

ICA-Multiclass SVM as a Monitoring System of Complex Processes

Jesús Alejandro Navarro Acosta¹, Juan Pablo Nieto González¹, and Víctor Manuel Cortés Figueroa²

¹ Corporación Mexicana de Investigación en Materiales, S.A. de C.V. COMIMSA, Mexico

² Vertex, S.A. de C.V., Mexico

jesus.navarro@comimsa.com, juan.nieto@comimsa.com, vmcortesf@gmail.com

Abstract. Nowadays due to advancement of technology, the processes have become increasingly complex. So perform effectively monitor and correct diagnosis of these, has become a challenging task. This because the current processes have in common presence of combinations and correlation between analog and digital variables, presence of noise, etc. This issued is overcome by multivariate statistical methods which can efficiently represent the different correlations between variables. One of the most popular methods for process monitoring is the Hotelling's T^2 distance but a large amount of data makes the process monitoring a difficult task. Principal Component Analysis (PCA) is used for the data reduction which means the extraction of few numbers representing the most variance of the analyzed data. However the use of these two methods is limited to normal multivariate data set and in industry rarely normal data set is presented, hence different methods are used for this purpose. Independent component analysis (ICA) has been used in no-normal multivariate processes. In this paper a monitoring and fault diagnosis system for complex multivariate processes is presented. The new proposal is based on Multiclass Support Vector Machines (MCSVM) and the case study was addressed to the automotive industry. A comparison against a similar technique is carried out. Simulation of the new proposal shows promisory results.

Keywords: Support vector machines, independent component analysis, process monitoring.

1 Introduction

Since today's customer's satisfaction has become a priority, the manufacturing of high quality products is what every industry is looking for. In order to achieve this, different statistical control tools have been applied such as $\bar{X} - R$, $\bar{X} - S$. However, these tools are limited because their one-dimensional nature besides the normality requirement over the data. In several practical cases these

methodologies are not suitable because of the correlation among the variables is not taken into consideration. The multivariate statistical process control provides certain advantages over the univariate models, since it considers the correlation between each pair of variables. Hotelling's T^2 distance is the most used tool in the multivariate statistical analysis, however large amounts of data makes the task difficult and time consuming. It is said that a big problem to tackle is the "information overload" [8] therefore a method that reduces the dimension of the data without losing a great deal of information is required. The Hotelling distance and PCA are combined for industrial process monitoring where the corresponding control limit is determined assuming that latent variables obey a Gaussian distribution. Unlike PCA, ICA has no restrictions of orthogonality which not only allows decorrelation of variables but also considers a higher order statistics to make independent latent variables. Therefore, ICA can be applied to a non-Gaussian process. On the other hand, support vector machines (SVM) based on statistical learning theory and principle of structural risk minimization, have been applied in the field of fault recognition for its excellent ability of generalization [10]. In this paper a system for monitoring and fault diagnosis for a multivariate process based on multiclass support vector machines (MCSVM) is presented, this methodology is applied in the automotive industry, the data were obtained by using the simulator VEHDYNA. The results of this system are compared with those obtained by implementing ICA as proposed in [6]. The organization of this paper is as follows: In Section 2 the techniques are explained. Section 3 presents the description of the systems to compare. In Section 4 the analysis and results are shown. Section 5 presents the conclusions.

2 Techniques

2.1 Independent Component Analysis

Independent Component Analysis (ICA) is a projection method which finds underlying components from multivariate statistical data. The difference between ICA and other projection methods is that ICA can handle with non-Gaussian or independent statistic data. ICA was originally proposed to solve the blind source separation problem, which involves recovering independent source signals after they have been linearly mixed by unknown matrix A .

