A COMPARATIVE STUDY ON THE PERFORMANCE OF SUPPORT VECTOR MACHINES WITH ARTIFICIAL NEURAL NETWORKS IN RAIN FALL FORECASTING

[1]KolluruVenkata Nagendra, [2]Dr. Maligela Ussenaiah

[1]Research Scholar, [2]Assistant Professor,
Department of Computer Science,
Vikrama Simhapuri University, Nellore, Andhra Pradesh, India

Abstract: **The scope of the Machine Learning techniques is rapidly changed and increased in the field of rain fall forecasting and weather prediction. The Rain fall prediction is essential for the agriculture dependent countries like India, for analyzing the crop productivity, use of water resources and pre-planning of water resources, rainfall prediction is important. Statistical techniques for rainfall forecasting cannot perform well for long term rainfall forecasting due to the dynamic nature of climate phenomena. A new regression technique called Support Vector Machines was developed. The basic principal of SVM is Structural Risk Minimization (SRM). This paper provides a detailed survey and comparison of SVM with ANN. Moreover, the paper also presents different accuracy measures used by researchers for evaluating performance of SVM**

IndexTerms **- Artificial Intelligence, Numerical Weather Prediction, Support Vector Machines, Risk minimization.**

I. INTRODUCTION

In the prevention of flood and the water resources management, the exert and accurate prediction of rainfall is needed and also it is very important for crop productivity countries like India, China, Australia, Pakistan, and Iran. The prediction of rain fall helps to minimize the risks and damages of floods and heavy rains. Classical statistical estimation methods are linear, model driven, and parametric in nature, which assumes strong a priori knowledge about the unknown dependency. The training data is used to estimate the parameter values. The previous Statistical methods and Numerical Weather Prediction (NWP) model may not generate good results for non-linear process because statistical methods are developed based on the assumption of linear time series. So that the statistical methods cannot identify irregularities in the time series and nonlinear pattern. Local weather situations such as cloud, fog, and peak of strong rush of wind can affect the rainfall generation process. NWP models cannot solve the local weather situations because the local weather situations are unstable. This is the general problem with weather prediction models. [1][2].

Recently, the support vector machine has attracted the attention of many researchers. It has a functional form (similar to model driven approach), the complexity of which is decided by the available data to be "learned" (similar to data driven approach), and although SVM has an underlying functional form, its exact nature is not assumed a priori (unlike model driven methods). In other words, this machine can be seen as a statistical tool that approaches the problem similar to a Neural Network with a novel way to train polynomial function, Radial Basis Function, or neural network regression estimators. More precisely, it is an approximate implementation of the method of structural risk minimization. This induction principle minimizes an upper bound on the error rate of a learning machine on test data (i.e. generalization error) rather than minimizing the training error itself (used in empirical risk minimization as in ANN). This helps in making generalizations on the unseen data.

Support vector classifiers have already become competitive with the best available techniques for Classification. In the recent past, its excellent performances on regression and time series prediction have been demonstrated in various studies.[3][4]. The intention of this paper is twofold to introduce SVM with its applications and to compare it with ANN. ANN has been chosen for comparison because SVM is essentially a data driven model although it has the final functional form similar to model driven approach. Further, ANN has been shown to perform better than many conventional regression methods. [5][6][7].

In the next section, a discussion on Structural Risk Minimization principle is presented followed by the background knowledge of Support Vector Machine for regression. Section III are the features of SVM. After that Section IV is a brief literature survey on SVM. Section V gives a qualitative discussion on the advantages of SVM over ANN report by American Society of Civil Engineering [ASCE] Task Committee (2000a and 2000b). The final section deals with conclusion followed by references.

