ISSN: 2414-1895 DOI: 10.6919/ICJE.202108_7(8).0051

A Classification Prediction based on RFE-GA-SVM for Concrete Duct Grouting Quality

Zhonghua Li, Jiaye Wu

College of Automation and Information Engineering, Sichuan University of Science & Engineering, Zigong, Sichuan, China, 643000, China.

Abstract

In order to better assist the manual quality inspection of concrete duct grouting quality and the auxiliary location and identification for grouting defects, the impact elastic wave nondestructive technique is used to obtain datasets of the concrete cube specimens, combined with support vector machine (SVM) can realize the function of two-classifiers with the small-sample and high-dimensional datasets. The recursive feature elimination (RFE) algorithm is introduced into the SVM to reduce the factor relationship for feature selection, At the same time, introduced into genetic algorithm (GA) to found the optimal value of SVM key parameter penalty factor C and kernel function parameter σ, and the classification prediction model of RFE-GA-SVM is proposed to predict the grouting quality of concrete duct. Compared with the classifiers constructed by grid search, particle swarm optimization (PSO) and other machine learning method, this classification prediction model has higher accuracy and stronger generalization ability on independent test set, which shows that it has good recognition effect on the prediction of concrete grouting quality.

Keywords

Nondestructive Technique; Concrete Duct Grouting; Support Vector Machine; Recursive Feature Elimination; Genetic Algorithm; Classification Prediction.

1. Introduction

Concrete duct grouting structure is widely used in expressways and prestressed bridges. As an important protective medium of prestressed tendons in prestressed system, duct grouting not only provided reliable bonding force between prestressed strand and concrete, but also prevented the corrosion of prestressed tendons by air and water. Therefore, grouting quality is a direct result of the safety and durability of the structure [1]. At present, nondestructive testing methods for grouting compactness including ultrasonic method, electromagnetic radar method, ray method and impact elastic wave method [2]. Among them, the impact elastic wave technique has the characteristics of small detection error rate and higher reliability, which can accurately locate the internal defects of prestressed duct [3]. However, on the one hand, the technology has strict requirements for testing equipment and technicians, but the high manual cost and waste time and energy, which limits the development of the method to a certain extent. On the other hand, as the detection work goes on, it will accumulate a large number of data, which has potential mining value and economic benefits and has yet to be further developed.

With the development of artificial intelligence and application in various industries, in view of the above problems, referring to some intelligent classification models commonly used in the field of concrete quality inspection in recent years, such as artificial neural network (ANN) and support vector machine (SVM) [4,5], this paper uses impact elastic wave detection technique to collect discrete signal data of reflected wave from concrete duct grouting specimen, the detected high-dimensional

ISSN: 2414-1895 DOI: 10.6919/ICJE.202108_7(8).0051

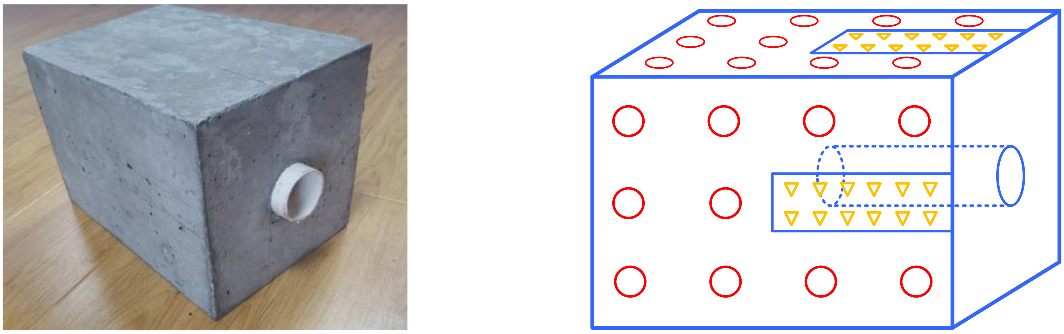
data contains more comprehensive information. Using it combined with SVM to build a machine learning classification model, explore the application value of data, and assist manual defect diagnosis. SVM is a machine learning method first proposed by Vapnik and others in 1995. Depending on the classification prediction of practical problems can be better solved by using SVM in the small-sample, nonlinear, high-dimensional datasets, it tally with requirements of concrete duct grouting classification model very well. Firstly, the feature is processed by recursive feature elimination(RFE) algorithm, Genetic algorithm(GA) is applied to determine the parameters to be optimized by SVM, and a concrete duct grouting classification model based on RFE-GA-SVM is established, which is used for rapid qualitative diagnosis of grouting defects and auxiliary location detection.