In the ICA algorithm, it is assumed that d measured variables x_1, x_2, \dots, x_d can be expressed as linear combinations of $m (\leq d)$ unknown independent components s_1, s_2, \dots, s_m . The independent components and the measured variables have means of zero. The relationship between them is given by Equation 1.

$$X = AS + E \quad (1)$$

where $X = [x_1, x_2, \dots, x_n] \in R^{d \times n}$ is the data matrix (ICA employs the transposed data matrix), $A = [a_1, \dots, a_m] \in R^{d \times m}$ is the unknown mixing matrix, $S = [s_1, s_2, \dots, s_n] \in R^{m \times n}$ is the independent component matrix, $E \in R^{d \times n}$ is the residual matrix, and n is the number of samples. Here, we assume $d \geq m$

(when $d = m$, the residual matrix E becomes the zero matrix). The main problem of ICA is to estimate both the mixing matrix A and the independents S from only the observed data X . Now we can define the objective of ICA as follows: to find a demixing matrix W whose form is such that the rows of the reconstructed matrix \hat{S} becomes as independent of each other as possible (see Equation 2).

$$\hat{S} = WX \quad (2)$$

This formulation is not really different from the previous one, since after estimating A , its inverse gives W when $d = m$. It will assume $d = m$, unless otherwise specified. For mathematical convenience, we define that the independent components have unit variance. This makes the independent components unique, up to their signs. The initial step of ICA is whitening, also known as sphering, which eliminates all the cross-correlation between random variables. Consider a d -dimensional random vector x_k at sample k with covariance $R_x = E(x_k x_k^T)$ where E represent expectations. The eigen decomposition of R_x is given by Equation 3.

$$R_x = U \Lambda U^T \quad (3)$$

The whitening transformation is expressed as:

$$z_k = Q x_k \quad (4)$$

where $Q = \Lambda^{-1/2} U^T$ [13]. One can easily verify that $R_z = E(z_k z_k^T)$ is the identity matrix under this transformation. After the transformation we have:

$$z_k = Q x_k = Q A s_k = B s_k \quad (5)$$

where B is an orthogonal matrix as verified by the following relation:

$$E\{z_k z_k^T\} = B E\{s_k s_k^T\} B^T = B B^T = I \quad (6)$$

Therefore, the problem is reduced to finding an arbitrary full-rank matrix A to the simpler problem of finding an orthogonal matrix B , since B has fewer parameters to estimate as a result of orthogonality constraint. Then, s_k from Equation 5 can be estimated as follows:

$$\hat{s}_k = B^T z_k = B^T Q x_k \quad (7)$$

The relation between W and B from Equations 2) and 7) can be expressed as:

$$W = B^T Q \quad (8)$$

Hence Equation (2) can be rewritten as:

$$\hat{s} = WX = B^T Z = B^T Q X = B^T \Lambda^{-1/2} U^T \quad (9)$$

To calculate B , a fast fixed-point algorithm for ICA (FastICA) is used. This algorithm calculates the column vector $b_i (i = 1, 2, \dots, m)$ of B through iterative

steps [7]. After obtaining B , can calculate \hat{s} by using equation (2) and W from equation (8). To divide W into two parts, dominant part(W_d) and excluded part (W_e). Lee in [9] propose three statistics for process monitoring as follows:

$$I^2 = \hat{S}_d^T \hat{S}_d \quad (10)$$

$$I_e^2 = \hat{S}_e^T \hat{S}_e \quad (11)$$

$$SPE = e^T e = (x - \hat{x})^T (x - \hat{x})$$

where $\hat{S}_d = W_d X$, $\hat{S}_e = W_e X$ and $\hat{x} = Q^{-1} B_d \hat{S} = Q^{-1} B_d W_d X$. In PCA monitoring, the latent variables are assumed to be Gaussian distributed, hence the upper control limit for T^2 can be determined by using Equation 4. In ICA, I^2 , I_e^2 and SPE depart from the normality assumption. Lee et al. therefore proposed to use KDE to determine the control limits for I^2 , I_e^2 and SPE statistics for ICA monitoring. Additionally, the variable contributions of $x(t)$ for $I^2(t)$ and $I_e^2(t)$ can be defined as:

$$X_{cd}(t) = \frac{Q^{-1} B_d \hat{S}_d(t)}{\|Q^{-1} B_d \hat{S}_d(t)\|} \|\hat{S}_d(t)\| \quad (12)$$

$$X_{ce}(t) = \frac{Q^{-1} B_e \hat{S}_e(t)}{\|Q^{-1} B_e \hat{S}_e(t)\|} \|\hat{S}_e(t)\| \quad (13)$$

After extracting essential features, it requires introducing a method for the classification problem. SVM is used in order to achieved this objective.

