IDENTIFICATION AND DETECTION OF THE PROCESS FAULT IN A CEMENT ROTARY KILN BY EXTREME LEARNING MACHINE AND ANT COLONY OPTIMIZATION

Ouahab KADRI¹, and Leila Hayet MOUSS²

ABSTRACT: The aim of this paper is to propose a new fault diagnosis method for complex manufacturing system. We have used an artificial neural network (ANN) and an Ant Colony Optimization (ACO) algorithm to diagnose the condition monitoring of a rotary cement kiln. The Ant Colony algorithm can find a small features subset from the original real time signals and the Extreme Learning Machine (ELM) enables a good accuracy with a limiting learning time. Many benchmark datasets have used to evaluate the performances of our algorithm and the result indicates its higher efficiency and effectiveness comparing to other methods.

KEY WORDS: Complex manufacturing system, Extreme Learning Machine, Artificial Neural Networks, Ant Colony Optimization.

1 INTRODUCTION

Currently, rotary kiln equipment is widely used in many manufacturing systems. It represents the most important component in the cement manufacturing process. The occurrence of fault in rotary kiln system can cause considerable economic damage and even destruction of the total drive system.

The rotary kiln consists of the following components: the shell, the refractory lining, support tyres and rollers, drive gear and internal heat exchangers (Huntzinger & Eatmon, 2009). We cannot use mathematical model to diagnose the faults of a complex system such as rotary kiln, because this traditional method cannot detect multivariate fault and the accuracy rate of diagnosing is low (Yin et al., 2014). Extreme Learning Machine (ELM) is new method. It can find the complex relationship between symptoms and fault classes (Huang et al., 2006). This paper discusses the method of diagnosing rotary kiln fault based on Extreme Learning Machine and Ant Colony Optimization algorithm (ACO). ELM is a new learning algorithm for Single Layer Feedforward Neural Network (SLFN) architecture proposed by Huang et al.

ELM overcomes the disadvantages caused by gradient descent based algorithms such as Back propagation applied in ANNs. Unlike other types of neural networks, ELM offers the possibility of accelerating the learning phase (Huang et al., 2006). We have used ACO algorithm to find the best parameters for the ELM algorithm.

The remainder of this paper is organized as follows. In section 2, the related work is surveyed and presented. Sections 3 and 4 briefly review Extreme Learning Machine and Ant Colony algorithm. In Section 5 and 6, we describe the rotary kiln system and the proposed diagnostic method. Section 7 describes the experiments conducted on both artificial and real-world datasets. Finally, Section 8 presents the conclusion of the work.

2 BACKGROUND

For rotary kiln fault diagnosis, there are two problems must be taken into account: several parameters and various faults, which makes the system of rotary kiln condition very complex (Yao & Pan, 2014). In our system, the number of parameters is more than forty. We need to monitor these parameters daily and correct the fault immediately. We cite among others the following parameter: Shell temperatures, which is a condition indicator of the refractory lining and slag/ash ring, accumulates inside the kiln. The most important problem in fault diagnosis is to find the relation between the faults and parameters subset.

These last years, several papers discuss the importance of using of neural networks in industrial diagnosis. Jaouher et al. developed a method based on the data-driven diagnosis approach. They combined a Simplified Fuzzy Adaptive Resonance Theory Map (SFAM) neural network and Weibull
distribution (WD) (Jaouher et al., 2015). In the training phase, they used WD to fit measurement and to avoid areas of fluctuation in the time domain. SFAM training process is based on fitted measurements at present and previous inspection time points as input.

Gang et al. developed online fault diagnosis method based on Incremental Support Vector Data Description (ISVDD) and Extreme Learning Machine with incremental output structure (IOELM) (Yin et al., 2014). ISVDD is used to find a new failure mode quickly in the continuous condition monitoring of equipment. The fixed structure of Extreme Learning Machine is changed into an elastic structure whose output nodes could be added incrementally to recognize the new fault mode efficiently.

Javed et al. introduced a predictability scheme to reduce the dimensionality of the data (Javed et al., 2015). The proposed diagnostic model is achieved by integrating two new algorithms namely, the Summation Wavelet-Extreme Learning Machine and Subtractive- Maximum Entropy Fuzzy Clustering to show evolution of machine degradation by simultaneous predictions and discrete state estimation. The diagnostic model is equipped with a dynamic failure threshold assignment procedure to estimate RUL in a realistic manner.

