International Journal of Computer and Electrical Engineering, Vol.3, No.1, February, 2011
1793-8163

Prediction of Key Symptoms of Learning Disabilities in School-Age Children Using Rough Sets

Julie M. David and Kannan Balakrishnan

***Abstract*— This paper highlights the prediction of learning disabilities (LD) in school-age children using rough set theory (RST) with an emphasis on application of data mining. In rough sets, data analysis start from a data table called an information system, which contains data about objects of interest, characterized in terms of attributes. These attributes consist of the properties of learning disabilities. By finding the relationship between these attributes, the redundant attributes can be eliminated and core attributes determined. Also, rule mining is performed in rough sets using the algorithm LEM1. The prediction of LD is accurately done by using Rosetta, the rough set tool kit for analysis of data. The result obtained from this study is compared with the output of a similar study conducted by us using Support Vector Machine (SVM) with Sequential Minimal Optimisation (SMO) algorithm. It is found that, using the concepts of reduct and global covering, we can easily predict the learning disabilities in children.**

***Index Terms*— Global Covering, Indiscernibility Relation, Learning Disability, Reduct and Core**

I. INTRODUCTION

In recent years the sizes of databases have increased rapidly. This has lead to a growing interest in the development of tools capable in the automatic extraction of knowledge from data. The term Data Mining or Knowledge Discovery in databases has been adopted for a field of research dealing with the automatic discovery of implicit information or knowledge within databases [19]. Knowledge Discovery in Databases (KDD) is the process of identifying useful information in data [16]. A widely accepted formal definition of data mining is given subsequently. According to this definition, data mining is the non-trivial extraction of implicit previously unknown and potentially useful information about data [6]. Conventionally, the information mined is denoted as a model of the semantic structure of the datasets. The model might be utilized for prediction and categorization of new data [5].

A majority of areas related to medical services such as prediction of effectiveness of surgical procedures, medical tests, medication and the discovery of relationship among clinical and diagnosis data also make use of data mining methodologies [3].

Manuscript received August 20, 2010.

Julie M. David is with MES College, Aluva, Cochin – 683 107, India, as Asst. Professor in the Department of Computer Applications. (phone: +91-9447104152/9447434303; fax : +91- 484- 2678587; e-mail : julieeldhosem @yahoo.com)

Kannan Balakrishnan is with Cochin University of Science & Technology, Cochin – 682 022, India, as Reader in the Department of Computer Applications. (e-mail:mullayilkannan@gmail.com)

Rough set theory is a new intelligent mathematical tool introduced by Z. Pawlak in 1982[16, 11, 7]. Rough set theory represents an objective approach to imperfections in data. As per this theory, there is no need for any additional information about data and hence no feed back from additional expert is necessary. All computations are performed directly on data sets. A rough set is an approximation tool that works well when in environments heavy with inconsistency and ambiguity in data or involving missing data [16]. Along the years, rough set theory has earned a well-deserved reputation as a sound methodology for dealing with imperfect knowledge in a simple though mathematically sound way [1].

This paper presents the basic idea of rough set theory and shows how these ideas may be utilized for data mining. The rough set approach seems to be of fundamental importance to artificial intelligence and especially in the case of machine learning, knowledge acquisition, decision analysis, knowledge discovery from databases, expert systems, inductive reasoning and pattern recognition [9]. Rough set theory has been successfully applied in many real life problems in medicine, pharmacology, engineering, banking, finance, market analysis, environment management and others [12]. The rough set approach of data analysis has much important advantage. It provides efficient algorithms for finding hidden patterns in data from database and has the following advantages [2].

- Find minimal set of data (data reduct).
- Evaluates significance of data.
- Generates set of decision rules from data.
- Offers straightforward interpretation of obtained results.
- Most algorithms based on the rough set theory are particularly suited for the parallel processing.
- It is easy to understand.

In RST, the datasets are represented in two forms viz. information tables and decision tables. The information table contains attributes and objects. The decision table describes decision in terms of conditions that must be satisfied in order to carry out the decision specified in the decision table. In both tables, the attributes called variables and cases called objects are presented in columns and rows respectively.