2.2 Support Vector Machines

Support Vector Machines (SVM) is a machine learning algorithm used for classification and regression of high dimensionality data sets. This provides great results due to their ability to deal with problems such as local minimum, few data samples and nonlinear problems. Standard SVM are used for binary classification problems [3]. The central idea of this technique is to determine a linear separation (separating hyperplane) which is oriented in a way that its distance to the nearest data points in each class is maximized. The nearest data points are known as support vectors [1].

The Linearly Separable Case The input vectors $\mathbf{x}_i \in \mathbb{R}^d$ ($i = 1, 2, \dots, n$) correspond with labels $\mathbf{y}_i \in \{-1, +1\}$. There exists a separating hyperplane and its function is as:

$$\mathbf{w} \cdot \mathbf{x} + b = 0 \quad (14)$$

where $\mathbf{w} \in \mathbb{R}^n$ is a normal vector, the bias b is a scale and $\frac{|b|}{\|\mathbf{w}\|}$ represents the perpendicular distance from the separating hyperplane to the origin. While the

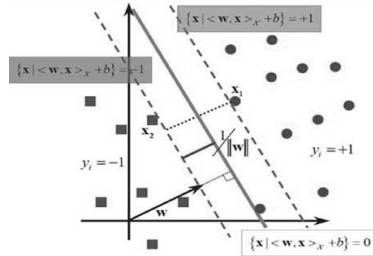


Fig. 1. Definition of geometric hyperplane and margin. Adapted from [2].

distance from hyperplane to the margin is given by $d = \frac{1}{\|\mathbf{w}\|}$. Figure 1 shows the separating hyperplane and the maximum margin between classes (dotted lines).

Two parallel hyperplanes can be represented as shown in Equation 15.

$$\mathbf{y}_i (\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 \quad (15)$$

As defined above SVM seeks to maximize the distance between two classes, where the amplitude of the margin between two parallel hyperplanes is $d = \frac{2}{\|\mathbf{w}\|}$. Hence for the linearly separable case can find the optimal hyperplane by solving the following quadratic optimization problem:

$$\min \frac{1}{2} \|\mathbf{w}\|^2 \quad (16)$$

$$\text{subject to: } \mathbf{y}_i (\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1$$

By introducing Lagrange multipliers $\alpha_i (i = 1, 2, \dots, n)$ for the constraint, the primal problem (17) becomes a task of finding the saddle point of Lagrange. Thus, the dual problem becomes:

$$\max L(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j \mathbf{y}_i \mathbf{y}_j \quad (17)$$

$$\text{subject to: } \sum_{i=1}^n \alpha_i \mathbf{y}_i = 0, \alpha_i \geq 0$$

By applying Karush–Kukn–Tucker (KKT) conditions, the following relationship holds:

$$\alpha_i [\mathbf{y}_i (\mathbf{w} \cdot \mathbf{x}_i + b) - 1] = 0 \quad (18)$$

If $\alpha_i > 0$, the corresponding data points are called support vectors (SVs). Hence, the optimal solution for the normal vector is given by:

$$\mathbf{w}^* = \sum_{i=1}^N \alpha_i \mathbf{y}_i \mathbf{x}_i \quad (19)$$

where N is the number of SVs. From Eq. (18), by choosing any SVs $(\mathbf{x}_k, \mathbf{y}_k)$, $b^* = \mathbf{y}_k - \mathbf{w}^* \cdot \mathbf{x}_k$. After (\mathbf{w}^*, b^*) is determined. From the equation (19) the optimal separating hyperplane can be written as in [4,11]:

$$f(\mathbf{x}) = \text{sign} \left(\sum_{i=1}^N \alpha_i \mathbf{y}_i (\mathbf{x} \cdot \mathbf{x}_i) + b^* \right) \quad (20)$$

$$\mathbf{x} \in \begin{cases} +1 & \text{si } f(\mathbf{x}) > 0 \\ -1 & \text{si } f(\mathbf{x}) < 0 \end{cases}$$

