

<http://ansinet.com/itj>

ITJ

ISSN 1812-5638

# INFORMATION TECHNOLOGY JOURNAL

**ANSI***net*

Asian Network for Scientific Information  
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

## Applying PCA and Fixed Size LS-SVM Method for Large Scale Classification Problems

Y.H. Hung and Y.S. Liao

Department of Industrial Engineering and Management, National Chin-Yi University of Technology,  
35, Lane 215, Section 1, Chung-Shan Road, Taiping, Taichung, 411, Taiwan, Republic of China

**Abstract:** In data mining, machine learning application is widely discussed and studied. Purpose of using machine learning method to study classification is to attempt to assist related personnel in decision-making. Success of machine learning lies in its classification accuracy and stability, thus improving classification accuracy and stability becomes hot topic of researchers. Support Vector Machine is an excellent tool at high accuracy rate of classification and prediction, but lacking the confidence at analyzing highly complex data. This study utilized Principal Component Analysis (PCA) and Fixed Size (FS) algorithm to reduce data freedom degree and delete data noise and find critical support vector to improve Least Squares Support Vector Machine (LS-SVM) classification accuracy. This study tested four UCI libraries. From experimental result of highly complex PIDD library, classification accuracy of two-stage PCA\_FS\_LS-SVM approach is 3.23% higher than single LS-SVM. With the application of FS\_LS-SVM algorithms, UCI datasets classification systems can produce classification accuracies above 96%.

**Key words:** Classification accuracy, principal component analysis, fixed size, least squares support vector machine

### INTRODUCTION

Machine learning has become one of the most popular classification tools worldwide in the field of decision support. It is connected to computer science and artificial intelligence and aims to allow systems improving performance on data input. Support Vector Machine (SVM) is a relatively new classifier based on strong foundations from the broad area of statistical learning theory (Mumtaz *et al.*, 2008). This class of algorithms proposed by Vapnik (1998) has shown good performance over real applications. SVM for classification and regression is a powerful method of machine learning. Computational advances are making SVM potentially applicable to large data mining problems.

SVM is the latest supervised learning technique from statistical learning theory and has shown to deliver state-of-the-art performance in many real-world applications (Zhao *et al.*, 2007), such as pattern recognition (Wang *et al.*, 1999), bioinformatics, image recognition (Lee *et al.*, 2007), clinical diagnosis (Tsumoto, 2000), text categorization and phoneme classification (Katz *et al.*, 2005). The main principle of SVM is that, each object can be mapped as a point in a high-dimension and it uses a nonlinear mapping to transform the primal training data into a higher dimension. Within this new

dimension, it searches for the linear optimal separating hyperplane. With an appropriate nonlinear mapping to a sufficiently high dimension, data from two classes can be separated by a hyperplane. The main advantage of SVM is that it can automatically derive a network structure that guarantees an upper bound on the generalization error. Recently, many SVM algorithms have been discussed, such as LIBSVM (Chang and Lin, 2001), BSVM (Hsu and Lin, 2002), MYSVM (Ruping, 2001), SVM light (Joachims, 1999) and Least Squares Support Vector Machine (LS-SVM) (Suykens *et al.*, 1999). Suykens introduced the quadratic cost function in SVM and proposed LS-SVM (Suykens *et al.*, 2002). Extensive empirical studies (Gestel *et al.*, 2004) have shown that LS-SVM is comparable to SVM in terms of generalization performance. This is very important in many real life classification problems (Suykens *et al.*, 2002). Therefore, SVM is usually considered the most accurate classification tool for many applications.

For large data sets, they are large dimensional input vectors and highly correlated. It is possible that the points of the two classes cannot be separated by a hyperplane in the primal space. Therefore, special and more efficient algorithms can be used before training LS-SVM. Recently, there has been considerable effort made in selecting features (performing dimensionality

