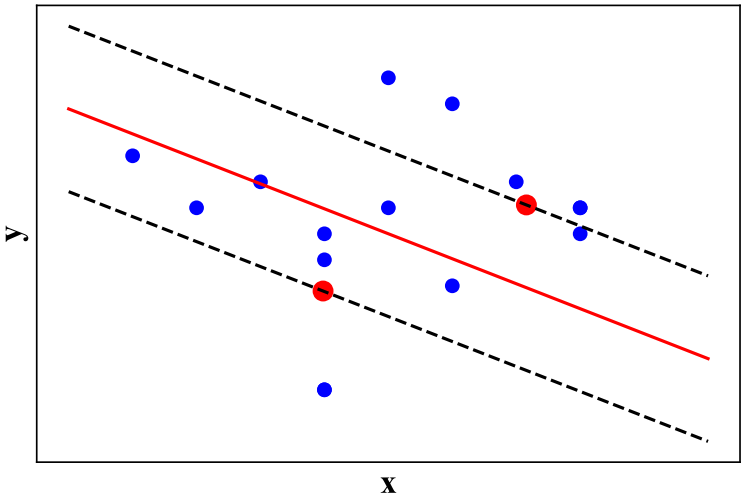

FIGURE 5 Schematic diagram of support vector regression.

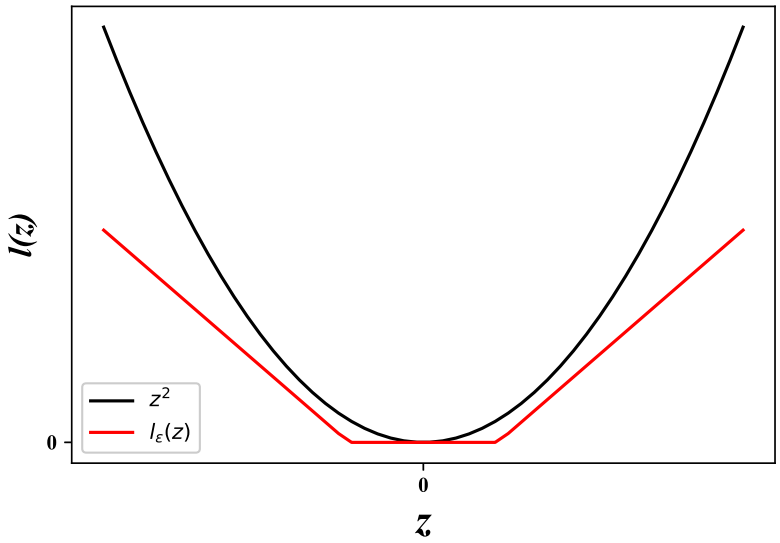

FIGURE 6 ε-insensitive loss function.

$$s.t.\ \ f(x_i) - y_i \leq \varepsilon + \xi_i,$$

$$y_i - f(x_i) \leq \varepsilon + \hat{\xi}_i,$$

$$\xi_i \geq 0, \hat{\xi}_i \geq 0, i = 1, 2, ..., m.$$

Therefore, the solution of SVR can be obtained by Lagrange multiplier method as follows

$$f(x) = \sum_{i=1}^{m} (\hat{\alpha}_i - \alpha_i) x_i^T x + b \tag{4}$$

Furthermore, the samples of $(\hat{\alpha}_i - \alpha_i) \neq 0$ in Equation (4) are the support vectors of SVR, which must fall outside the ϵ -interval. Obviously, SVR's support vector is only a part of the training sample, that is, its solution is still sparse.

If the feature mapping form $f(x) = w^T \phi(x) + b$ is considered, then the corresponding solution can be obtained by Lagrange multiplier method

$$w = \sum_{i=1}^{m} (\hat{\alpha}_i - \alpha_i) \phi(x_i) \tag{5}$$

Substituting Equation (5) into $f(x) = w^T \phi(x) + b$, SVR can be expressed as

$$f(x) = \sum_{i=1}^{m} (\hat{\alpha}_i - \alpha) \kappa(x, x_i) + b \tag{6}$$

Where, $\kappa(x_i, x_j) = \phi(x_i)^T \phi(x_j)$ represent the kernel function. The kernel function used in this paper is Gauss kernel function, whose expression is $\kappa(x_i, x_j) = \exp(-\frac{\|x_i - x_j\|^2}{2\sigma^2})$. Where, $\sigma > 0$ stand for the width of Gaussian kernel.

