Prediction of Learning Disabilities in Children: Development of a New Algorithm in Decision Tree

1Julie M. David, 2Shereena V. B., 3Suny Raja
Dept. of Computer Applications
MES College, Marampally, Aluva, Cochin-683 107, India
Email: julieeldhosem@yahoo.com

Abstract—Learning Disability (LD) is a neurological condition that affects a child's brain and impairs his ability to carry out one or many specific tasks. It is a classification including several disorders in which a child has difficulty learning in a typical manner, usually caused by an unknown factor or factors. LD affects about 15% of children enrolled in schools. An affected child can have normal or above average intelligence. A child with a learning disability is often wrongly labelled as being smart but lazy. Learning disability is not indicative of intelligence level. Rather, children with a learning disability have trouble performing specific types of skills or completing tasks if left to figure things out by themselves or if taught in conventional ways. As there is no cure for learning disabilities and they are life-long, the problems of children with specific learning disabilities have been a cause of concern to parents and teachers for some time. This paper proposes a systematic approach for identification of LD in school-age children using a modified supervised learning algorithm in decision tree, viz. Modified J48 algorithm. Decision trees with J48 are powerful and popular tool used for classification and prediction in data mining, but it is a failure in handling inconsistent data. So, in this paper, we are introducing the Modified J48 algorithm, which imputes the missing values in the pre-processing stage, and then the tree is pruned. The different rules extracted from the tree have used in prediction of learning disabilities.

Index Terms—Closest Fit, Data Mining, Decision Tree, Learning Disability (LD), Modified J48.

I. INTRODUCTION

Data mining is a technology that blends traditional data analysis methods with sophisticated algorithms for processing large volume of data. It has also opened up exciting opportunities for exploring and analyzing new types of data and for analyzing old types of data in new ways [1]. Data mining is an integral part of knowledge discovery in databases, which is the overall process of converting raw data into useful information. It is the non-trivial extraction of implicit previously unknown and potentially useful information about data [2]. In brief, data mining is the process of extracting patterns from data. In recent years the sizes of databases has 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 [3]. It is currently used in a wide range of profiling, practices, such as marketing, surveillance, fraud detection, scientific discovery, engineering, medicine, expert prediction, web mining, mobile computing etc. [4]. Databases are rich with hidden information, which can be used for intelligent decision making. A majority of areas related to medical services such as prediction of effectiveness of surgical procedures, medical tests, predication and the discovery of relationship among clinical and diagnosis data also make use of data mining methodologies [5].

Classification is a data mining (machine learning) technique used to predict group membership for data instances. Machine learning refers to a system that has the capability to automatically learn knowledge from experience and other ways [6]. Classification and prediction are two forms of data analysis that can be used to extract models describing important data classes or to predict future data trends [7]. Classification predicts categorical labels whereas prediction models continuous valued functions. Classification is the task of generalizing known structure to apply to new data while clustering is the task of discovering groups and structures in the data that are in some way or another similar, without using known structures in the data.

Supervised learning is the machine-learning task of inferring a function from supervised training data. A supervised learning algorithm analyzes the training data and produces an inferred function, which is called a classifier. The inferred function should predict the correct output value for any valid input object. This requires the learning algorithm to generalize from the training data to unseen situations. Decision trees are supervised
algorithms which recursively partition the data based on its attributes, until some stopping condition is reached [7]. This recursive partitioning, gives rise to a tree-like structure. Decision trees are white boxes as the classification rules learned by them can be easily obtained by tracing the path from the root node to each leaf node in the tree [8]. Decision trees very efficiently classify even large volumes data. This is because of the partitioning nature of the decision tree algorithm. The decision tree classifier, each time works on these smaller and smaller pieces of the data set formed by partitioning. This partitioned data set otherwise known as simple attribute-value data, on which the classifier works with, is easy to manipulate [9]. The most important feature of Decision Tree Classifier, one of the possible approaches to multistage decision-making, is its capability to break down a complex decision making process into a collection of simpler decisions, thus providing a solution, which is often easier to interpret [1].

