A Novel Feature Selection method for Fault Detection and Diagnosis of Control Valves

Binoy B. Nair[1], Vamsi Preetam M. T.[2] , Vandana R. Panicker[3] , Grishma Kumar V.[4] and Tharanya A.[5]

[1] Assistant Professor (Sr.),
Department of Electronics and Communication Engineering
Amrita Vishwa Vidyapeetham,
Coimbatore, Tamilnadu, India

[2] Department of Electronics and Communication Engineering
Amrita Vishwa Vidyapeetham,
Coimbatore, Tamilnadu, India

[3] Department of Electronics and Communication Engineering
Amrita Vishwa Vidyapeetham,
Coimbatore, Tamilnadu, India

[4] Department of Electronics and Communication Engineering
Amrita Vishwa Vidyapeetham,
Coimbatore, Tamilnadu, India

[5] Department of Electronics and Communication Engineering
Amrita Vishwa Vidyapeetham,
Coimbatore, Tamilnadu, India

Abstract

In this paper, a novel method for feature selection and its application to fault detection and Isolation (FDI) of control valves is presented. The proposed system uses an artificial bee colony (ABC) optimized minimum redundancy maximum relevance (mRMR) based feature selection method to identify the important features from the measured control valve parameters. The selected features are then given to a naïve Bayes classifier to detect nineteen different types of faults. The performance of the proposed feature selection system is compared to that of six other feature selection techniques and the proposed system is found to be superior.

Keywords: *Feature Selection, Control Valves, Fault Detection and Diagnosis, Artificial bee colony, Feature selection, naïve Bayes.*

1. Introduction

Control valves are extensively used in industry to control various parameters such as flow, temperature, pressure, liquid level etc. For this reason it is of vital importance that its condition is monitored continuously and deviations in its function be noted to prevent and control hazardous consequences that may follow. Timely fault detection and diagnosis in control valves can be used to develop maintenance strategies and consequently, the plant's overall downtime and hence, the resulting maintenance costs can be brought under control.

Fast Fourier Transforms (FFTs) are one of the oldest methods for FDI and have been widely used in fault detection. FFTs have been used for fault detection in gas turbine engines [1], rotor bars [2], [3] and induction motors [4]. Statistical techniques like multivariate statistical projection method (MSPM) [5] and dynamic PCA have been used [6] with varying degrees of success. Signal processing techniques like the Kalman filter [7], [8] have also been employed. Application of data mining techniques like support vector machines (SVMs) [9], [10], [11], [12], artificial neural networks (ANNs) have also been widely used [13], [14], [15], [16], [17], [18], [19], [20], [21], [22].

However, it is seen that the features required for the purpose of classification of faults are usually selected based on expert knowledge rather than automatically. A suboptimal feature set can compromise the accuracy of the classification system, leading to poor performance of the system. In the present study, a feature selection mechanism which can identify the importance of features from a large feature set is proposed. The performance of the proposed system is compared to that of six other feature selection techniques and the proposed system was found to outperform all the other techniques considered, by producing highest classification accuracy for the smallest number of faults. The dataset used for validating the proposed system is the Development and Application of Methods for Actuator Diagnosis in Industrial Control Systems (DAMADICS) standard benchmark dataset.

The rest of the paper is organized as follows: section 2 presents an overview of the DAMADICS benchmark, section 3 presents the design of the proposed system and the results are presented in section 4.

2. DAMADICS

This section presents the overview of the DAMADICS benchmark dataset used in the present study.

DAMADICS benchmark was developed for real time training of an actuator system [23],[24]. This benchmark has become a standard for analyzing wide range of FDI methods in terms of standard performance. The DAMADICS benchmark data was designed for comparing various FDI methods by real time testing on industrial actuators in the Lublin sugar factory in Poland. The benchmark is based on the complete working of electro – pneumatic valve actuator used in almost all industrial applications. The testing was performed by inducing abrupt (sudden) and incipient (gradually developing) faults to the actuators and recording the data.