2. Datasets source of concrete duct grouting

2.1 Experimental materials and data collection

The experimental materials are three concrete grouting cubes with defects, and the material, proportion and size are the same. Two of them are used for data acquisition of model training dataset, and the other one is used for data acquisition of test dataset. The image of a specimen block is shown in Figure 1(a), its size is 300 * 200 * 200 mm, the cylindrical hole defect areas is located on the side of the cube, with a length of 150 mm and a diameter of 50 mm.

Since the area of sampling area corresponding to defects only accounts for 1/8 of the whole sampling area, in order to avoid the imbalance of two types of samples caused by uniform interval sampling, which affects the classification prediction performance of the model [6], this paper adopts the sampling strategy that large interval in sound area and small interval in defect area to ensure that the sampling data is a balanced dataset, and the schematic diagram of sampling is shown in Figure 1(b). There are 400 samples in the data set, and the ratio of samples in the training and test sets is 3:1.



(a) Physical image of a specimen block, (b) Schematic diagram of sampling

Fig. 1 image of concrete specimen and sampling schematic

2.2 Experimental equipment and principle of impact elastic wave nondestructive testing

The experimental inspection equipment is "SPS-MATS prestressed concrete beam multi-function nondestructive inspection instrument" designed and produced by Sichuan Central Inspection Technology Co., Ltd(SCIT). The equipment is composed of shock hammer, communication cable, sensor and host. It holds the characteristics of high efficiency, real-time reliability and output images visually.

The principle of concrete duct defect detection by shock elastic wave is shown in Figure 2 below. The shock elastic wave is excited by knocking the detection point with an appropriate size shock hammer. When the wave encounters defects such as holes and cracks in the propagation process, the reflection, diffraction and refraction phenomena will occur. Some of the reflections can only be reflected by the bottom interface by bypassing the internal defects, resulting in the increase of reflection propagation path at the defects and the longer propagation time [7], By receiving the change of reflected signal, the sensor can indirectly detect the grouting defects in concrete.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202108_7(8).0051

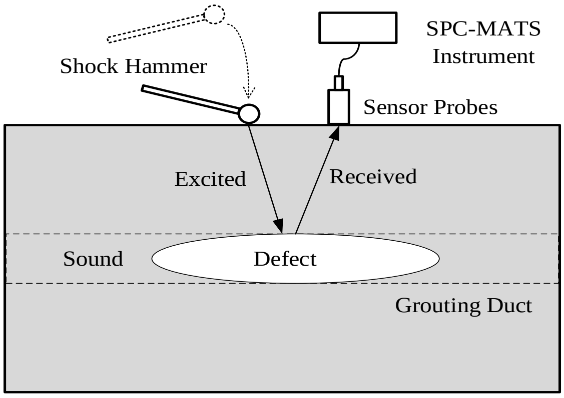
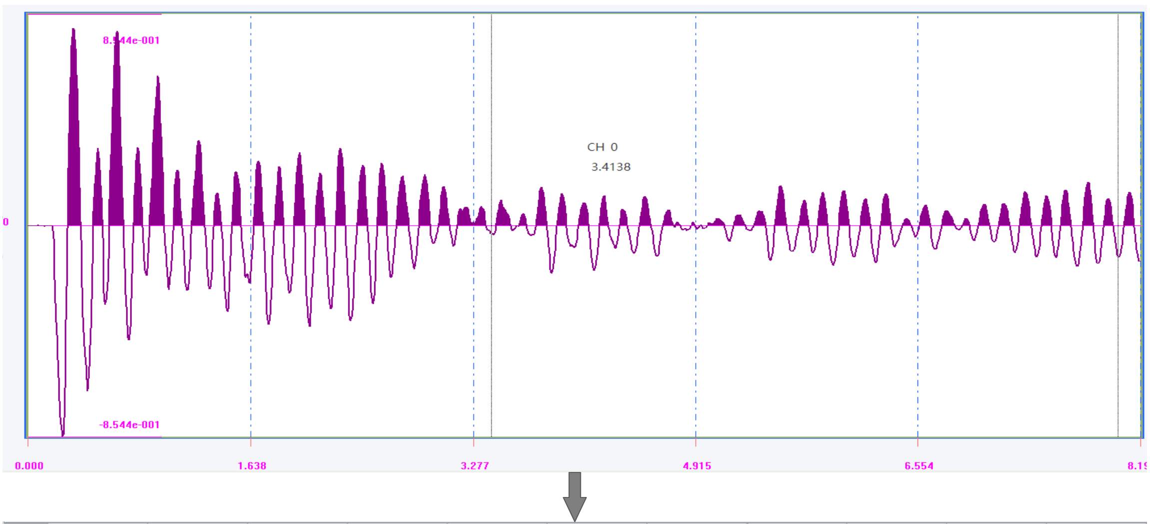


Fig. 2 principle diagram of impact elastic wave detection method

The detection equipment will convert the reflected signal into waveform after a series of processing, which is convenient for technicians to analyze and judge. In this paper, the grouting dataset comes from these waveform data. The discrete waveform data are taken according to the sampling interval and normalized. Then a classification label (SOUND for sound and DEFECT for defective) is added at the end of sampling and saved as a dataset. Figure 3 shows the voltage waveform of a certain sample and a part of the dataset (in the label, SOUND means sound, DEFECT means defective).