The Linearly Non-Separable Case SVM can also handle cases where the data are not linearly separable. These attempt to map the input vector $\mathbf{x}_i \in \mathbb{R}^d$ on a higher dimensional space. This process is based on the selection of a kernel function. Some of the most used kernel (\mathbf{K}) functions are [6]:

- Linear Kernel
- Polynomial Kernel
- Radial basis function
- Sigmoid kernel

Hence, the optimal hyperplane (eq. (21)) takes the form as.

$$f(\mathbf{x}) = \text{sgn} \left(\sum_{i=1}^N \alpha_i \mathbf{y}_i \cdot \mathbf{K}(\mathbf{x} \cdot \mathbf{x}_i) + b^* \right) \quad (21)$$

Multiclass SVM As mentioned previously, SVM is able to classify into two classes only. Among the most applied methods for achieving multi classification with SVM are known as one against all (OAA) which builds M SVM where M is equal to the number of classes to be classified, after each of these SVM separates one class from the rest. The i th MSV is trained with all the training samples of the i th class with a different label to the other. Another method is known as one against one (OAO), which builds $M * (M - 1)/2$ MSV. These combine their classification function to determine the class to which the test sample belongs, that predictions by accumulating round (votes), prediction with the most votes shall be the final classification [5,12].

2.3 Confusion Matrix

In the area of artificial intelligence, the confusion matrix is a visualization tool used in supervised learning. Each column of the matrix represents the number of predictions for each class, and each row represents each real class. The confusion matrix facilitates observe if the system is confusing two classes.

3 System Description

3.1 Independent Component Analysis for Monitoring Process

The feature extraction based on ICA is used to project the high dimension dataset into a lower one $m \leq d$. The extracted independent components are then used to calculate the systematic part statistic (in this paper will be used only (I^2) statistic), thus performs a process monitoring. Detailed explanation and steps of the algorithm can be reviewed at [9].

3.2 MCSVM for Fault Diagnosis

The monitoring system based on Support Vector Machines requires databases both normal operation and failure mode for learning. As shown in Figure 2, the system does not have any pre-processing. Once generated these databases serve as input for training and system validation. If system performance is suitable, this is ready for test data.

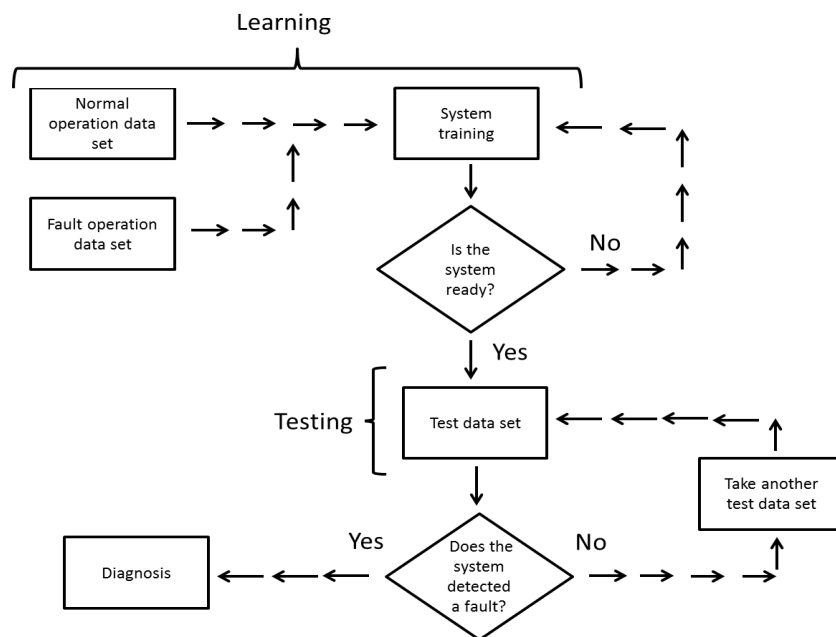


Fig. 2. Methodology based on MCSVM.