The structure of benchmark actuator system [23], [24] is given in Fig. 1. For designing the benchmark data, five available measurements and one control value signal have been considered (measurements being made at every second). They are: process control external signal CV, values of liquid pressure on the valve inlet P1' and outlet P2', stem displacement X', liquid flow rate F' and liquid temperature T'. The apostrophe denotes signals that are measured. The set of main variables used in benchmark, as given in Fig. 1 is as follows: CV (process control external signal), CVI (internal current acting on E/P unit), E/P (electro-pneumatic transducer), F (main pipeline flow rate), Fv(control valve flow rate), Fv3(actuator by-pass pipeline flow rate), FT (flow rate transmitter), P (positioned), P1,P2 (pressures on valve: inlet and outlet), Ps E/P (transducer output pressure), PSP (positioner supply pressure unit), PT (pressure transmitter), Pz (positioner air supply pressure), S (pneumatic servo-motor), T1 (liquid temperature), TT (temperature transmitter), V (control valve),V1,V2 andV3 (cut-off valves), X (valve plug displacement), ZC (internal controller), ZT (stem position transmitter).

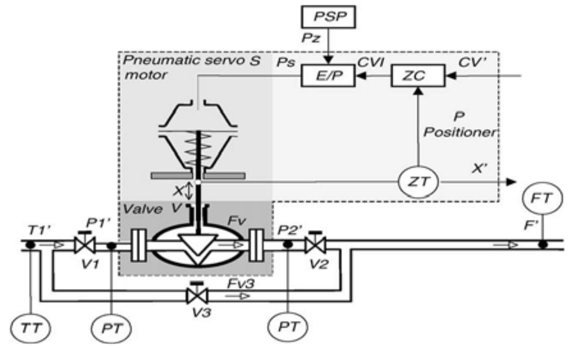

Fig.1 Structure of benchmark actuator system [23],[24]

There are 19 actuator faults which have been considered in the benchmark study [40]. These faults are:
Valve clogging (f_1), Valve plug or Valve seat sedimentation (f_2), Valve plug or Valve seat erosion (f_3), increase of valve friction (f_4), external leakage (f_5), internal leakage (f_6), medium evaporation or critical flow (f_7), twisted servomotor stem (f_8), servomotor housing or terminal tightness (f_9), servomotor diaphragm perforation (f_{10}), servomotor spring fault (f_{11}), electro pneumatic transducer fault (f_{12}), stem displacement sensor fault (f_{13}), pressure sensor fault (f_{14}), positioner spring fault (f_{15}), positioner supply pressure drop (f_{16}), unexpected pressure change across valve (f_{17}), fully or partly opened bypass valve (f_{18}) and flow rate sensor fault (f_{19}).

The faults are grouped based on the severity of the fault as abrupt (large, medium and small) and incipient [23],[24]. The dataset consists of the data for the following simulated fault groups: abrupt-large for f_1, f_2, f_7, f_8 , f_{10}, f_{11}, f_{12}, f_{13}, f_{14}, f_{15},f_{16}, f_{17} and f_{18}; abrupt-medium for f_1, f_7, f_8 , f_{10}, f_{12}, f_{13}, f_{14}, f_{16}, and f_{18} ; abrupt-small for f_1, f_7, f_8 , f_{10}, f_{12}, f_{13}, f_{14}, f_{16}, and f_{18} and f_{19} ; incipient for f_2, f_3, f_4 , f_5, f_6, f_9, f_{11} and f_{13}. It must be noted that same fault can manifest itself with different levels of severity under different circumstances. Hence some faults, for example, f_1 have been simulated at different fault severity levels, resulting in distinct measured data.

3. Design of the Proposed System

The system proposed in the present paper performs FDI in three steps:

Step 1: Extract statistical parameters (average, median, minimum, maximum, standard deviation, kurtosis, skew and Variance) using moving windows, from each of the six measured parameters.
Step 2: Select important features for fault classification.
Step 3: Use the selected features for identification of the fault and its type using Naïve Bayes classifier.

As can be seen, extracting eight parameters from six initial features creates a feature set with (6*8) = 48 features and when taken together with the initial feature set, the total number of features in the feature set becomes 54. Also, the number of measurements made is large (a total of 65535 measurements for each of the initial features) as well. This makes the task of feature selection quite challenging.

3.1 Feature Selection

The proposed feature selection system is derived from ABC [25],[26] and mRMR [27] algorithms. This method was developed using principles of the ABC and mutual information (MI) [28].