	A	B	C	D	E	F	G	EWP	EWQ	EWR	EWS
1	data0	data1	data2	data3	data4	data5	data6	data3993	data3994	data3995	class
2	−0. 134	−0. 18	−0. 226	−0. 268	−0. 304	−0. 328	−0. 339	0. 087	0. 098	0. 109	SOUND
3	−0. 104	−0. 139	−0. 175	−0. 211	−0. 246	−0. 278	−0. 305	−0. 117	−0. 134	−0. 149	SOUND
4	−0. 087	−0. 107	−0. 128	−0. 144	−0. 156	−0. 164	−0. 168	−0. 009	−0. 003	0. 002	SOUND
5	−0. 183	−0. 236	−0. 286	−0. 332	−0. 369	−0. 401	−0. 423	−0. 055	−0. 057	−0. 059	DEFECT
6	−0. 108	−0. 159	−0. 216	−0. 278	−0. 339	−0. 398	−0. 449	−0. 082	−0. 096	−0. 107	DEFECT
7	−0. 142	−0. 17	−0. 196	−0. 219	−0. 239	−0. 254	−0. 261	−0. 223	−0. 228	−0. 235	SOUND
8	−0. 291	−0. 347	−0. 408	−0. 463	−0. 521	−0. 567	−0. 616	−0. 239	−0. 245	−0. 247	SOUND

Fig. 3 schematic illustration of a waveform and partial dataset

The data label is encoded by One-Hot Encoding. The negative sample(DEFECT) is encoded as 0, and the positive sample(SOUND) is encoded as 1. Finally, after data cleaning, it saves as the original data set.

2.3 Evaluation index of classification prediction model

10 fold cross validation (CV) is used in model training, and the mean accuracy (MACC) of training set cross validation is used as an index to evaluate the training effect of the classification model. The test set uses the confusion matrix table shown in Table 1 to construct test indicators.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202108_7(8).0051

Table 1. Confusion matrix of test set

Confusion matrix		Predicted values	
		0	1
True values	0	TN	FP
	1	FN	TP

As the risk of negative (defect) prediction error is higher than that of positive, the test set index is slightly biased to the prediction ability of the negative. The following four indexes are used to evaluate the generalization performance of the classification model: Accuracy(*ACC*), Precision(*P*), specificity (*SP*), F1 Score(*F1*).

$$\begin{cases} ACC = \dfrac{TP+TN}{TP+FN+TN+FP} \\ P = \dfrac{TP}{TP+FP} \\ SP = \dfrac{TN}{TN+FP} \\ F1 = \dfrac{2TP}{2TP+FP+FN} \end{cases} \tag{1}$$

The accuracy rate represents the sample proportion of correct prediction, and the opposite is the error rate (E): E=1-ACC. The precision rate represents the correct proportion of the positives, which can indirectly reflect the ability of the model to distinguish the negatives from the wrong. Specificity reflects the ability of the model to correctly predict negative samples. F1 index reflects the robustness of the model.

In addition, the test set uses error rate (E), false alarm rate (FA) and miss rate (M) to further evaluate the misclassification of the model.

$$\begin{cases} Fa = \dfrac{FP}{FP+TN} \\ M = \dfrac{FN}{TP+FN} \end{cases} \tag{2}$$

3. RFE-GA-SVM model construction and testing

Hardware and software configuration: The hardware environment for all classifier models is i5, the memory is 8GB, and running for Windows 10. The software environment is based on pycharm2019.3.5.

3.1 SVM and kernel function selection

Support vector machine (SVM) is a machine learning method based on Vapnik-Chervonenkis (VC) Dimension theory and structured risk minimize (SRM) principle. It maximizes the interval between datasets by finding support vector, so as to find the optimal segmentation hyperplane [8]. Compared with the artificial neural network, SVM has lower requirements on the number of samples, can effectively avoid the "dimension disaster", improve the nonlinear fitting ability, and make the model more universal in test set [9]. It is suitable to deal with the classification prediction problem of concrete grouting data which is small-samples, high-dimension and nonlinear.