II. LEARNING DISABILITY

Learning disability is a general term that describes specific kinds of learning problems. Learning disabilities are formally defined in many ways in many countries. However, they usually contain three essential elements: a discrepancy clause, an exclusion clause, and an etiologic clause [13]. The discrepancy clause states there is a significant disparity between aspects of specific functioning and general ability; the exclusion clause states the disparity is not primarily due to intellectual, physical, emotional, or environmental problems; and the etiologic clause speaks to causation involving genetic, biochemical, or neurological factors. The most frequent clause used in determining whether a child has a learning disability is the difference between areas of functioning. When a person shows a great disparity between those areas of functioning in which she or he does well and those in which considerable difficulty is experienced, this child is described as having a learning disability [13]. A learning disability can cause a child to have trouble in learning and using certain skills. The skills most often affected are: reading, writing, listening, speaking, reasoning and doing math [14]. Learning disabilities vary from child to child. One child with LD may not have the same kind of learning problems as another child with LD. There is no *cure* for learning disabilities [18]. They are life-long. However, children with LD can be high achievers and can be taught ways to get around the learning disability. With the right help, children with LD can and do learn successfully.

As many as 1 out of every 10 children in the United States has a learning disability. Almost 3 million children (ages 6 through 21) have some form of a learning disability and receive special education in school [3]. In fact, over half of all children who receive special education have a learning disability [4]. There is no *one sign* that shows a child has a learning disability. Experts look for a noticeable difference between how well a child does in school and how well he or she could do, given his or her intelligence or ability. There are also certain clues, most relate to elementary school tasks, because learning disabilities tend to be identified in elementary school, which may mean a child has a learning disability [13]. A child probably won't show all of these signs, or even most of them. However, if a child shows a number of these problems, then parents and the teacher should consider the possibility that the child has a learning disability. If a child has unexpected problems in learning to read, write, listen, speak, or do math, then teachers and parents may want to investigate more. The same is true, if the child is struggling to do any one of these skills [15]. The child may need to be evaluated to see if he or she has a learning disability.

When a LD is suspected based on parent and/or teacher observations, a formal evaluation of the child is necessary. A parent can request this evaluation, or the school might advise it. Parental consent is needed before a child can be tested [13, 14]. Many types of assessment tests are available. Child's age and the type of problem determines the tests that child needs. Just as there are many different types of LDs, there are a variety of tests that may be done to pinpoint the problem [15]. A complete evaluation often begins with a physical examination and testing to rule out any visual or hearing impairment [4]. Many other professionals can be involved in the testing process.

The purpose of any evaluation for LDs is to determine child's strengths and weaknesses and to understand how he or she best learns and where they have difficulty [14]. The information gained from an evaluation is crucial for finding out how the parents and the school authorities can provide the best possible learning environment for child [13].

III. PROPOSED APPROACH

This study is based on rough set approach. The application of rough set approach enables reduction of superfluous data in the information system and generation of classification rules showing relationships between the description of objects and their assignment to classes of a technical state.

Rough set theory is useful for rule induction from incomplete data sets. Using this approach we can distinguish between three types of missing attribute values: lost values, attribute-concept values and *do not care* conditions [20].

Rough set is defined in the following way. Let $X \subset U$ be a target set that we wish to represent using attribute subset P; that is, we are told that an arbitrary set of objects X comprises a single class, and we wish to express this class, i.e., this subset, using the equivalence classes induced by attribute subset P. In general, X cannot be expressed exactly, because the set may include and exclude objects, which are indistinguishable on the basis of attributes P.

The target set X can be approximated using only the information contained within P by constructing the P-lower and P-upper approximations of X:

$$\underline{P}_{X} = \{ x \mid [x]_P \subseteq X \} \quad (1)$$

$$P_{X} = \{ x \mid [x]_P \cap X \neq \emptyset \} \quad (2)$$

The P-lower approximation, or positive region, is the union of all equivalence classes in $[x]_P$ which are the subsets and are contained by the target set. The P-upper approximation is the union of all equivalence classes in $[x]_P$ which have non-empty intersection with the target set.

The lower approximation of a target set is a conservative approximation consisting of only those objects, which can positively be identified as members of the set. The upper approximation is a liberal approximation, which includes all objects that might be members of target set.