II.STRUCTURAL RISK MINIMIZATION PRINCIPLE

The problem of learning from data (examples) is to choose from the given set of functions fβ, β€Δ the one that best approximates the measured output based on a training exemplars of n examples (x_1,y_1),. . (x_n, y_n),. Each generated from an unknown probability distribution P(x, y). The best approximation implies the smallest possible value of the following risk, R(β),

$$R(\beta) = \int \left(y - f_\beta\right)^2 dP(x, y) \qquad (1)$$

The problem is that R(β) is unknown, since P(x, y) is unknown. Therefore an induction principle for risk minimization is necessary. The straightforward approach is to minimize the empirical risk given by

$$R_{emp}(\beta) = \frac{1}{n}\sum_{i=1}^{n}\left(f_\beta - y\right)^2 \qquad (2)$$

However, this approach does not guarantee a small actual risk (test set error) for a small error on training exemplars, if the number of training examples, n, is limited. To obtain the best possible actual risk from limited data, novel techniques have been developed in the last two decades based on statistical learning theory. According to this theory, the generalization ability of learning machines depends on capacity concepts that are more sophisticated than merely the dimensionality of the space or the number of free parameters of the loss function (as espoused by the classical paradigm of generalization). One such technique is the Structural Risk Minimization principle (Vapnik, 1999). It is based on the fact that for the above learning problem, for any β€Δ, the bound on test error is of the form,

$$R(\beta) \le R_{emp}(\beta) + \Omega\left(\frac{n}{h}\right) \qquad (3)$$

Where the first term is an estimate of the risk and the second term is the confidence interval for this estimate. The parameter h is called VC-dimension (named after the authors) of a set of functions. It can be seen as the capacity of a set of functions implementable by the learning machine. For ANN, determining h corresponds with choosing appropriate network architecture for a given training set. During the training phase, the network tries to minimize the first term in Equation (3). If the chosen architecture happens to be too complex for the given amount of training data, the confidence interval term will be large. So, even if one could minimize the empirical risk, the actual risk still remains large, thus resulting in poor generalization.

According to the Structural Risk Minimization principle (SRM), one can control the actual risk by controlling the two terms in Equation (3). Thus, for a given set of observations (x_1,y_1),. . (x_n,y_n), the SRM principle chooses the function fβ* in the subset {fβ : β€Δ }, for which the guaranteed risk bound as
given by Equation (3) is minimal.

III. BACKGROUND KNOWLEDGE OF SVM

SVM is the state-of-the-art neural network technology based on statistical learning [3][4]. The basic idea of SVM is to use linear model to implement nonlinear class boundaries through some nonlinear mapping of the input vector into the high-dimensional feature space. The linear model constructed in the new space can represent a nonlinear decision boundary in the original space. In the new space, SVM constructs an optimal separating hyper plane. If the data is linearly separated, linear machines are trained for an optimal hyper plane that separates the data without error and into the maximum distance between the hyper plane and the closest training points. The training points that are closest to the optimal separating hyper plane are called support vectors. The basic concept of SVM illustrates in the below Fig.1. There exist uncountable decision functions, i.e. hyper planes, which can effectively separate the negative and positive data set (denoted by ‘x’ and ‘o’, respectively) that has the maximal margin. This indicates that the distance from the closest positive samples to a hyper plane and the distance from the closest negative samples to it will be maximized.

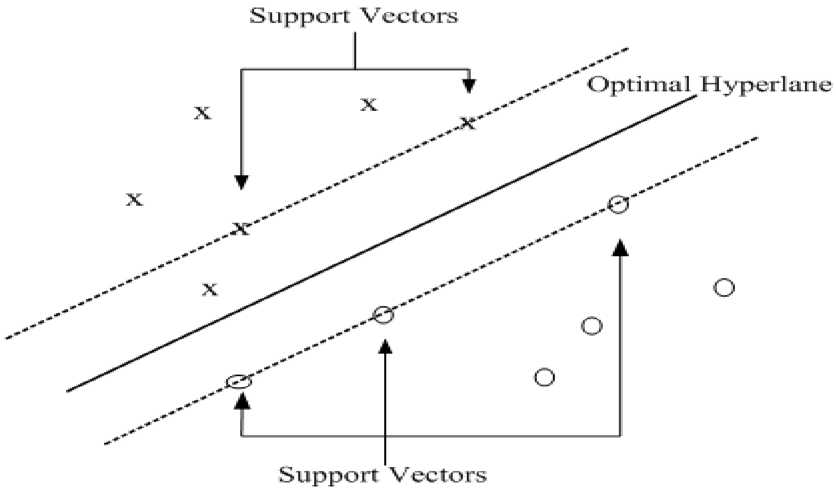

Fig. 1. The basis of the support vector machines

The basic idea of SVM (support Vector Machines) is to map the original data X into a feature space F with high dimensionality through a non-linear mapping function and construct an optimal hyper-plane in new space. SVM can be applied to both classification and regression. In the case of classification, an optimal hyper-plane is found that separates the data into two classes. Whereas in the case of regression a hyper-plane is to be constructed or developed that lies close or near to as many points as possible.