Directed towards the linear inseparable binary classification grouting sample set, Ω:{($\boldsymbol{x}_1$, y_1), ($\boldsymbol{x}_2$, y_2), …, ($\boldsymbol{x}_N$, y_N)}, $\boldsymbol{x}_i \in R^d$, $y_i \in \{0,1\}$, By introducing kernel function into SVM, a linear constrained quadratic programming problem is solved, and the global optimal solution is obtained, which can achieve better classification effect in the case of a small number of samples. The optimization problem can be expressed as the objective function of the generalized optimal classification surface.

$$\begin{aligned} &\min_{w,b,\xi_i} \frac{1}{2}(\boldsymbol{w}g\boldsymbol{w}) + C(\sum_{i=1}^{N}\xi_i) \\ &s.t. \quad \begin{cases} y_i\left[(\boldsymbol{w}g\boldsymbol{x}_i)+b\right]+\xi_i \geq 1 \\ \xi_i \geq 0 \end{cases} \end{aligned} \tag{3}$$

ISSN: 2414-1895 DOI: 10.6919/ICJE.202108_7(8).0051

Where $\sum_{i=1}^{N} \xi_i$ is the sum of the relaxation factors of the samples, which reflects the degree of misclassification in the training set. The introduction of the penalty coefficient C can artificially regulate the tolerance of misclassification samples, so that the value should be as small as possible in a certain range.

By solving the dual problem of generalized optimal classification surface, the solution of the original problem can be obtained indirectly:

$$S(\boldsymbol{x}) = \operatorname{sgn}\left(\sum_{i=1}^{N} \alpha_i^* y_i (\boldsymbol{x}_i \cdot \boldsymbol{x}) + b^* \right) \tag{4}$$

If an appropriate kernel function is selected to replace the inner product operation in high-dimensional space, i.e. $K(\boldsymbol{x}, \boldsymbol{x}') = \phi(\boldsymbol{x}) \cdot \phi(\boldsymbol{x}')$, the objective function of the original optimal classification surface can be transformed into:

$$S(\boldsymbol{x}) = \operatorname{sgn}\left\{ \sum_{i=1}^{N} \alpha_i^* y_i \left[\phi(\boldsymbol{x}) \cdot \phi(\boldsymbol{x}') \right] + b^* \right\} \tag{5}$$

The performance of SVM classification model is determined by both the kernel function parameters and the penalty coefficient C. the kernel functions of linear inseparable problems are polynomial kernel function, radial basis function (RBF) and sigmoid kernel function. RBF has strong linear mapping ability and high efficiency in dealing with multivariate problems. Therefore, RBF is selected as the kernel function of SVM. The formula of RBF is as follow:

$$K(\mathrm{x}, \mathrm{x}') = exp\left(-\frac{\|x-\mathrm{x}'\|^2}{\sigma^2}\right) \tag{6}$$

Where, parameters σ controls the recognition ability of kernel function.

From the above analysis, it can be seen that the penalty coefficient C and the RBF parameter σ in the SVM model jointly affect the efficiency and classification performance of the model.

3.2 RFE algorithm optimizes input features

The feature dimension of the original high-dimensional datasets is 3996, which contains a large number of redundant and noise features. The accuracy of the model will be affected if the dataset is used as initial input directly. Using dimension reduction or feature selection technology to process high-dimensional datasets can improve the efficiency and performance of model construction to a certain extent [10].

In this paper, the recursive feature elimination (RFE) algorithm based on support vector machine classifier (SVC) is used for feature selection. This method is a sequential backward selection (SBS) process, the goal is to find the feature subset of *j* dimension in *i* dimension features ($j < i$), so that it has the best learning performance in internal SVC [11], and the complexity of the algorithm is independent of the dimension of samples. SVC is used to train the dataset in RFE algorithm, and then sorts the features according to the weight of the features obtained by the training. Next, deleting the feature with the smallest weight, the remaining data are used to train the classifier, iterating step by step, and finally the required feature [12] is selected to realize the dimension reduction of the high-dimensional grouting data.

The feature dimension is reduced to 1997 by RFE. The training set after feature selection and the original training set are used as the input of the model respectively. Gaussian naive Bayes (GNB) model, k-nearest neighbor classification (KNN) model and SVM classifier model are constructed for training, and the improvement of feature selection on model training is evaluated is evaluated by comparison with unprocessed dataset at different classifiers. The effect of RFE feature selection on the performance of the model is shown in Figure 4.

It can be seen from Figure 4 that SVM classifier has more advantages in grouting data classification compared with other classifiers, and RFE algorithm can improve the training accuracy of different

ISSN: 2414-1895 DOI: 10.6919/ICJE.202108_7(8).0051

classifiers to a certain extent. Compared with other training models, RFE-SVM classification model has the higher training accuracy.