4 Analysis and Results

Five variables are monitored (A,B,C,D,E) these correspond to five sensors which monitor different parts of a vehicle. The simulations were carried out on the vehicle maneuver known as chicane. Chicane refers to a change in the straight path of the vehicle, for example, when the vehicle performs an overtake maneuver. The aim is detect and diagnose failures in the sensors. Databases both in normal operation and fault modes are generated: the first one consists of 4400 observations for the 5 variables and the second database is constructed from data for each of the five possible faults. 1600 observations for each variable into failure mode were taken, forming a 12400×5 size matrix. It is considered that a fault exists when in the sensor occur changes its face value outside a range of $\pm 5\%$, the Fig 3 shows an example of a failure in the variable A, starting at observation 21 and ending at observation 32.

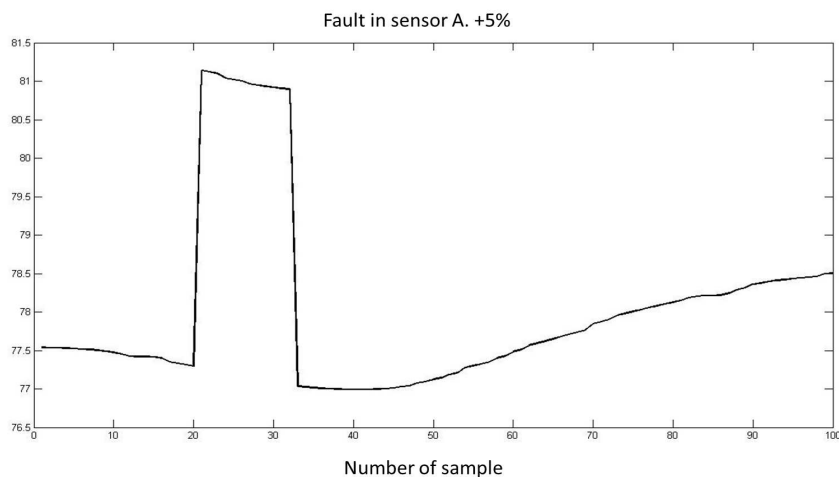


Fig. 3. Fault in sensor A.

4.1 ICA Performance

ICA is implemented to obtain the independent components and thereby can compute the statistics described in [9]. Then, the univariate kernel density estimator is used to estimate the density function of the normal statistics values and therefore, it was able to compute the control limit of normal operating data. At this stage the normal operation database is used, the control limit value is 4.6685. In the next stage the system is ready to monitor new samples. First data under control are monitored. Figure 4 shows state of the process after analyze

100 samples. It can be seen that the control chart show the process under the control limit computed by ICA algorithm.

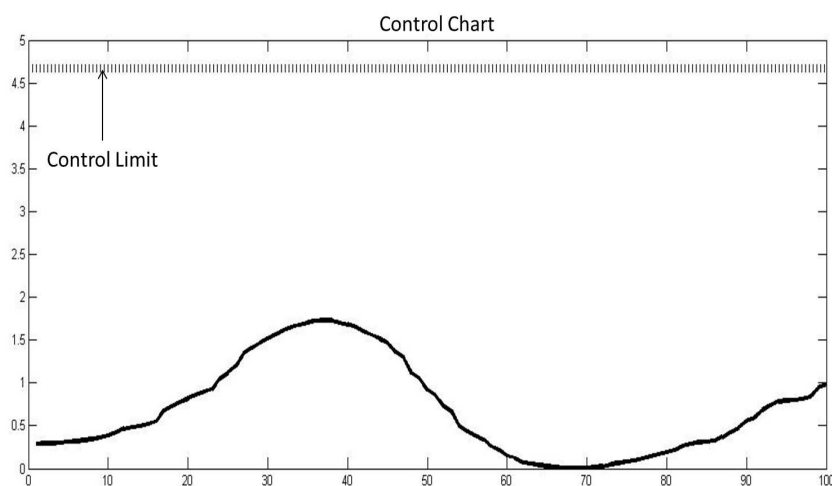


Fig. 4. Control chart for normal operation sample.

