

# Diagnosis Model of Pipeline Cracks According to Metal Magnetic Memory Signals Based on Adaptive Genetic Algorithm and Support Vector Machine

Lihong Gong<sup>\*1,2</sup>, Zhuxin Li<sup>1</sup> and Zhen Zhang<sup>1</sup>

<sup>1</sup>Department of Military Oil Supply Engineering, Logistic Engineering University of PLA, Chongqing 401331, China

<sup>2</sup>Department of Military Electric Power Engineering, Chongqing Communication Institute, Chongqing 400035, China

**Abstract:** Metal magnetic memory (MMM) signals can reflect stress concentration and cracks on the surface of ferromagnetic components, but the traditional criteria used to distinguish the locations of these stress concentrations and cracks are not sufficiently accurate. In this study, 22 indices were extracted from the original MMM signals, and the diagnosis results of 4 kernel functions of support vector machine (SVM) were compared. Of these 4, the radial basis function (RBF) kernel performed the best in the simulations, with a diagnostic accuracy of 94.03%. Using the principles of adaptive genetic algorithms (AGA), a combined AGA-SVM diagnosis model was created, resulting in an improvement in accuracy to 95.52%, using the same training and test sets as those used in the simulation of SVM with an RBF kernel. The results show that AGA-SVM can accurately distinguish stress concentrations and cracks from normal points, enabling them to be located more accurately.

**Keywords:** Adaptive genetic algorithm, metal magnetic memory, pipeline crack recognition, support vector machine.

## 1. INTRODUCTION

Oil and gas pipelines are important channels of energy delivery in many countries. However, these pipelines are often damaged during transportation or general use that lead corrosion or crack defects. Therefore, there are a large number of non-destructive testing (NDT) methods such as magnetic particle testing (MPT), magnetic flux leakage (MFL), eddy current testing (ECT), magneto-acoustic emission (MAE), magnetic Barkhausen noise (MBN), and metal magnetic memory (MMM) to detect pipeline defects [1].

Among these methods, MFL, MBN, and MMM are vital magnetic NDT methods. MFL and MBN are active magnetic test methods, which require the application of a strong artificial field to magnetize the test objects. In MMM, unlike MFL and MBN, the geomagnetic field is applied as the stimulus source instead of an external magnetic field.

Thus, MMM testing is a new highly effective passive magnetic flux leakage NDT technique, with low requirements in terms of both testing equipment and operational complexity, and thereby meeting desirable engineering requirements. MMM techniques measure the self-magnetic flux leakage (SMFL) signal of a ferromagnetic material. These materials generate SMFL signals in their stress concentration zones under the combined effect of the geomagnetic field and their operational load [2-5].

MMM techniques can be used to assess stress concentration or the location of defects by detecting the changes in the surface SMFL signal of a ferromagnetic component. This technique is therefore widely used to detect areas of stress concentration and different kinds of microscopic and macroscopic cracks caused by stress concentration. According to experimental studies, two primary criteria are used in MMM testing to identify the stress concentration and crack locations: the tangential component of the SMFL signal,  $H_p(x)$ , which reaches its maximum value; and the normal component of the SMFL signal,  $H_p(y)$ , which passes through zero and changes its polarity [6, 7].

However, it has been reported that these two criteria are not sufficient to distinguish between the three statuses of pipelines: normal, stress concentration, and cracks [8]. Therefore, it is necessary to find an improved method that can be used to diagnose defects and areas of stress concentration. This paper describes a series of experiments in which pipelines with prefabricated cracks were tested using MMM. The MMM signals obtained from the experiments were analyzed, and a mathematical model based on support vector machine (SVM) and adaptive genetic algorithm (AGA) was proposed to diagnose the cracks and stress concentration points in selected test specimens.

