2009 American Control Conference
Hyatt Regency Riverfront, St. Louis, MO, USA
June 10-12, 2009

Application of combined support vector machines in process fault diagnosis

Esmaeil Tafazzoli and Mehrdad Saif

Abstract— **The performance of Combined Support Vector Machines, C-SVM, is examined by comparing it's classification results with k-nearest neighbor and simple SVM classifier. For our experiments we use training and testing data obtained from two benchmark industrial processes. The first set is simulated data generated from Tennessee Eastman process simulator and the second set is the data obtained by running experiment on a Three Tank system. Our results show that the C-SVM classifier gives the lowest classification error compared to other methods. However, the complexity and computation time become issues, which depend on the number of faults in the data and the data dimension. We also examined Principal Component Analysis, using PC scores as input features for the classifiers but the performance was not comparable to other classifiers' results. By selecting appropriate number of variables using contribution charts for classification, the performance of the classifiers on Tennessee Eastman data enhances significantly. Therefore, using contribution charts for selecting the most important variables is necessary when the number of variables is large.**

I. INTRODUCTION

Support vector machine is a well known technique in the field of machine learning which is used for classification. Implementing nonlinear kernels in the SVM structure enables classification of nonlinear data which can not be classified by simple linear classifiers. In SVM classification method, an optimum hyperplane is defined which maximizes the separation between data point classes [3].

In many works on fault detection and diagnosis, the SVM classifier is combined with another method such as Principal Component Analysis (PCA), Independent Component Analysis (ICA), Fisher Discriminant Analysis (FDA), etc, to reduce the data dimension and to accomplish the detection part of the Fault Detection and Identification (FDI) process. Therefore, the diagnosis part is carried out by SVM classifier. Mostly, the SVM classifier operates on the processed data or features, resulting from other methods (PCA scores for example) [1]. In [4], ICA projection coefficients were used as feature data for training the SVM classifiers. In [5], authors compared the performance of FDA, SVM, and PSVM (proximal support vector machines). They showed that support vector machines perform better than FDA in classifying TE data. In general, SVM is a two-class classifier. A Two-class classification means that data points are assigned to only one of the two class labels in the data set while in multiclass classifiers, there are multiple class labels and the classifier assigns each point to one of the classes. Multiple classification problems can be turned to multiple two-class classification problem. The number of required classifiers depends on the number of faults to be classified. As a result, most SVM classifiers are multiple-SVM classifiers. The term Committee is referred to the combination of classifiers in machine learning area. A committee is built by combining several models (classifiers). Usually the outcome of the committee is better than individual models [8]. Averaging, boosting, and adaptive boosting are some of the methods of combining the models [3].

K-Nearest Neighbor (KNN) is one of the simplest classification algorithms in machine learning. K-Nearest neighbor classification method was first introduced by Cover and Hart [2], in which the class of each sample point is determined by its K neighboring points in the training set. The point is assigned to the class with the majority of votes amongst the K-neighbor points. Several types of KNN algorithm have been suggested and applied to different data sets in the fields of data mining and machine learning. Many papers can be found on KNN or combination of KNN with other methods for improving data classification. For more information on KNN algorithm and its application the following references would be helpful [9]-[17]. In this paper we use averaging method for the combined classifiers. Considering the idea of committee classifier, we develop a combined- SVM,C-SVM, classifier and investigate its performance compared to individual classifiers on the data generated from Tennessee-Eastman (TE) simulator and the Three Tank System which are well known benchmark experimental processes used for control , monitoring, and fault diagnosis experiments. We also examine the performance of a K-nearest neighbor classifier in comparison with C-SVM when applied to this set of data.

II. TWO CLASSIFICATION METHODS

A. Support Vector Machines

SVM algorithm is usually used for two-class separation problems [3]. The algorithm finds the maximum margin for a separating boundary between two classes of data. Suppose we have a set of data that can be separated into two classes. The data is separated by training a linear model

$$y(\mathbf{x}) = w^T \varphi(\mathbf{x}) + b \tag{1}$$

Equation (1) is the mathematical representation of the linear model. In this model the training data matrix is an $n \times m$ matrix where each row of the matrix represents an observed data point, $\mathbf{x}_i$, which is a vector of length m. So, n is the number of data points and m is the number of variables. Each data point's class is determined by its target value. The corresponding target values are stacked in a vector $\mathbf{t}$