---

**Corresponding Author:** Y.H. Hung, Department of Industrial Engineering and Management,  
National Chin-Yi University of Technology, 35, Lane 215, Section 1, Chung-Shan Road,  
Taiping, Taichung, 411, Taiwan, Republic of China Tel: +886-04-23924505/7628 Fax: 886-4-23934620

reduction) which are most relevant to a SVM. An effective procedure for performing this operation is the Principal Component Analysis (PCA). This linear transform has been widely used to reduce the dimensionality by calculating the eigenvectors of the covariance matrix of the primal input vectors (Cao and Chong, 2002; Draper *et al.*, 2003). Besides, a new technique of fixed size LS-SVM is presented. The fixed size LS-SVM uses an entropy based iterative method to select a predefined size subset of the training samples as support vectors. It could solve the primal problem after mapping to a feature space based on the eigenfunctions obtained from the Kernel Principal Component Analysis (KPCA). The training of the fixed size LS-SVM can be placed by incremental chunk, which avoids computing large-scale matrix inverse but maintaining the precision when training and testing data (Espinoza *et al.*, 2006). This study integrated PCA and fixed size LS-SVM algorithm to improve the classification accuracy of larger scale datasets.

### INTRODUCTION OF FIXED SIZE LS-SVM (FS\_LS-SVM)

**The nonlinear SVMs classification method:** Given the training data  $\{(x_1, y_1), \dots, (x_i, y_i), \dots, (x_n, y_n)\}$  where  $x_i \in \mathbb{R}^m$  is the  $i$ th input pattern and  $y_i \in \mathbb{R}$  the  $i$ th output pattern and  $y_i \in \{+1, -1\}$ . In a standard SVM, primal weight space formulation is shown as:

$$f(x) = \text{sign} [w^T \Phi(x) + b] \quad (1)$$

$$[w_1, w_2, \dots, w_p] \quad \Phi = [\Phi_1, \Phi_2, \dots, \Phi_p]$$

where,  $\Phi(\cdot)$  is a nonlinear function that maps the input space into a higher dimensional feature space. The optimal separating hyperplane  $w$  for the feature can be found with:

$$\min_{w, b, \epsilon} D(w, b, \epsilon) = \frac{1}{2} \|w\|^2 + \gamma \sum_{i=1}^n \epsilon_i \quad (2)$$

$$\text{s.t. } y_i (w^T \Phi(x_i) + b) \geq 1 - \epsilon_i \text{ for } i = 1, \dots, n \text{ and } \epsilon_i \geq 0$$

where,  $\gamma$  is a penalty, trading off margin size for the number of misclassifications, larger  $\|w\|^2 = w^T w$  results in a smaller margin and vice versa. Maximizing the margin is equivalent to minimizing the magnitude of the weights. The SVM can also handle data that is not separable by introducing slack variables  $\epsilon_i$  to the optimization problem. Data points are penalized if they are misclassified. In nonlinear function estimation and nonlinear modeling problems the kernel trick has been applied with  $K(x_i, x_j) = \Phi(x_i)^T \Phi(x_j)$  for  $i, j = 1, \dots, n$ .  $K(x_i, x_j)$  is a

kernel function satisfying Mercer's conditions. Mercer's theorem specifies how to test whether a kernel is a valid innerproduct in some space (Mercer, 1909). It can be used to solve a quadratic optimization problem in a  $\Phi$ - space. The choice of kernel function includes the linear kernel, polynomial kernel or Radial Basis Function (RBF) kernel. Thus, the LS-SVM classifier  $f(x)$  is defined as:

$$f(x) = \text{sign} \left( \sum_{i=1}^n y_i \alpha_i K(x_i, x) + b \right) = \text{sign} \left( \sum_{i=1}^n d_i K(x_i, x) + b \right) \quad (3)$$

where,  $\alpha_i$  is the solution to the quadratic optimization problem and  $b$  follows from the complementarity Karush-Kuhn-Tucker (KKT) conditions. The solution for Eq. 3 is a linear classifier in a feature space to create a nonlinear separating hypersurface in the primal input space.

**The fixed size LS-SVM:** For larger data sets, the components of the input vectors may be highly correlated (redundant). It is possible that the points of the two classes cannot be separated by a hyperplane in the primal space. The reduced problems are fixed size and can be solved using a standard QP code. Convergence proofs are difficult, but this approach seems to converge to an optimal solution in practice. An important issue in the decomposition is selecting the working set. In short, SVMs may pose heavy computational challenges for large data sets. Therefore, a method of fixed size LS-SVM (FS\_LS-SVM) is proposed, which uses an entropy based iterative method to select a predefined size subset of the training samples as support vectors. The estimation is done in the primal space in relation to a Nyström method sampling with active selection of support vectors (Suykens *et al.*, 2002). Nyström method can be use to estimate the feature space mapping. The link between Nyström sampling, kernel PCA and density estimation has been discussed by Williams and Seeger (2001). KPCA is suitable for extracting interesting nonlinear structures in the data. It is a type of nonlinear PCA developed by generalizing the kernel method into PCA (Scholkopf *et al.*, 1998). The FS\_LS-SVM requires the definition of a subset of size  $M$  to build a finite-dimensional approximation of the feature map that maximizes the quadratic Renyi entropy. Renyi's entropy for an input vector  $x$  with probability function (pdf)  $p(x)$  is given by (Girolami, 2002):