The ABC algorithm is an optimization algorithm that uses the behavior of the bees while searching for food [25]. A bee colony is an organized team work system where each bee contributes significant information to the system. There are three types of worker bees which involve in collecting nectar viz. employed bees, onlooker bees and scout bees. The ABC algorithm considers the position of food source as the possible solution of the optimization problem and the food source corresponds to the quality (fitness) of the associated solution [26]. The number of the employed bees or the onlooker bees is equal to the number of solutions in the population. The initial population of N solutions is randomly generated. Each solution is a *D*-dimensional vector where D is the number of parameters to be optimised. They are relevance and redundancy in this case. The population of solutions is subject to repeated search processes by the employed bees, onlooker bees and scout bees. A solution is randomly chosen and compared with the current solution. The objective function used here will be the mRMR function. The fitness function of each solution is given by

$$fit_i = \frac{1}{1+f(i)} \quad (1)$$

where *f(i)* is the objective function of the i^{th} solution. If the fitness function of the new chosen solution is greater than the existing one, then the new solution is memorized and the old one is discarded. The employed bees share the information i.e, fitness value of the solutions in their memory with the onlooker bees.

The probability of each solution based on its fitness, is calculated by

$$p_i = \frac{fit_i}{\sum_{n=1}^{N} fit_n} \quad (2)$$

where fit_i is the fitness value of the solution i and N is the number of solutions in the population. Candidate solutions are produced using the formula

$$v_{ij} = x_{ij} + \varphi_{ij}(x_{ij} - x_{kj}) \quad (3)$$

where k ∈ {1, 2, . . . , N} and j ∈ {1, 2, . . . ,D} and φ_{ij} is a random number ranging between -1 and 1. This ensures that values generated are different from those already existing. And also the newly generated solutions lie within the defined boundary. The parameter exceeding its limit is set to its limit value.

The performance of each candidate solution is compared with that of the existing solution. If the new solution has equal or better fitness value than the old solution, the old one is discarded with the new one occupying its position. Else, the old one is retained. In other words, a greedy selection mechanism is used for the selection process.

If an optimal solution cannot be obtained from a population within the predefined number of cycles i.e. limit then, that population is abandoned and replaced with a new population. ABC algorithm is used here to optimize the redundancy and relevance parameters of mRMR function. The mRMR method proposed in [27] uses the principle of mutual Information. The mutual information between two variables A and B can be defined as

$$I(A;B) = \log\left(\frac{p(A|B)}{p(A)}\right) \quad (4)$$

Maximum Relevance orders features based on the mutual information between individual features x_i and target class h such that the feature with the highest mutual information is the most relevant feature. The relationship is expressed as follows:

$$\max V_I, V_I = \frac{1}{|S|}\sum_{i \in S} I(h,i) \quad (5)$$

Max Relevance often shows a high inter-dependence among the features. When two features are highly dependent on one another, the class-discriminative power of these two features would not change much if either one of them were to be removed and if not removed they become redundant as they convey the same characteristics. The minimal redundancy condition can be added to select mutually exclusive features of the dataset. The following relationship helps establish the minimum redundancy measure.

$$\min W_I, W_I = \frac{1}{|S|^2}\sum_{i,j \in S} I(i,j) \quad (6)$$

The criterion combining the above two parameters is called "minimal-redundancy-maximal-relevance". It was seen that the two measures could be used together to form two combinations for the purpose of improving the feature selection process [26]. The two combinations considered were:

$$\max(V_I - W_I) \tag{7}$$

$$\max(V_I / W_I) \tag{8}$$

Here Eq. (7) forms MID: Mutual Information Difference criterion and Eq. (8) forms MIQ: Mutual Information Quotient criterion. It was observed in [30] that MID gave a better performance when compared to MIQ. This was found to be the case in the present study, as well.

Redundancy is often a matter of concern when dealing with large datasets. It was noticed that redundancy caused a negative effect on the accuracy of the classifying system. But it cannot be presumed that the relevance factor only facilitated the increase in accuracy. The conditions are seen to be purely situational. That is, depending on the dataset under study, either of the two, relevance or redundancy may drastically affect the percentage of accuracy.

The following expression defines the proposed optimization criterion

$$\max(a.V_I - b.W_I) \tag{9}$$

where a and b are constants describing the weightage to be given to relevance and redundancy for selecting the optimal feature set, V_I is relevance and W_I is redundancy.

The value of the constants a and b are arrived at from the ABC algorithm. This algorithm was noticed to be applicable only for discrete datasets. And so, when continuous datasets are to be analyzed, discretization has to be done. Discretization comes with the disadvantage of loss of data which will further reduce the accuracy. This can be alternated by scaling the data and employing logarithmic functions. Thus, the relationship is modified for continuous datasets as follows:

$$\max(a \times \log(V_I) - b \times \log(W_I)) \tag{10}$$

The output of the proposed feature selection mechanism is the set of features in the decreasing order of importance.