The purpose of this research paper is to introduce a new supervised learning algorithm in decision tree, which is modified from the existing J48 algorithm, with an emphasis in data mining, for prediction of LD, in such a view to overcome the drawbacks of the existing algorithm. The remaining paper is organized as follows. A detailed literature review is made in Section 2. Section 3 describes about LD. The existing and proposed approaches in the classification methodology are explained in Section 4. Section 5 presents the result analysis and findings followed by the comparison of results in Section 6. Finally, Section 7 deals with conclusion and future research work.

II. LITERATURE REVIEW

The field of learning disabilities is undergone in studies during the recent times only. In this section, we discuss the literature survey conducted on the fields of learning disabilities as well as on various soft computing methods used for classification, prediction and data pre-processing.

In a study to identify specific learning disabilities Kenneth A. Kavale, in the year 2005, has developed an alternative model for making decision about the presence or absence of special learning disabilities [10]. Benjamin J. Lovett conducted a study, in 2010, on extended time testing accommodations for students with disabilities-answers to five fundamental questions [11]. This study reviews a wide variety of empirical evidence to draw conclusions about the appropriateness of extended time accommodations. The evidence reviewed raises concerns with the way that extended time accommodations are currently provided, although the same literature also points to potential solutions and best practices. The matter of students with reading and spelling disabilities, peer groups and educational attainment in secondary education, to investigate whether the members of adolescents’ peer groups are similar in reading and spelling disabilities and whether this similarity contributes to subsequent school achievement and educational attainment were studied by Noona Kiuru et. al., in the year 2011 [12].

Regarding study in decision tree, Hongcui Wang and Tatsuya Kawahara, in the year 2008 studied on effective error prediction using decision tree for automatic speech recognition grammar network in call system [13]. Anupama Kumar and Vijayalakshmi, M., studied on Efficiency of decision trees in predicting student’s academic performance in 2011 [14]. Many other researchers focused on the topic of imputing missing values. Chen and Chen [15] presented an estimating null value method, where a fuzzy similarity matrix is used to represent fuzzy relations, and the method is used to deal with one missing value in an attribute. Chen and Huang [16] constructed a genetic algorithm to impute in relational database systems. The machine learning methods also include auto associative neural network, decision tree imputation, and so on. All of these are pre-replacing methods. Embedded methods include case-wise deletion, lazy decision tree, dynamic path generation and some popular methods such as C4.5 and CART [17,18,19]. But, these methods are not a completely satisfactory way to handle missing value problems. First, these methods are only designed to deal with the discrete values and the continuous ones are discretized before imputing the missing value, which may lose the true characteristic during the converting process from the continuous value to discretized one. Secondly, these methods usually studied the problem of missing covariates or conditional attributes.

Apart from the above, there are little studies available, as indicated below, in the area of LD prediction with knowledge based theories. Tung-Kuang Wu, Shian Chang Huang and Ying Ru in 2008, studied two well-known artificial intelligence techniques, artificial neural network and support vector machine, to the LD diagnosis problem [20]. To improve the overall identification accuracy, they applied GA-based feature selection algorithms as the pre-processing step in the study. This study is based on the formal assessment of LD whereas the present paper relates to informal assessment of LD, which is more tedious compared to that of formal assessment. In another study, Maitrei Kohli and Prasad T.V., in 2010 proposed an approach for identification of dyslexia and to classify potential cases accurately and easily by ANN [21]. As dyslexia is only a type of LD, the present research paper, on general assessment of LD, is entirely different from their study.

