Diagnosis of Mental Handicap in School Children using Machine Learning Techniques Support Vector Machines and Artificial Neural Network

J.Macklin Abraham Navamani 1, A. Kannammal 2

Abstract--- Mental Handicap is a neurological disorder that affects the brain's ability to receive process, store and respond to information. A mental handicap can cause a child to have trouble in learning certain skills. In India around 13% to 14% of all school children suffer from mental handicap. There is no cure for mental handicap. Many children in India from the age of 6 years to 16 years and 11 months now have some form of mental handicap. Mental handicap can be diagnosed using psychometric intelligence scale test that is WISC IV (Wechsler Intelligent Scale Test for children) test. The information gained from evaluation helps the parents and school teachers provide a good learning environment for these students. Machine learning techniques are used for diagnosis. Machine learning is a subfield of artificial intelligence (AI) concerned with the development of algorithms and techniques that allow computers to learn. This paper proposes two classification methods based on SVM with two kernels and ANN for the diagnosis of mental handicap and compares the accuracy and sensitivity.

Index Terms— Data Mining - Machine Learning Techniques - Mental Handicap - Artificial Neural Network - MLP - Support Vector Machines- kernel

I INTRODUCTION

Popularity of data mining today is because of which data is captured with every transaction through on-line transaction processing (OLTP) systems, capacity to store large (measured in terabytes) amounts of data and finally, the continuously falling cost of computing power. Data Mining is the process of discovering new meaningful correlations, non-trivial, implicit, previously unknown and potentially useful information from large volumes of data. Society produces huge amounts of raw data from various sectors including business, science, medicine, web mining sports. The raw data which has been stored is useless, it needs techniques or programs to automatically detect patterns, regularities and extract information from it. Data mining uses statistical or mathematical techniques (referred to as pattern recognition technologies) to bring out the relationships hidden in the data. Data mining uses two strategies: supervised and unsupervised learning.

In supervised learning, a training set is used to learn model parameters whereas in unsupervised learning no training set is used (e.g., k-means clustering is unsupervised) [15]. A wide variety of fields associated with health care services such as diagnosis of effectiveness of medical tests, predication and the assessment of relationship among clinical data also make use of various data mining methodologies [3]. This paper proposes two classification methods based on SVM and ANN for the diagnosis of mental handicap and compares the accuracy and sensitivity. SVM is regarded as the first choice for classification problems and was first introduced in 1992. SVM is now regarded as an important example of "kernel methods", one of the key areas in machine learning. The SVM approach of data classification and analysis to a great extent has important advantage and hence it has been effectively applied in various existent real world problems. Sequential minimal optimization (SMO) is one of the most popular algorithms for large-margin classification by SVM. Artificial neural networks (ANN) are stimulated by early models of sensory processing by the brain. An artificial neural network can be produced by simulating a network of representation neurons in a computer. Neural networks use learning algorithms that are inspired by our understanding of how the brain learns, but they are evaluated by how well they work for practical applications such as speech

J.Macklin Abrahm navamani is with the Karunya University, Coimbatore, Tamilnadu, India. working as Asst. Professor in the Department of Computer Applications (e-mail: macklin@karunya.edu)
A.Kannammal is with Coimbatore Institute of technology, Coimbatore, Tamilnadu, India. She is working as professor in the department of Computer technology and applications. (e-mail: kannaphd@yahoo.co.in)

recognition, hand-written character recognition laboratory medicine and pathology. Neural network algorithms are used in the parallelization technique to speed up the computation process. All these things help in the usefulness of neural network for classification and prediction in data mining [5]. The purpose of this research paper is to use two well known machine learning techniques Support vector machine (SVM) and Artificial Neural Network (ANN) in diagnosis of mental handicap in school children and evaluation classifier accuracy and sensitivity. The rest of the paper is structured as follows. Section 2 briefly describes about mental handicap. Section 3 briefly describes about machine learning techniques SVM and ANN. Section 4 describes about Experimental settings and results comparison between the performance and results obtained for the classifiers SVM and ANN and merits and demerits of two classifiers using mental handicap data set are discussed. Section 5 describes about conclusion and future work.