4 | PREDICTION OF CRACK LENGTH BY SVR

4.1 | Data standardization

As we all know, the input data to be normalized is required in SVR algorithm. Therefore, it is necessary to normalize the data to be processed. The data to be used needs to be determined before normalization. According to the data exploration in the early stage, we can know from the visual graph that the crack length data is in the state of non expansion when the number of cycles is small. That is, a large number of zero values in the original data have no impact on the prediction results, so it can be discarded. After a certain length that the crack length reaches, the crack length remains unchanged when the number of cycles in the later period is large. Therefore, this part of the data can be discarded and only part of the data with changing rules can be retained. The influence of point A and point D of two-hole plate and point A and point F of three-hole plate on interpore crack propagation is ignored in this test. Therefore, the following investigate contents only consider the crack growth of between holes. We also add the initial crack length of each crack to each row of crack data, so that the crack data has a certain physical significance.

For convenience, standardization methods in data standardization as follows:

Transform sequence $x_1, x_2, ..., x_n : x_i^* = \frac{x_i - \min\limits_{1 \le j \le n} x_j}{\max\limits_{1 \le j \le n} x_j - \min\limits_{1 \le j \le n} x_j}$, the new sequence $x_1^*, x_2^*, ..., x_n^* \in [0, 1]$ is dimensionless. The data can be standardized first for data preprocessing.

Using the Matplotlib Library in Python to draw the image and get the relevant statistics for each test specimen. Data preprocessing images of between holes cracks of the above specimen are presented in this paper. Each image drawn below does not consider the zero values data in the front part of the data file, and the number of cycles is processed by the above normalization method. The abscissa is the number of cycles after normalization, and the ordinate is the crack length in mm. So as expected shown in Figures 7-10 for the specific image:

After the above standardized treatment, the crack length is treated as follows: The crack data of point B and point C are added to get the cumulative crack length, which is recorded as BCI, and the cumulative crack length is about 12 mm for the two holes crack specimen. Meanwhile, the crack data of points B, C, D and E are added and record them as BCDEI to get the cumulative crack length, which is about 25 mm for the three holes crack specimen.

After the above-mentioned processing, the pre-processing image of crack growth data of each test piece is obtained. The abscissa is the number of cycles, and the ordinate is the sum of crack length between holes BCI or BCDEI data which unit is mm in all the following figures. See Figures 11-14 for the specific image:

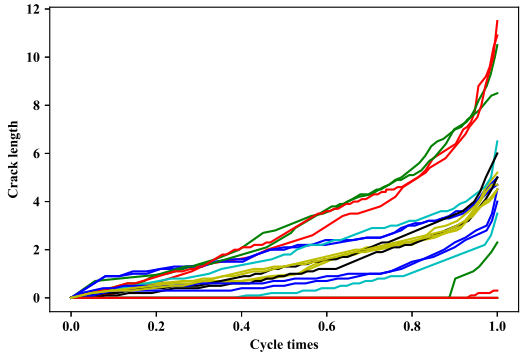

FIGURE 7 Normalized data of two holes with constant amplitude.

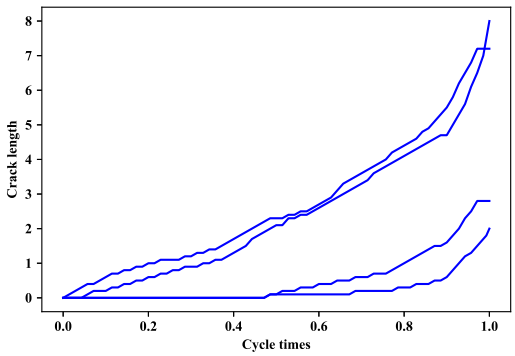

FIGURE 8 Normalized data of three holes with constant amplitude 18.