The authors are both with the School of Engineering Science, Simon Fraser University, 8888 University Drive, Vancouver, BC, V5A 1S6, Canada. Corresponding Email: saif@ensc.sfu.ca

with $t_i \in \{-1, 1\}$ as it's elements. $\varphi(\mathbf{x})$ is called feature space transformation function and b is bias. w's are weights which affect the separating plane direction. Function $y(\mathbf{x})$ has the property that $y(\mathbf{x}_i) > 0$when $t_i = 1$ and $y(\mathbf{x}_i) < 0$ when $t_i = -1$. Therefore, $t_i y(\mathbf{x}_i) > 0$ for all i . In SVM algorithm, the distance between the closest data points to the decision boundary which is called the boundary margin, is maximized (see Fig.1). Therefore, in SVM the hyperplane which maximizes the margin is chosen as the decision boundary. The maximization criterion is:

$$arg \max_{w,b}\{\frac{1}{\|w\|} \min_{i=1,...,n}[t_i(w^T\varphi(\mathbf{x}_i) + b)]\}$$

and the points with minimum distance are known as Support Vectors. Fig (1) illustrates the location of support vectors and the decision boundary.

The model parameters, w and b, are found by solving a constrained optimization problem as

$$arg \min_{w} \frac{1}{2}\|w\|^2$$
$$s.t., \forall i, t_i(w^T(\varphi(\mathbf{x}_i) + b) \geq 1$$

This problem is solved by using Lagrange multipliers. The lagrangian is

$$L(w, b, a) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{n} a_i\{t_i(w^T\varphi(\mathbf{x}_i) + b) - 1\}$$

where a_i are Lagrange multipliers. As a result, the weights and bias are found and the decision function, becomes

$$y(\mathbf{x}) = w^T\varphi(\mathbf{x}) + b = \sum_{i=1}^{n} a_i t_i k(\mathbf{x}, \mathbf{x}_i) + b$$

The data classification task is carried out by computing the $sign(y(\mathbf{x}))$ for each test point. Using nonlinear kernels allows linear classification of nonlinearly separable data in higher dimension of the kernel space. The two well known kernels are RBF kernel and polynomial kernel which are defined as

$$RBF : k(\mathbf{x}_i, \mathbf{x}_j) = exp(\frac{-\|\mathbf{x}_i - \mathbf{x}_j\|}{\delta})$$
$$Polynomial : k(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i\mathbf{x}_j + 1)^d$$

In many problems data points in different classes have overlap which causes problem for classification. This happens when data is not linearly separable in the feature space. In this case, support vectors can not classify the points' class properly and give poor result. To overcome this problem, SVM constraint is relaxed from

$$t_i y(\mathbf{x}_i) \geq 1$$

to

$$t_i y(\mathbf{x}_i) \geq 1 - \zeta_i \quad (2)$$

where ζ_i $i = 1, ..., n$ is called the slack variable. Fig (2) shows the concept of slack variables.

By using slack variables, some points can be misclassified which gives flexibility to classifier. In this way some data points are misclassified but there will be a penalty which increases the error function. Therefore, the algorithm maximizes the margin while minimizes the penalty for the points in the wrong side of the boundary. So the criterion becomes

$$\min\{C\sum_{i=1}^{n}\zeta_i + \frac{\|w\|^2}{2}\} \quad (3)$$

where C is the controlling parameter, which controls the trade off between the model complexity and minimizing classification error. High value of C results in over-fitting the data and in the limit, the SVM model is the same as the SVM for separable data.

The optimization problem now turns to minimizing (3) with constrains in (2). The Lagrangian is given by

$$L(w, b, a) = \frac{\|w\|^2}{2} + C\sum_{i=1}^{n}\zeta_i \quad (4)$$
$$-\sum_{i=1}^{n} a_i\{t_i y(\mathbf{x}_i) - 1 + \zeta_i\} - \sum_{i=1}^{n}\mu_i\zeta_i$$

where $a_i > 0$ and $\mu_i > 0$ are Lagrangian multipliers [3].

