Diagnosis of Learning Disability using WISC-R: A Comparative Study between ANN and SVM Algorithm

Varun Ramachandran, Tisha, J Anuradha, B K Tripathy

ABSTRACT

Learning disability (also called **learning disorder** or **learning difficulty**), is a disorder in which a person has difficulty in learning a typical manner, usually caused by an unknown factors. The unknown factors are the disorder that affects the brain's ability to receive and process information. This disorder can make it challenging for a person to learn as quickly as the same way the others learn. Learning disability is not indicative of intelligence level, it describes specific kinds of learning problems. It can cause difficulty for a person in learning or using certain skills like reading, writing, listening, speaking, reasoning, and doing maths. A learning disability cannot be cured or fixed. In this paper we try to diagnose the problem of Learning Disability by carrying out a test on school children. The test used by us is the WISC-R or the **Wechsler Intelligence Scale for Children** (**WISC**), developed by David Wechsler. In this paper, we compare the diagnosis results produced by two artificial intelligence techniques namely, neural networks using the back propagation algorithm and SVM (Support Vector Machine) Algorithm.

Index Terms

Back Propagation Algorithm, Learning Disability, Support Vector Machine, WISC.

1. INTRODUCTION

Learning disability is a disorder which is characterised by trouble in learning or by problem in using certain skills or both. Unawareness among people makes the job of diagnosing LD among children more difficult. In this paper we try to develop an expert system that could help the doctor in screening the LD children. Thus by saving time, money and reach ability for the patients.

1.1 LEARNING DISABILITY

A LD child do not have self motivation, cannot try harder and they are inattentive; they need help in participating in learning activities. These challenged children do not have trouble in their intelligence. They may be intellectual also.

Varun Ramachandran B.Tech IT, School of Information Technology, VIT University, Vellore email Varun.ramachandran89@gmail.com
Tisha B.Tech IT, School of Information Technology, VIT University, Vellore email tisha.dhiman88@gmail.com
J.Anuradha Assistant professor(Senior), School of Computing Science, VIT University, Vellore email j.anuradha@vit.ac.in
DR BK TRIPATHI Senior professor, School of Computing Science, VIT University,Vellore

Learning disorders are caused by a variation in the brain that affects how information is received, processed, or communicated. Children and adults with learning disabilities have complexity in processing sensory information because they see, hear, and understand things differently.

The categories of learning disabilities are formulated based on school–area skill set or cognitive weakness. It will be apparent to find if he or she is struggling with reading, writing, or math among school going children, and also be easier in narrowing down the type.

1.1.1Motor difficulties and learning disabilities

Motor difficulty refers to clumsiness in hand movement and coordination whether it is with fine motor skills (cutting, writing) or gross motor skills (running, jumping). A motor clumsiness is also referred as an “output” activity means that it transmits (output) the information from the brain. In order to perform run, jump, write or cut something, the brain must be able to correspond its output to the essential limbs. Signs of the child’s coordination disability might include trouble with physical abilities that require hand–eye coordination, like holding a pencil or tying the shoe lace.

1.1.2 Math difficulties (Dyscalculia)

Learning disabilities in math also called as dyscalculia highly depends on his strengths and weaknesses in other area. For example, a child’s ability to do math will be affected by a language learning disability, or memorization or a difficulty with sequencing, memory or organization.

A child with a mathematics learning disorder may struggle with memorization and association of numbers, operators, signs, and basic number “facts” e.g 7+7=14 or 7x7=49. Children with these disorders might also face difficulty in counting principles (such as counting by 2s or counting by 5s) or have trouble in identifying time.

1.1.3Language difficulties (Dysphasia)

Language and communication learning disabilities entitle the ability to comprehend or produce spoken language. Language is also considered an output activity because it requires systematize thoughts in the brain and recollect upon the appropriate words to verbally explain and communicate with others.

Indication of a language–based learning disorder bring out problems like verbal language skills, like the ability to retell a story , the fluency of communication, as well as the ability to understand the meaning of words, etc.