## 2. EXPERIMENTAL DETAILS

Experimental data were obtained from a series of MMM testing experiments using an MFL-4032 magnetic flux leakage/magnetic memory detector, which was researched and developed by our project team and Xiamen Eddysun

\*Address correspondence to this author at the Department of Military Oil Supply Engineering, Logistic Engineering University of PLA, Chongqing 401331, China; Tel: +862368775967, +8613983678633; Fax: +862368775967; E-mail: [glh1130@aliyun.com](mailto:glh1130@aliyun.com)

**Table 1. Trace metal composition of the pipeline steel.**

Chemical Comp.	C	S	Mn	V	Si	P	Nb	Ti
Content (%)	0.04-0.06	≤0.010	1.24-1.30	0.015-0.03	≤0.40	≤0.020	0.04-0.05	0.006-0.008

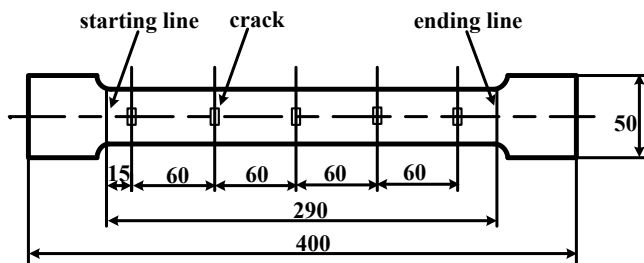
**Table 2. Mechanical properties of the pipeline steel.**

Properties	Elastic Modulus (GPa)	Tensile Strength (MPa)	Compressive Strength (MPa)	Yield Strength (MPa)	Poisson's Ratio
Content	200	520	496	420	0.26

Company. The experiments used metal pipe constructed of ferromagnetic material with a wall thickness of 2.3 mm. The trace metal composition and mechanical properties are listed in Tables 1 and 2, respectively. Fifteen samples were prepared, each of which was 400 mm long and 50 mm wide. Each sample contained five prefabricated cracks of different sizes, making a total of 75 cracks to be detected. The cracks were numbered in advance according to a predesigned ordering scheme. A schematic diagram of a sample is shown in Fig. (1). For each sample, the start and end locations were marked with a detection line and the cracks were inserted at distances of 15, 75, 135, 195, and 255 mm from the starting point.

The experiments were conducted as follows.

- (1) Prepare the test objects, including processing samples, determining the sizes of the cracks, prefabricating the crack defects in the samples and eliminating the stress in the processed samples.
- (2) Choose test instruments and equipment. For these experiments, an RGM-4100 electronic tensile testing machine and an MFL-4032 magnetic flux leakage/magnetic memory detector were chosen.
- (3) Compute the loads according to the operational conditions of the test objects.
- (4) Stretch the samples one by one with the computed loads, test the samples and collect the MMM signals.
- (5) Analyze the signals and extract data for the indices.



**Fig. (1).** Schematic diagram of sample with prefabricated cracks.

### 3. METHODS

#### 3.1. Indices and Data Set

Analysis determined that the 15 samples contained 42 stress concentration points. In addition to the 75 prefabricated crack points, 150 normal points were chosen randomly for comparison purposes. An accurate diagnosis model was established by producing additional indices based

on the original signals, including 22 indices, such as the peak value, peak-to-peak value, and waviness width, thereby obtaining a new data matrix with dimensions  $267 \times 22$ . The classifications of all the data in the matrix were marked as 0 (normal points, NPs), 1 (stress concentration points, SCPs), or 2 (crack points, CPs).

Then, all the samples were randomly divided into training and testing sets. The training set contained 75 NPs, 21 SCPs, and 37 CPs, while the testing set contained 75 NPs, 21 SCPs, and 38 CPs.

#### 3.2. Support Vector Machine

SVM has been widely applied since its initial development by Vapnik *et al.* [9]. The classical SVM is a classification learning algorithm for solving two-category problems. It is used for structural health monitoring, damage detection, and the classification of different engineering structures and machineries [10-12]. The core theory of SVM is to build the maximum interval hyperplane (as shown in Fig. 2). Given a linearly separable training sample  $S = ((\mathbf{x}_1, y_1), \dots, (\mathbf{x}_l, y_l))$ , the mathematical model is the hyperplane  $(\mathbf{w}, b)$  that solves the optimization problem

$$\min \frac{1}{2} \|\mathbf{w}\|^2 = \frac{1}{2} \langle \mathbf{w} \cdot \mathbf{w} \rangle$$

$$s.t. \quad y_i (\langle \mathbf{w} \cdot \mathbf{x}_i \rangle + b) \geq 1, \quad i = 1, \dots, l$$
(1)

where  $\mathbf{w}$  is a weight vector and the training goal of the classifier. A calculation enables the hyperplane  $(\mathbf{w}, b)$  to be found, with a maximum interval of  $\frac{1}{\|\mathbf{w}\|}$ . As soon as these two quantities are obtained, the SVM classifier is determined. It is then possible to classify a new unknown sample by inserting it into the classifier, which produces the classification as its output.