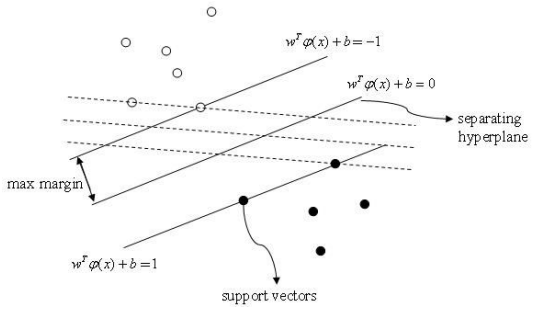

Fig. 1. support vectors illustration

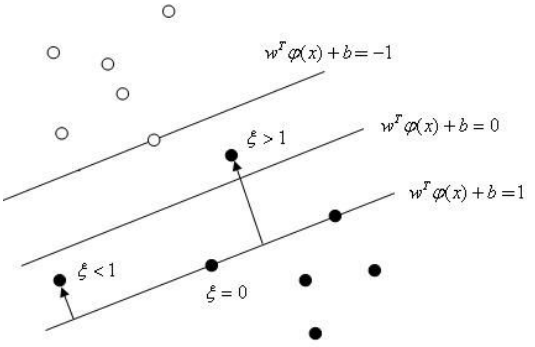

Fig. 2. Illustration of slack variables used for non-separable data

B. K-Nearest Neighbor Classification

In K-nearest neighbor classification, the class of each sample point is determined by its K neighboring points in the training set. The point is assigned to the class with the majority of votes for class label amongst the K-neighbor points. The classifier is defined by its parameters. Setting parameter *K* depends on the data and effects the performance of the classifier. *K* must be large enough to reduce missclassification

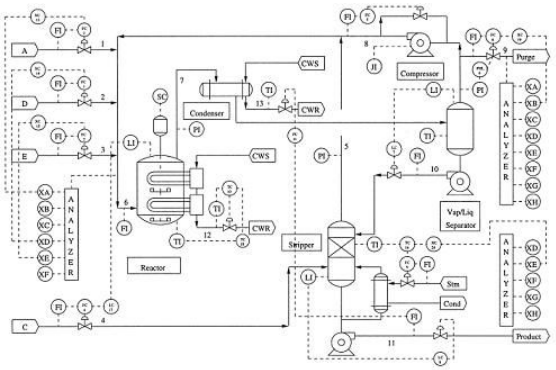

Fig. 3. Tennessee Eastman process simulator diagram[5]

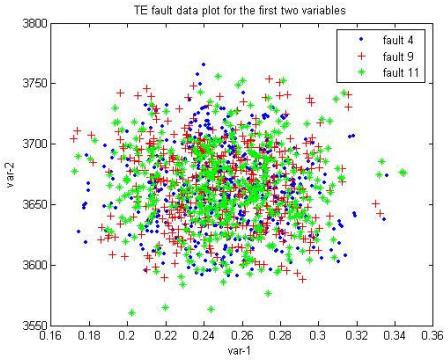

Fig. 4. Test data plot of variables 1 and 2 for fault 4, 9, and 11

of an example point and must be small enough so that the sample point is close to the neighboring points which results in better estimation of the point's class [2].

III. EXPERIMENT DATA

A. Tennessee Eastman process

The Tennessee Eastman process (TE) which is a chemical plant involving four exothermic gas reactions was proposed and modeled by Downs and Vogel as a plant-wide control challenge problem [6]. The process has been used for many research experiments in fault detection and control. It has fifty two variables including measured and manipulated variables and twenty one faults that have been defined for the process. In this work, faults $4, 9$, and 11 are chosen as the training and testing data which have overlap between each other [7]. Fault 4 is defined as a step change in the reactor cooling water temperature. Fault 9 is a random variation in one of the reactants (reactant D) feed temperature and fault 11 is a random variation in the reactor cooling water temperature. The data is taken from *http://brahms.scs.uiuc.edu*. Each set of training and testing data contain 480×52 and 960×52 points respectively, observed every three min of simulation and faults occur after 1 hour and 8 hour of simulation respectively. Figure (3) illustrates the TE plant simulation diagram. Figure (4) shows the plot of faulty data in first and second variable space and figure (5) shows the plot of faulty data in the two dimensions where the data has the most separability.