A Support Vector Machine (SVM) [7] performs classification by constructing an N-dimensional hyper plane that optimally separates the data into two categories. SVM models are closely related to neural networks. In fact, a SVM model using a sigmoid kernel function is equivalent to a two-layer, perceptron neural network.

Support Vector Machine (SVM) models are a close cousin to classical multilayer perceptron neural networks. Using a kernel function, it acts as an alternative training method for polynomial, radial basis function and multi-layer perceptron classifiers. The vectors near the hyper plane are the support vectors. The following figures explains [8] the support vectors, hyper plane, margin and related equations. The sequence of figures shows the how support vectors are defined and their equations, margins and noisy data, Outliers and slack variables ξi. Apart from that possible solutions along with hyper planes are shown.

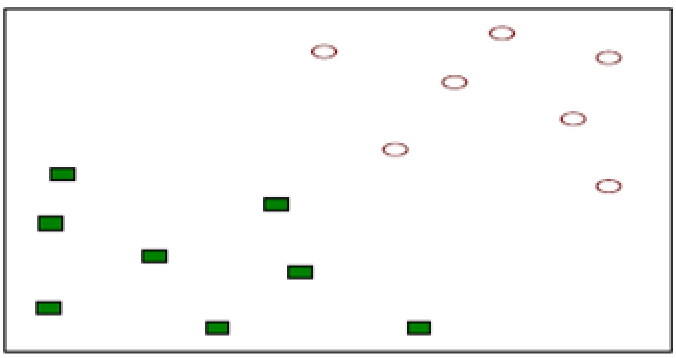

Fig:2. Find a Linear Hyper Plane (Decision Boundary)