Now for the following step, the fault mode database is used to test the ICA algorithm's performance. Figure 5 shows state of the process after analyze the described failure above on the sensor A. It can see that the control chart show the process over control limit calculated by the ICA algorithm.

4.2 MCSVM Performance

A matrix size of 12400×5 and a vector of $size = 12400 \times 1$ are used for training and validation. Such vector represents six different labels for the six possible states of the system (0,1,...,5 for normal, fault in sensor A,...,E). The approximate computation time for training was 6.3 seconds, while the performance of the classifier was 99.95%. The model for the classification used a radial kernel and found a total of 1612 support vectors. Table 1 shows the confusion matrix.

Figure 6 shows the analysis of failure in the variable A of the MCSVM based system. We can observe that system has detected, located and identified the fault at 100%.

The results obtained by performing 200 simulations of single failures (one variable at a time) in different locations and at different magnitudes of deviation from the nominal values of the monitored variables are presented. Table 2 shows the performance of MCSVM system for fault diagnosis. Table 3 shows the performance of ICA system for fault diagnosis (only use I^2 statistic). Table 4 presents the comparison of qualitative features of both methods.

Table 1. Confusion matrix.

	0	1	2	3	4	5
0	899	0	0	0	0	0
1	0	313	0	0	0	0
2	0	0	324	0	0	0
3	0	0	0	331	0	0
4	1	0	0	0	322	0
5	0	0	0	0	0	290

Table 2. Performance of MCSVM system for fault diagnosis.

Deviation	Faulty Samples	Detection	Identification	Location
5%	15	100%	100%	100%
5%	10	100%	99.6%	100%
5%	5	100%	100%	100%
3%	15	100%	99.4%	100%
3%	10	100%	100%	100%
3%	5	100%	100%	100%

Table 3. Performance of ICA system for fault diagnosis.

Deviation	Faulty Samples	Detection	Identification	Location
5%	15	100%	100%	100%
5%	10	100%	100%	100%
5%	5	100%	100%	100%
3%	15	0%	0%	0%
3%	10	0%	0%	0%
3%	5	0%	0%	0%

Table 4. Features of both methods.

MSVMC system	ICA process monitoring
Need databases in normal operation and failure mode	Only needs database in normal operation
Requires training time	Only requires computation time
Able to operate online	Able to operate online
By nature of the algorithm is able to detect failure variable	Requires to build the contribution charts
Capable of handling noise	Requires the calculation of other statistics to sensitize

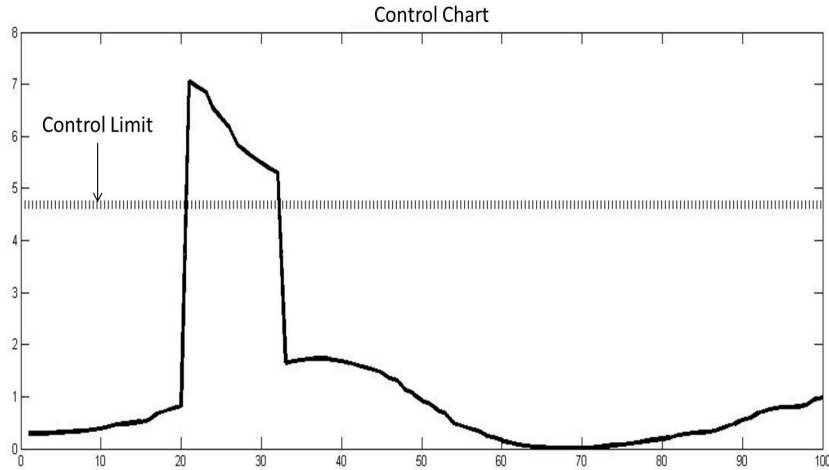


Fig. 5. Control chart for fault operation sample.

```
> predict(model, x)
 1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20
 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40
 1  1  1  1  1  1  1  1  1  1  1  1  1  0  0  0  0  0  0  0
41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60
 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80
 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100
 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
Levels: 0 1 2 3 4 5
```

Fig. 6. MCSVM analysis of failure in the variable A.