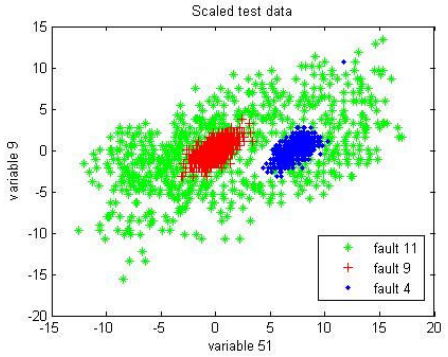

Fig. 5. Test data plot of variables 9 and 51 for fault 4, 9, and 11

B. Three Tank System

As a benchmark control problem, the Three Tanks System (3TS) is used in many different research projects. The basic structure of the system contains three tanks which are connected to each other by pipes. Two tanks are filled with 2 pumps while the third one is filled only through the pipes connected to the other two. Our experimental setup is an AMIRA DTS200 in which the water level is measured with three piezo-resistive difference pressure sensor [19]. DTS200 contains 6 valves which are used to emulate clogging and leakage in the system. Figure (6) shows the system flow sheet. The system has the following specifications:
Tank cross section area, $A = .0154m^2$
Connecting pipes cross section area, $az = 5 \times 10^{-5} m^2$
Highest liquid level, $H_{max} = 62cm$
Maximum pump flow rate, $Q_{max} = 100mltr/sec.$
The system is equipped with a disturbance module which allows simulating 11 types of faults for fault detection research purposes including three sensor faults, two actuator faults, leakage for each of the three tanks ,clogging between the tanks, and clogging in the outflow. The training and testing data size are 500×5 for each case of fault with water levels and flow rates as variables. Faults are instigated at sample 55 in each case. We assume that only one fault occurs at a time and there are no simultaneous faults. Figures (7) and (8) show two example plots of data when leak and sensor fault occur in the system.

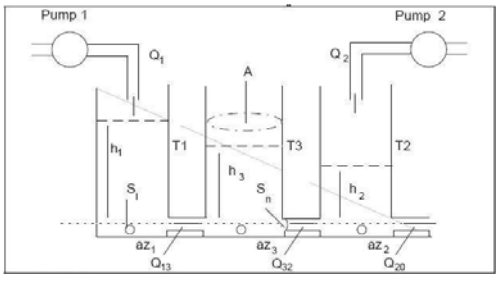

Fig. 6. Three Tank system structure[19]

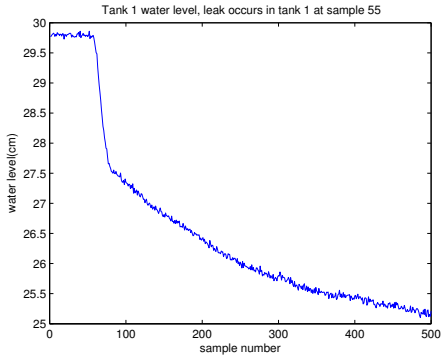

Fig. 7. Example plot of level sensor in three tank system

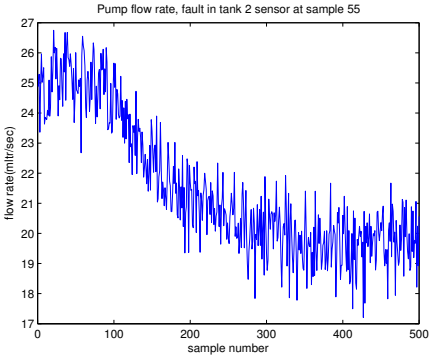

Fig. 8. Example plot of flow rate for three tank system

IV. CLASSIFICATION PROCEDURE AND RESULTS

In every fault detection and diagnosis system, the FDI process includes detecting the fault in the process and then identifying the type of the fault. Here, we focus on the diagnosis part of the FDI process and assume that fault detection has been accomplished. After fault detection stage in FDI, we use SVM for fault classification. It should be noted that using this method for fault diagnosis requires prior knowledge about different faults because classifiers are trained and structured based on this knowledge. Here, we examine the performance of the C-SVM compared to simple SVM with different kernels and also to K-nearest neighbor classifier. Hence, a training and a testing data set is collected from the processes.

The choices of different SVM depend on their parameters. Type of the kernel, value of C, width of the RBF kernel, polynomial kernel degree, and number of SVM to be used in the committee are such example parameters.

Since there are many different combinations to choose, we only restrict our experiment to a simple case with three different kernels to be used in the SVM-classifier.

