Journal of Harbin Engineering University
ISSN: 1006-7043

Vol 45 No. 02
February 2024

A Comprehensive Survey on Twin Support Vector Machine Learning in Big Data

[1] **Sajja Suneel**, [2] **Dr. Ch N. Santhosh Kumar**, [3] **Dr. Saurabh Pal**

[1]Research Scholar, Veer Bahadur Singh Purvanchal University, Jaunpur-UP.

[2]Research Supervisor, Professor, Anurag Engineering College, Kodada.

[3]Research Supervisor, Professor & Head, VBS Purvanchal University, Jaunpur -UP

Abstract

During the last two decades, a substantial amount of research efforts has been intended for support vector machine at the application of various data mining tasks. Data Mining is a pioneering and attractive research area due to its huge application areas and task primitives. Support Vector Machine (SVM) is playing a decisive role as it provides techniques those are especially well suited to obtain results in an efficient way and with a good level of quality. In this paper, we survey the role of SVM in various Big data tasks like classification, Map-Reduce Processing, prediction, forecasting and others applications. In broader point of view, we have reviewed the number of research publications that have been contributed in various internationally reputed journals for the data mining applications and also suggested a possible no. of issues of SVM. The main aim of this paper is to extrapolate the various areas of SVM with Big Data Processing.

Keywords: Support Vector Machine (SVM), Data Mining, Artificial Neural Network (ANN)

1. Introduction

In recent years, the Artificial Neural Networks (ANNs) have been playing a significant role for variants of data mining tasks which is extensively popular and active research area among the researchers. The intend of neural network is to mimic the human ability to acclimatize to varying circumstances and the current environment. Starting from Mc Culloch-Pitts network, the research is highly popular to some of the higher order neural network. The methodology of an Artificial Neural Network, intended to imitate few capabilities of the human brain and has demonstrated great prospective for various low level computations and embodies prominent features such as learning, fault tolerance, parallelism etc. ANN is a popular technique for the application in most of the data mining fields including classification [1-3], forecasting [4-12], functional approximation [13-15], rule extraction [16-19], pattern recognition and medical applications [20-22].

In the present day of research, ANN has stood forward as a strong alternative to the traditional recognition models. The computer science research is already in young age for the implementation technique of some popular ANN models like Hopfield network, Multilayer Perceptron, Self-Organizing Feature Map, Learning Vector Quantization, Radial Basis Function, Cellular Neural Network, Adaptive Resonance Theory Networks, Counter Propagation Networks, Back Propagation Network and Support Vector Machines *etc*. Due to the presence of all these neural networks, it is as a reaction to this complexity, we designed a new abstraction that allows us to express the simple computations we were trying to perform but hides the messy details of parallelization, fault-tolerance, data distribution and load balancing in a library. Our abstraction is inspired by the map and reduce primitives present in Lisp and many other functional languages.

We realized that most of our computations involved applying a map operation to each logical “record” in our input in order to compute a set of intermediate key/value pairs, and then applying a reduce operation to all the values that shared the same key, in order to combine the derived data appropriately.

Our use of a functional model with user specified map and reduce operations allows us to parallelize large computations easily and to use re-execution as the primary mechanism for fault tolerance. reasonable to expect a rapid increase in our understanding of artificial neural networks leading to improved network paradigms and a host of application opportunities.

The insidious use of Support Vector Machine (SVM) in various data mining applications makes it an obligatory tool in the development of products that have implications for the human society. SVMs, being computationally powerful tools for supervised learning, are widely used in classification, clustering and regression problems. SVMs have been successfully applied to a variety of real-world problems [23] like particle identification, face recognition, text categorization, bioinformatics, civil engineering and electrical engineering *etc*.