2. MENTAL HANDICAP

Mental Handicap is a neurological chaos that affects the brain's capacity to receive process, store and respond to information. It is an invisible handicap that was difficult to identify [15]. Since it is an invisible handicap children show virtually no significant, external symptoms or features resulting in problem of child's learning. A mental handicap can be a reason in school children to have problem in learning and using positive skills. The U.S office of Education in 1977 defined mental handicap a permanent-information processing deficit (disorder) that affects the manner in which individuals with average to above average intelligence learn[4]. Mental Handicap can affect a person's capability in the areas of , speaking, listening, writing, reading and solving mathematical problems and is suspected when there is a apparent and mysterious gap between an individual's level of performance, expected and actual levels of achievement[2] [7]. The main cause for mental handicap in a child can be classified as a) heredity b) problems during pregnancy and birth c) incidents after birth. One child with mental handicap may possibly not have the identical kind of learning problem as another child with mental handicap. Metal handicap is not an ailment so there is no cure for mental handicap it is a lifelong process there are ways to triumph over the challenges it poses. If identified early they can become high achievers and come out of this problem. With right help they can become successful in learning and society.

In Taiwan the identification process roughly separated in to four steps: 1) screening by parents general education teachers. 2) Identification by junior level evaluation personal. 3) Diagnosis by senior level evaluation personal 4) final confirmation by special education specialist. The first issue with the above procedure is the more number of manpower and resources are required. Second issue is the lack of standard testing method.

According to National Information Center for Children and Youth with Disabilities 1 out of every 5 school children in the U.S. has a mental handicap [4]. In India, around 13 to 14 per cent of all school children go through learning disorders or mental handicap [17]. Many children in India (ages 6 – 16) now have some form of Mental Handicap. We have to carefully look for the difference in child behavior and their learning abilities. The best place to identify mental handicap and various resources to overcome it are to be found in the school. Teachers also are challenged to identify mental handicap in children due to the lack time and resources in schools. The most common treatment for mental handicap or (LD) is appropriate education in school and hard work and support in family. Psychiatrists or special education teachers may perform analytic edifying assessment on child's academic and logical potential and level of academic performance. Many type of assessments test exist to diagnose. The test that child needs depend on Child's age and the type of problem. In India as the school class strength is above 60 the time and staff needed for the assessment of Mental Handicap in children is very high and might not be accurate .The identification of Mental Handicap has become

critical considering the importance of education and academic advancement. Current research avenues focus on developing Machine Learning and AI (Artificial Intelligence) techniques to diagnose and predict Mental Handicap in school children and help them to get special education.
For assessing Mental Handicap we are using Wechsler Intelligence Scale (WISC-IV). The Wechsler Intelligence Scale for Children – Fourth Edition (WISC-IV) is the most widely used intelligence test for assessing children between the ages of 6 to 16 years and 11 months. WISC-IV contains 16 symptoms of Mental Handicap which are the attributes of this study. The data was collected through a questioner from various schools in and around Coimbatore.

Sl. No	Attribute	Symptoms of LD
1.	DVC	Difficulty with Vocabulary
2.	DSI	Difficulty with Similarities
3.	DIN	Difficulty with Information
4.	DCO	Difficulty with Comprehension
5.	DAR	Difficulty with Arithmetic
6.	DDS	Difficulty with Digit Span
7.	DLN	Difficulty with Letter Number Sequencing
8.	DPC	Difficulty with Picture Completion
9.	DPA	Difficulty with Picture Arrangement
10.	DBD	Difficulty with Block Design
11.	DMR	Difficulty with Matrix Reasoning
12.	DFW	Difficulty with Figure Weights
13.	DOA	Difficulty with Object Assembly
14.	DSC	Difficulty with Digit Symbol Coding
15.	DSS	Difficulty with Symbol Search
16.	DC	Difficulty with Cancellation