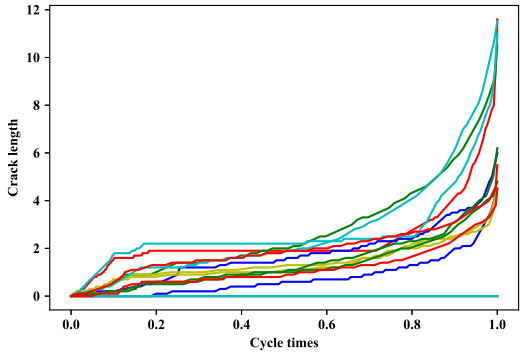

FIGURE 9 Normalized data of two holes variable amplitude.

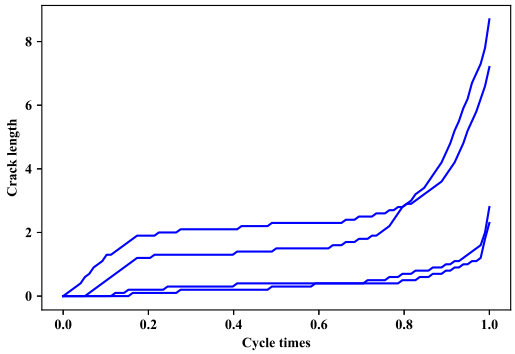

FIGURE 10 Normalized data of three holes variable amplitude 09.

4.2 | Prediction steps of SVR

Obviously, the steps of using support vector regression to predict the crack length are as follows: first, establish the support vector regression model for the normalized BCI crack data of the constant amplitude two holes specimen, obtain the training and testing errors, and then test the effect of the model on the constant amplitude two holes specimen. Then, support vector regression model was established and predicted for BCDEI crack data of normalized constant amplitude three holes specimen, observe the error and deviation verify the prediction effect of the model. Finally, the prediction model is established for the variable amplitude specimen similar to the above prediction steps.

The SVR algorithm in scikit-learn is used to build the SVR model for crack prediction of aviation aluminum alloy plate by adjusting the parameters such as *kernel* (i.e., specify the kernel type to be used in the algorithm), *gamma* (i.e., for kernel coefficient) and *C* (i.e., for penalty parameter *C* of error term) in this paper.

Furthermore, this paper adopts the grid search optimization algorithm when adjusting the parameters. Grid search optimization algorithm is a kind of exhaustive search algorithm which optimizes the parameters by cross validation, and then obtains the optimal learning algorithm. That is to say, to traverse and search the corresponding permutation and combination values of each parameter in a given range, and each group of combination results constitutes a "grid". Then, each combination is applied to the training learning algorithm, and the evaluation results are obtained by cross validation. After traversing and training all the parameter combinations, the grid search algorithm will automatically return the best parameter combination with the highest score, and its corresponding learning model is the optimal regression model.

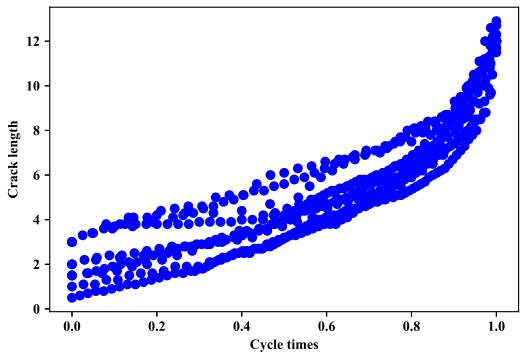

FIGURE 11 Cumulative length of two holes with constant amplitude.

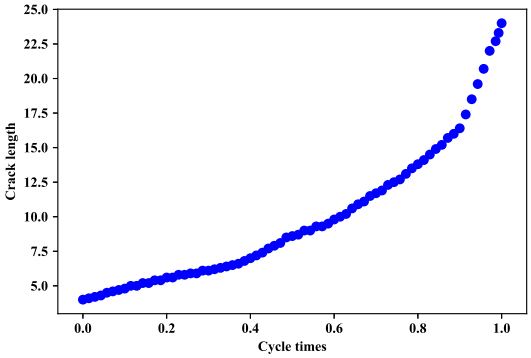

FIGURE 12 Cumulative length of three holes with constant amplitude 18.