In this work, a detailed survey is performed in almost majority of the data mining fields starting from the year 2001 to 2014. Section 2 gives the fundamental ideas about the Support Vector Machines, Section 3 outlines a road map to SVM, Section 4 gives the statistical analysis about SVM research and Section 5 concludes our work with some of the major issues of SVM technique. The majority of the work cited in this paper is journal articles. The reason for this is that, we want to report on data mining applications that are extensibly used through SVM. Specifically, we have considered papers written in the English language and are of Journals and Conference proceedings types. This paper has two major objectives which are to provide a short background to the SVMs that are relevant to data mining tasks and Provide a review of the state of the art in the application of SVM methods in data mining.

2. Support Vector Machine

Support Vector Machines (SVMs) as originally proposed by Vladimir Vapnik[24] within the area of statistical learning theory and structural risk minimization, have demonstrated to work successfully on various classification and forecasting problems. SVMs have been used in many pattern recognition and regression estimation problems and have been applied to the problems of dependency estimation, forecasting and constructing intelligent machines [25]. SVMs have the prospective to capture very large feature spaces, due to the generalization principle which is based on the Structural Risk Minimization Theory (SRM) *i.e.*, the algorithm is based on guaranteed risk bounds of statistical learning theory[26].

In MLP classifiers, the weights are updated during the training phase for which the total sum of errors among the network outputs and the desired output is minimized. The performance of the network strongly degrades for small data sizes, as the decision boundaries between classes acquired by training are indirect to resolute and the generalization ability is dependent on the training approach. In contrast to this, in SVM the decision boundaries are directly determined from the training data set for which the separating margins of the boundaries can be maximized in feature space.

A SVM is a maximum fringe hyperplane that lies in some space and classifies the data separated by non-linear boundaries which can be constructed by locating a set of hyperplanes that separate two or more classes of data points. After construction of the hyperplanes, the SVM discovers the boundaries between the input classes and the input elements defining the boundaries (support vectors [27]). From a set of given training samples labeled either positive or negative, a maximum margin hyperplane splits the positive or negative training sample, as a result the distance between the margin and the hyperplane is maximized. If there exist no hyperplanes that can split the positive or negative samples, a SVM selects a hyperplane that splits the sample as austerely as possible, while still maximizing the distance to the nearest austerely split examples.

Figure 1 indicates a linearly separable hyper plane, where there are two groups of data points represented by '□' and '□'. There may be possibility of an infinite no. of hyperplanes but in the described figure, only one hyper plane represented by solid line optimally separates the sample points and is situated in between the maximal margins.

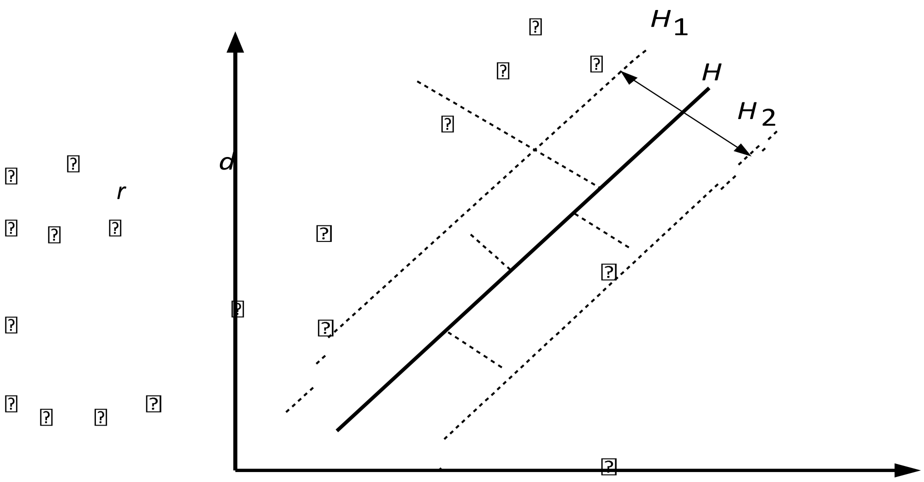

Figure 1. Linearly Separable Samples Indicated in a Hyperplane

Suppose we have N training samples like $\{(p_1, q_1), (p_2, q_2) \ldots (p_N, q_N)\}$ where $p_i \in R^d$ and $q_i \in \{-1, 1\}$. Equation (1) represents the equations of a hyper plane used for data portioning in SVM.