#### 3.3. Classification Algorithm

SVM can only be used to solve two-category problems, so multi-category problems required a new algorithm, which was designed using the following steps.

Step 1: Establishing classifiers. Use the data with the classifications 0 and 1 in the training set to establish

classifier 1, those with 0 and 2 to establish classifier 2, and those with 1 and 2 to establish classifier 3.

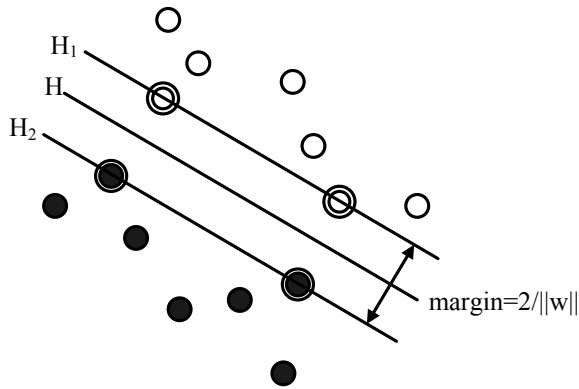


Fig. (2). Maximum interval hyperplanes.

Step 2: Testing. Insert all the input variable data belonging to the testing set into the three classifiers to obtain three classifications for every sample.

Step 3: Classifying. Of the three classifications that were obtained for every sample, select the highest classification to be its real classification. For example, a sample found to be classified as 1, 2, 2 by the three classifiers would be classified as 2, that is to say, the sample would be considered to be a crack.

#### 4. RESULTS AND DISCUSSION

##### 4.1. Comparison of the Kernel Functions

SVM has a number of different kernel functions, including linear kernel or dot product (linear), quadratic kernel (quadratic), polynomial kernel (polynomial), and radial basis function (RBF). A simulation was conducted to compare the accuracy of these four functions, with results of 93.28%, 91.79%, 91.79%, and 94.03% diagnosis accuracy, respectively, confirming RBF as the most accurate kernel function. Therefore, RBF was selected for more detailed attention.

Diagnosis results for RBF SVM are mainly determined by two factors: a penalty factor (denoted by C), and a parameter of the Gaussian kernel function (denoted by  $\sigma$ )

$$K(x, y) = e^{-\sigma \|x - y\|^2} \tag{2}$$

Different combinations of C and  $\sigma$  may influence the accuracy of the diagnosis differently. The effects of C and  $\sigma$  on the accuracy of the diagnosis were analyzed, and the results are shown in Fig. (3), with C ranging from 10 to 100 in steps of 10, and  $\sigma$  ranging from 1 to 16. The best result, 94.03% accuracy, was achieved with C = 100 and  $\sigma = 9$ . Table 3 lists the results for all three classifications: NPs, SCPs, and CPs.

Fig. (3) enables three conclusions to be drawn. First, the same value of C can result in a big difference in the predicted accuracy. Second, generally speaking, the greater the value of  $\sigma$ , the higher the accuracy. Third, a change in the value of C does not markedly influence the diagnosis accuracy. However, a prediction accuracy higher than 94.03% may be possible, because there are many combinations of C and  $\sigma$  that were not investigated.

Table 3 shows that most NPs and CPs can be distinguished easily but that it is more difficult to differentiate CSPs from NPs.

##### 4.2. Improvement of SVM Based on AGA

Achieving the best possible diagnosis result relies on finding the best possible combination of C and  $\sigma$  while the training set is being determined. However, as  $C > 0$  and  $\sigma > 0$ , there are many possible values of C and  $\sigma$ , so adaptive genetic algorithm (AGA) was used to find the best combination of C and  $\sigma$  before running SVM. Genetic algorithm (GA) has been widely used as a stochastic optimization method for solving optimization problems; however, AGA is known to outperform GA [13-15]. The flow chart of the combined AGA-SVM method is shown in Fig. (4).

The steps of the AGA-SVM method are as follows.

Step 1: Generate a number of initial populations. Every individual value is expressed as a genetic code of the chromosome, which is translated into a binary number.