III. LEARNING DISABILITY

Learning disability is a general term that describes specific kinds of learning problems. 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. A learning disability cannot be cured or
fixed [22]. The problems of children with specific learning disabilities have been a cause of concern to parents and teachers for some time. Paediatricians are often called on to diagnose specific learning disabilities in school-age children. Learning disabilities affect children both academically and socially [9]. Specific learning disabilities have been recognized in some countries for much of the 20th century, in other countries only in the latter half of the century, and yet not at all in other places [23]. These may be detected only after a child begins school and faces difficulties in acquiring basic academic skills. A learning disability can cause a child to have trouble learning and using certain skills. The skills most often affected are: reading, writing, listening, speaking, reasoning and doing math [24]. If a child has unexpected problems or struggling to do any one of these skills, then teachers and parents may want to investigate more. The child may need to be evaluated to see if he or she has a learning disability.

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. 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. A child probably won't show all of these signs, or even most of them [25]. Individuals with learning disabilities can face unique challenges that are often pervasive throughout the lifespan. Depending on the type and severity of the disability, interventions may be used to help the individual learn strategies that will foster future success. Some interventions can be quite simplistic, while others are intricate and complex. Teachers and parents will be a part of the intervention in terms of how they aid the individual in successfully completing different tasks. School psychologists quite often help to design the intervention and coordinate the execution of the intervention with teachers and parents. Social support can be a crucial component for students with learning disabilities in the school system and should not be overlooked in the intervention plan. With the right support and intervention, people with learning disabilities can succeed in school and go on to be successful later in life.

The most frequent clause used in determining whether a child has a learning disability is the difference between areas of functioning. When a child 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 [26]. 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 [27]. 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. A complete evaluation often begins with a physical examination and testing to rule out any visual or hearing impairment [28]. 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 [24]. 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 the child.

IV. CLASSIFICATION METHODOLOGY

The proposed classification methodology of this research work, for predicting the learning disability, is detailed below;

A. Data sets

Data sets are differ in a number of ways and may have different characteristics. The main characteristics of data sets having a significant impact on the data mining techniques are dimensionality, sparsity and resolution [1]. The dimensionality of a dataset is the number of attributes that the objects in the data set possess. Data with a small number of dimensions tend to be qualitatively different than moderate or high dimensional data. The difficulties associated with analyzing high dimensional data are sometimes referred to as the curse of dimensionality. Because of this, an important motivation in preprocessing the data is dimensionality reduction. In our learning disability prediction dataset, each case has 16 attributes with class LD. The attributes used in this study are, the same signs and symptoms of learning disabilities used in LD clinics. These attributes are shown in the attribute list given at Table 1. For classification, we used an attribute selection method for finding the selected attribute. Data set is carrying Boolean values. In this study, we have used 513 real world data sets for the implementation of decision tree.

TABLE 1. LIST OF ATTRIBUTES

<table>
<thead>
<tr>
<th>Sl. No.</th>
<th>Attribute</th>
<th>Signs &amp; Symptoms of LD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>DR</td>
<td>Difficulty with Reading</td>
</tr>
<tr>
<td>2</td>
<td>DS</td>
<td>Difficulty with Spelling</td>
</tr>
<tr>
<td>3</td>
<td>DH</td>
<td>Difficulty with Handwriting</td>
</tr>
<tr>
<td>4</td>
<td>DWE</td>
<td>Difficulty with Written Expression</td>
</tr>
<tr>
<td>5</td>
<td>DBA</td>
<td>Difficulty with Basic Arithmetic skills</td>
</tr>
<tr>
<td>6</td>
<td>DHA</td>
<td>Difficulty with Higher Arithmetic skills</td>
</tr>
<tr>
<td>7</td>
<td>DA</td>
<td>Difficulty with Attention</td>
</tr>
<tr>
<td>8</td>
<td>ED</td>
<td>Easily Distracted</td>
</tr>
<tr>
<td>9</td>
<td>DM</td>
<td>Difficulty with Memory</td>
</tr>
</tbody>
</table>