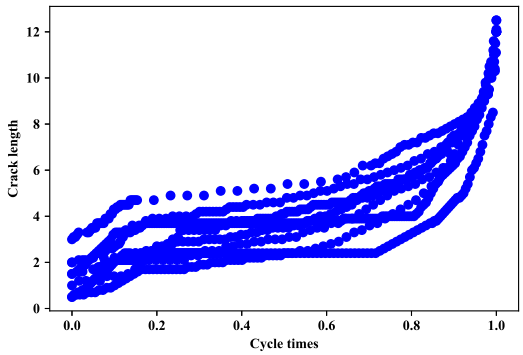

FIGURE 13 Cumulative length of two holes with variable amplitude.

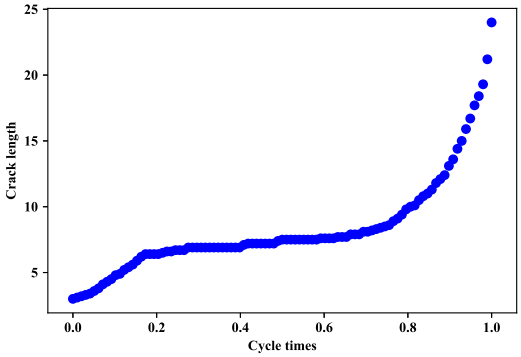

FIGURE 14 Cumulative length of three holes with variable amplitude 09.

It is known that the kernel, *gamma* and C parameters need to be adjusted in SVR. For kernel, i.e., the type of the kernel function, we specify to use radial basis kernel function. For *gamma* and C parameters, we use GridSearchCV method in Scikit-learn to optimize the grid search parameters and get the optimal parameters in this paper. Where, the value range of *gamma* is set $\{0.01, 0.1, 1, 10, 100\}$, that is $gamma \in \{0.01, 0.1, 1, 10, 100\}$, the value range of C is set $\{1, 10, 100, 1000\}$, that is $C \in \{1, 10, 100, 1000\}$. According to the range of *gamma* and C from the above, there are 20 grid points in this method.

4.3 | Evaluating indicator

Data set partition standard: For the division of data sets under different load spectra, this paper uses two-hole data under the corresponding load spectrum as training data and three-hole data as test data.

The commonly used cross validation method in machine learning is used for the aspect of model evaluation in data set partition. In view of the fact that the half-fold cross validation method[25] is specifically used in this paper. The reliable evaluation of the model effect is obtained by making full use of the data set, so as to verify the performance of the established SVR model for crack prediction of aviation aluminum alloy plate.

Additionally, error analysis indicators: mean square error[26], root mean square error[27] and other error analysis indicators are commonly used in regression problems. Variance, standard deviation, mean square error and root mean square error are the error evaluation indicators used in this paper. The calculation formulas of each indicator are as follows:

Variance (S^2):

$$S^2 = \frac{1}{m}\sum_{i=1}^{m}(y_i - \bar{y})^2 \tag{7}$$

Standard deviation (S):

$$S = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(y_i - \bar{y})^2} \tag{8}$$

Mean Square Error (MSE):

$$MSE = \frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2 \tag{9}$$

Root Mean Square Error ($RMSE$):

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2} \tag{10}$$

Where, m, y_i, $\hat{y}_i$ and $\bar{y}$ are the total number of samples, the predicted value of the model, the real value, the sample mean value, respectively.

It is significant to compare variance, standard deviation with MSE and $RMSE$. For example, if the MSE of prediction error is almost equal to the variance of the target (or $RMSE$ is almost equal to the standard deviation of the target), this indicates that the prediction algorithm is not effective. By simply averaging the target value to replace the prediction algorithm, almost the same effect can be achieved. If the prediction error $RMSE$ is about half of the actual target standard deviation, this is already a pretty good performance. Because the variance and MSE comparison and the standard deviation and $RMSE$ comparison have a certain correlation. And thus, the comparison between standard deviation and $RMSE$ is the main error indicator used in this paper.