ISBN: 978-988-17012-0-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

1.1.4Reading difficulties (Dyslexia)

Reading disorder can be categorised into two types viz. basic reading and comprehension reading problem. i) Basic reading problems occur when there is difficulty in perceiving the relationship between sounds, letters and words. ii) Reading comprehension problems occur when there is incapability in understanding the meaning of words, phrases, and paragraphs.
Some of the signs of reading disorders are

- letter and word recognition
- understanding words and ideas
- reading speed and fluency
- general vocabulary skills

1.1.5 Writing difficulties (Dygraphia)

The category of writing disabilities under LD involves the physical act of writing or the mental activity of understanding and processing of information. Basic writing disorder refers to difficulty in forming words and letters. Expressive writing disability refers to the trouble in systematize thoughts on paper.

Symptoms of a written language LD encompass the art of writing. They include problems with:

- neatness and consistency of writing
- accurately copying letters and words
- spelling consistency
- writing organization and coherence

1.1.6 Auditory and visual processing:

The eyes and the ears are the most important means of delivering information to the brain. This process is sometimes called as “input” activity. Though impaired eyes and ears may affect auditory and visual process, professionals will not consider those cases for LD diagnosis.

Professionals refer the ability to hear as “auditory processing skills” or “receptive language.” The good auditory skill has greater impact on the ability to read, write and spell. Inabilities to differentiate the sounds that they hear and perceive at the wrong speed makes it difficult understand the basic concepts of reading and writing.

Visual processing refers to the work of eye. Visual perception is difficulty in identifying shapes and reverse letters. Problems in visual perception include leaving out words, skipping lines, misperceiving depth or distance, or having trouble eye–hand coordination. Visual perception may affect gross and fine motor skills, reading comprehension, and math.

1.2 BACKPROPAGATION ALGORITHM

An artificial neural network has to be trained to perform the required task. The **Backpropogation**, or **propagation of error**, is a one of the well-known methods of teaching the network. [10] It was given by Arthur E. Bryson and Yu-Chi Ho in 1969.

It implements the delta rule and is used for supervised learning. The user needs to know how to calculate the required output for the input data. It is mostly used for feed- forward networks i.e the networks that do not require any feedback or that do not contain a connection that loops. The term is short for "backwards propagation of errors" [10]. It requires the activation function to be differentiable by the artificial nodes.

X = input training vector($x_1 \ldots x_n$)
t = target ouput vector($t_1 \ldots t_n$)

α = learning rate parameter
xii= Input unit i
v_{0j} = bias on jth hidden unit
w_{0k} = bias on kth output unit z_j = hidden unit j.

the net input to z_j is:

$$z_{inj} = v_{0j} + \sum x_i v_{ij} \quad (1)$$

and output is

$$z_j = f(z_{inj}) \quad (2)$$

y_k = output unit k
net output to y_k is:

$$y_{ink} = w_{0k} + \sum z_j w_{jk} \quad (3)$$

and output is $y_k = f(y_{ink})$ (4)

δ_k = error correction for w_{jk}
δ_j = error correction v_{ij}

The error back-propogation algotirhm consists of two main phases Feed Forward phase and Back-propagation of error phase [10].

Feed forward phase(Phase I)

Each input unit receives input signal xi and sends it to the hidden unit (i=1 to n). Each hidden unit z_j (j = 1 to p) sums its wighed input signals to calculate net input.

$$z_{inj} = v_{0j} + \sum x_i v_{ij} \quad (5)$$

For each output unit yk (k=1 to m) calculate net input

$$y_{ink} = w_{0k} + \sum z_j w_{jk} \quad (6)$$

Back-propogation of error Phase (II)

Each output unit y_k receives a target pattern corresponding to the input training pattern and computes the error correction term.

$$\delta_k = (t_k - y_k) f`(y_{ink}) \quad (7)$$

On the basis of error correction term

$$\Delta W_{jk} = \alpha \delta_k z_j \quad (8)$$

$$\Delta W_{0k} = \alpha \delta_k \quad (9)$$

Each hidden unit (z_j , j = 1 to p) sums it delta unit from its output unit. Each output unit (y_k , k = 1 to m) update the bias and wieghts:

$$W_{jk}(\text{new}) = w_{jk}(\text{old}) + \Delta w_{jk} \quad (10)$$

$$W_{0k}(\text{new}) = w_{0k}(\text{old}) + \Delta w_{0k} \quad (11)$$

Each hidden unit (z_j , j = 1 to p) updates bias and weights

$$V_{ij}(\text{new}) = v_{ij}(\text{old}) + \Delta v_{ij} \quad (12)$$

$$v_{0j}(\text{new}) = v_{0j}(\text{old}) + \Delta v_{0j} \quad (13)$$

1.3 SVM ALGORITHM

Support vector machines (SVMs) are a set of related supervised learning methods used for classification and regression [3]. One or more hyperplanes are constructed in a high dimensional space by the support vector machine algorithm in order to perform classification, regression or other tasks.
When the SVM Algorithm is used we consider to have a good separation by the hyper plane that has the largest distance to the

ISBN: 978-988-17012-0-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

nearest training data points of any class. In general, the margin is directly proportional to the generalization error of the classifier [6].