Step 2: Determine the fitness of every individual value using the strategy of roulette. Then, judge whether it conforms to the optimization criteria. If sufficient iterations have been completed, compare the results of all the

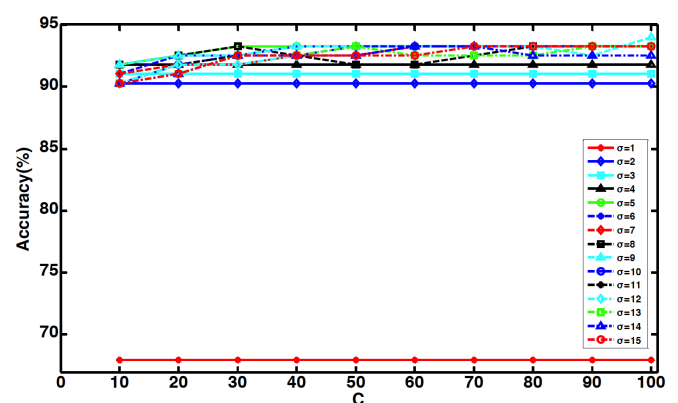
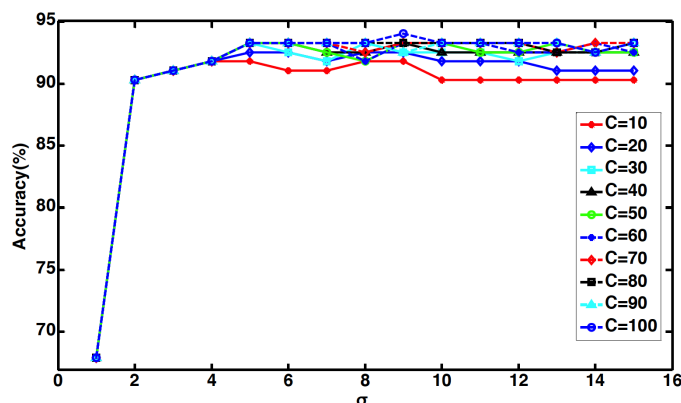


Fig. (3). Accuracy for different values of C and  $\sigma$ .

Table 3. Diagnosis results for RBF SVM (number and percentage).

Classification	Judged as NP	Judged as CSP	Judged as CP	The Total Result
NP	74 (98.67%)	1 (1.33%)	0 (0.00%)	126(94.03%)
CSP	5 (23.81%)	16 (76.19%)	0 (0.00%)	
CP	2 (5.26%)	0 (0.00%)	36 (94.74%)	

generations, output the best individual result and its optimal solution, and end. Otherwise, move on to step 3.

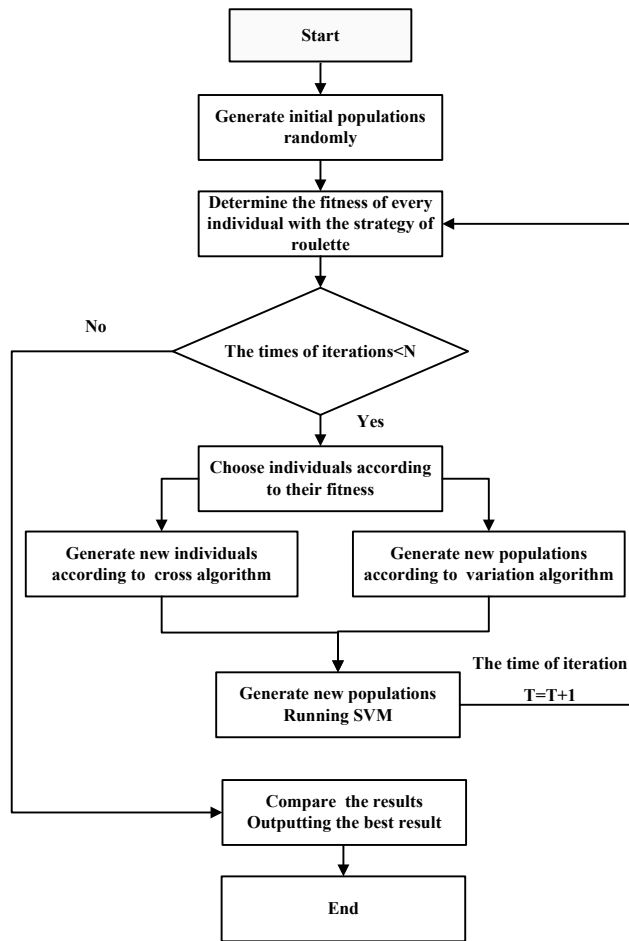


Fig. (4). Flow chart of AGA-SVM.