4.4 | Result analysis

According to the previous solution, this paper presents a visual image of the results of predicting the crack length of three holes aluminum alloy plate specimens under the same load spectrum. In the visualization image, Figure 15 and Figure 17 are the images of test data and prediction data. Besides, Figure 16 and Figure 18 are the learning curve images. In the images in Figure 15 and Figure 17, the red line is the line graph of test data, and the blue line is the line graph of support vector regression prediction data.

However, through the optimization of SVR parameters by grid search algorithm, we get the optimal SVR model for aluminum alloy plate under constant amplitude. The parameters of it are as follows: SVR (C=100.0, $cache\_size$=200, $coef0$=0.0, $degree$=3, $epsilon$=0.1, $gamma$=10.0, $kernel$="RBF", max_iter=-1, $shrinking$=true, tol=0.001, $verbose$=false). Among them,
$kernel$="RBF" is the setting parameters, C=100.0 and $gamma$=10.0 are the optimization parameters of grid search algorithm. And other parameters are the default values. A simple interpretation can be given as follows.

Figures 15-16 show the results of constant amplitude two holes specimen 18:

The standard deviation is about 10.11, the $RMSE$ of this model is about 0.64 which is far less than the standard deviation. It is easily found that the effect of this model is very well. The results of half-fold cross validation are: 0.99942329, 0.99940067, 0.9994369, 0.9991129, 0.9994331, the mean value is: 0.9993613737263342, which also shows that the prediction effect of the model is good. In this way, the model can be used for this problem.

Moreover, through the optimization of SVR parameters by grid search algorithm, we get the optimal SVR model for aluminum alloy plate under variable amplitude. The parameters of it are as follows: SVR (C=1000.0, $cache\_size$=200, $coef0$=0.0, $degree$=3, $epsilon$=0.1, $gamma$=1.0, $kernel$="RBF", max_iter=-1, $shrinking$=true, tol=0.001, $verbose$=false). Among them, $kernel$="RBF" is the setting parameters, C=1000.0 and $gamma$=1.0 are the optimization parameters of grid search algorithm. And other parameters are the default values. A simple interpretation can be given as follows.

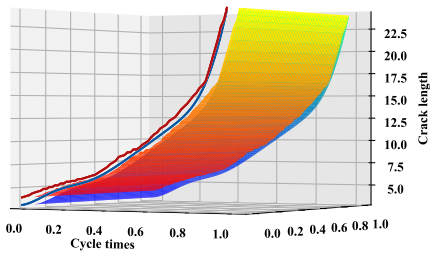

FIGURE 15 Prediction of crack length with constant amplitude 18 by SVR.

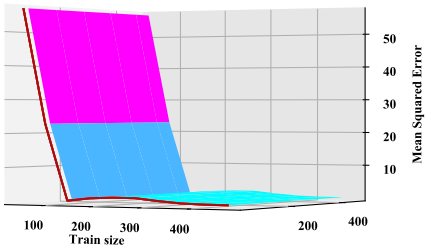

FIGURE 16 Learning curve of constant amplitude 18.

Figures 17-18 show the results of constant amplitude three holes specimen 09:

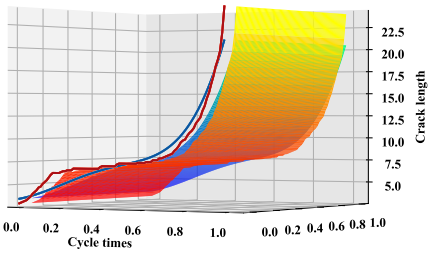

FIGURE 17 Prediction of crack length with variable amplitude 09 by SVR.

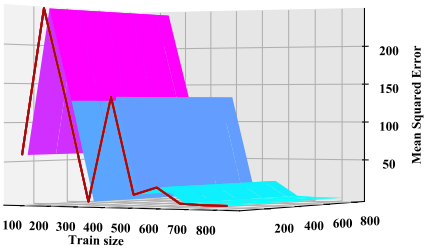

FIGURE 18 Learning curve of variable amplitude 09.