$$W \cdot p + b = 0 \quad (1)$$

where W is a weight, p is the training sample and b is the bias of the hyper plane. The margin between two classes are to be maximized, for which W should be minimized [28] subject to the condition

$$q_i(W \cdot p_i + b) \geq 1 \quad (2)$$

The optimization problem can be defined as

$$\min_{W,b} \frac{1}{2} W^T W \text{ with respect to } q_i(W * p_i + b) \geq 1, \text{ for } i = 1, 2, \ldots N \quad (3)$$

for the minimization of the value $\frac{1}{2} W^T W$, p_i and $q_i \in \{-1, 1\}$.

By introducing the Lagrange multiplier $\alpha_1, \alpha_2, \ldots \alpha_N \geq 0$ for solving this problem,

$$L(W, b, \alpha) = \frac{1}{2} W^T W - \sum_{i=1}^{N} \alpha_i q_i (W * p_i + b) + \sum_{i=1}^{N} \alpha_i \quad (4)$$

Hence, the problem becomes [29] $\max L(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{N} \alpha_i \alpha_j q_i q_j p_i \cdot p_j$ subject to

$$\sum_{i=1}^{N} \alpha_i q_i = 0, \quad \alpha_i \geq 0 \quad (5)$$

Now, if α_i^* is an optimal solution, the corresponding optimal bias and weight values can be updated as;

$$W^* = \sum_{i=1}^{N} \alpha_i^* q_i p_i \tag{6}$$

$$b^* = -\frac{1}{2} W^* [p_m + p_n]$$

where p_m, p_n are support vectors.

For linearly non-separable data (Figure 2) we introduce a non-negative variable called slack variable $V_i \geq 0$

Now, equation (2) becomes $q_i(W^* p_i + b) \geq 1 - V_i$ (7)

So, equation (3) will have the form,

$$\min_{W,b,V_i} \frac{1}{2} W^T W + R\left(\sum_{i=1}^{N} V_i\right) \tag{8}$$

subject to $q_i(W^* p_i + b) - 1 + V_i \geq 0$

where, $V_i \geq 0$, i = 1, 2, ..., N and R is a positive parameter [30] if the sample p_i is in the correct region *i.e.* $0 \leq V_i \leq 1$ otherwise, $V_i > 1$.

In the same way by introducing Lagrange multiplier $\alpha_1, \alpha_2, \ldots, \alpha_N \geq 0$. For solving the problem

$$\max L(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{i,j=1}^{N} \alpha_i \alpha_j q_i q_j \, p_i \cdot p_j$$

Subject to $\sum_{i=1}^{N} \alpha_i q_i = 0, \; \alpha_i \in R$, where N represents the number of support vectors.

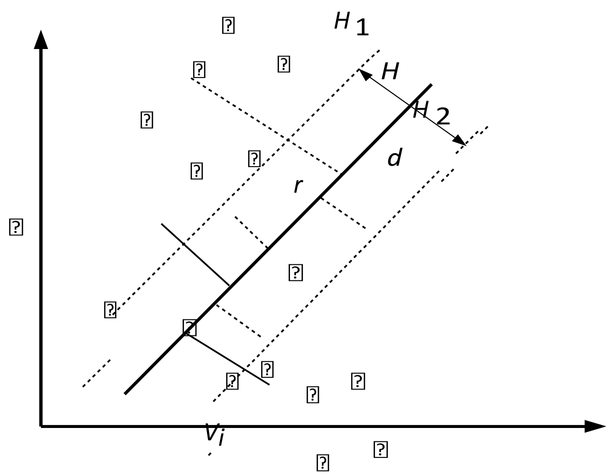

Figure 2. Linearly Non-Separable Samples Indicated in a Hyperplane

3. Literature Review

In this section, let us survey some major contributions towards SVM and its successful applications in various Big data Algorithms.