In the above situation we have three hyper-planes given and we need to classify these planes in such a way that their inter-plan space is maximum which helps us in distinguishing the normal values with the abnormal ones. Thus, after applying SVM algorithm we get the result as given below.

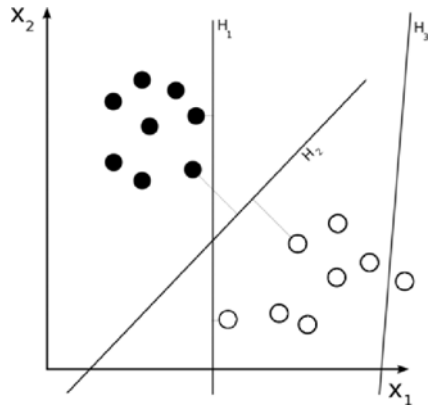

Fig 1 Support Vector Machine Representation

The training data used is in the form :

$$D = \{(x_i,c_i) \mid x_i \ \varepsilon R^p, c_i \ \varepsilon \{-1,1\}\}_{i=1 \text{ to } n} \quad ...(14)$$

The SVM is given by :

$$(w.x_1) + b = +1 \qquad ... (15)$$

$$(w.x_1) + b = -1$$
$$\Rightarrow (w.(x_1 - x_2)) = 2$$
$$\Rightarrow ((w/||w||).(x_1-x_2)) = 2/||w|| \qquad (16)$$

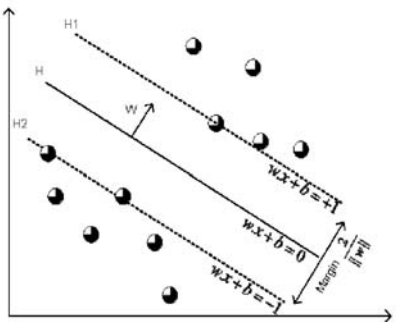

Fig 2 Output Of SVM

Parameter *C* determines the trade off between the model complexity (flatness) and the degree to which deviations larger than ε are tolerated in optimization formulation. For example, if *C* tends to infinity, then the main aim is to minimize the empirical risk only, without considering the model complexity in the optimization formulation. Parameter ε controls the width of the ε -insensitive zone, used to fit the training data [6].

2. RELATED WORKS

Soft computing (SC) is biological inspired evolution technique. Nowadays, it is pretty popular in the fields of computer science and information technology. Nonetheless, SC is a consortium of methodologies in which the synergistic performance of SC techniques has capability of handling the real life problems and can exceed the performance of other techniques. Its aim is to exploit the tolerance for imprecision, uncertainty, approximate reasoning and partial truth in order to achieve tractability, robustness and low-cost solutions. The three major areas of SC include fuzzy logic (FL), neural networks (NNs), and genetic algorithm (GA). Hybrid computing on these areas proved to produce good results. Hybrid computing in SC, combines these methodologies as FL and NN (FL–NN), NN and GA (NN–GA) and FL and GA (FL–GA). Recent years have witnessed the phenomenal growth of bio-informatics and medical.

Soft Computing is used extensively in medicine by "Ahmet Yardimci" where they prove that 71% of diagnosis in medicine is now done using Soft Computing [1].

Another paper about a practical application of Soft Computing, published in the Indian Journal of Pediatrics, cites the academic performance of 60 children (45 boys, 15 girls) at the SSC board examination with benefit of chosen provisions was compared with their performance at their last annual school examination before diagnosis of LD [8]**.**There was a significant improvement in their mean percentage (%) total marks scored at the SSC board examination as compared with the mean % total marks scored by them at their last annual school examination before the diagnosis. Children with LD who availed the benefit of provisions showed a significant improvement in their academic performance at the SSC board examination. This proves that if we can provide software to the doctors and psychologists which eases their effort and provides accurate and dependable results, it will in turn benefit the students suffering from this disorder.