Table 1: List of Attributes

3. Support Vector Machine and Artificial Neural Network

Vladimir Vapnik invented Support Vector Machine in 1979 [12]. SVM has received more and more attention from researchers in recent years [10, 11], and its great feature is the capability of dealing with linear and nonlinear questions and is free from the limitations of data size. SVM can be applied to problems of classification and regression. A distinctive two category problem as the one shown in figure 1 is similar to the problem of diagnosing children as children with MH or not. For a classification problem, it is necessary to first try to estimate a function f : R N→ {± 1} using training data, which are l N-dimensional patterns Xi and class labels Yi , where (x1 , y1), …. , (xl ,yl)∈ R N×{±1} Where f will classify new samples (X, y) correctly. Support Vector Machines select a small number of critical boundary instances called support vectors for each class.

SVM training with few parameters can be conducted without having sufficient data. In classification, the primary concept of SVM is to construct an optimal hyper-plane that serves as an interface for classification decisions. Through this interface we can effectively separate positive examples and negative examples.

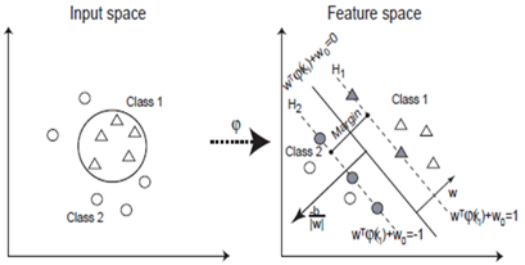

Fig. 1: Basic concept of SVM

Among all the classification techniques, artificial neural network (ANN) has received lots of attentions due to their demonstrated performance and has gained widely acceptance beginning from the 1990s [6]. Artificial neural network is information processing prototype inspired by biological nervous systems, such as our brain where large number of highly

interconnected processing elements (neurons) work together. Artificial Neural networks are used in specific application, such as pattern recognition or data classification, through a learning process [1] to discover knowledge. Neural networks are also called machine learning algorithms, because changing of its connection weights causes the network to learn and find solution for various problems. The strength of relationship between various neurons is calculated and stored as a weight-value for the specific connection. This type of classification system learns new information by adjusting these connection weights. The learning capability of a neural network is found out by its architecture and the algorithmic method selected for training. Neural Network can be broadly divided in to feed-forward networks, feed-back networks and self-organization network. In data mining mostly for classification and predicting multilayer perceptron (MLP) neural network architecture with back propagation is used [8]. Backpropagation MLP is supervised ANN. This means desired output for training examples. The backpropagation algorithm adapts the weights and the thresholds of the neurons in a way that minimizes the error function E.

$$E = \frac{1}{2}\sum_{p=1}^{n}(d_p - y_p)^2$$

Where yp is the actual output and dp the desired output for input pattern p.

ANN is composed of a collection of perceptrons grouped in layers.

A typical structure is shown in Fig.2

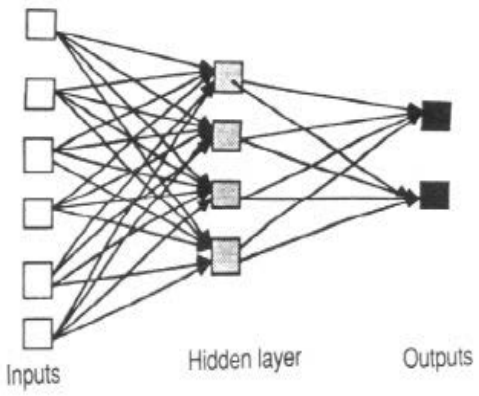

Fig.2: The architecture of the MLP network (input layer, hidden layer and output layer).

The input layer represents the input data (the input data is described in section 4.1). The usage of a hidden layer enables the representation of data sets that are not linearly separable. The output layer represents the classification result.