R. Burbidge *et. al.*, [09] have shown that the support vector machine (SVM) classification algorithm, proves its potential for structure–activity relationship analysis. In a benchmark test, they compared SVM with various machine learning techniques currently used in this field. The classification task involves in predicting the inhibition of dihydrofolate reductase by pyrimidines, using the data obtained from the UCI machine learning repository. Among three tested artificial neural networks, they found that SVM is significantly better than all of these.

Giorgio Valentini [13] have proposed classification methods, based on non-linear SVM with polynomial and Gaussian kernels, and output coding (OC), ensembles of learning machines to separate normal from malignant tissues, to classify different types of lymphoma and to analyze the role of sets of coordinately expressed genes in carcinogenic processes of lymphoid tissues. By using gene expression data from ‘‘Lymphochip’’, he has shown that SVM can correctly separate the tumoural tissues, and OC ensembles can be successfully used to classify different types of lymphoma.

Shutao Li *et. al.*, [15] have applied SVMs by taking DWFT as input for classifying texture, using translation-invariant texture features. They used a fusion scheme based on simple voting among multiple SVMs, each with a different setting of the kernel parameter, to alleviate the problem of selecting a proper value for the kernel parameter in SVM training and performed the experiments on a subset of natural textures from the Brodatz album. They claim that, as compared to the traditional Bayes classier and LVQ, SVMs, in general, produced more accurate classification results.

A training method to increase the efficiency of SVM has been presented by Yiqiang Zhan [14] for fast classification without system degradation. Experimental results on real prostate ultrasound images show good performance of their training method in discriminating the prostate tissues from other tissues and they claim that their proposed training method is able to generate more efficient SVMs with better classification abilities.

Yuchun Tang *et. al.*, [16] have developed an innovative learning model called granular support vector machines for data classification problems by building just two information granules in the top-down way. The experiment results on three medical binary classification problems show that granular support vector machines proposed in their work provides an interesting new mechanism to address complex classification problems, which are common in medical or biological information processing applications.

Bo-Suk Yang *et. al.*, [17] have presented a novel scheme to detect faulty products at semi-product stage in an automatic mass product line of reciprocating compressors for small-type refrigerators used in family electrical appliances. They presented the classification accuracy using the ANNs, SVM, LVQ, SOFM and SOFM with LVQ (SOFM-LVQ) and found SOFM-LVQ gives high accuracy and are the best techniques for classifying healthy and faulty conditions of small reciprocating compressors. The result shows SOFM with LVQ can improve the classification performance of SOFM but cannot eliminate the classification error, indicated in the concluding remarks.

Rung-Ching Chen [18] has proposed a web page classification method for extraction of feature vectors from both the LSA and WPFS methods by using a SVM based on a weighted voting schema. The LSA classifies semantically related web pages, offering users more complete information. The experimental results show that the anova kernel function yields the best result of these four kernel functions. The LSA-SVM, BPN and WVSVM were then compared and demonstrated that the WVSVM yields better accuracy even with a small data set.

Shu-Xin Du *et. al.*, [19] have developed a Weighted support vector machines for classification where penalty of misclassification for each training sample is different. Two weighted support vector machines, namely weighted C-SVM and V- SVM, have been developed for experimenting on breast cancer diagnosis which shows the effectiveness of the proposed methods. They have indicated that, the improvement obtained at the cost of the possible decrease of classification accuracy for the class with large training size and the possible decrease of the total classification accuracy.

Chih-Fong Tsai [20] has presented a two-level stacked generalization scheme composed of three generalizers having color texture of support vector machines (SVMs) for image classification. He has mainly investigated two training strategies based on two-fold cross-validation and non-cross-validation for the proposed classification scheme by evaluating their classification performances, margin of the hyperplane and numbers of support vectors of SVMs. The results show that the non- cross-validation training method performs better, having higher correct classification rates, larger margin of the hyperplane and smaller numbers of support vectors.