3. IMPLEMENTATION & RESULTS

The entire paper is divided into different modules so that it would be easy to work on:-

3.1 THE QUESTIONAIRRE & WISC-R

We sought each and every subtest of the WISC-R test and developed a questionnaire based on them. These basically tested the child's skills and would also work as our input data. Training of the net was done on the basis of the subtest results as the answers given by the child were compared with the correct answers and hence a binary 1 or 0 was assigned to each correct and wrong answer respectively. The main advantage of the WISC-R is that a child answering all correct or all wrong does not guarantee the usual every subtest has its own weightage and the mean is calculated by taking into account the number of questions answered correctly, the total number of questions, the number of subtests, the weights of each subtest and finally we get the threshold value [5].The questionnaire had all the usual questions of the WISC-R test testing the child on vocabulary, arithmetic, block building etc. The questionnaire had 5 questions in each subtype hence making a total of 55 questions and all of the answers were taken into consideration to find out the threshold values [5].

ISBN: 978-988-17012-0-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

Table 1 snippet of the Training Data given to the neural network

SB1 -14		SB2 -5		SB3 -33		SB4 -15		SB5 -4		SB6 -47		SB7 +37		SB8 +27		SB9 +26		SB10 +45		SB11 -25		SS	SS*w	LDI	Z_{LDI}	Diagnosis
0	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	20	+45	+59	+0.196	0
0	0	1	1	1	0	1	1	1	1	0	0	1	1	1	0	1	1	1	1	1	1	16	+79	+90.2	+0.300	0
0	1	0	1	1	1	0	0	0	0	0	1	0	1	1	1	0	0	0	0	0	0	08	-46	-40.4	-0.134	1
0	1	1	0	0	0	1	1	1	1	0	1	1	0	0	0	1	1	1	1	1	1	14	+25	+34.8	+0.116	1
1	0	0	1	1	1	0	0	0	0	1	0	0	1	1	1	0	0	0	0	0	0	08	-41	-35.4	-0.118	1
1	0	1	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	04	-29	-26.2	-0.087	1
1	1	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	04	-122	-119.2	-0.396	0
1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	20	-10	+4	+0.013	1

The WISC-R test is used by the psychologists to diagnose LD. To do so we begin by calculating a Learning Disability Index (LDI) by using the answers to the 11 sub-tests. Each sub -test has a weight associated with it as given in the table 2 below [7]. The LDI is the factorially weighted sum of scaled deviations obtained from the subject's own means scaled score from his or her performance on all of the subtests of the WISC-R administered.
We find the LDI by using the given formula:

LDI = Sum (Scaled Score * W) + 0.7 * Sum(Scaled Score) (17)
Where : w is the weight of the subtest

Since we are dealing with the normal distribution of score according to sub-tests we need to calculate the value of Z. This is done by using the formula given below:

$$Z_{LDI} = LDI/301 \quad (18)$$

For convenience, in the table1 and table 3 we have illustrated our evaluation for only 2 questions per subtest.

3.2 IMPLEMENTING BPA ALGORITHM

We use the above described formula to create a dataset, with a diagnosis value assigned based on the formula. This data is used to train the network and hence to diagnose the test data.

The Table 1 shows the training data given to the BPA Algorithm. The SBn columns where n [1.11] represents each of the 11 subtests in the same order. Their weights considered for all calculation purposes is also mentioned in the table. The 0's and 1's refer to the answers for the WISC-R test, where a 0 represents a correct answer and a 1 represents a wrong answer.

After all the calculations we get the threshold value to be equal to ± 0.17. Thus, all Z_{LDI} values below this threshold represent kids possibly with LD.