The accuracy of the rough-set representation of the set X can be given by the following [20]

$$\alpha_P(X) = \frac{|\underline{P}X|}{|\overline{P}X|} \quad (3)$$

In order to consider the features of rough set used to predict the important signs and symptoms of learning disability, we are using the concept of information table, decision table, global covering and data reduct. An information table consists of different variables called attributes and cases called objects. Variables are present in columns and cases in rows. The attributes contained in the information table are the signs and symptoms of learning disabilities. In this paper, we are using a checklist containing the same 16 most frequent signs & symptoms (attributes) generally used for the assessment of LD [15]. This attribute list is shown at Table 1 below.

Table 1. List of Attributes

Sl. No	Attribute	Signs & Symptoms of LD
1	DR	Difficulty with Reading
2	DS	Difficulty with Spelling
3	DH	Difficulty with Handwriting
4	DWE	Difficulty with Written Expression
5	DBA	Difficulty with Basic Arithmetic skills
6	DHA	Difficulty with Higher Arithmetic skills
7	DA	Difficulty with Attention
8	ED	Easily Distracted
9	DM	Difficulty with Memory
10	LM	Lack of Motivation
11	DSS	Difficulty with Study Skills
12	DNS	Does Not like School
13	DLL	Difficulty Learning a Language
14	DLS	Difficulty Learning a Subject
15	STL	Slow To Learn
16	RG	Repeated a Grade

However, for convenience, we are presenting only six attributes and five cases in the sample information table given at Table 2 below, for illustration. For this study, we have collected more than 500 datasets (cases) from the learning disability clinics/schools in and around Cochin. By using the real time datasets for assessing the LD in children, we have identified the type of LD belongs to each child. Since such identification of LD in each child using all the attributes is a very difficult task, we are using certain rules which enable us to easily identify different symptoms which are causing LD. Based on these, we are strictly assessed the symptoms of LD in each child. The mined rules are used for finding the relationship between the symptoms of learning disability.

Table 2. Sample Information Table

Cases	Attributes					
	DR	DS	DH	DWE	DBA	DHA
1	Y	Y	N	Y	N	Y
2	Y	N	N	N	Y	N
3	Y	Y	N	N	N	N
4	Y	Y	Y	Y	Y	N
5	Y	Y	Y	Y	Y	N

Let U denotes the set of all cases. A be the set of all attributes and V be the set of all attribute values. The information table defines an information function ρ: $U \times A \rightarrow V$. For example, ρ (1, DR) =Y. Let x Є U and B≤A. An elementary set of B containing x is denoted by $[x]_B$. Elementary sets are subsets of U consisting all cases from U. Elementary set may be defined in another way, through the notion of an indiscernibility relation [15]. The indiscernibility relation IND (B) is a binary relation on U defined for x, y Є U as follows.

(x,y)Є IND (B) if and only if ρ(x, a)=ρ(y, a) for all aЄB

Obviously, IND(B) is an equivalence relation. Equivalence relation is present through partitions [10]. Partition relation is a family of mutually disjoint nonempty sets of U, called blocks. So the union of all blocks is U. The partition induced by IND (B) will be denoted by B*. Blocks of B* are called elementary set of B.

IV. DETERMINATION OF REDUCT AND CORE

There is subsets of attributes, which can, by itself, fully characterize the knowledge in the database; such an attribute set is called a reduct [17]. The reduct of an information system is not unique: there may be many subsets of attributes, which preserve the equivalence-class structure expressed in the information system.

The set of attributes which is common to all reducts is called the core: the core is the set of attributes which is possessed by every legitimate reduct, and therefore consists of attributes which cannot be removed from the information system without causing collapse of the equivalence-class structure [17]. It is possible for the core to be empty, which means that there is no indispensable attribute.

In our study, determination of the core attributes of LD is important. Normally we can create different attribute reducts. But, the minimum number of reducts has to be determined. From the sample information table, we first take a single attribute to compare with the set of all attributes, viz. A*= {1},{2},{3},{4,5}. Then we take two attributes, then three and then four for similar comparison, as shown below.