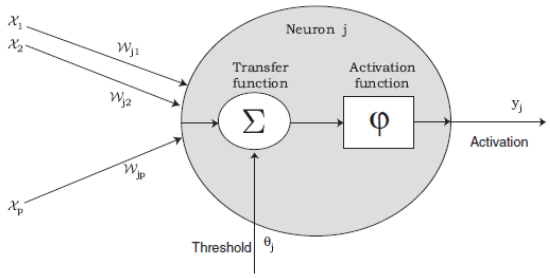

Fig. 3: A neuron in the hidden layer in MLP.

A single neuron in the MLP is able to linearly separate its input space into two subspaces by a hyper plane defined by the weights and the threshold.

4. Experimental Settings and Results

Proposed Methodology Deign

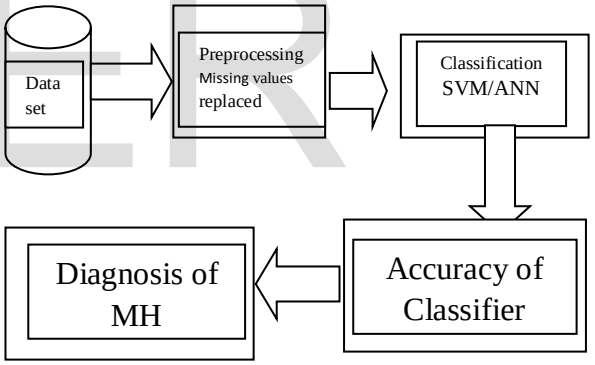

Fig. 4: System Flow Chart

4.1 Data Preparation

For assessing Mental Handicap we are using Wechsler Intelligence Scale (WISC-IV). The Wechsler Intelligence Scale for Children – Fourth Edition (WISC-IV) is the most widely used intelligence test for assessing children between the ages of 6 to 16 years and 11 months. WISC-IV contains 16 symptoms of Mental Handicap which are the attributes of this study. The data was collected through a questioner from various schools in and around Coimbatore. After conducting interview with the children with the help of special education teachers the questioner

is filled, which is ultimately used for preparing the data for conducting our study. The dataset used in this study consists of nearly 1085 real world data.

4.2 DATA PREPROCESSING PHASE

Before doing classification we have to do preprocessing in order to increase the accuracy of the output and to facilitate learning process. Data preprocessing is a broad area and various stage in preprocessing is dimensionality reduction, feature subset selection, removal of noise and imputing missing values [9]. In data set used in this study contains 1085 tuples of 17 attributes each. Since our data set is collected from school children it contained many missing values. The missing data of each attributes and their percentage is given below in table 2. These missing data are handled using regression algorithm to obtain complete data. The regression method estimates missing values using multiple linear regressions. The means, the covariance matrix, and the correlation matrix of the predicted variables are displayed. Error terms are chosen randomly from the observed residuals of complete cases to be added to the regression estimate SPSS statistical tool is used for imputing missing values. After this process is over new data set without any missing values is available for classification using SVM and ANN.

	N	Mean	Std. Deviation	Missing Count	Missing Percent
DV	1080	.51	.500	5	.5
DS	1082	.51	.500	3	.3
DI	1083	.50	.500	2	.2
DCO	1083	.51	.500	2	.2
DA	1081	.54	.499	4	.4
DDS	1070	.52	.500	15	1.4
DLNS	1060	.50	.500	25	2.3
DPC	1058	.52	.500	27	2.5
DPA	1060	.55	.498	25	2.3
DBD	1060	.53	.499	25	2.3
DMR	1062	.53	.499	23	2.1
DFW	1066	.53	.499	19	1.8
DOA	1079	.53	.499	6	.6
DSC	1074	.53	.499	11	1.0
DSS	1082	.53	.499	3	.3
DC	1084	.52	.500	1	.1

Table 2: Shows the missing count and percentile of each attributes.

4.3 CLASSIFICATION OF MENTAL HANDICAP USING SVM AND ANN

As a result we recommend a hybrid classification system combining ANN and SVM classifiers (SVM is tested with two RBF Kernel and Normalized Polykernel). The objective is not only to institute comparison among all of them but also to gain from the good classification accuracies of each classifier.