The Table 3 on the other hand represents the test data provided to the neural network. To validate the results provided by the Back propagation Algorithm (BPA) we manually calculate the expected result and then compare it to the diagnosis given by the BPA and found that currently our algorithm can provide a diagnosis with an accuracy of **75%.**

Table 2 Weights taken for all subtests [7]

Sub-Tests	**Weights (w)**
Information	-14
Similarities	-05
Arithmetic	-33
Vocabulary	-15
Comprehension	-04
Digit Span	-47
Picture Completion	+37
Picture Arrangement	+27
Block Design	+26
Object Assembly	+45
Coding	-25

Learning Disability Index(LDI) = Sum(SS * W) + 0.7*Sum(SS)
Where, SS : Scaled Score

Z_{LDI} = LDI/301

3.3 IMPLEMENTING SVM ALGORITHM

The same inputs were taken for SVM algorithm and again first the network was trained with the training data and then using the test data results were obtained. For the SVM algorithm different types of functions can be used and hence different results are generated for each. This depends largely upon the kernel choice and SVM type we are choosing. In this particular case we use nu-svm as it basically minimizes the error function [11] and also because it is easier and more intuitive in use [3].
We checked with the different types of kernel to check which suites our cause the best and compared each.

3.3.1Radial Basis kernel

This is the most widely used kernel because it is localized and it gives finite responses across the entire range of real x-axis [11].

Exp (-gamma*|u-v|^2)... (19)

ISBN: 978-988-17012-0-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

Table 3 snippet of the Testing data given to the neural network

SB1 -14		SB2 -5		SB3 -33		SB4 -15		SB5 -4		SB6 -47		SB7 +37		SB8 +27		SB9 +26		SB10 +45		SB11 -25		SS	SS*w	LDI	ZLDI	By calc	Acc to BPA
0	0	0	1	1	1	1	1	1	0	0	0	1	1	1	1	1	1	1	1	1	1	16	+115	+126.2	+0.419	0	1
0	0	1	1	1	1	1	0	0	0	0	1	1	1	1	1	0	0	0	0	0	0	10	-10	-3	-0.001	1	1
0	0	0	0	1	1	1	1	0	1	0	1	1	1	1	1	1	1	0	0	0	1	11	+8	+15.7	+.052	1	1
0	1	0	1	1	1	0	1	1	0	1	0	1	1	1	0	1	1	1	1	1	1	17	+42	+53.9	+0.179	0	1
0	1	1	0	0	0	1	0	0	0	1	1	0	0	0	1	0	0	0	0	0	0	06	-101	-96.8	-0.321	0	0
1	0	1	1	0	1	0	0	0	1	0	1	1	0	1	0	0	0	0	0	0	0	08	-47	-41.4	-0.138	1	1
1	1	1	1	1	1	0	0	0	1	1	1	1	1	1	0	0	0	0	0	0	0	12	-101	-92.6	-0.308	0	0
0	0	0	0	1	1	1	1	0	1	0	1	1	1	1	1	1	1	0	0	0	0	10	+33	+40	+.132	1	1

This gave **75%** accuracy which is at par with the output given by the back-propagation algorithm.

3.3.2Sigmoid function kernel

It is used when there is a lack of complex information a sigmoid function is used. [12]

tanh(gamma*u'*v + coef0)....(20)

Accuracy of **62.5%** was achieved.

3.3.3 Polynomial function kernel

The polynomial kernel function is directional, i.e. the output depends on the direction of the two vectors in low-dimensional space. The efficiency was **50%.**

(gamma *u' * v + coef0)^degree.... (21)

4. CONCLUSION

Diagnosis of students with learning disability has never been an easy job. In this paper we have tried to diagnose Learning Disability (LD) by applying one of the techniques of artificial Neural Networks (ANN) to the problem.

Table 4 comparisons

TYPE	ACCURACY
BPN	75%
SVM(RBF)	75%
SVM(SIGMOID)	62.5%
SVM(POLYNOMIAL)	50%

According to our results this makes the diagnosing accurate, less time consuming and a less tedious job. If proven feasible this model will not only save time but also manpower and other resources and avoid the possible human bias.

Our study about diagnosing LD using Back propagation algorithm has shown a percentage **75%** success in diagnosing. There is scope for development in this field and hence increasing the efficiency and decreasing the error rate. Fig 3 and Table 4 describe the accuracy given by each algorithm.

When SVM algorithm was considered according to the different kernel type's different accuracy was obtained with RBF kernel having maximum accuracy at par with BPN at **75%**. Hence in our case of diagnosing Learning Disability we can use the BPN algorithm or the SVM with RBF kernel and get satisfactc '' results.