Chin-Teng Lin *et. al.*, [21] have proposed a support-vector-based fuzzy neural network (SVFNN) to minimize the training and testing error for better performance. They have developed a learning algorithm consisting of three learning phases is to construct the SVFNN in which the fuzzy rules and membership functions are automatically determined by the clustering principle. To investigate the effectiveness of the proposed SVFNN classification, they applied the corresponding model to various datasets from the UCI Repository and Statlog collection. Experimental results show that the proposed SVFNN for pattern classification can achieve good classification performance with drastically reduced number of fuzzy kernel functions.

Kemal Polat [22] has developed a medical decision making system based on Least Square Support Vector Machine (LSSVM) which was applied on the task of diagnosing breast cancer and the most accurate learning methods was evaluated. He conducted the experiment on the WBCD dataset to diagnose breast cancer in a fully automatic manner using LSSVM. The results strongly suggest that LSSVM can aid in the diagnosis of breast cancer. In his conclusion he has claimed that on the exploration of large data sets the accuracy level may increase.

Sandeep Chaplot *et. al.*, [24] have proposed and implemented a novel approach for classification of MR brain images using wavelet as an input to self-organizing maps and support vector machine .They have noticed classification percentage of more than 94% in case of self organizing maps and 98% in case of support vector machine. They have applied the method only to axial T2-weighted images at a particular depth inside the brain. The same method can be employed for T1- weighted, T2-weighted, proton density and other types of MR images. Also they claim that with the help of above approaches, one can develop software for a diagnostic system for the detection of brain disorders like Alzheimer's, Huntington's, Parkinson's diseases *etc*.

Jin-Hyuk Hong *et. al.*, [25] proposed a novel fingerprint classification method which effectively integrates NBs and OVA SVMs, which produces better accuracy than previously reported in the literature contained in the NIST-4 database. In their proposed method, several popular fingerprint features such as singularities, pseudo codes and the Finger Code were used, and the combination of methods described in the experimental analysis produced better results (90.8% for the five-class classification problem and 94.9% for the four-class classification problem with 1.8% rejection during the feature extraction phase of the Finger Code) than any of the component classifiers.

4. Analytical Discussions, Limitations & Suggestions

This literature review surveys the applications of SVM in diversified fields in connection with the author's background, the application interest and expertise knowledge in the particular field. Some authors have been repeated for different applications. The paper discusses the SVM method applied in a mixture of application areas including medical, engineering, pattern classifications, nuclear component classification, classification problems, prediction, science and other applications, which were extracted from the databases like Elsevier, IEEE X-plore, Springer Link, Taylor Francis and Inderscience.

As the popularity of SVM is increasing day by day, so different research applications relying on SVM must be published, to facilitate the wide broaden scope of SVM, in the academic and practical fields. However, many researchers have pointed some limitations of SVM on which work must be carried out like: (1) The selection of kernel for a problem (2) The functional speed of the machine in training and testing, (3) Slower Convergence rate at testing phase,(4) Choosing good quality kernel parameters ,(5) Large requirements of memory space to implement the model (6) Choosing either parametric or non-parametric method for implementation. This integration of methodologies and cross-disciplinary research may put forward new insights for problem solving with SVM. This paper reviews a no. of major applications using SVM, but still inclusion of some other areas like social, statistical, and behavioral science *etc.*, are needed. Also, the qualitative and quantitative aspects of SVM technique are to be included in our future work.

As shown in Table 1, support vector machine have been applied in almost every application domain including classification, prediction/forecasting, image analysis, pattern recognition, rule extraction and optimization problems. The majority of the applications that we have reviewed, about 54% are in the area of classification. Works in the areas of clustering and forecasting account for 9% and 13%, respectively and others represent 24% each of the reviewed work (Figure 3). The majority of the work review used the MATLAB software for implementing their models.