(i) {DR}*={1,2,3,4,5}; comparing with A* {DR}*≠A*, therefore {DR} is not a reduct.
(ii) {DR, DS}*={1,3,4,5},{2}; comparing with A*, {DR, DS}*≠A*, therefore {DR, DS} is not a reduct.
(iii) {DR, DS, DH}*={1,3},{2},{4,5};comparing with A*, {DR,DS,DH}*≠A*, therefore {DR, DS, DH} is not a reduct.
(iv) {DR, DS, DH, DWE}*={1},{2},{3}, {4,5}; But here, {DR,DS,DH,DWE}*=A*, therefore {DR, DS,DH,DWE} is a reduct.

Reducts are important subsets of attributes. A subset B of the set A is called a reduct, if and only if (i) B* = A* and (ii) B is minimal with the property (B-{a})* ≠ A* for all a Є B [10]. Based on these properties, only {DR, DS, DH, DWE} is reduct. Similarly, by considering another set of attributes, we are also getting {DH, DWE, DBA, DHA} as reduct.

Computing of all reducts, by this method, is time consuming with respect to the number of attributes considered. In such cases, computation of all the reducts is a complex task. So, we restrict to compute a single reduct using a heuristic algorithm LEM1 [8]. The first step of this algorithm is elimination of the leftmost attribute from the set and check whether the remaining set is reduct or not. If the set is not reduct, put the attribute back into that set and eliminate the next attribute for similar checking. Like, we are eliminating until the last attribute for reduct checking.

As explained, we have already two properties, {DR, DS, DH, DWE} and {DH, DWE, DBA, DHA}, as reducts from our sample information table. Now, the left most attribute, DR is eliminated from the reducts and check whether the remaining combined set is reduct or not. Then we are getting {DS, DH, DWE, DBA, DHA}* = A*; therefore {DS, DH, DWE, DBA, DHA} is a reduct. Then we are eliminating the next left most attribute, DS and check whether the remaining set is reduct or not. Now we are getting {DH, DWE, DBA, DHA}* = A*, therefore {DH, DWE, DBA, DHA} is a reduct. Similarly, after eliminating DH we are getting {DWE, DBA, DHA}* = A*, therefore {DWE, DBA, DHA} is also a reduct. But after eliminating DWE, the remaining set {DBA, DHA}* ≠ A*, therefore, the set {DBA, DHA} is not a reduct. Hence put the attribute DWE into this set and eliminate next attribute DBA, getting the set {DWE, DHA} for reduct checking. Now, {DWE, DHA} ≠ A*, hence {DWE, DHA} is also not a reduct. Finally, after eliminating all other attributes other than the last one, we are getting {DHA}* ≠ A* resulting {DHA} is also not a reduct. From these steps, we are arriving at the conclusion that, the LEM1 algorithm forms the set of attributes {DWE, DBA, DHA} as the core reducts.

The determination of reducts from the real world data set using LEM1 algorithm, as explained above, is tedious, time consuming and complex in nature. Hence, we are using another algorithm viz. Johnson's Reduction Algorithm. This algorithm is applied by using the rough set tool kit, Rosetta, for analysis of data, on our 513 real datasets (cases) with 16 attributes and we are obtaining the set of core reducts as {DH, DBA, LM, DSS, STL} with a length of 5 as shown in Table 4.

V. DECISION TABLE

One of the important aspects in the analysis of decision tables is the extraction and elimination of redundant attributes. The identification of the most important attribute from the data set is also an equally important aspect. Redundant attributes are attributes that could be eliminated without affecting the degree of dependency between the remaining attributes and decision [11]. The degree of dependency is a measure used to convey the ability to discern objects from each other. In a decision table, variables are presented in columns. But it contains two categories- attributes and decisions. Decision table has only one decision Y or N, i.e. LD yes or LD no. Rows of decision table, like information tables, are labeled by case names.

A checklist, containing signs & symptoms of LD, ie. attributes, is used for evaluating LD. In the Sample Decision Table given at Table 3 below, there are two elementary sets - {LD}:{1,4,5} for LD has value yes (Y) and {LD}:{2,3} for LD has value No (N). Elementary sets of decisions are called concepts. Decision table contains the cases, which are diagnosed by experts. Decision tables are crucial to data mining. Based on RST, there are two approaches of data mining from complete data sets. They are Global Covering and Local Covering [9]. In our study, we are considering only global covering of consistent data in which the entire attributes are used for analysis.