The standard deviation is about 9.56, the $RMSE$ of this model is about 0.86 which is far less than the standard deviation. It is easily found that the effect of this model is very well. The results of half-fold cross validation are: 0.99930434, 0.99930836, 0.99873057, 0.999125, 0.9991911, with the mean value of 0.9991318738985194, which also shows that the prediction effect of the model is good. In this way, the model can be used for this problem.

Additionally, through the establishment, implementation and evaluation of the above models, it can be seen that the constant amplitude data and the variable amplitude data are two different types of models. The crack propagation law between the holes of two-hole aluminum alloy plate and three-hole aluminum alloy plate has roughly the same trend in each type of model. Therefore, three holes crack data can be predicted by two holes crack data under the same load form. In practical application, we can refer to the crack configuration form of the specimen in this paper, find a crack configuration form model similar to the actual situation. In addition, we can get the crack growth prediction results through the model.

5 | CONCLUSIONS

The SVR model can be used to predict the crack between holes in similar aluminum alloy plates. Using different models under different conditions can make the prediction more accurate: use the model with higher score in constant amplitude model under constant amplitude condition; use the model with higher score in variable amplitude model under constant amplitude condition. According to the results, we can know that two holes and three holes aluminum alloy plate have roughly the same crack propagation law under the same load spectrum. It leads us to predict the crack law between three holes with constant amplitude by the crack law between two holes with constant amplitude. And at the same time, we also can predict the crack law between three holes with variable amplitude by the crack law between two holes with variable amplitude. Alternatively, the data-driven SVR algorithm model for crack growth prediction is a useful supplement to the existing methods for crack growth prediction in this paper. The accuracy of the model can meet the accuracy requirements of the aviation crack growth problem. In view of this, the established model can be modified into the aviation crack growth prediction method to a certain extent, and can guide the actual prediction of the law of the aviation cracks between holes.

ACKNOWLEDGMENTS

The supports from the National Natural Science Foundation of China (Nos. 51674121 and 61702184), the Tangshan Team Project (No. 18130209B) and the Aviation Fund Project (No. 2017ZD41006) are gratefully acknowledged.

References

1. Maosheng YANG Jian REN Haidong WANG, JIANG Wenhao. Study on Propagation Life of Fretting Fatigue Crack of Aerial Aluminum Alloy. *Equipment Environmental Engineering.* 2012;9(5):1–5.
2. ZHENG Zhisha, WANG Zhongguang. Historical Review of Fatigue Study. *Journal Materials Science and Engineering.* 1993;2.
3. Paris P C Gomez M P, P Anderson W. A rational analytic theory of fatigue. *The Trend in Engineering.* 1961;13:9–14.
4. Thompson N Wadsworth N J, N Louat. The origin of fatigue fracture in copper. *Philosophical Magazine.* 1956;1:113–126.
5. YANG Xinhua, CHEN Chuanyao. *Fatigue and Fracture, 2nd ed.* Wuhan, China: Huazhong University of Science and Technology Press; 2018. ISBN 978-7-5680-4367-0.
6. Xianggui NI Xinliang LI, WANG Xiuxi. General Modification and Application of the Paris Law for Fatigue Crack Propagation. *Pressure Vessel Technology.* 2006;23(12):8–15+19.
7. Alshamma Fathi A., Jassim Omar Ali. Dynamic crack propagation in nano-composite thin plates under multi-axial cyclic loading. *Journal of Materials Research and Technology.* 2019;8(5):4672–4681.
8. Vasileva Gürlebeck K., Silvestrov V. Mixed antiplane boundary-value problem for a piecewise-homogeneous elastic body with a semi-infinite interfacial crack. *Math. Meth. Appl. Sci..* 2016;39:4419–4427.
9. ZHURAVLOVA Z.. Stress analysis near the tips of a transverse crack in an elastic semi-strip. *Applied Mathematics and Mechanics.* 2017;38(7):935–956.
10. Wang Yang Y Bi H. Multiscale finite element discretizations based on local defect correction for the biharmonic eigenvalue problem of plate buckling. *Math Meth Appl Sci.* 2019;42:999–1017.
11. Griso Orlik J. Homogenization of contact problem with Coulomb's friction on periodic cracks. *Math Meth Appl Sci.* 2019;42:6435–6458.