3 EXTREME LEARNING MACHINE

In this section, we briefly describe the essence of ELM. ELM was originally developed by Huang et al., (2006). They proposed a simple learning algorithm for SLFNs called Extreme Learning Machine (ELM) whose learning speed can be faster than other network learning algorithms while obtaining better classification accuracy.

![Figure 1. ELM for Feature learning, Clustering, Regression, and Classification](image)


We consider a set of N arbitrary distinct training data \((x_i, t_i)\), where \(x_i = [x_{i1}, x_{i2}, \cdots, x_{im}] \in \mathbb{R}^m\) and \(t_i = [t_{i1}, t_{i2}, \cdots, t_{in}] \in \mathbb{R}^n\), standard SLFNs with N’ hidden neurons as shown in Fig. 1.

The activation function \(g(x)\) is mathematically modelled as

\[
\sum_{i=1}^{N'} \beta_i g(w_i \cdot x_j + b_i) = o_j, \quad j = 1, \ldots, N
\]

where \(w_i = [w_{i1}, w_{i2}, \cdots, w_{im}]^T\) is the weight vector connecting the ith hidden neuron and the input neurons, \(\beta_i = [\beta_{i1}, \beta_{i2}, \cdots, \beta_{im}]^T\) is the weight vector connecting the ith hidden neuron and the output neurons, and \(b_i\) is the threshold of the ith hidden neuron. \(w_i \cdot x_j\) denotes the inner product of \(w_i\) and \(x_j\).

That standard SLFNs with N’ hidden neurons with activation function \(g(x)\) can approximate these N training data with zero error means that \(\sum_{i=1}^{N} \text{abs}(o_i - t_i) = 0\), i.e., there exist \(\beta, w_i\) and \(b_i\) such that

\[
\sum_{i=1}^{N'} \beta_i g(w_i \cdot x_j + b_i) = t_j, \quad j = 1, \ldots, N
\]

The above N equations can be written compactly as:

\[
H[\beta] = T
\]

where

\[
H(w_{i1}, \ldots, w_{N'}, b_{i1}, \ldots, b_{N'}, x_{i1}, \ldots, x_{N}) =
\begin{bmatrix}
g(w_{i1} \cdot x_j + b_{i1}) & \cdots & g(w_{i1} \cdot x_N + b_{i1}) \\
\vdots & \ddots & \vdots \\
g(w_{iN} \cdot x_j + b_{iN}) & \cdots & g(w_{iN} \cdot x_N + b_{iN})
\end{bmatrix}_{N \times N}
\]

\[
\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_N^T \end{bmatrix}_N \quad \text{and} \quad T = \begin{bmatrix} t_1^T \\ \vdots \\ t_N^T \end{bmatrix}_N
\]

\[
H is the hidden layer output matrix of the neural network; the i\text{th} column of \(H\) is the i\text{th} hidden neuron’s output vector with respect to inputs \(x_1, x_2, \cdots, x_N\).

In most cases the number of hidden neurons is much less than the number of distinct training samples, \(N' \ll N\), \(H\) is a nonsquare matrix and there may not exist \(w_n, b_n, \beta_i (i = 1, \cdots, N')\) such that \(H[\beta] = T\). Thus, instead one may need to find specific \(w_{i_n}^*, b_{i_n}^*, \beta^* (i = 1, \cdots, N')\) such that

\[
\|H(w_{1}, \ldots, w_{N'}, b_{1}, \ldots, b_{N'})[\beta^*] - T\| = \min_{w_i, b_i, \beta} \|H(w_{1}, \ldots, w_{N'}, b_{1}, \ldots, b_{N'})[\beta] - T\|
\]

which is equivalent to minimizing the cost function
\[ E = \sum_{i=1}^{N} \left( \sum_{m=1}^{N'} \beta_m g(w_i x_j + b_j) - t_j \right)^2 \]  

When H is unknown gradient-based learning algorithms are generally used to search the minimum of abs(Hβ−T) in the minimization procedure by using gradient-based algorithms, vector w which is the set of weights (w_i, b_i) and biases (b_i) parameters w is iteratively adjusted as follows:

\[ W_k = W_{k-1} - \eta \frac{\partial E(W)}{\partial W} \]  

Here \( \eta \) is a learning rate. The popular learning algorithm used in feedforward neural networks is the backpropagation learning algorithm where gradients can be computed efficiently by propagation from the output to the input (LeCun et al., 2012).