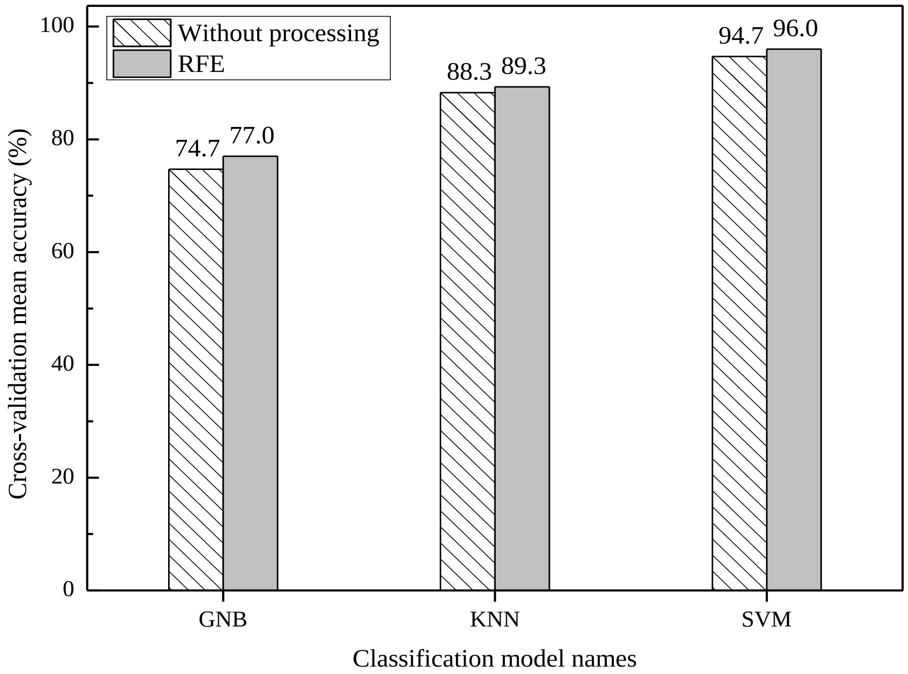


Fig. 4 Effect of RFE feature extraction optimization model input in training set

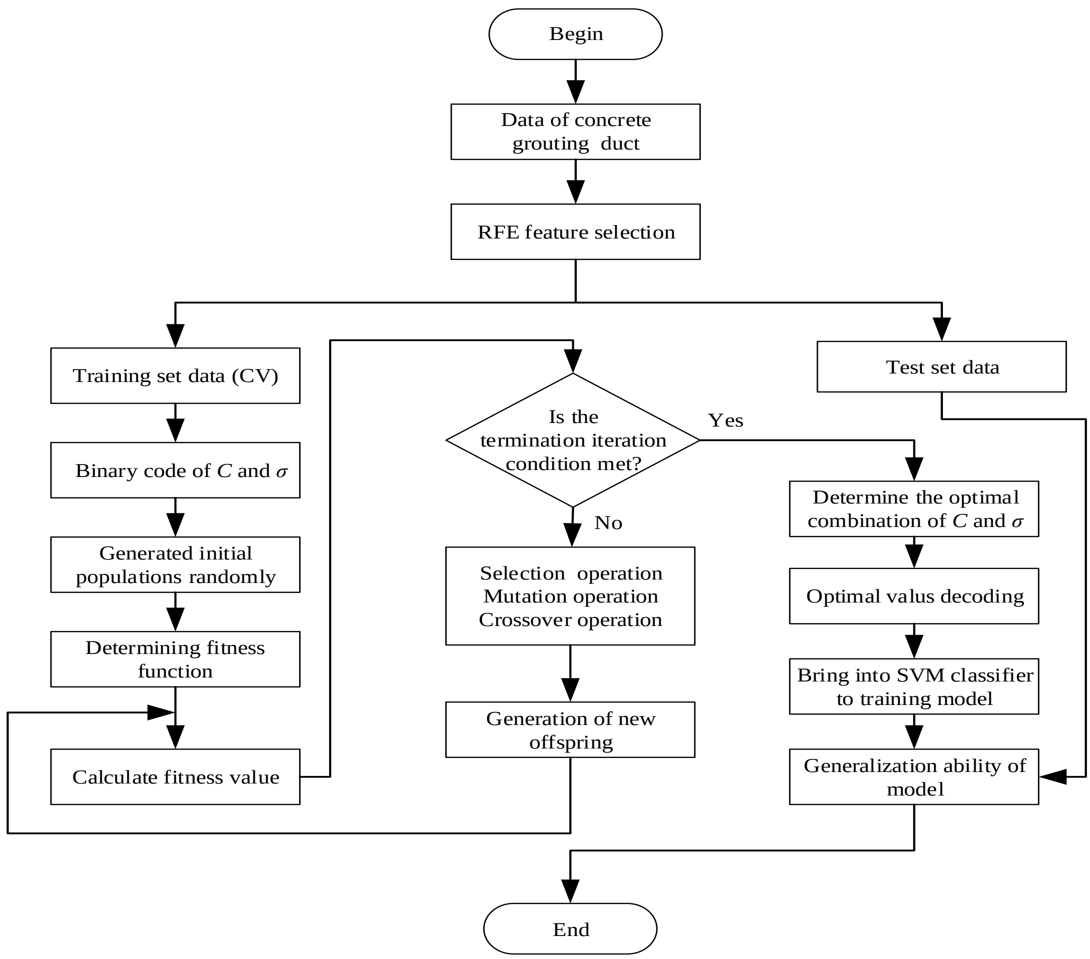


Fig. 5 Flow diagram of RFE-GA-SVM model construction

ISSN: 2414-1895 DOI: 10.6919/ICJE.202108_7(8).0051

3.3 Parameter optimization and model construction

In SVM model, the values of both parameters C and σ directly affect the recognition performance of the classification. People adjusted the parameters of the support vector machine by the means of artificial regulation early. At present, some intelligent optimization algorithms are commonly used, such as genetic algorithm (GA), particle swarm optimization (PSO) and grid search algorithm [13]. Genetic algorithm is selected to determine the best combination values of C and σ in this paper.

Genetic algorithm is an intelligent optimization algorithm that simulates the process of biological evolution. It generates the better solution at the next generation through chromosome inheritance, crossover, mutation and natural selection, and evolves to the direction with higher fitness function value. The