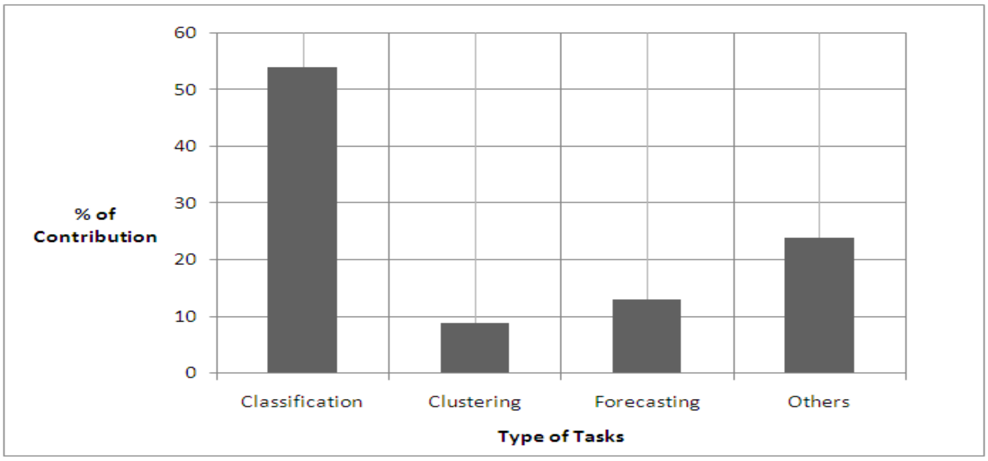

Figure 3. Application of SVM in Data Miningasks

Table 1. Summary of Various SVM Models used in Various Applications

Referen	Model Name	Type of Basis Function used	Application Area
[31]	SVM	Gaussian	Drug Design
[32]	SVM	Gaussian	Gene Classification
[33]	SVM	Polynomial	Classification
[34]	SVM	Gradient Descent	Classification
[35]	GSVM	Polynomial	Classification
[36]	SVM	Gaussian	Classification
[37]	WVSVM	Polynomial	Classification
[38]	WC-SVM	Polynomial	Classification
[39]	SVM	Polynomial	Image Classification
[40]	SVFNN	Polynomial	Pattern Classification
[41]	LSSVM	Polynomial	Medical Diagnosis

5. Conclusion

Support Vector Machine is a rapidly increasing field with promise for greater applicability in all domain of research. In this paper the review of different applications of support vector machine is being given focusing on data mining tasks from the year 2001 to 2014. Although, this paper is by no means a meticulous review of the literature in the application of support vector machine to application of data mining, we hope that we have given a passable overview of what is currently happening in this evolving and dynamic area of research. Also the fundamentals of the support vector machines (SVMs) have been discussed along with the different formulations of the optimization problem resulting from the training of such machines. A review of the literature concerning the amalgamation of prior knowledge into SVMs has been exposed. In the last section, a detailed survey statistics report on various papers related to SVM applications published in the standard journals of IEEE, Elsevier, Inderscience, Springer, Taylor Francis are presented. After analyzing the literature survey, the following are some of the issues recognized that can be taken further to do the research. Even though various techniques have been used in the literature survey, still there is a need of best techniques to solve the following research issues.

References

[1] H. Gish, "A probabilistic approach to the understanding and training of neural network classifiers", IEEE international conference acoustic, speech signal process, vol. 3, **(2015)**,