ISSN (Online): 2347 - 2812, Volume-2, Issue -5, 2014
B. Data preprocessing

Incomplete, noisy and inconsistent data are commonplace properties of large real world databases and data warehouses [7]. Incomplete data can occur for a number of reasons. On assessment of learning disability, relevant data may not be recorded due to misunderstanding. Our aim is to apply the preprocessing step to make the data more suitable for data mining. Data preprocessing is a broad area and consists of a number of different strategies and techniques that are interrelated in complex ways [1]. Among many approaches to handle missing attribute values, we are adopted an approach based on the closest fit idea. The closest fit algorithm for missing attribute value is based on replacing a missing attribute value by existing values of the same attribute in another case that resembles as much as possible the case with the missing attribute values. In searching for the closest fit case, we need to compare two vectors of attribute values of the given case with missing attribute values and of a searched case. In a case where any attribute values are missing, we may look for the closest fitting case within that case or among all cases, and then the algorithms are called concept closest fit or global closest fit respectively. On another way, the search can be performed on cases with missing attribute values or among cases without missing attribute values. During the search, the entire training set is scanned and for each case a distance is computed. The case for which the distance is the smallest is the closest fitting case. That case is used to determine the missing attribute values. We have implemented the closest fit algorithm using the Mathworks Software, MatLab.

Let e and e’ be the two cases from the training set. The distance between cases e and e’ is computed as follows;

\[ \text{distance}(e_i, e_i') = \begin{cases} 0 & \text{if } e_i = e_i' \\ 1 & \text{if } e_i \neq e_i' \land e_i \neq ? \land e_i' \neq ? \land e_i \neq e_i' \land e_i \neq ? \quad \text{and} \\ \frac{a_i}{b_i} & \text{if } e_i \neq e_i' \land e_i \neq ? \land e_i' \neq ? \text{ where,} \\ \end{cases} \]

where

- \(a_i\) is the maximum of values of \(A_i\)
- \(b_i\) is the minimum of values of \(A_i\)
- \(A_i\) is an attribute.

In a special case of closest fit algorithm, instead of comparing entire vectors of attribute values, the search is reduced to just one attribute, for which the case has a missing value. The missing value is replaced by the most frequent value within the same concept to which belongs the case with a missing attribute value for symbolic attributes and by the average of all existing values within the same concept for numerical attributes.

C. Classification by decision tree

Data mining techniques are useful for predicting and understanding the frequent signs and symptoms of behaviour of LD. There are different types of learning disabilities. If we study the signs and symptoms of LD, we can easily predict which attribute from the data set is more related to learning disability. The first task, to handle learning disability, is to construct a database consisting of the signs, characteristics and level of difficulties faced by those children having LD. Data mining can be used as a tool for analyzing complex decision tables associated with the learning disabilities. Our goal is to provide concise and accurate set of diagnostic attributes, which can be implemented in a user friendly and automated fashion. After identifying the dependencies between these diagnostic attributes, rules are generated and these rules are then be used to predict learning disability. In this paper, we are using a checklist containing the same 16 most frequent signs & symptoms, in our study - attributes, generally used for the assessment of LD [9] to investigate the presence of learning disability. This checklist is a series of questions that are general indicators of learning disabilities. It is not a screening activity or an assessment, but a checklist to focus our understanding of learning disability.

Based on the information obtained from the checklist, a data set is generated. This set is in the form of an information system containing cases, attributes and class. A complete information system expresses all the knowledge available about objects being studied.