We can summarize ELM as follows:

Algorithm ELM : Given a training data \((x_i, t_i)\), where \(x_i = [x_{i1}, x_{i2}, \ldots, x_{in}]^T \in \mathbb{R}^n\) and \(t_i = [t_{i1}, t_{i2}, \ldots, t_{in}]^T \in \mathbb{R}^n\), activation function \(g(x)\), and hidden neuron number \(N\).

step 1: Assign arbitrary input weight \(w_i\) and bias \(b_i\), \(i = 1, \ldots, N\).

step 2: Calculate the hidden layer output matrix H.

step 3: Calculate the output weight \(\beta = H^T\) where \(H^T\) is the Moore–Penrose generalized inverse of matrix H. The orthogonal projection method can be used to calculate the Moore–Penrose generalized inverse of H in two cases: when \(H^TH\) is nonsingular and \(H^T = (H^TH)^{-1}H^T\), or when \(HH^T\) is nonsingular and \(H = (HH^T)^{-1}H\), where H, \(\beta\) and T are defined as formula (3) and (4).

4 ANT COLONY OPTIMIZATION

Ant colony optimization is based on the cooperative behavior of real ant colonies, which are able to find the shortest path from their nest to a food source (Maniezzo et al., 1996), (Dorigo et al., 1999). ACO algorithms can be applied to any optimization problems that can be characterized as follows:

1. A finite set of components \(C = \{c_1, c_2, \ldots, c_N\}\) is given.

2. A finite set of L of possible connections/transition among the elements of C is defined over a subset \(C'\) of the Cartesian product \(C \times C\), \(L = \{CG_i\}|(c_i, c_j) \in C'\}, |L| \leq N^2_e\).

3. For each \(i_{CG} \in L\) a connection cost function \(J_{i_{CG}} = J(l_{CG}, t)\), possibly parameterized by some time measure t, and is defined.

4. A finite set of constraints \(\Omega \equiv \Omega(C, L, t)\) is assigned over the elements of C and L.

5. The states of the problem are defined in terms of sequences \(S = (c_1, c_2, \ldots, c_k, \ldots)\) over the elements of C or of L. \(S\) is a subset of S. The elements in \(S\) define the problem’s feasible states.

6. A neighborhood structure is assigned as follows: the state \(s_2\) is said to be a neighbor of \(s_1\) if \(s_1\) and \(s_2\) are in \(S\) and the state \(s_2\) can be reached from \(s_1\) in one logical step, that is, if \(c_1\) is the last component in the sequence determining the state \(s_1\), it must exists \(c_2 \in C\) such that \(i_{CG} \in L\) and \(s_2 \equiv (s_1, c_2)\).

7. A solution \(\Psi\) is an element of \(S\) satisfying all the problem’s requirements. A solution is said multi-dimensional if it is defined in terms of multiple distinct sequences over the elements of C.

8. A cost \(J_\Psi(L, t)\) is associated to each solution \(\Psi\). \(J_\Psi(L, t)\) is a function of all the costs \(J_{i_{CG}}\) of all the connections belonging to the solution.

It is worth mentioning that ACO makes probabilistic decision in terms of the artificial pheromone trails and the local heuristic information. This allows ACO to explore larger number of solutions than greedy heuristics. Another characteristic of the ACO algorithm is the pheromone trail evaporation, which is a process that leads to decreasing the pheromone trail intensity over time. Pheromone evaporation helps in avoiding rapid convergence of the algorithm towards a suboptimal region (Kadri et al., 2012).

In the next section, we present our proposed Extreme Learning Machine/Binary ACO algorithm, and explain how it is used for selecting an appropriate subset of features to diagnosis an industrial system.

5 PROPOSED APPROACH

We propose in this research a new approach. It based on Extreme Learning Machine and ACO algorithm applied to fault diagnosis of a rotary kiln system. Where the best number of features is determined automatically. In this approach, each ant searches the same routine, and pheromone is left on each edge. As an intelligent object, each ant just chooses one edge of the two. The intelligent behavior of ant is very simple, and the incidence matrix traversed by each ant needs only 2 × n’s
space, which to some extent solves the descriptive difficulty generated from long coding and the reduction of solution quality.