chromosome individuals with high fitness function value will be evolved and found when the pre-set iterations of evolutions is satisfied, so as to achieve the purpose of optimization [14]. The flow chart of the program of model construction based on RFE-GA-SVM and optimization by GA is illustrated in Figure 5.

The parameters of GA algorithm are set as follows: The optimized values range from zero to 100, the population size is 80, the number of iterations is 300, the crossover probability and mutation probability of chromosome were 0.7 and 0.01 respectively. The evolutionary process fitness curve of the GA algorithm at the end of the program run is shown in Figure 6.

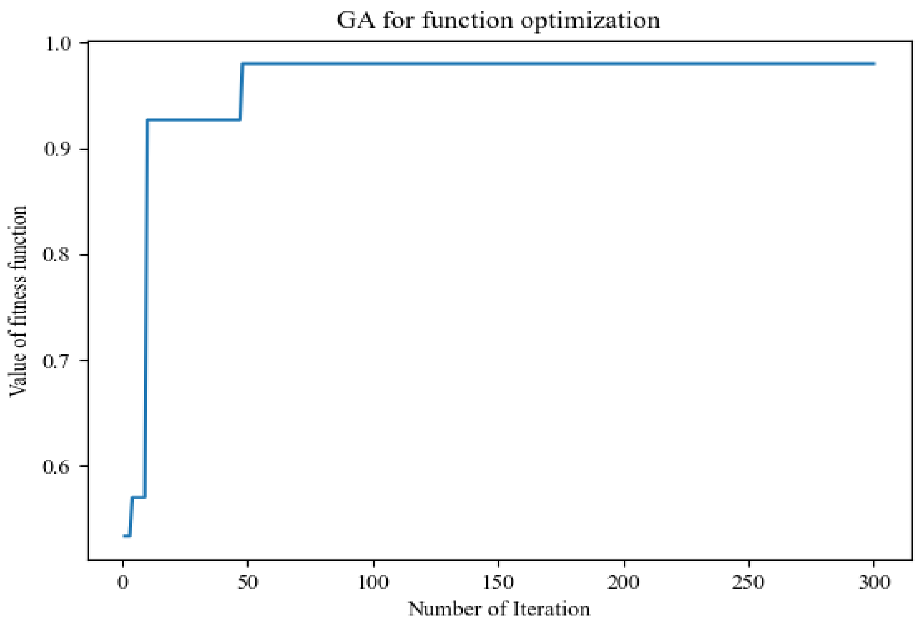


Fig. 6 Fitness curve of genetic algorithm evolution process

As can be seen from Figure 6, with the increase in the number of iterations, the average training accuracy of the cross validation (fitness function value) keeps improving until it maintains 98.33% convergence and stability at around 65 generations, corresponding to the optimal combination of parameters of the SVM solution: C=38.8986, σ=0.00023.