Decision tree induction is the learning of decisions from class labelled training tuples [26]. Given a data set \(D = \{t_1, t_2, \ldots, t_n\}\) where \(t_i = <t_{i1}, \ldots, t_{in}>\). In our study, each tuple is represented by 16 attributes and the class is LD. Then, Decision or Classification Tree is a tree associated with D such that each internal node is labelled with attributes DR, DS, DH, DWE, etc. Each arc is labelled with predicate, which can be applied to the attribute at the parent node. Each leaf node is labelled with a class LD. The basic steps in the decision tree are building the tree by using the training data sets and applying the tree to the new data sets. Decision tree induction is the process of learning about the classification using the inductive approach [7]. During this process, we create a new decision tree from the training data. This decision tree can be used for making classifications. Here we are using the J48 algorithm, which is a greedy approach in which decision trees are constructed in a top-down recursive divide and conquer manner. Most algorithms for decision tree approach are following such a top down approach. It starts with a training set of tuples and their associated class labels. The training set is recursively partitioned into smaller subsets as a tree is being built. This algorithm consists of three parameters – attribute list, attribute selection method and classification. The attribute list is a list of...
attributes describing the tuples. Attribute selection method specifies a heuristic procedure for selecting the attribute that best discriminate the given tuples according to the class. The procedure employs an attribute selection measure such as information gain that allows a multi-way splits. Attribute selection method determines the splitting criteria. The splitting criteria tells as which attribute to test at a node by determining the best way to separate or partition the tuples into individual classes. In this study, we are using the Mathwork tool, MatLab, for attribute selection and classification.

1) Existing approach

J48 algorithm is used for classifying the Learning Disability. The procedure consists of three steps viz. (i) attribute list (ii) attribute selection method based on information gain and (iii) Classification. To illustrate this method, first we partition the datasets into two subsets and choose one of the subsets for training and other for testing. Then swap the roles of the subsets so that the previous training set becomes the test set and vice versa. The Information Gain Ratio for a test is defined as follows. IGR (Ex, a) = IG / IV, where IG is the Information Gain and IV is the Gain Ratio [26]. Information gain ratio biases the decision tree against considering attributes with a large number of distinct values. So it solves the drawback of information gain.

Based on the algorithm J48, we have constructed a decision tree in MatLab with 513 real data sets in which some of the attribute values are missing. After preprocessing the data, the dataset is classified using decision tree and then some rules are extracted for the prediction of LD. These rules show accuracy in the tune of 70%. The number of leaves and size of the tree are 7 and 13 respectively. From the 16 attributes, 6 attributes, viz. DR, DA, DBA, DLS, DSS and DHA are selected for rule formation. The pruned tree and extracted rules are given at (a) and (b) below;

a) J48 pruned tree

\[
\begin{align*}
\text{DR} &\leq 0 \\
&| \quad \text{DA} \leq 0: F (19.0/4.0) \\
&| \quad \text{DBA} \leq 0 \\
&| \quad | \quad \text{DLS} \leq 0 \\
&| \quad | \quad | \quad \text{DSS} \leq 0: F (7.0/1.0) \\
&| \quad | \quad | \quad | \quad \text{DLS} > 0: T (13.0/3.0) \\
&| \quad | \quad | \quad | \quad | \quad \text{DLS} > 0: T (21.0/2.0) \\
&| \quad | \quad \text{DA} > 0 \\
&| \quad | \quad | \quad \text{DHA} \leq 0: F (14.0/3.0) \\
&| \quad | \quad | \quad | \quad \text{DHA} > 0: T (7.0/2.0) \\
&| \quad | \quad | \quad | \quad | \quad \text{DBA} > 0: T (19.0) \\
&| \quad | \quad \text{DA} > 0 \\
&| \quad | \quad | \quad \text{DHA} \leq 0: F (14.0/3.0) \\
&| \quad | \quad | \quad | \quad \text{DHA} > 0: T (7.0/2.0) \\
&| \quad | \quad | \quad | \quad | \quad \text{DBA} > 0: T (19.0) \\
\end{align*}
\]

Number of Leaves: 7
Size of the tree: 13

b) Rules extracted from J48 decision tree;

R1: (DR=N, DA=N) => (LD, N)
R2 : (DR=N, DA=Y, DBA=N) => (LD, Y)
R3 : (DR=N, DA=Y, DHA=N) => (LD, N)
R4 : (DR=Y, DBA=N, DLS=N, DSS=N) => (LD, N)
R5 : (DR=Y, DBA=N, DLS=Y) => (LD, Y)
R6 : (DR=Y, DBA=N, DLS=N, DSS=Y) => (LD, Y)
R7 : (DR=Y, DBA=Y) => (LD, Y)