A decision table may contain more than one reduct and any of these reducts can be used to replace the original table. We can define the number of reducts from decision table. Selecting the best reduct, from a decision table, is important in this study. In this paper, we are adopted a criteria that the best reducts are those with minimum number of attributes. Here, we are getting such a type of reduct for the prediction of LD. Hence, based on the sample decision table, we are evolving to a solution that, a single attribute is enough for the prediction of LD.

Table 3. Sample Decision Table

Cases	Attributes						Decisions (LD)
	DR	DS	DH	DWE	DBA	DHA	
1	Y	Y	N	Y	N	Y	**Y**
2	Y	N	N	N	Y	N	**N**
3	Y	Y	N	N	N	N	**N**
4	Y	Y	Y	Y	Y	N	**Y**
5	Y	Y	Y	Y	Y	N	**Y**

VI. GLOBAL COVERING

A minimal subset of the set of all attribute, such that the substitution partition depends on it, is called global covering. It may be selected on the basis of lower boundaries. In the case of inconsistent data the system computes lower and upper approximations of each concept [9]. In global approach, each concept is represented by the substitution partition.

Relative reducts or rule sets may be induced using Global Coverings. We start from the definition of a partition being finer than another partition. Let α and β be the partitions of U. α is finer than β, denoted α ≤ β, if and only if, for each block X of α, there exists a block Y of β such that X ≤ Y. Let d be a decision. Then, a subset B of the attribute set A is a global covering if and only if (i) B* ≤ {d}* and (ii) B is minimal with the property (B-{a})* ≤ {d}* is false for any a Є B [10]. Based on these properties, we are checking all subsets of A in the sample decision table, with {LD}* = {1,4,5},{2,3}, with cardinality equal to one.

(i) {DR}* = {1,2,3,4,5}; then {DR}* ≤ {LD}* is false.
(ii) {DS}*={1,3,4,5},{2}; then {DS}* ≤ {LD}* is false.
(iii) {DH}*={1,2,3,},{4,5}; then {DH}*≤{LD}* is false.
(iv) {DBA}*={1,3},{2,4,5};then{DBA}*≤{LD}* is false.
(v) {DHA}*={1},{2,3,4,5};then {DHA}*≤{LD}* is false.
(vi) {DWE}*={1,4,5}{2,3};then {DWE}*={LD}* is true.

Since in the cases (i) to (v) above, the attribute sets {A}* is not finer than {LD}*, they are not in global covering. The algorithm used for computing all reduct is similar to the algorithm for global covering and local covering. Here, first we have to check whether {A}* ≤ {d}*, where d is the decision. But, for the case (vi) above, A* is finer than {LD}*. Therefore, there is only one global covering of size one, ie. {DWE}.

Then we are checking all subsets of A with the cardinality equal to two.

(i) {DR, DS} = {1,3,4,5},{2};
then {DR, DS}*≤{LD}* is false
(ii) {DS, DH} ={1,3},{2,4,5};
then {DS, DH}*≤ {LD}* is false
(iii) {DH, DWE} = {1},{2},{3},{4,5};
then {DH, DWE}*≤ {LD}* is false
(iv) {DBA, DHA} ={1},{2,4,5},{3};
then{DBA, DHA}*≤{LD}* is false

Hence, there is no global covering of size two since in all the above cases A* is not finer than {LD}*. Then we are checking all subsets of A with the cardinality equal to three.

(i) {DR, DS, DH}*={1,3}{2}{4,5};
then {DR, DS, DH}* ≤ {LD}* is false.
(ii) {DWE, DBA, DHA}* = {1}{2}{3}{4,5};
then {DWE, DBA, DHA}*≤{LD}* is true.

Hence, there is only one global covering of size three, ie. {DWE, DBA, DHA}. Then we are checking all subsets of A with the cardinality equal to four.

(i) {DR, DS, DH, DWE}*= {1}{2}{3 }{4,5},
then {DR, DS, DH, DWE}*≤{LD}* is true.

Hence, there is only one global covering of the size four, ie. {DR, DS, DH, DWE}.