However, we selected the parameter C, by testing the SVM performance on different values of C ranging in $[.1, 10^5]$. The parameter values used in the experiment are: $C = 100$, $\delta = 1$ (RBF kernel parameter as suggested in [5]), and $Poly - degree = 3$

In [5], it is pointed out that for TE data in this case

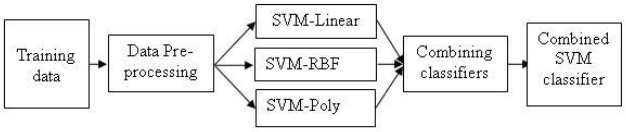

Fig. 9. SVM training procedure

TABLE I
CLASSIFICATION ERROR FOR DIFFERENT CLASSIFIERS APPLIED TO TE PROCESS DATA

Classifier	Classification error %
SVM(linear kernel)	26.7
SVM(RBF kernel)	8.3
SVM(Polynomial kernel)	7.3
C-SVM	6.7
KNN classifier	8.4

(fault 4,9, 11) only two variables are important and the other fifty variables do not show significant changes caused by the faults. They used contribution charts to find the most important variables for this case. These variables are $var-51$ (reactor cooling water valve position) and $var - 9$ (reactor temperature). We use these two variables to train and test our classifier in fault classification on TE data set.

The algorithms was implemented in MATLAB using SVM toolbox from [18]. The procedure for building the classifier is as follows: For every two faults we train a C-SVM classifier. Each classifier is a combination of three SVM with different kernels (linear, RBF, polynomial), trained with data that are a mixture of the two fault class data set. The output is simply the average of the three. Fig (9) depicts the training procedure for C-SVM. In this figure, data pre-processing includes scaling and selecting appropriate variables for classification which has to be done before training SVM's. When SVM is trained the final classifier is tested with test data to determine the classification error and to evaluate the performance of the classification system. The error is simply defined as the percentage of misclassified points in the whole data set. Here, misclassification indicates a point whose class is determined incorrectly. Fig (10) shows the block diagram of the test data classification process. The data class is determined by selecting the maximum vote for data from the classifiers. If there is a tie between classifiers' vote then the fault class is chosen randomly. The TE test data for class 4, 9, and 11 were applied to the classifier. Classification is based on one-against-one classifier which means for every two faults a classifier is trained. So we have three SVM classifier for fault $4-9$, $9-11$ and $4-11$ shown as C-SVM 1, 2, and 3 in Fig (10).

When all variables were included in the data for training and testing, the classification error was 43.1%. Using selected data variables (variables 9 and 51) in the training and testing data sets resulted in 6.7% error which shows about 36.3% decrease. Classification error for applying SVM on the first two PCA scores gave very poor performance with 64% error which is not an acceptable result. Table(I)

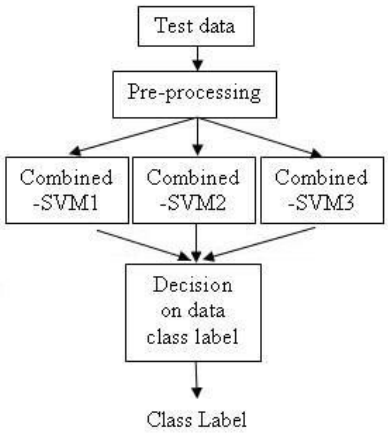

Fig. 10. Classification procedure for TE data

TABLE II
CLASSIFICATION ERROR FOR DIFFERENT CLASSIFIERS APPLIED TO THREE TANK SYSTEM DATA

Classifier	Classification error %
SVM(linear kernel)	14.03
SVM(RBF kernel)	13.74
SVM(Polynomial kernel)	30.53
C-SVM	12.17
KNN classifier	14.57

presents the results for different classifiers applied to TE data. In the second experiment with real data from the Three Tank System(3TS), the procedure is modified to enhance the computation time and complexity of classification. We first train a classifier to separate faults based on their type into four classes, i.e., leakage, clogging, sensor fault, and pump fault. When the type of the fault is determined then the location is determined by using another classifier which is trained for that specific category, e.g., leak in tank 1. The classification results are shown in table (II). The C-SVM gives the best result for classification with 12.17% classification error. SVM classifiers with linear and RBF kernel also give slightly better results than KNN and SVM with polynomial kernel.