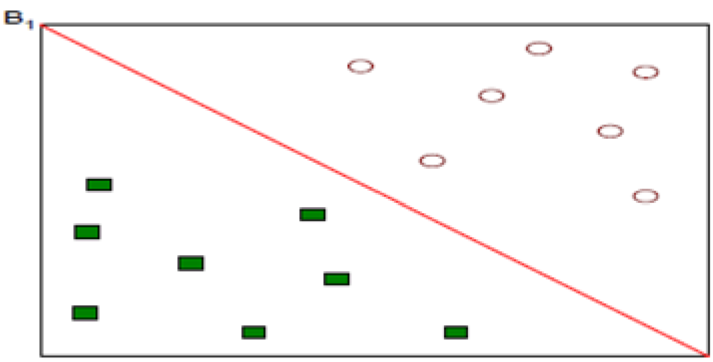

Fig:3. One possible solution that will Separate the data

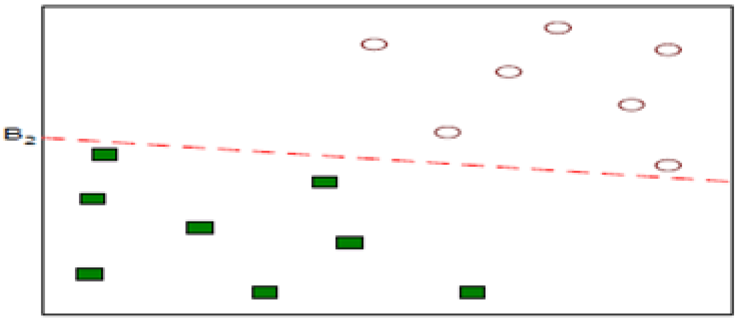

Fig:4.Another possible Line

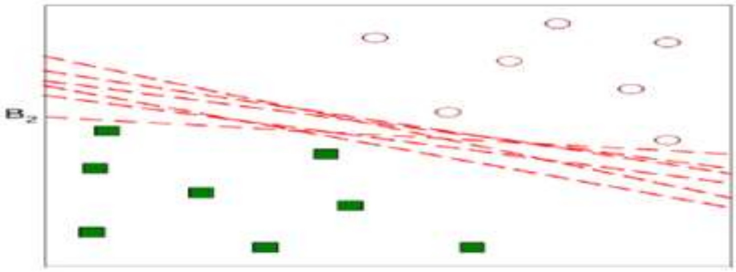

Fig:5. Other Possible Solutions

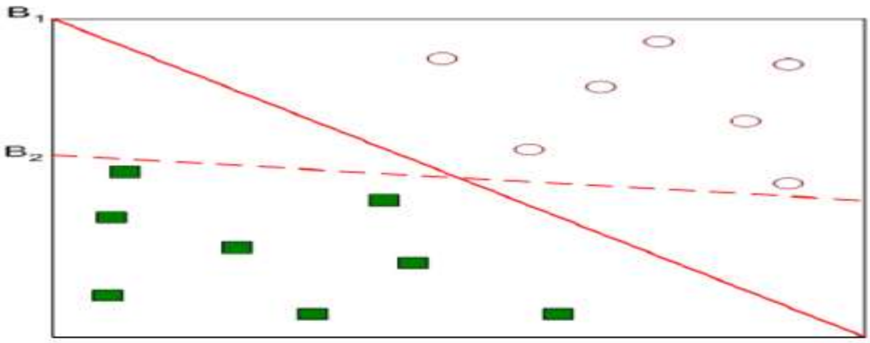

Fig:6 Which one between B1 and B2 defines better

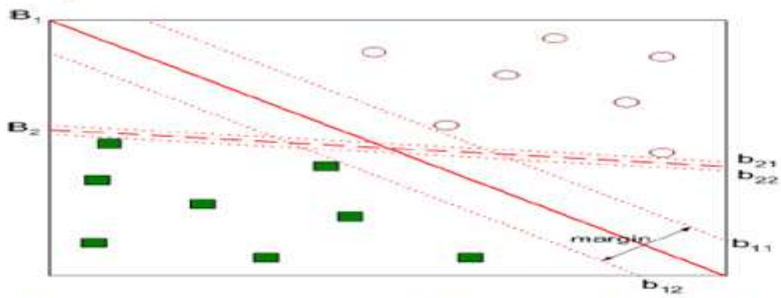

Fig:7. Find Hyper Plane maximizes the margin: B1 is better than B2

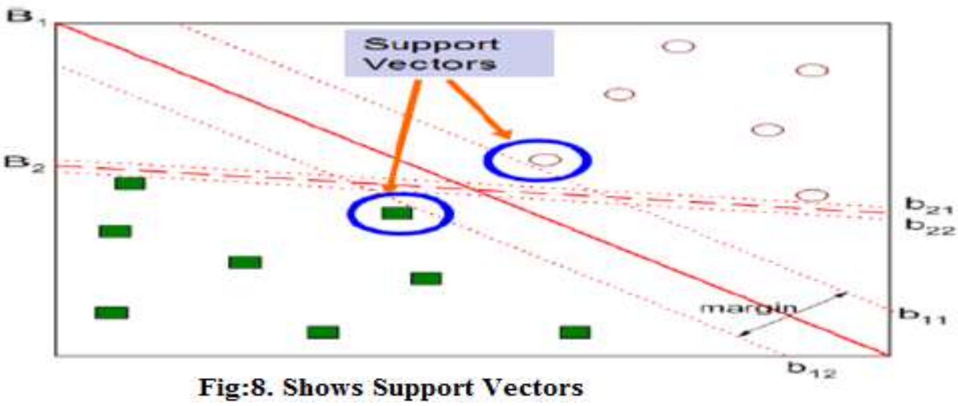

Fig:8. Shows Support Vectors

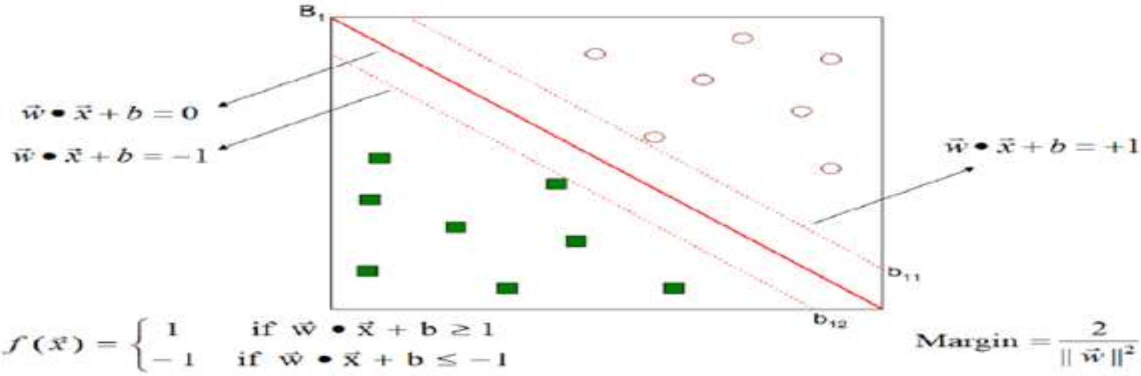

Fig:9. Shows Equation and Margin

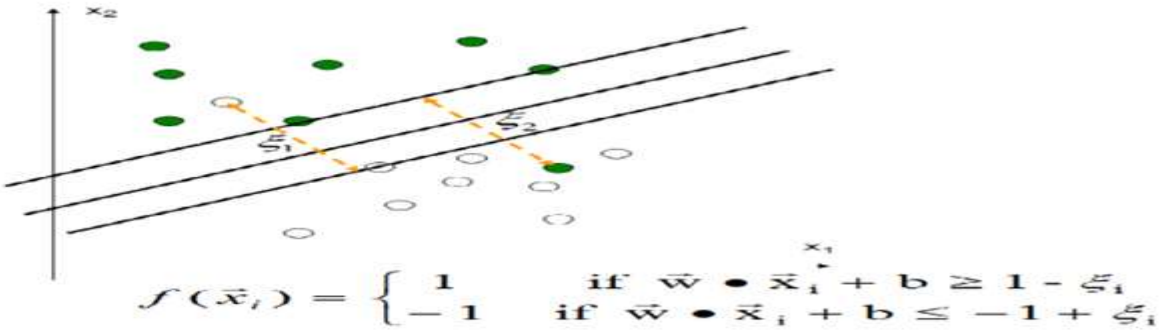

Fig:10.Noisy data, Outliers etc., Slack Variable ξi

IV.FEATURES OF SVM MODELS

Support Vector Machine is [9, 10, and 11] is the outgrowth of ANN.

- Highly accurate models.
- Classification and Regression analyses.
- Automatic grid search and pattern search for optimal parameters.
- Model building performance.
- Continuous, categorical and non-numeric variables
- Missing value substitution
- V-fold cross validation

V. LITERATURE SURVEY OF SVM

The foundation of Support Vector Machines (SVM) was given by Vapnik, a Russian mathematician in the early 1960s [3], based on the Structural Risk Minimization principle from statistical learning theory and gained popularity due to its many attractive features and promising empirical performance. SVM has been proved to be effective in classification by many researchers in many different fields such as electric and electrical engineering, civil engineering, mechanical engineering, medical, financial and others [12]. Recently, it has been extended to the domain of regression problems [13]. In the river flow modeling field, Liong & Sivapragasam (2002) [14] compared