From the above, we are getting 3 sets of attributes, viz. {DWE}, {DWE, DBA, DHA} and {DR, DS, DH, DWE} as global covering, considering our sample decision table. Obviously, the worst time complexity of the algorithm for computing all global covering is the same as the algorithm for computing all reduct. Thus, we should restrict our attention for computing a single global covering. For this, we are using the same procedure of elimination of left most attribute, one by one, and checking the condition {A}* ≤ {d}*, until the last element is eliminated. A single global covering is used for rule induction [10]. We restrict our attention to attributes from the global covering and check the cases in the decision table. If such a rule condition is not exists in the decision table, it is not consistent and this rule condition can be dropped. From this concept, we can induce certain rules.

As derived from the global covering, the mined rule (DR, Y) (DS, Y) (DH, N) (DWE, Y) = (LD, Y) is consistent and which is existing as first case in the decision table. So we simplify by removing the left most attribute from the mined rule. Then, we get {DS, DH, DWE} as not consistent. By applying the same process of elimination, we are getting {DH, DWE} and {DWE} as consistent. From the above, the following rules can be mined.

(DR,Y) (DS,Y) (DH,N) (DWE,Y) = (LD,Y) **(R1)**
(DH,N) (DWE,N) = (LD, N) **(R2)**
(DH, Y) (DWE, Y) = (LD, Y) **(R3)**
(DH, N) (DWE, Y) = (LD, Y) **(R4)**
(DWE, Y) = (LD, Y) **(R5)**
(DWE, N) = (LD, N) **(R6)**

VII. RESULT ANALYSIS AND FINDINGS

The reduct (core attributes) and classification results on the 513 real data sets with 16 attributes are obtained from Rosetta, the rough set tool kit for analysis of data is shown in Table 4 and Table 5 below respectively. In Rosetta tool, Johnson's reduction algorithm is used for obtaining the reduct results and Naive Bayes Batch classifier is used for obtaining the classification results.

Table 4. Reduct Results

No name

	Reduct	Support	Length
1	{DH, DBA, LM, DSS, STL}	97	5

Table 5 . Classification Results

No name

		Predicted		
		t	f	
Actual	t	287	30	0.905363
	f	4	192	0.979592
		0.986254	0.864865	0.933723
ROC	Class	t		
	Area	0.985048		
	Std. error	0.004927		
	Thr. (0, 1)	0.136		
	Thr. acc.	0.136		

This study consists of two parts. The first part explains the features of rough set using LEM1 algorithm and in the second part LD in children is predicted using the Rosetta tool is well explained. The major findings from this study are the determination of core attributes of LD, the accuracy of rough set classification and the importance of rule mining for LD prediction in children.

As a pre-processing before data mining, a subset of original data, which is sufficient to represent the whole data set, is generated from the initial detailed data contained in the information system. This subset contains only minimum number of independent attributes for prediction of LD. This attribute is used to study about the original large data set. It is common to divide the database into two parts for creating training set and test set. One of these parts, for instance 10% of the data, is used as training set and examined by the data mining system. The rest of the original database is used as test set for checking whether the knowledge acquired from the training set is general or not. By examining the 513 data in the database, the system tries to create general rules and descriptions of the patterns and relations in database to gain knowledge, which is valid not only in the specific database considered but also for other similar data.

The knowledge is tested against the test set. It is then clearly seen that the patterns found in the training set are valid also for other data. Therefore, if the knowledge gained from the training set is the general knowledge, it is correct for most parts of the test set as well.

The learning disability detection process can be considered as a decision making process. The rules generated by considering the original data set give a strong platform for making decisions. We are interested in applying these rules for making decisions.

VIII. COMPARISON OF RESULT

In this study, we are used the algorithm LEM1 for forming rough set knowledge for prediction of LD in children. We also used the rough set tool Rosetta for obtaining the reduct and classification results. The result obtained from this study is compared with the output of a similar study conducted by

IACSIT
International Association of Computer Science and Information Technology
WWW.IACSIT.ORG

us using Support Vector Machine (SVM) with Sequential Minimal Optimisation (SMO) algorithm. The accuracy of SVM classification results, we obtained, is shown in Table 6 below.