3.4 Model classification prediction test and analysis

In order to verify the parameter optimization capability and classification prediction performance of RFE-GA-SVM(RGS) model, a comparative experiment is conducted with the model of unoptimized parameters (RFE-SVM, RS), the grid search optimization algorithm model (RFE-Grid search-SVM, RSS) and the particle swarm optimization algorithm model (RFE-PSO-SVM, RPS). The comparison of the optimal parameters combination results and the classified evaluation index of the training set and the test set is shown in table 2, it can be seen that the model classification effect of parameter optimization by GA algorithm is better.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202108_7(8).0051

Table 2. Comparison of SVM models by different parameters optimization algorithms

Model	Best parameters (C, σ)	Classified evaluation index(%)		
		MACC	ACC	FA
RS	—	95.33	84.00	16.00
RSS	(64.00, 0.0002)	97.00	85.00	14.00
RPS	(40.17, 0.0001)	98.00	87.00	10.00
RGS	(38.89, 0.00023)	98.33	88.00	10.00

In order to further evaluate the comprehensive performance of the model, the comparison and analysis of the accuracy of cross validation on training set and the generalization ability on the test set is shown in Figure 7, and the misclassification error comparison of models is shown in Figure 8. The nine classification prediction models mentioned above and their abbreviations are as follows: GNB(G), RFE-GNB(RG), KNN(K), RFE-KNN(RK), SVM(S), RFE-SVM(RS), RFE-Grid Search-SVM(RSS), RFE-PSO-SVM(RPS), RFE-GA-SVM(RGS).

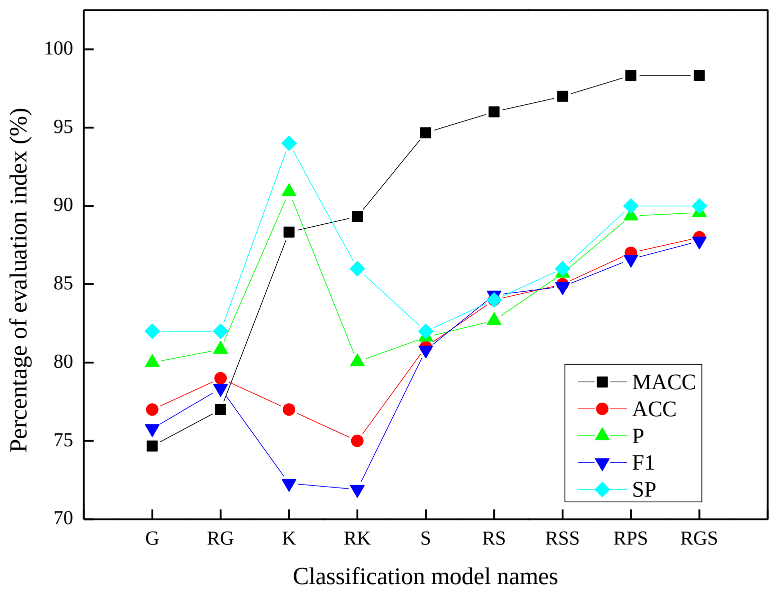


Fig. 7 Classification and generalization ability of the models

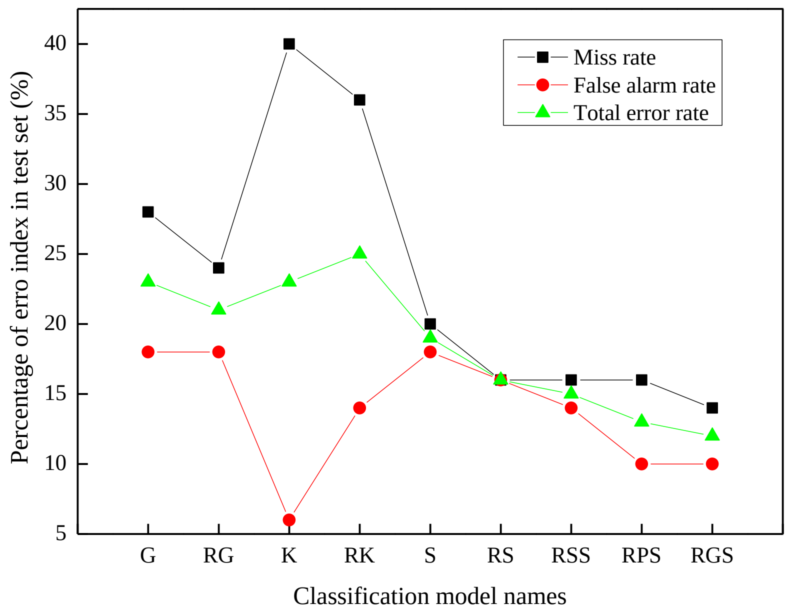


Fig. 8 Misclassification errors comparison of models on test set

ISSN: 2414-1895 DOI: 10.6919/ICJE.202108_7(8).0051

It can be seen from the comparative analysis of the Figure 7 and Figure 8, compared with the lower accuracy of GNB and the serious polarization of KNN, the SVM classifiers is more stable and efficient. RFE algorithm and different optimization algorithms can improve the accuracy of SVM at different degrees in the training set. In the independent test set, the classification prediction effect of RGS model is significantly higher than the other eight classification models. It can also be seen from Figure 8 that the total error rate and false detection rate of RGS model are lower, especially in the case of higher-risk false detection in practical engineering, RGS model has stronger robustness. Therefore, the RFE-GA-SVM classification prediction model can be used in the concrete grouting quality detection to get more precise predicted results, and assist the manual judgment and positioning of grouting defects, which has a definite application prospect. [15]

4. Summary

(1) For the shortcomings that the manual detection method is wasting time and energy, this paper combines the theory of impact elastic wave nondestructive detection with machine learning algorithm, and applies it to concrete duct grouting classification predictions so that assist manual judgment and analysis.