pp. 1361–1364.
[2] G. P. Zhang, "Neural networks for classification: a survey", IEEE Trans Syst Man Cybern C, vol. 30, vol. 4, **(2014)**, pp. 451-462.
[3] D. Michie, D. J. Spiegel Halter and C. C. Taylor, "Machine learning, neural and statistical classification", Ellis Horwood, NewYork, **(2016)**.
[4] M. Adya and F. Collopy, "How effective are neural networks at forecasting and prediction? A review and evaluation", Forecast, vol. 17, **(2014)**, pp. 481-495.
[5] J. L. Callen, C. C. Y. Kwan, P. C. Y. Yip and Y. Yuan, "Neural network forecasting of quarterly accounting earnings", Forecast, vol. 12, **(2015)**, pp. 475-482.
[6] K. B. Church and S. P. Curram, "Forecasting consumers expenditure: a comparison between econometric and neural network models", Forecast, vol. 12, **(2019)**, pp. 255–267.
[7] J. T. Connor, R. D. Martin and L. E. Atlas, "Recurrent neural networks and robust time series prediction", IEEE Trans Neural Netw., vol. 51, no. 2, **(2015)**, pp. 240–254.
[8] M. Cottrell, B. Girard, Y. Girard, M. Mangeas and C. Muller, "Neural modeling for time series: a statistical stepwise method for weight elimination", IEEE Trans Neural Netw., vol. 6, no. 6, **(2014)**, pp. 1355–1364.
[9] J. J. Faraway and C. Chatfield, "Time series forecasting with neural networks: a comparative study using the airline data", Appl Stat., vol. 47, **(2012)**, pp. 231–250.
[10] D. Fletcher and E. Goss, "Forecasting with neural networks— an application using bankruptcy data", InfManag, vol. 24, **(2016)**, pp. 159–167.
[11] W. L. Gorr, "Research prospective on neural network forecasting", Forecast, vol. 10, **(2020)**, pp. 1–4. [12] H. S. Hippert, C. E. Pedreira and R. C. Souza, "Neural networks for short-term forecasting: a review and evaluation", IEEE Trans Power Syst., vol. 16, no. 1, **(2011)**, pp. 44-55.
[13] M. R. Belli, M. Conti, P. Crippa and C. Turchetti, "Artificial neural networks as approximators of stochastic processes", Neural Netw., vol. 12, no. 4-5, **(2019)**, pp. 647–658.
[14] J. L. Castro, C. J. Mantas and J. M. Benitez, "Neural networks with a continuous squashing function in the output are universal approximators", Neural Netw., vol. 13, no. 6, **(2010)**, pp. 561–563.
[15] K. Funahashi, "On the approximate realization of continuous mappings by neural networks", Neural Netw., vol. 2, **(2019)**, pp. 183-192.
[16] R. Andrews, J. Diederich and A. B. Tickle, "Survey and critique of techniques for extracting rules from trained artificial neural networks", Knowl Based Syst., vol. 8, no. 6, **(2019)**, pp. 373–389.
[17] J. L. Castro, I. Requena and J. M. Benitez, "Interpretation of artificial neural networks by means of fuzzy rules", IEEE Trans Neural Netw., vol. 13, no. 1, **(2016)**, pp. 101-116.
[18] R. Setiono, W. K. Leow and J. Zurada, "Extraction of rules from artificial neural networks for nonlinear regression", IEEE Trans Neural Network, vol. 13, no. 3, **(2018)**, pp. 564–577.
[19] R. Setiono and J. Y. L. Thong, "An approach to generate rules from neural networks for regression problems", Eur J Oper Res., vol. 155, **(2014)**, pp. 239-250.
[20] P. J. G. Lisboa, "A review of evidence of health benefit from artificial neural networks in medical intervention", Neural Netw., vol. 15, **(2012)**, pp. 11-39.
[21] L. G. Protney and M. P. Watkins, "Foundations of clinical research: applications to practice", Prentice- Hall, Princeton, **(2010)**.
[22] S. M. Hosseini-Nezhad, T. S. Yamashita, R. A. Bielefeld, S. E. Krug and Y. H. Pao, "A neural network approach for the determination of inter hospital transport mode", Comput Biomed Res., vol. 28, no. 4, **(1995)**, pp. 319-334.
[23] C. J. Burges, "A tutorial on support vector machines for pattern recognition", Data Mining and Knowledge Discovery, vol. 2, **(2012)**, pp. 1-43.
[24] V. Vapnik, "Statistical Learning Theory", John Wiley and Sons, **(2013)**.
[25] E. Sami, "Support Vector Machines for classification and locating faults on transmission lines", Applied Soft Computing, vol. 12, **(2012)**, pp. 1650–1658.
[26] T. Joachims, "Learning to classify text using support vector machines", Kluwer Academic Publishers, **(2015)**.
[27] S. N. Sivanandam, S. Sumathi and S. N. Deepa, "Introduction to Neural Networks using MATLAB 6.0", Tata McGraw Hill Education Pvt. Ltd., **(2016)**.