3.2 Naïve Bayes Classification

Naïve Bayes classifier is a simple technique for supervised learning based on probability theory and is highly suitable for datasets containing large number of attributes [30]. Also small amounts of noise in the data do not affect the system. It works on Bayes theorem and the relation is given as follows

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \tag{11}$$

where X is a tuple belonging to class C and H is some hypothesis under consideration. If two or more features are highly correlated, then the weightage for that feature is made high by the system and the result of classification is biased towards values with higher weightage thus pulling down the accuracy values.

4. Results

From the DAMADICS benchmark data six measurements have been considered, viz. process control external signal (CV), pressure on the valve inlet (P1), pressure on the valve outlet(P2), stem displacement (X), liquid flow rate (F), liquid temperature (T). As the first step, statistical parameters were extracted from each of the six initial features listed above. These parameters are average, median, minimum, maximum, standard deviation, kurtosis, skew and variance. Hence, a total of 54 features including the original six features are obtained.

The features were extracted for different moving averages (MA) and the corresponding accuracies using a naïve Bayes classifier were obtained (see Table 1). Use of MA is equivalent to passing the measured signal through a low pass filter. Increase in the number of points used for calculating the MA implies reduction in the cutoff frequency of the filtered signal. This is used to reduce the noise (which usually manifests itself as high frequency signals) in the measured signal and hence improve the fault detection accuracy of the system. However, the number of points used for calculating the MA cannot be arbitrarily high since, for larger moving averages (or equivalently, lower cut off frequencies), the filtered signal will begin to lose not only the high frequency noise but also the lower frequencies that may be useful. Also, processing time increases with higher moving points which results in slower classification.

Table 1 shows the result of the effort at identifying the optimum number of points for computing MA. An exhaustive search method to maximize the accuracy was carried out. It was observed that 100 point MA is optimal for abrupt medium, abrupt small and incipient faults.

Table 1: Accuracy (%) for faults from statistical features.

Moving Average (points)	Abrupt Small (%)	Abrupt medium (%)	Abrupt large (%)	Incipient (%)
40	72.1	65.0	78.5	32.9
60	70.4	75.9	80.8	39.5
70	73.5	82.8	94.0	49.4
80	79.0	79.0	81.4	37.8
90	72.6	88.9	82.8	65.5
100	**89.4**	**94.3**	**96.3**	**69.9**
110	81.1	89.4	82.3	60.5
120	87.6	89.0	94.5	49.7

The 54 features extracted using the 100 point MA are subjected to feature selection using following feature selection methods: Relief-F, MI, χ^2 , MID, MIQ, information gain (info gain), gain ratio and the proposed feature selection algorithm. Various feature selection algorithms applied to 100 point MA window large abrupt fault are shown in Table 2.

Table 2: Accuracy in % vs. no.of features for abrupt large faults (coefficients for proposed method, a=0.9106; b=0.0131)

no. of features	Proposed method	MIQ	MID	Info gain	Gain ratio	chi^2	Relief -F
3	78.1	80.7	85.0	87.5	30.8	87.5	24.8
6	80.8	85.9	85.7	93.3	84.5	93.3	87.3
9	86.8	87.2	88.0	94.0	86.4	94.0	89.6
12	89.1	88.3	88.1	94.7	91.8	94.2	89.9
15	89.2	89.0	89.0	94.5	94.4	94.5	90.3
18	95.1	89.8	83.9	95.1	95.0	95.1	92.3
21	95.4	83.9	84.1	95.2	95.2	95.5	96.1
24	95.9	84.5	84.1	95.8	95.4	95.6	96.6
27	95.9	84.5	84.6	95.7	95.8	95.7	96.4
30	**96.8**	84.7	84.6	96.3	95.9	96.3	96.4
33	96.3	84.7	84.7	96.3	96.1	96.3	96.3
36	96.3	95.9	84.7	96.3	96.3	96.3	96.3
39	96.3	96.0	95.9	96.3	96.3	96.3	96.3
42	96.3	96.2	96.2	96.3	96.3	96.3	96.3
45	96.3	96.2	96.2	96.3	96.3	96.3	96.3
48	96.3	96.3	96.2	96.3	96.3	96.3	96.3
51	96.3	96.3	96.3	96.3	96.3	96.3	96.3
54	96.3	96.3	96.3	96.3	96.3	96.3	96.3

For Abrupt medium and small faults also the proposed system showed relatively better accuracy of 95.0% for 36 features and 89.5% for 39 features respectively. The accuracies of the various feature selection algorithms for abrupt medium faults and abrupt small faults are shown in Table 3 and Table 4 respectively.