4.3 A. SUPPORT VECTOR MACHINE (SVM WITH RBF KERNELS)

The classification is performed through SVM. It is implemented in weka, a machine learning workbench. In these sections we test the mental handicap data set by using SVM with RBF kernel to test their accuracy. Sequential Minimal Optimization (SMO) algorithm is a faster training method for SVMs. SVMs have empirically shown good performance on a wide variety of problems. Training a support vector machine requires the clarification of a very large quadratic programming (QP) optimization problem. SMO breaks this large quadratic program problem into a series of smallest possible problems. The amount of memory

required for SMO is linear in the training set size, which allows SMO to handle very large training sets[13].

Definition of the confusion matrix with the values for every measure of the SVM classifier and RBF kernel.

Table 3: SVM with RBF Kernel

Actual	Predicted	
	Positive	Negative
Positive	True Positive (TP)= 780	False Negative (FN)= 2
Negative	False Positive (FP)= 152	True Negative (TN)= 151

Classifier:

Accuracy (%)

= ((TP + TN) /(TP + FP + FN + TN)) * 100 = 85.80 %

Sensitivity (%)

= TP/(TP + FN) * 100
= 99.74%

Specificity (%)

= TN/(FP + TN) * 100
= 49.83%

Positive Predictive Value(%)

= TP / (TP + FP) * 100
= 83.69%

Negative Predictive Value(%)

= TN / (FN + TN) * 100
= 98.69 %

4.3 B. Support Vector Machine (SVM with Normalized PolyKernel)

The classification is performed through SVM. It is implemented in weka, a machine learning workbench. In these sections we test the mental handicap data set by using SVM with Normalized PolyKernel to test their accuracy. This new method improves the accuracy of our system. From the outcome of these examinations, the Normalized Polykernel is robust and has an equal or even stronger mapping power as compared to the standard kernel functions leading to an equal or better generalization performance of SVMs.

Definition of the confusion matrix with the values for every measure of the SVM classifier and Normalized Polykernel.

Table 4: SVM with normalized Polykernel

Actual	Predicted	
	Positive	Negative
Positive	True Positive (TP)= 779	False Negative (FN)= 3
Negative	False Positive (FP)= 83	True Negative (TN)= 220

Classifier:

Accuracy (%)

= ((TP + TN) /(TP + FP + FN + TN)) * 100 = 92.73%

Sensitivity (%)

= TP/(TP + FN) * 100
= 99.61%

Specificity (%)

= TN/(FP + TN) * 100
= 72.60%

Positive Predictive Value(%)

= TP / (TP + FP) * 100
= 90.37%

Negative Predictive Value(%)

= TN / (FN + TN) * 100
= 98.65%

4.3 C. ARTIFICIAL NEURAL NETWORK (MLP)

The classification is performed through ANN. It is implemented in weka, a machine learning workbench. In these sections we test the mental handicap data set by using ANN with Multi Layer back Propagation algorithm to test their accuracy. For the construction of the architecture of the MLP we proceed as follows:

a) Layer 1(input layer) corresponds directly to the input i.e attributes of our study.

b) Layer 2 (the hidden layer). The number of hidden neurons for this layer is the most elaborated question in the network's architecture. This number represents a trade of between performance and the risk of over fitting.

c) Layer 3 (output layer) corresponds to output i.e Mental Handicap true or false.
The architecture of the neural network used in the study is the multilayer back propagation algorithm with 16 input nodes, 8 hidden layers, and 2 output nodes.
The architecture of MLP obtained in this study is shown in Fig.