$$H_R(x) = \frac{1}{1-\eta} \log \int p(x)^\eta dx, \quad \eta > 0 \quad \eta \neq 1 \quad (4)$$

when  $\eta = 2$ , the Eq. 4 can be rewritten as:

$$H_R(x) = -\log \int p(x)^2 dx \quad (5)$$

For optimization purposes, it is simpler to work with the argument of the logarithm, thus  $V(x) = \int p(x)^2 dx = E\{p(x)\}$  named Information Potential could be approximated by using:

$$\int \hat{p}(x)^2 dx = \frac{1}{n^2} \mathbf{1}_v^T \Psi \mathbf{1}_v$$

where,  $\mathbf{1}_v = [1; 1; \dots; 1]$  and a normalized kernel is assumed with respect to density estimation. The use of this active selection procedure can be important for large scale problems, as it is related to the underlying density distribution of the sample (Karsmakers *et al.*, 2007a).

However, since nonlinear SVM has no expression for  $\Phi(x)$ , the dual problem can only be solved in terms of the related kernel function. In the method of fixed size LS-SVM the Nyström method is used to estimate eigenfunctions. After obtaining estimates for  $\Phi(x)$  and linking primal-dual formulations, the computation of  $w, b$  is done in the primal space. In nonlinear function estimation and nonlinear modeling problems, the most widely used Mercer kernel is the RBF  $K(x_i, x_j) = \exp\{-\|x_i - x_j\|^2 / 2\sigma^2\}$  where,  $\sigma$  is a scale parameter which controls the width of the RBF. A RBF kernel corresponds to an infinite-dimensional Mercer kernel feature space, since the RBF has an infinite number of eigenfunctions. This is because an RBF only has effect on a certain region of the space which makes it prone to the curse of dimensionality. Fixed size algorithm is described as follows:

- Give a training dataset for data standardization  $\{(x_1, y_1), \dots, (x_i, y_i), \dots, (x_n, y_n)\}$ ,  $x_i \in \mathbb{R}^m, y_i \in \mathbb{R}$
- Select a support vector whose working set is  $M, M \ll n$
- Randomly select support vector  $x^*$  from working set  $M$
- From training set, randomly select sample  $x_i^*$  to substitute for  $x_i^*$  to be input to support vector set; compute the working set entropy after substitution. If the entropy increases, then  $x_i^*$  is kept in support vector set, otherwise, its possibility of joining support vector set is eliminated, using primal  $x^*$  as support vector
- Compute entropy of substituted working set; if the entropy is too small or iteration number exceeds the limit, then stop, otherwise, return to Step 3
- Use eigenfunction estimated in Nyström approximation to estimate  $w, b$  values in the primal space

In the last step, a regression is employed in the primal space which makes the method suitable for solving large scale nonlinear function estimation and classification problems.

Espinoza *et al.* (2006) applied FS\_LS-SVM on short-term power forecast, primarily using plant power consumption record as the training data to forecast power consumption in the upcoming 24 h period. The support vector setting of 1 and 4% yielded good results in the training samples. Hoegaerts *et al.* (2004) compared LS-SVM, FS\_LS-SVM and sparse KPLS in large scale problems, indicating that both FS\_LS-SVM and sparse KPLS have good performance in massive data computation. Karsmakers *et al.* (2007b) studied classification result of combined classification model of Fixed Size and KLR in TIMIT phonetic database. The result indicated that fixed size classification model is better than the KLR classification model.