Table 3: Accuracy in % vs. no.of features for abrupt medium faults (coefficients for proposed method, a=0.3462; b=0.9386)

no. of features	Proposed method	MIQ	MID	Info gain	Gain ratio	chi^2	Relief -F
3	83.9	80.3	83.9	75.2	20.0	75.2	82.4
6	84.2	80.6	84.2	84.0	77.7	84.0	83.2
9	70.1	87.3	87.2	91.2	92.3	91.2	84.6
12	72.4	87.7	87.7	92.1	92.6	92.1	86.2
15	73.6	89.2	89.2	92.1	92.2	92.1	92.8
18	73.9	76.2	76.2	93.2	92.9	94.0	93.3
21	75.7	76.4	76.2	94.0	94.0	94.0	93.4
24	76.7	77.1	77.1	94.0	94.1	94.2	94.6
27	78.4	77.7	77.3	94.3	94.0	94.3	94.6
30	85.1	81.2	78.5	94.3	94.3	94.3	94.5
33	86.2	86.2	85.2	94.3	94.3	94.3	94.5
36	**95.0**	95.0	86.9	94.3	94.3	94.3	94.5
39	95.0	95.0	95.0	94.3	94.3	94.3	94.5
42	95.0	95.0	95.0	94.3	94.3	94.3	94.5
45	95.0	95.0	95.0	94.3	94.3	94.3	94.3
48	95.0	95.0	95.0	94.3	94.3	94.3	94.3
51	94.3	94.3	94.3	94.3	94.3	94.3	94.3
54	94.3	94.3	94.3	94.3	94.3	94.3	94.3

Table 4: Accuracy vs. no.of features for abrupt small faults (coefficients for proposed method, a=0.0021; b=0.9386)

no. of features	Proposed method	MIQ	MID	Info gain	Gain ratio	chi^2	Relief -F
3	53.9	73.2	78.2	74.5	29.2	74.5	76.1
6	64.6	73.9	78.3	82.7	73.2	82.7	77.7
9	76.1	81.1	81.2	81.7	78.7	81.7	77.7
12	77.2	81.9	81.9	85.7	83.5	85.7	80.0
15	88.0	83.0	83.0	86.1	85.6	86.1	81.8
18	89.3	74.3	74.3	89.0	89.3	89.0	82.2
21	89.4	74.8	74.5	89.3	89.1	89.3	82.3
24	89.4	76.2	76.1	89.3	89.1	89.3	83.4
27	89.1	76.6	76.3	89.4	89.1	89.4	83.4
30	89.1	78.9	77.1	89.4	89.4	89.4	85.3
33	89.1	84.8	84.3	89.4	89.4	89.4	85.3
36	89.1	89.1	85.0	89.5	89.4	89.5	85.9
39	**89.5**	89.2	89.1	89.4	89.4	89.4	89.4
42	89.5	89.4	89.4	89.4	89.4	89.4	89.4
45	89.4	89.4	89.4	89.4	89.4	89.4	89.4
48	89.4	89.4	89.4	89.4	89.4	89.4	89.4
51	89.4	89.4	89.4	89.4	89.4	89.4	89.4
54	89.4	89.4	89.4	89.4	89.4	89.4	89.4

Results of feature selection for incipient fault are shown in Table 5. It can be observed that the proposed method shows accuracy of 70.7% for 36 features. Information gain shows better accuracy of 71.6% for 9 features. Also other methods like chi square, Relief-F and gain ratio give slightly better results when compared to the proposed method.