SVM with Artificial Neural Networks (ANN) and concluded that SVM's inherent properties give it an edge in overcoming some of the major problems in the application of ANN [15]. Nonlinear modeling of river flows of the Bird Creek catchment in the USA with SVM was reported to have its limitations (Han & Yang 2001; Han et al. 2002). Dibike et al. (2001) [16][17] presented some results showing that Radial Basis Function (RBF) is the best kernel function to be used in SVM models. However, the author of [18], found linear kernel outperformed other popular kernel functions (radial basis, polynomial, sigmoid). Bray & Han (2004) [19] illustrated the difficulties in SVM identification for flood forecasting problems. It is clear that, due to its short history, there are still many knowledge gaps in applying SVM in flood forecasting and some conflicting results from different researchers are a good indication that this technique is still in its infancy and more exploratory work is necessary to improve our understanding of this potentially powerful tool from the machine learning community.

VI. STRENGTHS OF SVM OVER ANN

In the earlier years, the researchers widely used ANNs in hydrologic applications like rainfall forecasting, stream flow prediction, ground water modeling, rainfall-runoff modeling, reservoir operations etc. In a study of the committee by ASCE it was pointed out that although ANN does have many attractive features, it suffers from some major limitations, inviting skeptical attitude towards the methodology. SVM seems to be a powerful alternative to overcome some of the basic lacunae in application of ANNs, while retaining all the strengths of ANN. The below table illustrates that the strengths of SVM over ANN.