**Performance evaluation:** After building the classifier model, researchers used a number of methods to evaluate classifier model performance and classification capability, including Receiver Operating Characteristic (ROC) curve, classification accuracy and F-Measure. ROC curve is able to identify the discrimination ability of classification system, while accuracy represents the ratio of discriminating class correctly under ROC. Higher F-Measure would produce better classifier identification ability. Data inputted into the classifier are divided into positive examples and negative examples and the classifier outputs are also divided into positive example and negative example. Table 1 shows the Confusion Matrix (Fawcett, 2006).

In Table 1, TP denotes that positive example is correctly classified as positive example (true positive); FP denotes that positive example is wrongly classified as negative example (false positive); FN is that negative example is wrongly classified as positive example (false negative); TN is that negative example is correctly classified as negative example (true negative). Prior to computing ROC, two indices, specificity and sensitivity, need to be calculated. Computational method is as shown in Eq. 6 and 7.

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{6}$$

$$\text{Specificity} = \frac{TN}{FP + TN} \tag{7}$$

Table 1: Confusion matrix

Actual predict	Positive	Negative
Positive	TP	FP
Negative	FN	TN

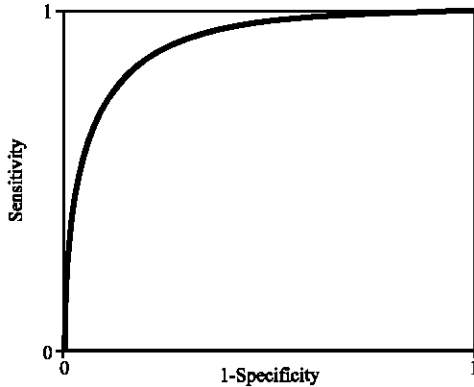


Fig. 1: ROC curve

Area under the ROC Curve (AUC) (Provost and Fawcett, 2001) is the area under curve (Fig. 1 black block line) plotted with X-axis (1-specificity) and Y-axis (sensitivity), 1.0 to the maximum. In general, with larger area under ROC curve, the result is better because larger area indicates that the classifier model can better identify data classes; otherwise, the discrimination ability is worse.

Accuracy is the ratio of discriminating class correctly under classifier-interpreted ROC and can be use to learn the number of class classifier can identify correctly and number of classes wrongly identified. Its computational method is shown in Eq. 8:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{8}$$

F-Measure is the harmonic mean of Recall and Precision (Landgrebe *et al.*, 2006), recall denotes accuracy of a real class; precision denotes accuracy of an estimated class. When recall and precision are higher, F-Measure is higher as well, indicating better classification model. This study used ROC, accuracy and F-Measure to evaluate the classifier performance.

$$Recall = \frac{TP}{TP + FN} \tag{9}$$

$$Precision = \frac{TP}{TP + FP} \tag{10}$$

$$F-Measure = \frac{2(Recall)(Precision)}{Recall+Precision} \tag{11}$$

**DATA SETS AND EXPERIMENT DESIGN**

For the experiments, data sets were obtained from the University of California at Irvine (UCI) machine learning repository and are Pima Indian diabetic database (PIDD),

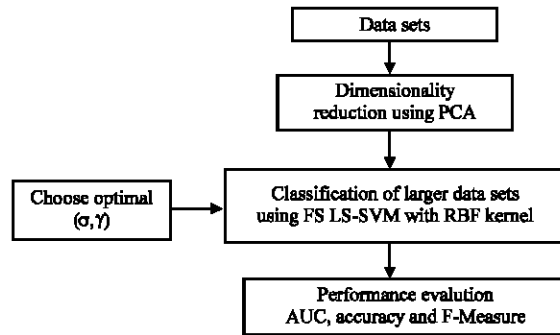


Fig. 2: The block diagram of proposed system

Table 2: Information about the experimental datasets

Dataset	Attributes	Size	Size	
			Class 1	Class 2
PIDD	8	768	500	268
Ionosphere	34	351	225	126
Vote	16	435	168	267
Tic-Tac-Toe	8	958	332	626

Ionosphere, Vote and Tic-Tac-Toe. The detailed information is shown in Table 2.