Step 3: Generate new populations according to a cross probability and a cross method.

Step 4: Generate new populations according to a variation probability and a variation method.

Step 5: Generate a new generation of populations with cross and variation, and run SVM to diagnose the populations. Return to step 2.

The method was programmed in MATLAB. The best result obtained was 95.52%, demonstrating the method's superior performance compared with traditional RBF SVM, for which the best result was 94.03%. Fig. (5) shows the variation in the accuracy of the diagnosis with the number of generations using AGA-SVM. It was concluded that the results stabilize after the 15<sup>th</sup> generation, and that the best result is reached with the 1<sup>st</sup> generation.

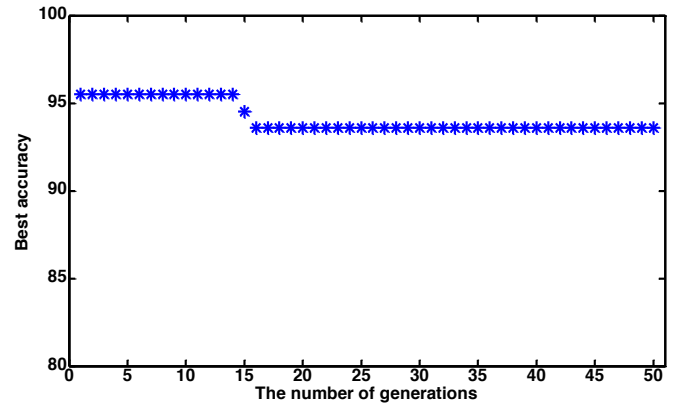


Fig. (5). Variation in the diagnosis accuracy with the number of generations using AGA-SVM.

The diagnostic capabilities of SVM and AGA-SVM were compared by selecting different training and test samples randomly, according to the method described in Section 3.1, and by performing 100 simulations. The results of the simulations are shown in Fig. (6). For every simulation, the accuracy obtained using AGA-SVM was higher than that obtained using SVM. Using AGA-SVM, the average and maximum accuracy values were improved from 92.71% and 97.76% to 93.88% and 98.51%, respectively. A paired-samples T-test was used to compare the diagnoses of SVM and AGA-SVM. The results were  $t = 9.36$  and  $p = 2.66 \times 10^{-15}$ , which indicates that the diagnosis accuracy of AGA-SVM is significantly higher than that of SVM. Based on these results, it can be concluded that the diagnostic capability of AGA-SVM is superior to that of SVM.

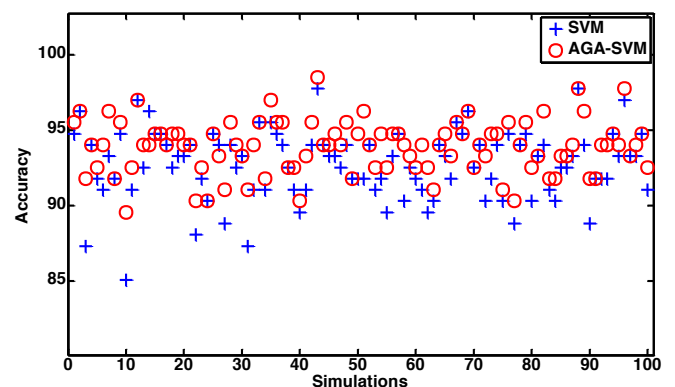


Fig. (6). Comparison of diagnosis accuracy between SVM and AGA-SVM.

CONCLUSION AND FUTURE WORK

The SFML signals of MMM are able to provide information about the pipeline's status: normal, stress

concentration, or crack, but the original signals are not able to clearly distinguish between these states. Analysis of the signals leads to the identification of selected indices, some of which have a high sensitivity. Although it is difficult to distinguish cracks from both normal and stress concentration points based on any particular index, a combination of indices can be used to diagnose the defects effectively. SVM is a perfect classifier for solving two-category problems; however, as this study has three categories, it was necessary to use a multi-category algorithm. Simulations of four kinds of kernel functions in SVM established the ability of RBF SVM to successfully diagnose pipeline cracks and it was further shown that these results could be improved by using a combination of SVM and AGA.