Table 5: Accuracy in % vs. no.of features for abrupt small faults (coefficients of the proposed method: a=0.9106; b=0.0131)

no. of features	Proposed method	MIQ	MID	Info gain	Gain ratio	chi^2	Relief-F
3	52.9	52.9	52.9	64.8	22.2	64.8	54.8
6	56.8	56.8	56.8	69.5	57.3	68.6	66.4
9	69.8	69.8	64.1	**71.6**	65.0	71.3	69.3
12	67.3	67.3	67.3	71.1	66.0	71.1	67.9
15	70.4	70.3	70.4	71.0	69.2	71.0	68.2
18	70.4	70.3	70.4	69.8	71.0	69.8	68.1
21	70.4	70.3	70.4	69.8	71.3	69.8	67.9
24	70.4	70.3	70.4	69.8	71.0	69.8	68.0
27	70.4	70.3	70.4	69.8	69.8	69.8	69.9
30	70.4	70.3	70.4	69.8	69.8	69.8	69.9
33	70.4	70.3	70.4	69.9	69.9	69.9	69.9
36	70.7	70.7	70.4	69.9	69.9	69.9	69.9
39	70.7	70.7	70.2	69.9	69.9	69.9	69.9
42	70.7	70.7	70.7	69.9	69.9	69.9	69.9
45	70.7	70.7	70.7	69.9	69.9	69.9	69.9
48	70.0	70.0	70.7	69.9	69.9	69.9	69.9
51	69.9	69.9	70.7	69.9	69.9	69.9	69.9
54	69.9	69.9	69.9	69.9	69.9	69.9	69.9

It can be seen from the above results that the proposed feature selection system is well capable of identifying the best features in a dataset and that the FDI system presented in this paper can be successfully used for identifying faults in actuators with a very high degree of accuracy.

References

[1] X. Dai, Z. Gao, T. Breikin and H. Wang, “Disturbance Attenuation in Fault Detection of Gas Turbine Engines: A Discrete Robust Observer Design”, IEEE Transactions on Systems, Man, And Cybernetics, 2009, Vol. 39, No.2, pp. 243-239.

[2] B. Ayhan, M. Y. Chow, H. J. Trussell and M.H. Song, “A Case Study on the Comparison of Non-parametric Spectrum Methods for Broken Rotor Bar Fault Detection”, in Proceedings of the 29th Annual Conference of the Industrial Electronics Society, 2003, Vol. 3, pp. 2835-2840.

[3] X. Wang and D. Zhang, “Optimization Method of Fault Feature Extraction of Broken Rotor Bar in Squirrel Cage Induction Motors”, in Proceedings of the IEEE International Conference on Information and Automation, 2010, pp. 1622-1625.

[4] G. King, M. Tarbouchi and D. McGaughey, “Rotor Fault Detection in Induction Motors Using the Fast Orthogonal Search Algorithm”, in Proceedings of the International Symposium on Industrial Electronics, 2010, pp. 2621-2625.

[5] J. Liang and N. Wang, “Faults Detection and Isolation Based on PCA: An Industrial Reheating Furnace Case Study”, in Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, 2003, Vol. 2, pp. 1193-1198.

[6] J. Mina and C. Verde, “Fault Detection Using Dynamic Principal Component Analysis by Average Estimation”, in Proceedings of the 2nd International Conference on Electrical and Electronics Engineering (ICEEE) and XI Conference on Electrical Engineering (CIE), 2005, pp. 374-377.

[7] N. Tudoroiu and M. Zaheeruddin, “Fault Detection And Diagnosis Of Valve Actuators In HVAC Systems”, in Proceedings of the IEEE Conference on Control Applications, 2005, pp. 1281-1286.

[8] N. Tudoroiu and M. Zaheeruddin, “Fault Detection and Diagnosis of Valve Actuators in Discharge Air Temperature (DAT) Systems, using Interactive Unscented Kalman Filter Estimation”, in Proceedings of the IEEE International Symposium on Industrial Electronics, 2006, pp. 2665-2670.

[9] K. Choi, S. M. Namburu, M. S. Azam, J. Luo, K. R. Pattipati, and A. Patterson-Hine, “Fault Diagnosis in HVAC chillers”, IEEE Instrumentation & Measurement Magazine, 2005, pp. 24-32.

[10] L. Hu, K. Cao, H. Xu and B. Li, “Fault Diagnosis of Hydraulic Actuator based on Least Squares Support Vector Machines”, in Proceedings of the IEEE International Conference on Automation and Logistics, 2007, pp. 985-989.

[11] J. Gao, W. Shi, J. Tan and F. Zhong, **“**Support vector machines based approach for fault diagnosis of valves in reciprocating pumps”, in Proceedings of the IEEE Canadian Conference on Electrical and Computer Engineering, 2002, pp. 1622-1627.

[12] F. He and W. Shi, “WPT-SVMs Based Approach for Fault Detection of Valves in Reciprocating Pumps”, in Proceedings of the American Control Conference, 2002, pp. 4566-4570.