V. DISCUSSION AND CONCLUSION

As presented in table(I), by comparing classification errors, the C-SVM outperforms all other classifiers. However, considering the computation time, the KNN classifier performs much faster than SVM based classifiers which is caused by using several SVM's, each of which containing kernel calculation that takes the computation time. This can be problematic when the data dimension is high. Therefore, data reduction techniques are highly recommended prior to using SVM. The number of SVM used in the combined classifier is also an important parameter in forming the classifier which has to be considered. The training time increases by the number of SVM. As presented above, the performance of the method is based on the results of the experiments performed on two benchmark systems. However, for further confirmation, the method should be tested on other different processes in order to achieve a comprehensive understanding of the proposed method.

REFERENCES

[1] C.M. Bishop, *Pattern Recognition and Machine Learning*, Springer, Singapore; 2006.
[2] X. Zhao, S. Huihe, "A Novel Combination Method for On-line Process Monitoring and Fault Diagnosis", *IEEE Tran. Industrial Electronics ISIE* , 4,2005, pp.1715- 1720.
[3] Y. Song et al., *IKNN: Informative K-Nearest Neighbor Pattern Classification*,PKDD 2007, Springer-Verlag, Berlin Heidelberg; 2007.
[4] M. Guo, L. Xie, S. Wang, J. Zhang, "Research on an Integrated ICA-SVM Based Framework for Fault Diagnosis", *IEEE Proc. Syst., Man, and Cybern.*, 3, 2003, pp. 2710-2715.
[5] Chiang L.H., M.E. Kotanchek, A.K. Kordon, "Fault diagnosis based on Fisher discriminant analysis and support vector machines", *Computer and Chemical Eng.*, 28, 2004, pp. 1389-1401.
[6] J.J. Downs, E.F. Vogel, "A Plant-Wide Industrial Process Control Problem", *Computers and Chemical Engineering*, 17(3), 1993,pp.245-255.
[7] L.H. Chiang, E.L. Russell, R.D. Braatz, "Fault diagnosis in chemical processes using Fisher discriminant analysis, discriminant partial least squares, and principal component analysis", Chemometrics and Intelligent Laboratory Systems 50, 2000, pp. 243-252.
[8] G. Mori, "Introduction to machine learning", lecture notes,[Online],available:http://www.cs.sfu.ca/ mori/courses/cmpt726, accessed Aug. 2008.
[9] C. Domeniconi, J. Peng, D. Gunopulos, "Locally adaptive metric nearest-neighbor classification",*IEEE Trans. Pattern Anal. Mach. Intell.*, 24(9), 2002, pp. 1281-1285.
[10] T. Cover, P.Hart,"Nearest neighbor pattern classification",*IEEE Trans. on Information Theory*, 13(1), 1967, pp. 21-27.
[11] V. Athitsos, S. Sclaroff,"Boosting nearest neighbor classifiers for multiclass recognition",*IEEE Compt. Society Conf. on Computer Vision and Pattern Recognition*, 3, 2005, pp. 45-45.
[12] T. Hastie, R. Tibshirani,"Discriminant adaptive nearest neighbor classification",*IEEE Trans. Pattern Anal. Mach. Intell.*, 18(6), 1996, pp. 607-616.
[13] H. Zhang, A.C. Berg, M. Maire, M. Malik,"Discriminative nearest neighbor classification for visual category recognition",*IEEE Compt. Society Conf. on Computer Vision and Pattern Recognition*, 2, 2006, pp. 2126- 2136.
[14] Y. Pingpeng, Y. Chen, H. Jin, L. Huang,"MSVM-kNN: combining SVM and k-NN for multi-class text classification",*IEEE Int. Workshop on Semantic Computing and Systems*, 2008, 133-140.
[15] W. Shu-Bin et al.,"Classification algorithm based on weighted SVMs and locally tuning kNN",*International Conference on Biomedical Engineering and Informatics*, 2008, pp. 240-244.
[16] L. Ping, L. Nan, W. Jian-yu, Z.Chun-Guang,"Combining weighted SVMs and spectrum-based kNN for multi-classification",*Proc. 4th Int. Symp. Neural Networks*, 2007, pp. 448-53.
[17] Q. He, J. Wang,"Principal component based k-nearest-neighbor rule for semiconductor process fault detection", *IEEE Trans. Semiconductor Manufacturing*, 20(4), 2008, pp. 345-354.
[18] S.R. Gunn,"Support Vector Machines for Classification and Regression", Technical Report, 1998, available: *http://www.isis.ecs.soton.ac.uk/resources/svminfo/*, accessed on July 2008.
[19] *DTS200 labratory setup Three Tank System* , AMIRA2002.