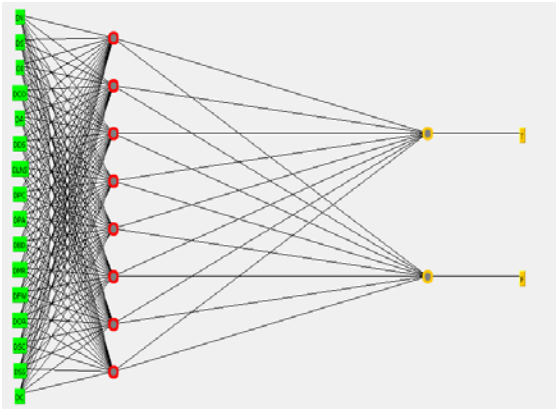

Fig. 5: Architecture of MLP with 16 Input Nodes 8 Hidden Layers and 2 Output Nodes

Definition of the confusion matrix with the values for every measure of the ANN classifier and MLP

Table 5: ANN with MLP

Actul	Predicted	
	Positive	Negative
Positive	True Positive (TP)= 775	FalseNegative (FN)= 7
Negative	False Positive (FP)= 14	True Negative (TN)= 289

Classifier:

Accuracy (%)
= ((TP + TN) /(TP + FP + FN + TN)) * 100 = 98.06%

Sensitivity (%)
= TP/(TP + FN) * 100
=99.10%

Specificity (%)
= TN/(FP + TN) * 100
= 95.37%

Positive Predictive Value(%)
= TP / (TP + FP) * 100
= 98.22%

Negative Predictive Value(%)
= TN / (FN + TN) * 100
=97.63%

One of the reasons why such a high degree of accuracy is obtained, is due to the data

preprocessing. Data preprocessing is necessary before applying them to artificial intelligence method.

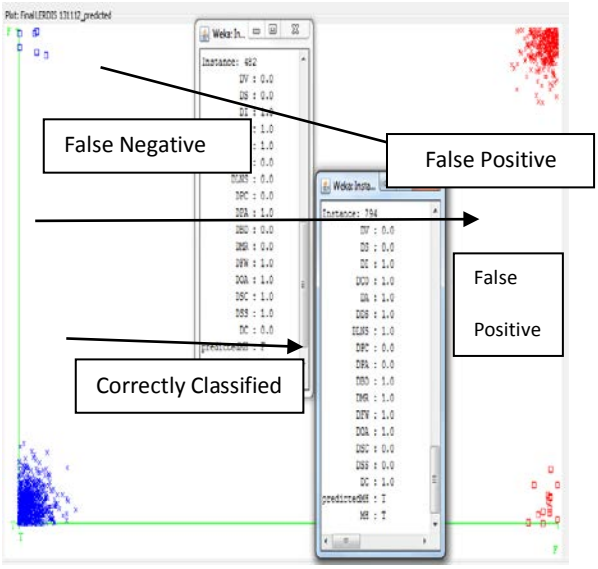

Fig 6: Shows the classification errors. The two errors showed in the figure are the two instances of the Mental Handicap. One of them is correctly classified (predicted MH and the real MH coincide). However, the other one does not. This is a false negative.

Classifiers	Accuracy (%)	Sensitivity (%)	Specificity (%)	Positive Predictive Value (%)	Negative Predictive Value(%)
SVM with RBF Kernel	85.80 %	99.74 %	49.83 %	83.69%	98.69 %
SVM with Normalized PolyKernel	92.73 %	99.61 %	72.60 %	90.37%	98.65 %
ANN with MLP	98.06 %	99.10 %	95.37 %	98.22%	97.63 %

Table 6 a: Comparison of Classification results obtained

Classifiers	Parameters				
	Total Instances	Correctly Classified Instances	Incorrectly Classified Instances	Time taken to build a model (in seconds)	ROC Area
SVM with RBF Kernel	1085	931	154	0.91	0.748
SVM with Normalized PolyKernel	1085	999	86	1.52	0.861
ANN with MLP	1085	1064	21	8.08	0.978