In the preprocessing, PCA method was used for dimensionality reduction before FS\_LS-SVM and LS-SVM training stage. In this stage, there are two criteria used to select new components: eigenvalue is greater than 1 and cumulative descriptive variation is greater than 80%. In setting training subset M,  $M = \{50, 100, 150, \dots, n\}$ , this study took RBF for core function in FS\_LS-SVM classifier training stage because its difference has significant effect on the classifier performance. In other words, factor setting difference would affect the classification model stability and accuracy, thus it is important to find appropriate factor setting for the classifier. In this study, Mean Squared Errors (MSE) is the evaluation index,

$$MSE = (\sum_{i=1}^n e_i^2) / n,$$

where, e is the real value- estimated value, n is the number of training samples. In setting factor  $\sigma$  and  $\gamma$ , the value domain searches for match between 0.001 and 1000. After selecting  $\sigma$  and  $\gamma$  suitable for training set data, a training data classification model is built. The block diagram of the proposed system is shown in Fig. 2.

**EXPERIMENTAL RESULTS AND DISCUSSIONS**

Based on above-mentioned research frame, this study used PIDD, Ionosphere, Vote and Tic-Tac-Toe databases for testing. Table 3 shows the experimental result of 4 test methods, which are LS-SVM, FS\_LS-SVM,

Table 3: Comparison of experimental performance of FS\_LS-SVM and PCA\_FS\_LS-SVM

PIDD dataset							
Method	Attributes	$\sigma^2$	$\gamma$	Subset	Accuracy (%)	AUC	F-measure
LS-SVM	8	35.0	6.5	N/A	78.06	0.825	0.671
FS_LS-SVM	5	125.0	1.5	100	79.35	0.836	0.685
PCA_LS-SVM	8	70.0	200.0	N/A	78.70	0.839	0.673
PCA_FS_LS-SVM	5	5.0	0.2	100	81.29	0.856	0.701
<b>Ionosphere dataset</b>							
FS_LS-SVM	34	2.0	6.0	50	97.18	0.996	0.979
PCA_FS_LS-SVM	12	1.5	15.0	50	98.59	1.000	0.989
<b>Vote dataset</b>							
FS_LS-SVM	16	100.0	0.1	50	98.85	0.998	0.991
PCA_FS_LS-SVM	9	15.0	0.1	50	96.55	0.991	0.967
<b>Tic-Tac-Toe dataset</b>							
FS_LS-SVM	9	20.0	20.0	300	99.47	0.999	0.995
PCA_FS_LS-SVM	7	10.0	10.0	300	88.02	0.867	0.909

Table 4: AUC of all the experimental result

Data set	Method	Accuracy (%)	AUC	F-Measure
PIDD	PCA_FS_LS-SVM	81.29	0.856	0.701
Ionosphere	PCA_FS_LS-SVM	98.59	1.000	0.989
Vote	FS_LS-SVM	98.85	0.998	0.991
Tic-Tac-Toe	FS_LS-SVM	99.47	0.999	0.995