Initially, the quantity of information in each routine is randomly generated. During the movement, ant $k$ shifts its direction according to the values of pheromone concentration $FP$ and the heuristic value $FH$ (Youn et al., 2010). The heuristic value $FH$ is computed using the Fisher discriminant criterion for feature selection, which determines the importance of each feature. The probability that an ant $k$ chooses the feature $X_i$ is given by:

$$PS_{ri} = \frac{FP_{r0} + \frac{FP_{r0}}{\text{Max}(FH_i)} FH_i}{FP_{r1} + FP_{r0}}$$  \hspace{1cm} (8)

After all ants have completed their solutions, pheromone evaporation on all nodes is triggered, and then according to Equation (2), pheromone concentration in the trails is updated.

$$FP \leftarrow (1 - \rho)FP + \Delta FP$$  \hspace{1cm} (9)

Where $\rho \in [0, 1]$ is the pheromone evaporation and $\Delta FP$ is the pheromone deposited on the trails by the ant $k$ that found the best solution for this tour:

$$\Delta FP = \frac{1}{1 + F(V) - F(V')}$$  \hspace{1cm} (10)

Where $F(V)$ represents the best solution built since the beginning of the execution and $F(V')$ represents the best solution built during the last iteration.

In this work, we have used a classification accuracy to evaluate our algorithm because it is the most universal parameter of criteria to determine the performance of the classifiers. The benefit of this measure is its simplicity; the drawback is that it can be deceptive. In the proposed technique, the classification accuracy was calculated by using the following formula:

Classification Accuracy = Correct classified patterns/Total number of patterns

The heuristic value is computed using the Fisher discriminant criterion for feature selection. Considering a classification problem with $M$ possible classes, the Fisher discriminant criterion is described as follow:

$$FH(\alpha) = \sum_{c=1}^{M} \sum_{r=1}^{C} \frac{m_c(\alpha) - m_r(\alpha)}{N_c \sigma_c^2 - N_r \sigma_r^2}$$  \hspace{1cm} (11)

Where:

$M$ represents the number of classes;

$m_c(\alpha)$ represents the center of gravity of the class number $c$ by considering only the parameter $\alpha$ it is calculated as follows:

$$m_c(\alpha) = \frac{1}{N_c} \sum_{i=1}^{N} X_{cv}(\alpha)$$  \hspace{1cm} (12)

With $X_{cv}$ is the number $v$ of the class number $c$, the value of NR equal to the number of vectors of the class in question is the vector.

$\sigma_r^2(\alpha)$ is the variance of the component $\alpha$ of the vectors of the class number $C$.

$$\sigma_r^2(\alpha) = \frac{1}{N} \sum_{i=1}^{N} \left[ X_{cv}(\alpha) - m_c(\alpha) \right]^2$$  \hspace{1cm} (13)

Algorithm 1 presents the description of the Binary ACO feature selection algorithm.

### Algorithm 1. Binary ACO feature selection algorithm

1. Initiate the phenomenon of the net;
2. Compute the $FH(\alpha)$ using (11);
3. Ants search using (8);
4. Evaluate the solutions founded by ant colony algorithm using classification Accuracy, and reserve the optimal;
5. Upgrade phenomenon of the net by optimal solution using (2);
6. Judge whether the stopping condition of the qualification is met, if qualified, ends; otherwise, goto 3;

The time complexity of proposed algorithm is $O(Im)$, where $I$ is the number of iterations, and $m$ the number of ants. This can be seen from algorithm 1. In the worst case, each ant selects all the features. As the objective function is evaluated after all ants have completed their solutions, this will result in $m$ evaluations. After $I$ iterations, the objective function will be evaluated $Im$ times.

### 6 ROTARY KILN SYSTEM

Cement is a soft grayish powder. It is made of ground gypsum and clinker which itself is produced from a mixture composed of 75% lime and 25% silica. Cement was primary invented by the Egyptians. It forms a thick paste when mixed with water.

Three different processes are used in the portland cement industry to accomplish the pyroprocessing step, as shown in Fig. 2.

Today, there are two types of Portland cement, which are Portland blast furnace slag cement and Portland pozzolana cement. Raw materials used in
cement production are Lime-containing materials, such as limestone and Clay and clay-like materials, such as shale. We used more than 1.5 ton of raw materials to make one ton of Portland cement. The blend of materials is ground in a raw mill. Then the mixture of limestone is burned in a chemical reactor called rotary kiln at temperatures around 4482 degrees Celsius to form clinker. The clinker nodules are then ground with about 3% gypsum to produce cement with a fineness typically of less than 90 micrometers.