2) Proposed approach

The algorithm developed by us, for our work, after incorporating modifications to the existing J48 algorithm for inducing the decision tree is given below;

Algorithm: Generate_decision_tree. Generate a decision tree from the training samples of data partition
Input: 
- Data preprocessing using closest fit algorithm;
- Data partition, which is a set of training samples and their associated class labels;
- attribute list, set of candidate attributes;
- Attribute_selection_method.
Output : A decision tree.
Method:
1) perform data preprocessing using closest fit algorithm;
2) create a node N;
3) if samples are all of the same class, C then
4) return N as a leaf node labeled with the class C;
5) if attribute list is empty then
6) return N as a leaf node labeled with the most common class in samples; //majority voting//
7) select test attribute, the attribute among attribute list with the highest information gain;
8) label node N with test attribute;
9) for each known value a of test attribute;
10) grow a branch from node N for the condition test_attribute = a;
11) let s_i be the set of samples in samples for which test_attribute = a_i // a partition/
12) if s_i is empty then
13) attach a leaf labelled with the most common class in samples;
14) else attach the node returned by Generate_decision_tree to node N;
endfor

In our proposed approach, the missing values are replaced with appropriate values by using the closest algorithm and then the tree is constructed. From this tree the rules for predicting LD are extracted. After imputing the missing values, which have a good impact in classification and prediction, five attributes viz. DR, DS, ED, DWE and DLS have selected from the 16 attributes. The number of leaves and size of the tree are same as that of the existing J48, but the selected attributes are different, hence the rules are also different. The important factor in the proposed approach is that these rules extracted from the tree shows accuracy in the tune of 95%. The pruned tree and extracted rules are given at (a) and (b) below.
is procedure for making a benefit of removing redundant tests. The obtained children. The presented a method for making mutually exclusive and exhaustive rules. We are using the MatLab tool for implementing this method. A decision tree can be expressed by a set of conjuncts of the rule antecedent, while the class label at every path from the root node to the leaf node of a tree. One rule is generated for each leaf. The antecedent of the rule includes a condition for every node on the path from the root to that leaf and the consequent of the rule is the class assigned by the leaf [6]. This procedure produces rules that are unambiguous in that the order in which they are executed is irrelevant. However in general, rules that are read directly off a decision tree are far more complex than necessary and rules derived from trees are usually pruned to remove redundant tests. The rules are so popular because each rule represents an independent knowledge. New rule can be added to an existing rule set without disturbing them, whereas to add to a tree structure may require reshaping the whole tree. In this section, we present a method for generating a rule set from a decision tree. In principle, every path from the root node to the leaf node of a decision tree can be expressed as a classification rule. The test conditions encountered along the path form the conjuncts of the rule antecedent, while the class label at the leaf node is assigned to the rule consequent. The expressiveness of a rule set is almost equivalent to that of a decision tree because a decision tree can be expressed by a set of mutually exclusive and exhaustive rules. We are using the MatLab tool for implementing the data pre-processing and classification. The classification algorithm J48 is now modified and it can be used for classifying and predicting the LD accurately.

### V. RESULT ANALYSIS, COMPARISON AND FINDINGS

The J48 and Modified J48 have used for the construction of tree. The rules generated from the data sets by applying these two algorithms are different. If the minimum confidence threshold is, say 90%, then the rules from the Modified J48 algorithm are more accurate and output sure. The certainty of the generated rules, ie. confidence of each rule, is calculated by considering the cases given in datasets. The accuracy obtained for the rules R1, R2, R3, R4, R5, R6 and R7 in both existing and proposed modified algorithm along with that of a rough set study, conducted by us recently, are given in Table 2. From this, it has been shown that, the rules extracted from the Modified J48 algorithm are more accurate and precise. By using the rules developed by us using this modified approach, we can easily and precisely predict the learning disability of any child. This study will be helpful for the parents, teachers and school authorities in diagnosing the child’s problem at an early stage. Hence these results will be helpful and beneficial for the educational as well as medical communities.