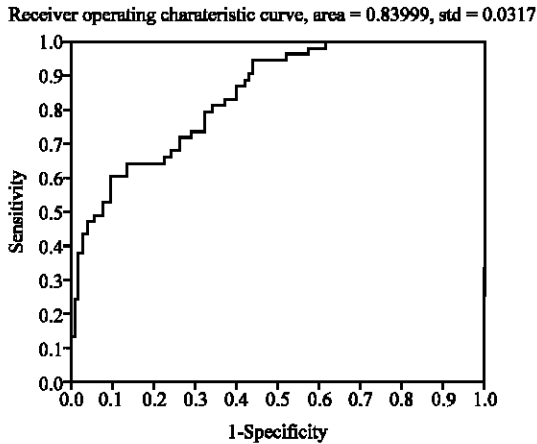


Fig. 3: ROC curve for PCA\_LS-SVM on 80-20% of training-test partition for PIDD

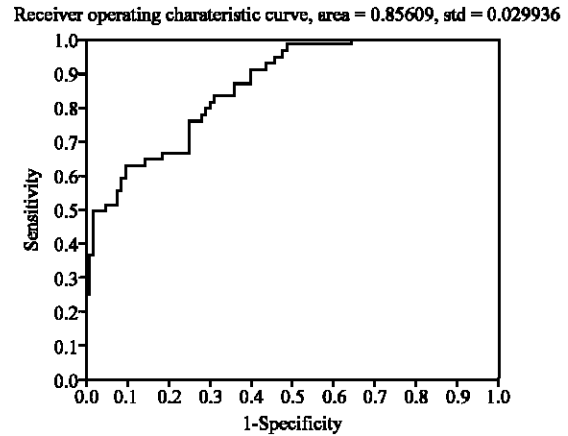


Fig. 4: ROC curve for PCA\_FS\_LS-SVM on fixed subset M = 100 for PIDD

PCA\_LS-SVM and PCA\_FS\_LS-SVM. As seen, when using FS\_LS-SVM and PCA\_FS\_LS-SVM to test the Ionosphere database, the test set accuracy was 97.18 and 98.59%, respectively; PCA\_FS\_LS-SVM (98.59%) showed better classification result. The research result of Vote dataset showed that, the test set accuracy of FS\_LS-SVM and PCA\_FS\_LS-SVM was 98.85 and 96.55%, respectively, such database reduction in PCA did not help to enhance the performance. Tic-Tac-Toe database was tested with FS\_LS-SVM and PCA\_FS\_LS-SVM and the test accuracy was 99.74 and 88.02%, respectively. FS\_LS-SVM classification accuracy was the best. The optimum classification results of 4 datasets are summarized in Table 4. Comprehensive research results showed that FS\_LS-SVM uses smaller subset to deal with larger scale

datasets, resulting in better performance than LS-SVM and PCA\_LS-SVM (such as Table 3 PIDD M = 100). Figure 3 and 4 show the PIDD testing ROC graphs of PCA\_LS-SVM and PCA\_FS\_LS-SVM. As for PIDD and Ionosphere datasets, two-stage classifier PCA\_FS\_LS-SVM after PCA dimensionality reduction would improve the classification accuracy.

### CONCLUSION

LS-SVM has been successfully applied to real application performance modeling considered as a non-linear/arbitrary function regression problem. This study applied PCA and fixed size method to improve the accuracy of LS-SVM classification algorithm in larger

scale datasets. The fixed size LS-SVM with an RBF kernel produced better results than a PCA\_LS-SVM nonlinear model estimated with the same variables and also better than a standard LS-SVM in dual space estimated using fewer data points. The simulation results indicated that fixed size LS\_SVM could shorten the training time significantly and with good predicting precision on different datasets. The experiment results presented that the FS\_LS-SVM with an RBF kernel is selected as the best classification method for the classification problems of larger scale datasets with an average test set accuracy (ACC) and an average test set AUC.

#### ACKNOWLEDGMENT

The authors would like to thank for the anonymous referees for their careful reading of the paper and several suggestions which improved the paper. The authors would also like to thank for the National Science Council of the Republic of China for financially supporting this research under contract No. 96-2221-E-167-022.