Table I. The advantages of SVM with ANN

S.No	Limitations	ANN	SVM
i	Black-Box Model.	The set of optimal weights and threshold values (after the training) does not reveal any information to the user.	SVM is not a "black box" model. It can be analyzed theoretically using concepts from computational Learning theory.
ii	Identifying Optimal Training Set.	ANN is data intensive, without proper quality and quantity of data; the generalization will be very poor.	SVM is based on Structural Risk Minimization principle (SRM), it offers a better generalization error as compared to ANN for a given training set.
iii	Optimal Architecture.	In ANN the architecture is usually determined by users past experience and preference, rather than physical aspect of the problem. The use of optimal network architecture is one of the major issues in the ANN applications.	In SVM, the final architecture is automatically obtained from the solution of the optimization problem that gives the support vectors. The number of support vectors can actually be seen as the number of hidden neurons in single hidden layer architecture.
iv	No Local Minima.	The optimization problem formulated for ANN is always more solvable and thus suffer from limitations of ways of regularization this lead them to a local minimum.	The optimization problem formulated for SVM is always uniquely solvable and thus does not suffer from limitations of ways of regularization.
v	Improving Time Series Analysis	ANN structure more complicated with a greater number of tunable parameters.	SVM can deal with the increase in the number of attributes with relatively much greater ease, since in dual representation the dot product of two vectors of any dimension can be easily estimated.
vi	Adaptive Learning	ANN learning is a "black box" learning, it is not data adaptive.	SVM learning is not a "black box" learning, it is data adaptive to some degree. In fact, since only the useful training vectors form the basis for defining the final decision function, SVM is expected to give a relatively good generalization performance for future hydrologic conditions also
vii	Exploiting Higher Dimensional Features	In ANN, efficient use of high dimensional feature space is not possible.	In SVM, efficient use of high dimensional feature space is possible through kernel functions.
viii	Learning Basis in Higher Dimension	Besides computational problems in the ANN, the danger of over-fitting inherent in high dimensions may result in poor generalization.	SVM provides a sophisticated learning bias from the statistical learning theory to account for the generalization problem in higher dimensional feature space.

VII. CONCLUSION

In Rain fall forecasting application the SVM is very strong and efficient. Even though the limited training set is available, the SVM gives a better generalization with the use of Structural Risk Minimization. The priori developed architecture is not required by SVM. In this paper we conclude that the support vector machines are the better and good approach for prediction of rain fall. With the use of the kernel functions SVM has efficient use in high dimensional feature space also. SVM solutions are identical, optimal and absent from local minima. In this paper Support Vector Machines are more efficient then Artificial Neural Networks.

REFERENCES

[1]. G.Geetha and R. S. Selvaraj, “Prediction of monthly rainfall in Chennai using Back Propagation Neural Network model,” Int. J. of Eng. Sci. and Technology, vol. 3, no. 1, pp. 211-213, 2011.
[2]. S. K. Nanda, D. P. Tripathy, S. K. Nayak, and S. Mohapatra, “Prediction of rainfall in India using Artificial Neural Network (ANN) models,” Int. J. of Intell. Syst. and Applicat., vol. 5, no. 12, pp. 1-22, 2013.
[3]. Vapnik, V., 1996. The Nature of Statistical Learning Theory. Springer-Verlag, New York, New York.
[4]. Karunanithi N., W. J. Grenney, D. Whitley, and K. Bovee, 1994. Neural Networks for River Flow Prediction. Journal of Computing in Civil Engineering, ASCE 8(2):201-220.
[5]. Tokar, A. S. and P. A. Johnson, 1999. Rainfall-Runoff Modeling Using Artificial Neural Networks. Journal of Hydrologic Engineering. ASCE 4(3):232-239.