Table 6 . Accuracy of SVM

TP Rate	*FP Rate*	*Pre-cision*	*Recall*	*F Mea-sure*	*ROC Area*	*Class*
0.990	0.030	0.981	0.990	0.986	0.980	T
0.979	0.016	0.974	0.979	0.977	0.978	F
Correctly Classified Instances					503 Nos.	98.05%
Incorrectly Classified Instances					10 Nos.	1.95 %
Time taken to build a model						2.78Sec

On comparison, it is found that, SMO is conceptually simple, easy to implement and generally faster. SVMs belong to the class of supervised learning algorithms in which the learning machine is given a set of examples with the associated labels.

In the present study, we used Naive Bayes Batch classifier for rough set classification and compared the results with the results obtained from the study with SMO algorithm in SVM for prediction of LD in children. In the case of large data sets, there may be chances of some incomplete data or attributes. In data mining concept, it is difficult to mine rules from these incomplete data sets. But in rough sets, the rules formulated will never influenced by any such incomplete datasets or attributes. Even though the result obtained in SVM method is slightly more accurate; the rough set method is best in accuracy of rule mining. It is found that LD can accurately be predicted by using both methods. If both SVM and rough sets approaches are compared, the data or the output in SVM is very complex while rough set method is much easier.

This study reveals that, out of the 513 real data sets, the SVM correctly classifies 503 instances in 2.78 seconds whereas Naive Bayes Batch classifier in rough sets correctly classifies 287 true-true instances. The accuracy of the classifiers can be determined by ROC curve. The area under ROC curve in both cases is nearer to 1, which means the accuracy of both classifiers is found to good. The other advantage of rough set concept is that it leads to significant advantages in many areas including knowledge discovery, machine learning and expert system. Also it may act as a knowledge discovery tool in uncovering rules for the diagnosis of LD affected children.

IX. CONCLUSION AND FUTURE RESEARCH

This paper highlights the application of rough set theory in LERS data mining system and use of Rosetta tool in rough set data analysis in particular emphasis to classification, in prediction of the learning disabilities in school age children. The extracted rules are very effective for the prediction. Obliviously, as the school class strength is 40 or so, the manpower and time needed for the assessment of LD in children is very high. But using the extracted rules, we can easily predict the learning disability of any child. Rough set approach shows, its capability in discovering knowledge behind the LD identification procedure. The main contribution of this study is the selection of the core attributes of LD, which has the capability in prediction. The discovered rules also prove its potential in correct identification of children with LD. In this paper, we are considering an approach to handle learning disability database and predicting the learning disability in school age children. Our future research work focuses on, fuzzy sets, to predict the percentage of LD, in each child.