The major disadvantage of a rotary kiln is not automatically turned off with the occurrence of a fault because the conventional automation system cannot detect all abnormal conditions. It can decrease productivity or even irreparable damage to the plant. Therefore, it is essential to use a diagnostic method to determine spoilage and avoid such damage Huntzinger & Eatmon, 2009) (Kääntee et al., 2004).

Figure 2. Cement Manufacturing Process

In this paper, in order to continue the previous attempts for fault detection in kiln, for the first time, in the process of identification, prediction and detection, we used a recent method of classification, which is ELM. This method can be implemented in a single learning step, so it runs fast. We have been collecting and analyzing the data of a thirteen-week period of operation of the kiln system. According to the collected, we found four types of fault condition, which are: Disrupted operation, Moving area, Inferior product, and Energy loosed. After removing invalid data and preprocessing on them that are divided to two parts: 50% of that is used as the training set and 50% as the test.

7 EXPERIMENTAL RESULTS

7.1 Datasets used

In this experiment, we have compared the results of our approach with those obtained by an ACO-SVM algorithm using two industrial datasets (rotary kiln System).

Figure 3. RCK1 dataset of clinkering system

The dataset RCK1 consists of 200 recordings which represent 4 classes. Each observation comprises 30 attributes (Fig. 3). The second dataset is RCK2 that consists of 500 recordings which represent 2 classes. Each observation comprises 47 attributes.

To prove the performance of our approach, tests are also conducted on other four datasets from the University of California, Irvine (UCI) machine learning repository. Specifications of the four classification datasets are shown in Table 1.

Table 1. Specifications of classification datasets

<table>
<thead>
<tr>
<th>Datasets</th>
<th>classes</th>
<th>attributes</th>
<th>instances</th>
<th>train</th>
<th>test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diabetes</td>
<td>2</td>
<td>8</td>
<td>768</td>
<td>576</td>
<td>192</td>
</tr>
<tr>
<td>Iris</td>
<td>3</td>
<td>4</td>
<td>150</td>
<td>102</td>
<td>48</td>
</tr>
<tr>
<td>Breastcan</td>
<td>2</td>
<td>30</td>
<td>569</td>
<td>301</td>
<td>268</td>
</tr>
<tr>
<td>Vowel</td>
<td>11</td>
<td>10</td>
<td>528</td>
<td>990</td>
<td>462</td>
</tr>
</tbody>
</table>

For ACO algorithm, in all simulation, the agents number N_A is 20; the random rate of behavior F_a was set to 0.2; the evaporation rate ρ was set to 0.3. These parameters are fixed after the execution of several simulations by using as entered a restricted dataset.

As mentioned, the heuristic factor FH was used to compute the probability PS. This factor didn’t affect the ants which have a random behavior. We have used this kind of ants to discover a new research spaces.
We have used kalman filter as heuristic factor. This filter is developed as a feature selection method and classifier for multivariate data. The feature has the biggest value of kalman filter will be appeared in all finals subsets.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Features</th>
<th>RCK1</th>
<th>RCK2</th>
<th>RCK1</th>
<th>RCK2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aco-svm</td>
<td>Generated subset</td>
<td>89%</td>
<td>0.7700</td>
<td>85%</td>
<td>0.0210</td>
</tr>
<tr>
<td></td>
<td>One feature</td>
<td>25%</td>
<td>0.032</td>
<td>54%</td>
<td>0.0061</td>
</tr>
<tr>
<td></td>
<td>All features</td>
<td>93%</td>
<td>0.787</td>
<td>90%</td>
<td>0.0218</td>
</tr>
<tr>
<td>Aco-elm</td>
<td>Generated subset</td>
<td>90%</td>
<td>0.780</td>
<td>86%</td>
<td>0.0214</td>
</tr>
<tr>
<td></td>
<td>One feature</td>
<td>26%</td>
<td>0.032</td>
<td>55%</td>
<td>0.0061</td>
</tr>
<tr>
<td></td>
<td>All features</td>
<td>95%</td>
<td>0.787</td>
<td>93%</td>
<td>0.0218</td>
</tr>
</tbody>
</table>

7.2 Results

We used as a classification quality evaluation criteria, the number of well classified observations on the total number of observations.

The following table shows the quality of classification while using:

a) The best discriminating feature;
b) The best subset of features generated;
c) All features.