In this study, by using the algorithm J48 we have developed a modified algorithm, for prediction of LD in children. The results on confidence of the rules obtained from this study is compared with that of a similar study [24] recently conducted by us using Rough Set Theory (RST) with LEM1 algorithm, which is also shown in Table 2. From these, we can see that, the rules generated based on Modified J48 decision tree, is also more powerful than those of rough set theory. From the comparison of results, we have noticed that, the decision tree algorithm modified from J48 has a number of advantages over RST with LEM1 algorithm for solving the similar nature of problems. For 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. In modified decision tree, the rules formulated will never influence by any such incomplete data sets or attributes. Because, all the missing values are imputed at the pre-processing stage itself, ie. before the construction of tree. Hence, LD can easily and more precisely be predicted by using the modified method developed by us. The other benefit of decision tree 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. The importance of this study is that, using the modified decision tree we can easily predict the key attributes (signs and symptoms) of LD and can predict whether a child has LD or not. This model would be highly beneficial to the parents of students to diagnose their child’s problem and then teachers/school authorities to identify the problems of children and provide a good solution and also it is beneficial to the young students to know their problems at an early stage and thereby avoid their academic losses.

### TABLE 2. CONFIDENCE OF RULES

<table>
<thead>
<tr>
<th>Rules</th>
<th>J48</th>
<th>Modified J48</th>
<th>Rough Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>75%</td>
<td>95%</td>
<td>80%</td>
</tr>
</tbody>
</table>

International Journal of Recent Advances in Engineering & Technology (IJRAET)
The extracted rules are very effective for the prediction. Obviously, 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 these rules, we can easily predict the learning disability of any child. Decision tree approach shows, its capability in discovering knowledge behind the LD identification procedure. The main contribution of this study is the selection of the best attributes that has the capability to predict LD and avoid the drawbacks in other classification methods. Usually all classification methods are ignoring the missing attribute values. But here we are providing a best method in the pre-processing stage itself for catering such drawbacks. The results show accuracy in the tune of 95%. In best of our knowledge, none of the study has conducted for prediction of learning disabilities in school children. The discovered rules also prove its potential in correct identification of children with learning disabilities.

VI. CONCLUSION AND FUTURE WORKS

In this paper, we developed a new approach in supervised learning algorithm in decision tree to effectively and accurately predict the learning disability in school age children. This study mainly focuses on removing the drawbacks of decision tree, because accuracy of decision-making can be improved by using our good method of missing value imputing and applying the extracted rules. This study has been carried out on more than 500 real data sets with most of the attributes takes binary values and more work need to be carried out on quantitative data as that is an important part of any data set. In future, more research is required to apply the same approach for large data set consisting of all relevant attributes. This study is a true comparison of the proposed approach by applying it to large datasets and analyzing the completeness and effectiveness of the generated rules.

Modified J48 Decision Tree application on discrete data shows that it is better than existing J48 and RST in terms of efficiency and complexity. J48 Decision Tree has to be applied on continuous or categorical data. Noise effects and their elimination have to be studied. The results from the experiments on these small datasets suggests that Modified J48 Decision Tree can serve as a model for classification as it generates simpler rules and remove irrelevant attributes and imputes the missing attribute values at a stage prior to the tree induction. Our future research work focus on, fuzzy-neuro methods, for finding the percentage of LD in each child.

REFERENCES


[4] Hsinchun Chen; Sherrilynne S. Fuller; Carol Friedman, William Hersh. Knowledge Discovery in Data Mining and Text Mining in Medical Informatics, 2005, 1, pp3-34.