REFERENCES

[1] Abraham, Ajith; Falcón, Rafael; Bello, Rafael (Eds.). "Rough Set Theory: A True Landmark in Data Analysis", Series: Studies in Computational Intelligence, Vol. 174, ISBN: 978-3-540-89920-4, 2009
[2] A. Kothari, A. Keskar : "Paper on Rough Set Approach for Overall Performance Improvement of an Unsupervised ANN - Based Pattern Classifier", Journal on Advanced Computational Intelligence and Intelligent Information, Vol. 13, No.4, 2009, pp 434-440
[3] Blackwell Synergy: Learning Disabilities Research Practices, vol. 22, 2006
[4] C. Carol, K. Doreen.: Children and Young People with Specific Learning Disabilities : Guides for Special Education, Vol. 9 UNESCO, 1993
[5] U.M. Fayyad, "From Data Mining to Knowledge Discovery: An Overview- Advances in Knowledge Discovery and Data Mining", 34, AAAI Press/MIT Press, ISBN 0-262-56097-6, 1996
[6] Frawley, Piaatetsky: Shaping Knowledge Discovery in Database; an Overview. The AAAI/MIT press, Menlo Park, 1996
[7] S. Greco, B. Matarazzo, R. Slowinski, "Dealing with missing data in rough set analysis of multi-attribute and multi-criteria decision problem", *Kluwer Academic Publishers*, Boston Dordrecht, London, 2000
[8] J. W. Grzymala-Busse, A New Version of the Rule Induction System LERS, Fundamenta Informaticae, 1997, 31: 27-39
[9] J.W. Grzymala-Busse, Knowledge Acquisition under Uncertainty-A Rough Set Approach. Journal of Intelligent & Robotic Systems, 1988, 1: 3-16
[10] J.W. Grzymala-Busse, Rough Set Theory with Applications to Data Mining, 2005
[11] Hameed Al-Qaheri, Aboul Ella Hassanien and Ajith Abraham. "Discovering Stock Price Prediction Rules using Rough Sets", 2008
[12] H. Chen, S.S. Fuller, C. Friedman, W. Hersh, *Knowledge Discovery in Data Mining and Text Mining in Medical Informatics*, 2005, pp. 3–34
[13] Julie M. David, Pramod K.V., "Paper on Prediction of Learning Disabilities in School Age children using Data Mining Techniques", In: Proceedings of AICTE Sponsored National Conference on Recent Developments and Applications of Probability Theory, Random Process and Random Variables in Computer Science, T. Thrivikram, P. Nagabhushan, M.S. Samuel (eds), 2008, pp 139-146
[14] 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.
[15] Julie M. David, Kannan Balakrishnan. "Paper on Prediction of Learning Disabilities in School Age Children using Decision Tree". In: Proceedings of the International Conference on Recent Trends in Network Communications- Communication in Computer and Information Science, Vol 90, Part - 3 N. Meghanathan, Selma Boumerdassi, Nabendu Chaki, Dhinaharan Nagamalai (eds), *Springer- Verlag Berlin Heidelberg*, ISSN:1865-0929(print) 1865-0937(online), ISBN 978-3-642-14492-9(print) 978-3-642-14493-6(online), DOI : 10.1007/978-3-642-14493-6_55, 2010, pp 533-542
[16] Matteo Magnani. Technical report on Rough Set Theory for Knowledge Discovery in Data Bases, 2003.
[17] Z. Pawlak, Rough Sets. Int. J. Computers and Information Sci., Vol 11, 1982, pp 341-356
[18] R. Paige, (Secretary): US Department of Education. In: Twenty-fourth Annual Report to Congress on the Implementation of the Individuals with disabilities Education Act-To Assure the Free Appropriate Public Education of all Children with Disabilities, 2002

[19] S.J. Cunningham, G Holmes, “Developing innovative applications in agricultural using data mining”. In: The Proceedings of the Southeast Asia Regional Computer Confederation Conference, 1999
[20] Z. Pawlak Rough Sets: “Theoretical Aspects of Reasoning About Data”. Dordrecht: *Kluwer Academic Publishing*. ISBN 0-7923-1472-7, 1991

Julie M. David, born in 1976, received her MCA degree from Bharathiyar University, Coimbatore, India in 2000, the M.Phil degree in Computer Science from Vinayaka Missions University, Salem, India in 2008 and is currently pursuing the Ph. D degree in the research area of Data Mining from Cochin University of Science and Technology, Cochin, India. During 2000-2007 she was with Mahatma Gandhi University, Kottayam, India as Lecturer in the Department of Computer Applications. Currently she is working as Asst. Professor in the Department of Computer Applications with MES College, Aluva, Cochin, India. She has published papers in International Journals and International and National conference proceedings. Her research interest includes Data Mining, Artificial Intelligence and Machine Learning. She is a member of International Association of Engineers. Also she is an International reviewer of Elsevier Knowledge Based Systems and International Journal of Computer Applications.

Dr. Kannan Balakrishnan, born in 1960, received his M.Sc and M. Phil degrees in Mathematics from University of Kerala, India, M. Tech degree in Computer and Information Science from Cochin University of Science & Technology, Cochin, India and Ph. D in Futures Studies from University of Kerala, India in 1982, 1983, 1988 and 2006 respectively. He is currently working with Cochin University of Science & Technology, Cochin, India, as an Associate Professor (Reader) in the Department of Computer Applications. He has visited Netherlands as part of a MHRD project on Computer Networks. Also he visited Slovenia as the co-investigator of Indo-Slovenian joint research project by Department of Science and Technology, Government of India. He has published several papers in international journals and national and international conference proceedings. His present areas of interest are Graph Algorithms, Intelligent systems, Image processing, CBIR and Machine Translation. He is a reviewer of American Mathematical Reviews. He is a recognized research Guide in the faculties of Technology and Science in the Cochin University of Science and Technology, Cochin, India. He has served in many academic bodies of various universities in Kerala, India. Also currently he is a member of the Board of Studies of Cochin, Calicut and Kannur Universities in India. He is also a member of MIR labs India.