5 Conclusions

In this paper a system for detection and fault diagnosis based on support vector machines has been presented. The basic idea of this system is to use the multiclass approach of support vector machines algorithm and to implement it on complex systems such as an automotive process. The results showed that the proposed system has promising results in the field of fault detection, identification and location. The performance of the proposal is improved due to the high efficiency of classification of support vector machines and their ability to work with large data. In addition, the ICA approach as process monitoring proposed in [6] was implemented, obtaining high performance on process monitoring. Because ICA can handle with non-Gaussian data, it is ideal for monitoring automotive processes. Both methods provide excellent results, but the simplicity in the design of the system based on MCSVM and the results presented in this

paper, concludes that the Support Vector Machines are a robust technique for classification tasks and has excellent results in the field of fault detection.

References

1. Andre, A.B., Beltrame, E., Wainer, J.: A combination of support vector machine and k-nearest neighbors for machine fault detection. *Applied Artificial Intelligence: An International Journal* Vol. 27, 36–49 (2013)
2. Daz, E.E.G.: Boosting support vector machines. Master's thesis, University of the Andes School of Engineering, Bogot Colombia (2010)
3. Elhariri, E., El-Bendary, N., Fouad, M.M.M., Platos, J., Hassanien, A.E.: Multi-class svm based classification approach for tomato ripeness. *Innovations in Bio-inspired Computing and Applications* pp. 175–186 (2014)
4. Gong Yanjie, Gao Xuejin, W.P.y.Q.Y.: Online modeling method based on dynamic time warping and least squares support vector machine for fermentation process. 8th World Congress on Intelligent Control and Automation. July 6-9, Jinan, China pp. 481–485 (2010)
5. Hassen Keskes, A.B., Lachiri, Z.: Broken rotor bar diagnosis in induction machines through stationary wavelet packet transform and multiclass wavelet svm. *Electric Power Systems Research*, Elsevier pp. 151–157 (2013)
6. Hsu Chun-Chin, C.M.C., Long-Sheng, C.: Intelligent icasvm fault detector for non-gaussian multivariate process monitoring. *Elsevier* 37, 3264–3273 (2010)
7. Hyvärinen, A., Köster, U.: Fastisa: A fast fixed-point algorithm for independent subspace analysis. *European Symposium on Artificial Neural Networks*, Bruges, Belgium, (2006)
8. Hyvrinen, A., Karhunen, J., Oja, E.: *Independent Component Analysis*. Wiley-Interscience publication (2001)
9. Jong-Min Lee, Chang, K.Y., Lee, I.B.: Statistical process monitoring with independent component analysis. *Journal of Process Control* Vol. 14, 467–485 (2004)
10. Liu, B., Hou, D., Huang, P., Liu, B., Tang, H., Zhang, W., Chen, P., Zhang, G.: An improved pso-svm model for online recognition defects in eddy current testing. *Nondestructive Testing and Evaluation* Vol. 28, No. 4, 367–385 (2013)
11. Nicholas I. Sapankevich, R.S.: Time series prediction using support vector machines. *IEEE computational intelligence magazine* pp. 24–37 (2009)
12. Oscar Castillo, L.X.y.S.L.A.: *Trends in intelligent systems and computer engineering*. Springer Science + Business Media. New York, USA (2008)
13. Roshan Joy Martis, U.R.A.: Ecg beat classification using pca, lda, ica and discrete wavelet transform. *Biomedical Signal Processing and Control*, Elsevier Vol. 8, 437–448 (2013)