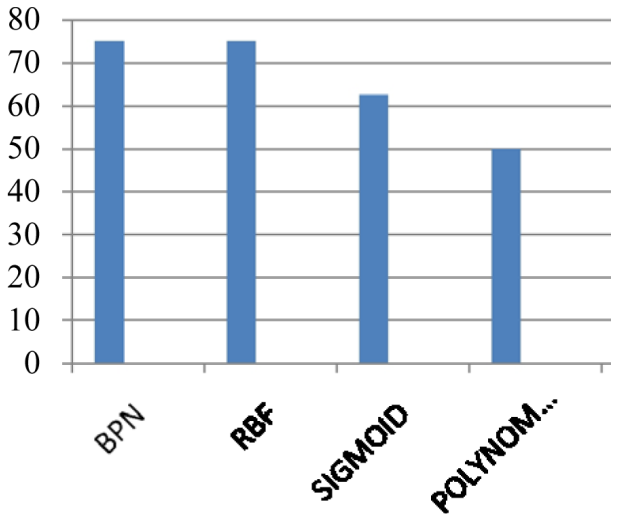

Fig 3 Accuracy v/s Artificial Intelligence Technique used

5. FUTURE WORKS

However, there is still work to be done in the future. The first most important one is to find an optimal feature set so as to develop a model that requires fewer features and has sufficiently high accuracy. Fewer features mean that parents and teachers (both regular and special education teachers) can concentrate on collecting variables that are relevant and essential. In addition, it is also necessary to collect much more data, not just in the number of samples, but the associated variables with each sample. More data samples and additional essential features help to build a well-

ISBN: 978-988-17012-0-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

supported AI classification model for future prediction. This is essential since there are 11 sub-types of learning disabilities, and chances are there could be more. As we have found in our experiment with data set 1, certain students manually diagnosed as LD are always classified as non-LD. Finally, most special education teachers or professionals we talked to tend to be sceptical to these kinds of predictor.

We will use fuzzy logic or neuro fuzzy to optimize our inputs and hence try and get a better success rate. By introducing this, the input that is being fed into the algorithm will be fuzzified and the results which we get from the algorithm will need to be again defuzzified. By combining ANN with fuzzy logic the tool which we will get will be highly efficient and error rate should be next to minimum.

REFERENCES

[1] Ahmet Yardimci, "Soft Computing in Medicine", Applied Soft Computing, Volume 9, 2009, Page No. 1029 – 1043

[2] A.S Kaufman, "Intelligence testing with WISC-R tests ", New York, Wiley, 1979

[3] A.S Kaufman," The WISC-R test and Learning Disabilities assessment: state of the art", Journal of Learning Disabilities, 1981, Volume 14, 520-526

[4] Bannatyne, "A Diagnosis: a note on recategorization of WISC-R tests", A Journal Of LD,1974, Vol 7,272-273

[5] David M. Schwartz, PhD," Wechsler Intelligence Test for Children", The Psychological Corporation, 2003

[6] J Anuradha, Tisha, Varun Ramachandran, Dr KV Arulalan, Dr. BK Tripathy, "Diagnosis of ADHD using SVM algorithm", ***Compute'10***, Jan 22-23, 2010, Bangalore, Karnataka, India Copyright 2010, Pg No. 29 ACM 978-1-4503-0001-8/00/0010

[7] J.S. Lawson and James Inglis, "The Psychometric Assessment of children with Learning Disabilities: An Index Derived from a Principal Components Analysis of the WISC-R", Journal of Learning Disabilities, 1984; 17; 517

[8] Madhuri Kulkarni, Sunil Karande, Anjana Thadani, "Educational Provision and Learning Disability", Indian Journal of Pediatrics, Volume 73, Sept 2006, Page No. 789-793

[9] Pai-Hsuen Chen,Chih-jen Lin, Nernhard Scholkopf, "A tutorial on nu-Support vector machines", Department of Computer Science and Information Engineering; National Taiwan University

[10] SN Sivanandan and SN Deepa," Principles of Soft Computing", 2007 Edition, Wiley India Publication

[11] T. Hill & Lewicki, P ,StatSoft, Inc. (2010). Electronic Statistics Textbook. Tulsa, StatSoft. WEB: http://www.statsoft.com/textbook/, (Printed Version). (2007), STATISTICS Methods and Applications. StatSoft, Tulsa

[12] Tung-Kuang Wu,Shian-Chang Huang,Ying-Ru Meng, "Evaluation of ANN and SVM classifiers as predictors to the diagnosis of students with learning disabilities", IEEE transactions on Neural Networks, 2008

ISBN: 978-988-17012-0-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)