Table 3. Description of selected feature subset

<table>
<thead>
<tr>
<th>Number</th>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A54P2</td>
<td>Cyclone Pressure A54t.</td>
</tr>
<tr>
<td>2</td>
<td>A50T1</td>
<td>Cyclone gas outlet temperature A50</td>
</tr>
<tr>
<td>3</td>
<td>U01T1</td>
<td>Clinker temperature</td>
</tr>
<tr>
<td>4</td>
<td>A54T2</td>
<td>Material temperature cyclone A54</td>
</tr>
<tr>
<td>5</td>
<td>TV</td>
<td>Kiln shell temperature</td>
</tr>
<tr>
<td>6</td>
<td>A53T1</td>
<td>Temperature gas cyclone A53</td>
</tr>
<tr>
<td>7</td>
<td>K01T1</td>
<td>Secondary air temperature</td>
</tr>
<tr>
<td>8</td>
<td>COC</td>
<td>Cyclone outlet CO content A50</td>
</tr>
<tr>
<td>9</td>
<td>A54T1</td>
<td>Temperature gas cyclone A54</td>
</tr>
<tr>
<td>10</td>
<td>A52P2</td>
<td>Cyclone Pressure A52.</td>
</tr>
<tr>
<td>11</td>
<td>V07P1</td>
<td>Primary air pressure</td>
</tr>
<tr>
<td>12</td>
<td>A53T2</td>
<td>Material temperature cyclone A53</td>
</tr>
<tr>
<td>13</td>
<td>O2C</td>
<td>Teneur O2 sortie cyclone A50</td>
</tr>
<tr>
<td>14</td>
<td>W01S1</td>
<td>Speed oven</td>
</tr>
<tr>
<td>15</td>
<td>COP</td>
<td>CO content smoke box</td>
</tr>
<tr>
<td>16</td>
<td>A53P1</td>
<td>Cyclone Pressure A53.</td>
</tr>
<tr>
<td>17</td>
<td>A52T1</td>
<td>Cyclone gas outlet temperature A52</td>
</tr>
<tr>
<td>18</td>
<td>V31F1</td>
<td>Gas flow</td>
</tr>
<tr>
<td>19</td>
<td>V01F1</td>
<td>Gas flow</td>
</tr>
<tr>
<td>20</td>
<td>W01X1</td>
<td>Oven time</td>
</tr>
</tbody>
</table>

Figure 4. The best solution in 20 iterations (RCK 1)

Using the parameters presented in previous section, the following results were obtained by taking the best solution after 20 trials. The Table 2 gives the best solutions obtained for each dataset (RCK1 and RCK2). For the two datasets, the FV of the best solution is indicated with the corresponding accuracy.

Table 2 shows that ELM classifier can diagnosis and detects the operation condition of the rotary kiln system with good accuracy using the subset generated by our algorithm. It is also noticed that the value of FV reflects well the quality of classification.

The results of the ACO-ELM method verify a 10% incorrect detection, which is equivalent to the 90% correct detection of fault condition. This result is better than those emerged through the ACO-SVM method. The selected subset feature, that produces those results, is presented in table 3.

The objective function of the ACO algorithm is shown in Fig. 4 for all iterations from the start to the end of the execution. It shows that for this dataset, after a few iterations, the optimal results can be obtained. The reason of this fast convergence to the
optimal results is the pheromone density which is updated at the end of each round.

The performance of ACO-ELM and ACO-SVM on the four datasets is listed in Table 4, including the execution time, parameters C and \( \gamma \) and the best classification accuracy.

It can be found from Table 4 and Fig.4 that the execution time of ACO-ELM is far less than that of ACO-SVM algorithm and the best classification accuracy obtained by ACO-ELM is even higher than that derived by ACO-SVM algorithm.

8 CONCLUSION

In this paper, a new method combined Ant Colony algorithm with Extreme Learning Machines is proposed to verify the operating condition of a rotary kiln System. The goal is to select the best subset that is sufficient to perform a good classification and obtain acceptable error rate. Comparing our technique with the approach based on ACO-SVM algorithm for fault diagnosis problem, the experimental results have shown that ACO-ELM achieves better performance and is less time-consuming compared with ACO-SVM. In future work, it will be interesting to evaluate this approach with an early diagnosis database and new tests oriented toward prognosis.

9 ACKNOWLEDGEMENTS

The research has been generously supported by Khenchela University, project number B*03520100014. The authors would like to express their sincere appreciation for all support provided.

10 REFERENCES