Table 6 b: Comparison of Classification results obtained

5. Conclusion and Future Works

In this paper we have evaluated the performance of a classifier constructed by means of ANN and SVM with two kernels. The diagnosis must be toughened and complemented in order to provide a better result, in order to obtain a final common diagnosis. The results obtained shows that MLP classifier has more classification accuracy (98.06%), specificity (95.37%) and Positive Predictive Value (98.22%) than SVM classifier with two kernels. But sensitivity (99.74%) and Negative Predictive Value (98.69%) was higher for SVM with RBF Kernel. MLP also had one disadvantage that time taken to build the model was higher than other classifiers. The study was carried on more than 1000 real world data set with most of the attributes take binary values, more work has to be carried out using quantitative data. Our system was able to correctly classify instances. But there were some incorrectly classified instances also. We have to

study why errors occurred and learn to avoid them. Since MLP classification accuracy is greater we can train the neural network classifier and it will help in diagnosing the school children with mental handicap which would save lots of time manpower and impact the society. Students with mental handicap can be easily be identified and make them overcome daily challenges by providing appropriate education making them work hard, support from family and teachers can make them lead a productive life. In our next work we will be using clustering techniques or meta learning algorithms to diagnose mental handicap which could help in finding the outliers and help in improving the classifier accuracy and diagnosis more correctly.

References

[1] Patil SB, Kumaraswamy YS (2009) Intelligent and effective heart attack prediction system using data mining and ANN. Eur J SciRes 31(4): 642–656. ISSN:1450-216X

[2]Julie MD, Kannan B (2010) Significance of classification techniques in prediction of learning disabilities in school age children. Int J Artif Intell Appl 1(4):111–120. doi:10.5121/ijaia.2010.1409

[3] Tzung-I Tang, Gang Zheng, Yalou Huang, Guangfu Shu, Pengtao Wang, "A Comparative Study of Medical Data Classification Methods Based on Decision Tree and System Reconstruction Analysis", IEMS, Vol. 4, No. 1, pp. 102-108, June 2005.

[4] Rochelle Kenyon, Ed.D. Project Director – Florida's Bridges to Practice:"Facts and statistics on learning disabilities and literacy" September 2003

[5] Han J, Kamber M (2008) Data mining: concepts and techniques, 2nd edn. Morgan Kaufmann—Elsevier Publishers. ISBN:978-1-55860-901-3

[6]. Razi, M. A. and K. Athappilly, A comparative analysis of neural network (NNs), nonlinear regression and classification and regression tree (CART) models, Expert Systems with Application, 29, pp. 65-74 (2005).

[7] Julie M. David, Kannan Balakrishnan: "Paper on Prediction of Frequent Signs of Learning Disabilities in School Age Children using Association Rules". In: Proceedings of the International Conference on Advanced Computing, ICAC 2009, MacMillion Publishers India Ltd., NYC, ISBN 10:0230-63915-1, ISBN 13:978-0230-63915-7,2009, pp. 202–207.

[8]NeuroDimension Incorporated, What is a Neural Network?, http://www.nd.com/neurosolutions/products/ns/whatisNN.html.

[9] Tan PN, Steinbach M, Kumar V (2008) Introduction to data mining, Low Price Edition. Pearson Education, USA. ISBN:978- 81-317-1472-0, 2008

[10] Vapnik, V., 1995, The Nature of Statistical Learning Theory, New York: Springer Verlag.

[11] Abe, S., 2002, "Analysis of Support Vector Machines", Proceeding of the 12th IEEE Workshop on Neural Networks for Signal Processing, Sept, pp 89-98.

[12] Vapnik V, Estimation of Dependences Based on Empirical Data, Springer- Verlag, 1982.

[13]J. Platt. Sequential minimal optimization: A fast algorithm for training support vector machines. Advances in Kernel Methods-Support Vector Learning, 208, 1999.

[14] Obenshain, M.K: "Application of Data Mining Techniques to Healthcare Data", Infection Control and Hospital Epidemiology, 25(8), 690–695, 2004.

[15] Hung, L., 1999. Education for individuals with learning disabilities. Taipei, Psychological Publishing Co., Ltd.

[16]http://www.articlesbase.com/education-articles/learning-disability-in-india-1126528.html