[13] J. Middleton, P. Urwin and M. Al-Akaidi, “Fault Detection And Diagnosis In Gas Control Systems”, Intelligent Measuring Systems for Control Applications, IEE Colloquium, 1995, pp. 8/1 - 8/3.

[14] D. Linaric and V. Koroman, “Fault Diagnosis of a Hydraulic Actuator using Neural Network”, in International Conference on Industrial Electronics (IClT), 2003,pp. 106-111.

[15] M. Karpenko, N. Sepehri and D. Scuse, “Neural Network Detection and Identification of Actuator Faults in a Pneumatic Process Control Valve”, in Proceedings of IEEE International Symposium on Computational Intelligence in Robotics and Automation, 2001, pp. 166-171.

[16] L. Hongmei, W. Shaoping and O. Pingchao, “Fault Diagnosis Based on Improved Elman Neural Network for a Hydraulic Servo System”, in Proceedings of the Conference on Robotics, Automation and Mechatronics, 2006, pp. 1-6.

[17] X. Z. Gao and S. J. Ovaska, “Motor Fault Detection Using Elman Neural Network with Genetic Algorithm-aided Training”, in Proceedings of the International Conference on Systems, Man, and Cybernetics, 2000,Vol. 4, pp. 2386-2392.

[18] B. Koppen-Seliger and P. M. Frank, “Fault Detection and Isolation in Technical Processes with Neural Networks”, in Proceedings of the 34th Conference on Decision & Control, 1995,Vol. 3, pp. 2414-2419.

[19] R. Sreedhar, B. Fernhndez and G. Y. Masada, “A Neural Network Based Adaptive Fault Detection Scheme”, in Proceedings of the American control conference, 1995, Vol.5, pp. 3259-3263.

[20] J. Tian, M. Gao, L. Cao and K. Li, “Fault Detection of Oil Pump Based on Fuzzy Neural Network”, in Proceedings of the Third International Conference on Natural Computation (ICNC), 2007, pp. 636-640.

[21] M. Y. Chow and P. V. Goode, “The Advantages and Challenges of Machine Fault Detection Using Artificial Neural Network and Fuzzy Logic Technologies”, in Proceedings of the 36th Midwest Symposium on Circuits and Systems, 1993, Vol. 1, pp. 708-711.

[22] A. Ramezanifar, A. Afshar and S. K. Y. Nikravesh, “Intelligent Fault Isolation of Control Valves in a Power Plant”, in Proceedings of the First International Conference on Innovative Computing, Information and Control (ICICIC), 2006, pp.1-4.

[23] M. Bartys, R. Patton, M. Syfert, S. de las Heras and J. Queveda, “Introduction to the DAMADICS actuator FDI benchmark study”, Control Engineering Practice, 2006, Vol.14, pp. 577-596.

[24] M. Bartys and M. Syfert, “Using DAMADICS Actuator Benchmark Library”, ‘DABLib’ Simulink Library Help File, ver.1.22, 2002, pp.1-32.

[25] M. Subotic, M. Tuba and N. Stanarevic, “Parallelization of the artificial bee colony (ABC) algorithm”, In proceedings of the 11th WSEAS International Conference on Neural Networks, 2010, pp. 191-196.

[26] S. M. Saab, N. K. T. El-Omari and H. H. Owaied, “Developing Optimization Algorithm using Artificial Bee Colony System”, UBiCC Journal, 2009,Vol.4, No 5, pp. 391-396.

[27] H. Peng, F. Long and C. Ding , “Feature Selection Based on Mutual Information: Criteria of Max-Dependency, Max-Relevance and Min-Redundancy”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, Vol. 27, No.8, pp. 1226-1236.

[28] R. Bose, “Information Theory, Coding and Cryptography”, Second Ed., Tata McGraw-Hill, New Delhi, 2008.

[29] C. Ding and H. Peng, “Minimum Redundancy Feature Selection From Microarray Gene Expression Data”, Journal for Bioinformatics and Computational Biology, 2005, Vol.3, No.2, pp. 185-205.

[30] C. A. Ratanamahatana and D. Gunopulos, “Scaling up the Naive Bayesian Classifier: Using Decision Trees for Feature Selection”, in proceedings of Workshop on Data Cleaning and Preprocessing (DCAP 2002), IEEE International Conference on Data Mining (ICDM 2002) , 2002, pp. 475-487.