This paper presented a method for distinguishing cracks and stress concentrations from normal points, and proposed an effective means of diagnosing stress concentrations and cracks based on MMM signals. A new model, AGA-SVM, was developed, taking into account more indices extracted from the MMM signals, leading to a fast, simple, and accurate method of stress concentration and crack diagnosis. Simulation results revealed that while SVM is quite effective in diagnosing crack defects, AGA-SVM offers improved diagnostic ability.

AGA-SVM is capable of successfully diagnosing the stress concentrations and cracks in pipelines. However, because of the principle of AGA, the populations are generated randomly, which may cause the accuracy of the results to vary across different simulations. This may be addressed by using sufficiently large initial population and generation sizes, to ensure that the best result is stable, but this would prolong the running time. Therefore, obtaining diagnoses of an acceptable accuracy within a running time of reasonable duration is an important consideration for future work.

## CONFLICT OF INTEREST

The authors confirm that there are no conflict of interest relating to the content of this article.

## ACKNOWLEDGEMENTS

We thank the editor for critical reading of the manuscript. This project was supported by the General Logistics department of the Chinese People's Liberation Army (No. 20040207).

## REFERENCES

- [1] S. Ghanei, M. Kashefi, and M. Mazinani, "Comparative study of eddy current and Barkhausen noise nondestructive testing methods in microstructural examination of ferrite-martensite dual-phase steel", *Journal of Magnetism and Magnetic Materials*, vol. 356, pp. 103-110, April 2014.
- [2] Z. D. Wang, Y. Gu, and Y. S. Wang, "A review of three magnetic NDT technologies," *Journal of Magnetism and Magnetic Materials*, vol. 324, pp. 382-388, February 2012.
- [3] J. L. Ren, K. Song, G. H. Wu, "Mechanism study of metal magnetic memory testing", In: *Proceedings of the 10<sup>th</sup> Asia-Pacific Conference on Non-Destructive Testing*, 2001, pp. 17-21.
- [4] D. C. Jiles, "Theory of the magnetomechanical effect", *Journal of Physics D: Applied Physics*, vol. 28, pp. 1537-1546, April 1995.
- [5] A.A. Doubov, "Diagnostics of metal items and equipment by means of metal magnetic memory", In: *7<sup>th</sup> Conference on NDT and International Research Symposium*, 1999, pp. 181-187.
- [6] A. Doubov, "Principal features of metal magnetic memory method and inspection tools as compared to known magnetic NDT methods", In: *Proceedings of the 16<sup>th</sup> Annual World Conference on Non-destructive Testing*, 2004, pp. 1-4.
- [7] M. Roskosz, and M. Bieniek, "Analysis of the universality of the residual stress evaluation method based on residual magnetic field measurements", *NDT & E International Journal*, vol. 54, pp. 63-68, March 2013.
- [8] L. H. Gong, Z. X. Li, Z. Zhang, and S. J. Liu, "Detection of cracks according to Bayesian wavelet threshold on metal magnetic memory signals", *International Journal of Applied Mathematics and Statistics*, vol. 52, no. 1, pp. 111-118, April 2014.
- [9] C. Nello, and S. T. John, "An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods", China Machine Press, 2005.
- [10] D. Duan, Y. Y. Liu, and Q. Gao, "Structure health monitoring using support vector data description and wavelet packet energy distributions", *Applied Mechanics and Materials*, vols. 135-136, pp. 930-937, October 2011.
- [11] H. X. He and W. M. Yan, "Structural damage detection with wavelet support vector machine: introduction and applications", *Structural Control and Health Monitoring*, vol. 14, pp. 162-176, February 2007.
- [12] S. Das, A. N. Srivastava, and A. Chattopadhyay, "Classification of damage signatures in composite plates using one-class SVMs", In: *Aerospace Conference, IEEE*, pp. 1-19, 2007.
- [13] D. Whitley, T. Starkweather, and C. Bogart, "Genetic algorithms and neural networks: optimizing connections and connectivity", *Parallel Computing*, vol. 14, pp. 347-361, August 1990.
- [14] M. O. Beatrice and V. Mario, "Local search genetic algorithms for the job-shop scheduling problem", *International Journal of Artificial Intelligence & Applications*, vol. 21, pp. 99-110, July 2004.
- [15] L. Wang and D. Tang, "An improved adaptive genetic algorithm based on hormone modulation mechanism for job-shop scheduling problem", *Expert Systems with Applications*, vol. 38, pp. 7243-7250, June 2011.