12. Singh Panja MM. Wavelet-based numerical techniques for 1D peristatic problems in infinite domain. *Math Meth Appl Sci.* 2020;43:4640–4657.

13. Lei Beibei. Study on fatigue crack growth behavior and fracture mechanism of Al-alloy subjected to tensile overloads. In: M.Sc. dissertation of Chang'an University; 2018; Chang'an, China.

14. ZHANG Shaoqin, YANG Weiyang. Prediction of mode I crack propagation direction in crabon-fiber reinforced composite plate. *Applied Mathematics and Mechanics (English Edition).* 2004;25(6):714–722.

15. HAJIMOHAMADI M., GHAJAR R.. An analytical solution for the stress field and stress intensity factor in an infinite plane containing an elliptical hole with two unequal aligned cracks. *Applied Mathematics and Mechanics (English Edition).* 2018;21(2):526–558.

16. BHARGAVA R. R., JANGID K.. Strip-coalesced interior zone model for two unequal collinear cracks weakening piezoelectric media. *Applied Mathematics and Mechanics (English Edition).* 2014;35(10):1249–1260.

17. S. SINGH K. SHARMA, BHARGAVA R. R.. Modified strip saturated models for two equal collinear cracks with coalesced zones in piezoelectric media. *Applied Mathematics and Mechanics (English Edition).* 2019;40(8):1097–1118.

18. Jinfang ZHAO Liyang XIE, LIU Jianzhong. Calculation of stress intensity factor for plate containing a number of symmetrical collinear holes. *Machinery Design & Manufacture.* 2010;12:12–14.

19. Zhenghong LI Xiaojing ZHANG, YU Yin. Experimental and analytical analyses of fatigue crack growth in sheets with multiple holes and cracks. *Acta Aeronautica et Astronautica Sinica.* 2018;39(7):154–162.

20. Drucker H Kaufman L, al . Support vector regression machines. In: Advances in Neural Information Processing Systems 9; 1996; Massachusetts.

21. Dalian Yang Songbai Li Jie Tao Chi Liu, Yi Jiuhuo. Fatigue crack growth prediction of 7075 aluminum alloy based on the GMSVR model optimized by the artifical bee colony algorithm. *Engineering Computations.* 2017;34(4):1034–1053.

22. Weizhen Song Zhansi Jiang, Jiang Hui. Predict the fatigue life of crack based on extended finite element method and SVR. In: CDMMS 2018, 6th International Conference on Computer-Aided Design, Manufacturing, Modeling and Simulation; 2018; American Institute of Physics, America.

23. ZHOU Zhihua. *Machine Learning, 1nd ed.* Beijing, China: Tsinghua University Press; 2016. ISBN 9787302423287.

24. LI Hang. *Statistical Learning Methods, 2nd ed.* Beijing, China: Tsinghua University Press; 2019. ISBN 978-7-302-51727-6.

25. Till Braun Charlotte Veltman Vera Hergesell Alexander Passow Gero Tenderich Martin Borggrefe, Koerner Michael M.. Detection of myocardial ischemia due to clinically asymptomatic coronary artery stenosis at rest using supervised artificial intelligence-enabled vectorcardiography – A five-fold cross validation of accuracy. *Journal of Electrocardiology.* 2020;59:100–105.

26. Haicheng YAO lin LIN Enkai XING Nantian HUANG, CHEN Yanwei. Solar radiation prediction method using bat algorithm optimized SVR. *Renewable Energy Resources.* 2018;36(11):1612–1617.

27. SHANG Zhiwei. Research on the Boarding Passengers Prediction Algorithm of Subway Trains Based on SVR. *Modern Computer (Professional).* 2019;7:42–44+48.

How to cite this article: Zhihang Wang, Jincai Chang, and Qinyu Zhu (2020), SVR Prediction Algorithm for Crack Propagation of Aviation Aluminum Alloy, *Q.J.R. Meteorol. Soc.*, *20xx;xx:xx–xx.*