#### REFERENCES

- Cao, L.J. and W.K. Chong, 2002. Feature extraction in support vector machine: A comparison of PCA, XPCA and ICA. *Neural Information Processing, ICONIP '02. Proceedings of the 9th International Conference*, Nov. 18-22, Singapore, pp: 1001-1005.
- Chang, C.C. and C.J. Lin, 2001. Training nu-support vector classifiers: Theory and algorithms. *Neural Comput.*, 13: 2119-2147.
- Draper, B., K. Baek, M. Bartlett and R. Beveridge, 2003. Recognizing faces with PCA and ICA. *Comput. Vision Image Understand.*, 91: 115-137.
- Espinoza, M., J.A.K. Suykens and B. De Moor, 2006. Fixed-size least squares support vector machines: A large scale application in electrical load forecasting. *Comput. Manage. Sci.*, 3: 113-129.
- Fawcett, T., 2006. An introduction to ROC analysis. *Pattern Recog. Lett.*, 27: 861-874.
- Gestel, T.V., J.A.K. Suykens, B. Baesens, S. Viaene and J. Vanthienen *et al.*, 2004. Benchmarking least squares support vector machine classifiers. *Mach. Learn.*, 54: 5-32.
- Girolami, M., 2002. Orthogonal series density estimation and the kernel eigenvalue problem. *Neural Comput.*, 14: 669-688.
- Hoegaerts, L., J.A.K. Suykens, J. Vandewalle and B. De Moor, 2004. Primal space sparse kernel partial least squares regression for large scale problems. *IEEE International Joint Conference on Neural Networks*, July 25-29, Budapest, Hungary, pp: 561-566.
- Hsu, C.W. and C.J. Lin, 2002. A comparison of methods for multi-class support vector machines. *IEEE Trans. Neural Networks*, 13: 415-425.
- Joachims, T., 1999. Making Large-Scale SVM Learning Practical. In: *Advances in Kernel Methods-Support Vector Learning*, Schölkopf, B., C.J.C. Burges and A.J. Smola (Eds.). MIT Press, Cambridge, MA., 0-262-19416-3, pp: 169-184.
- Karsmakers, P., K. Pelckmans and J.A.K. Suykens, 2007a. Multi-class kernel logistic regression: A fixed-size implementation. *International Joint Conference on Neural Networks*, Aug. 2-17, Orlando, Florida, USA., pp: 1756-1761.
- Karsmakers, P., K. Pelckmans, J.A.K. Suykens and H. Van Hamme, 2007b. Fixed-size kernel logistic regression for phoneme classification. *Proceedings Interspeech 2007-Eighth Annual Conference of the International Speech Communication Association*, Aug. 27-31, Antwerp, Belgium, pp: 78-81.
- Katz, M., M. Schaffner, E. Andelic, S. Krüger and A. Wendemuth, 2005. Sparse kernel logistic regression for phoneme classification. *Proceedings of 10th International Conference on Speech and Computer (SPECOM)*, October 05, Patras, Greece, pp: 523-526.
- Landgrebe, T.C.W., P. Paclik and R.P.W. Duin, 2006. Precision-recall operating characteristic (P-ROC) curves in imprecise environments. *Proceeding of the 18th International Conference Pattern Recog. (ICPR2006)*, Hong Kong, August 20-24, IEEE, USA., pp: 123-127.
- Lee, J., Y.M. Cheon, S.Y. Kim and E.J. Park, 2007. Emotional evaluation of color patterns based on rough sets. *International Symposium on Information Technology Convergence, ISITC 07*, November 23-24, IEEE Computer Society, Jeonju, Korea, pp: 325-332.
- Mercer, J., 1909. Functions of positive and negative type and their connection with the theory of integral equations. *Philos. Trans. Roy. Soc., Lond. Ser. A*, 209: 415-446.
- Mumtaz, A., S.A.M. Gilani, K. Hameed and T. Jameel, 2008. Enhancing performance of image retrieval systems using dual tree complex wavelet transform and support vector machines. *J. Cmput. Inform. Technol.*, 16: 57-68.
- Provost, F. and T. Fawcett, 2001. Robust classification for imprecise environments. *Mach. Learn.*, 42: 203-231.
- Ruping, S., 2001. Incremental learning with support vector machines. *Proceedings IEEE International Conference on Data Mining*, November-December 2, IEEE Computer Society, San Jose, California, USA., pp: 641-642.

- Scholkopf, B., A.J. Smola and K.R. Muller, 1998. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Comput.*, 10: 1299-1319.
- Suykens, J.A.K., L. Lukas, P.V. Dooren, B. de Moor and J. Vandewalle, 1999. . Least squares support vector machine classifiers: A large scale algorithm. *Neural Process. Lett.*, 9: 293-300.
- Suykens, J.A.K., T.V. Gestel, J.D. Brabanter, B. De Moor and J. Vandewalle, 2002. Least Squares Support Vector Machines. 1st Edn., World Scientific, Singapore, 978-981-238-151-4.
- Tsumoto, S., 2000. Automated discovery of positive and negative knowledge in clinical databases. *IEEE Eng. Med. Biol. Mag.*, 19: 56-62.
- Vapnik, V., 1998. *Statistical Learning Theory*. 1st Edn., Wiley-Interscience Publication, New York, ISBN: 0471030031.
- Wang, H., D. Bell and F. Murtagh, 1999. Axiomatic approach to feature subset selection based on relevance. *IEEE Trans. Pattern Anal. Mach. Intell.*, 21: 271-277.
- Williams, C. and M. Seeger, 2001. Using the Nyström Method to Speedup Kernel Machines. In: *Advances in Neural Information Processing Systems*, Leen, T.K., T.G. Diettrich and V. Tresp (Eds.). MIT Press, Cambridge, Massachusetts, ISBN: 0-262-56145-X, pp: 682-688.
- Zhao, K., Y.J. Tian and N.Y. Deng, 2007. Robust unsupervised and semisupervised bounded c-support vector machines. 7th IEEE International Conference on Data Mining Workshops, October 28-31, IEEE Computer Society, Omaha, NE, USA., pp: 331-336.