[6]. Thth, E., A. Brath, and A. Montanan, 2000. Comparison of Short- Tei-m Rainfall Prediction Models for Real-Time Flood Forecasting. Journal of Hydrology 239: 132-147.
[7]. http://www.dtreg.com/mlfn.htm
[8]. Yasmin Anwar, Support vector machines CS 795/895 – summer 2011, Old Dominion University.
[9]. http://www.imbe.med.uni-erlangen.de/links/ensemblews/talks/meyer.pdf.
[10]. Rajasekhar, N. ; Babu, S.J. ; Rajinikanth, T.V.-Magnetic resonance brain images classification using linear kernel based support vector machines on 3rd IEEE Nirma University International Conference on Engineering (NUiCONE), 2012 ISBN : 978-1-4673-1720-7.
[11]. Rao, T. Rajasekhar, N. ; Rajinikanth, T.V. Drainage water level classification using support vector machines on 3rd IEEENirma University International Conference on Engineering (NUiCONE), 2013 ISBN:978-1-4799-0726-7.
[12]. Vapnik, V., 1999. An Overview of Statistical Learning Theory. IEEE Transactions on Neural Networks l0(5):988-999.
[13]. Kecman, V. 2001 Learning and Soft Computing: Support Vector Machines, Neural Networks, and Fuzzy Logic Models. MIT Press, Cambridge, MA.
[14] Liong, S.Y. and Sivapragasam, C., 2002. Flood stage forecasting with support vector machines. Journal of the American Water Resources Association, 38(1): 173-186.
[15]. Han, D., Chan, L. & Zhu, N. 2007 Flood forecasting using support vector machines. J. Hydroinformatics 9 (4), 267–276.
[16] Han, D. & Yang, Z. 2001 River flow modelling using support vector machines. In XXIX IAHR Congress, Beijing, China,17–21 September, pp. 494–499. Qinghua University Press,
China.
[17]. Dibike, Y. B., Velickov, S., Solomatine, D. & Abbott, M. B.2001 Model induction with support vector machines:introduction and applications. J. Comput. Civil Engng.15 (3), 208–216.
[18]. Bray, M. 2002 Identification of Support Vector Machines for Runoff Modelling. MEng thesis, Department of Civil Engineering,University of Bristol.
[19]. Bray, M. & Han, D. 2004 Identification of support vector machines for runoff modelling. J. Hydroinf. 6, 265–280.
[20]. Arthur, W. T., C. M. Tam, and D. K. Liu, 2001. Comparative Study of ANN and MRA for Predicting Hosting Times of Tower Cranes. Building and Environment. 36:457-467.
[21]. ASCE Task Committee, 2000a. Artificial Neural Networks in Hydrology. 1. Preliminary Concepts. Journal of Hydrologic Engineering, ASCE 5(2):115-123.
[22]. ASCE Task Committee, 2000b. Artificial Neural Networks in Hydrology. 2. Hydrologic Applications. Journal of Hydrologic Engineering, ASCE 5(2):124-137.
[23]. Cristianini, N. and J. Shawe-Taylor, 2000. An Introduction to Support Vector Machines. Cambridge University Press, United Kingdom.
[24].DHI, 1993. MIKE 11: User Manual. Horshoim, Denmark.
[25]. Drucker, H., C. Burges, L. Kaufman, A. Smola, and V. Vapnik, 1996. Linear Support Vector Regression Machines. NIPS 96.
[26]. Gunn, S. R., 1997. Support Vector Machines for Classification and Regression. Technical Report. Image Speech and Intelligent Systems Research Group, University of Southampton, United Kingdom.
[27]. Haykin, S., 1999. Neural Networks: A Comprehensive Foundation. Prentice Hall International Inc., New Jersey.
[28]. Hoque, M. M., 1994. Evaluating Design Characteristics for Floods in Bangladesh. In: Proceedings of 2nd International Conference on River Flood Hydr., W. R. White and J. Watts (Editors). Wiley, Chichester, England and New York, New York, pp. 15-26.
[29]. Hsu, K. L., H. V. Gupta, and S. Sorooshian, 1995. Artificial Neural Networks Modelling of the Rainfall-Runoff Process. Water Resources Research 31(10):2517-2530.
[30]. Liong, S. Y., W. H. Lim, and G. N. Paudyal, 1999. River Stage Forecasting in Bangladesh: Neural Network Approach. Journal of Computing in Civil Engineering, ASCE 14(1): 1-8.
[31]. Smola, A. J. and B. Schdolkopf, 1998. A Tutorial on Support Vector Regression. Technical Report NeuroCOLT Technical Report NCTR- 98-030, Royal Holloway College, University of London, United Kingdom.