(2) According to the reflected wave data obtained from the detection which has small-sample, high-dimensional and nonlinear characteristics, a SVM classifier is proposed to construct a classification model. RFE algorithm is introduced to classifiers for selecting features, which effectively reduce information redundancy and improve the accuracy of classification. The parameters in the SVM are optimized by GA algorithm with the characteristics of global optimization. Through the comparison between and Analysis of the RFE-GA-SVM model and other models, the results show that it has good stability and high accuracy in the classification of concrete duct grouting, and it can be able to make rapid and effective response to detection and location of grouting defects.

(3) The author tries to use the waveform signal data of impact elastic wave method as the input to build a classification model based on machine learning by experiments. In practical engineering, the collected data will show different signal characteristics and working conditions since the influence of a series of complex factors, such as reinforcement location, inspection surface, defect degree, pipe material. These prediction models remain to be further studied according to the actual situation.

References

[1] Wu Jiaye. Civil engineering inspection and testing[M]. Beijing: Higher Education Press, 2015: 215-217.

[2] Zhao Jianhua. Research on Technical Status and Development Trend of Prestressed Hole Density Testing[J]. Construction Quality, 2020,38 (S1): 104-109.

[3] ZHANG Song-lei, TIAN Jian-hui, ZHANG Pen, et al. Research on Quality Testing Technology of Prestressed Grouting Based on Shock Elastic Wave Method[J]. Adhesion, 2020, 43(07): 108-111.

[4] HAN Xiao, TANG Hesheng, ZHOU Deyuan. A Damage Detection Method for Grout Sleeve Splicing at Assembly Column Based on the Sym Wavelet and BP Neural Network[J]. Structural Engineers, 2018, 34(06): 21-28.

[5] Xue Zhongjun, Zhang Jianlong, Wei Wenzhe, et al. Pile defect severity identification method based on wavelet packet analysis and support vector machine[J]. Highway transportation technology (Application Technology Edition), 2018,14 (11): 216-220.

[6] ZHANG Zhong-lin, WU Dang-ping. An imbalanced data classification method based on probability threshold Bagging[J]. Computer Engineering & Science, 2019, 41(06): 1086-1094.

[7] Li Yanlong, Wang Junzhong, Chen Junhao, et al. Experimental study on the detection of internal damage of concrete by impact-echo method [J], Chinese Journal of Applied Mechanics, 2020, 37(1): 149-154.

[8] QIAO Feng-juan, GUO Hong-li, LI Wei, et al. Research on Deep Learning Classification Based on SVM: A Review[J]. Journal of Qilu University of Technology, 2018, 32(05): 39-44.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202108_7(8).0051

[9] Li Jing, Xu Lulu. Comparison and Analysis of Research Trend Prediction Models Based on Machine Learning Algorithm—BP Neural Network, Support Vector Machine and LSTM Model[J]. Journal of Modern Information, 2019, 39(4):23-33.

[10] LI Zhiqin, DU Jianqiang, NIE Bin, et al. Summary of feature selection methods[J]. Computer Engineering and Applications, 2019, 55(24):10-19.

[11] LIANG Wuqi, WANG Ronghua, LIU Keli, et al. A Research Review of Feature Selection Algorithm[J]. Journal of Anhui Radio & TV University, 2019(04): 85-91.

[12] Li Cheng, Li Qiang, Zhang Qichao, et al. Partial discharge characteristic optimization of cables based on support vector machine-recursive feature elimination[J]. Electrical Engineering, 2020,21(01): 67-71.

[13] Qin Yuan, Zhang Binbin, Xie Gang. SVM Model to Redict Strength of Filling of Cemented Tailings Backfill [J]. Jiangxi Building Materials, 2020(07): 10-12+14.

[14] MAN Chun-tao, LIU Bo, CAO Yong-cheng. The Application Based on Support Vector Machine Optimized by Particle Swarm Optimization and Genetic Algorithm[J]. Journal of Harbin University of Science and Technology, 2019, 24(03): 87-92.

[15] LI Xingyu, CHAI Wenhao, YANG Yaxun. Analysis on Influential Factors of Grouting Quality in Concrete Pipe Based on Impact-echo Method[J]. Railway Engineering, 2020, 60(08):53-57.