Soft Comput (2014) 18:1093–1112
DOI 10.1007/s00500-013-1129-0

METHODOLOGIES AND APPLICATION

Learning disability prediction tool using ANN and ANFIS

Julie M. David · Kannan Balakrishnan

Published online: 24 September 2013

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. LD affects about 15 % of children enrolled in schools. The prediction of LD is a vital and intricate job. The aim of this paper is to design an effective and powerful tool, using the two intelligent methods viz., Artificial Neural Network and Adaptive Neuro-Fuzzy Inference System, for measuring the percentage of LD that affected in school-age children. In this study, we are proposing some soft computing methods in data preprocessing for improving the accuracy of the tool as well as the classifier. The data preprocessing is performed through Principal Component Analysis for attribute reduction and closest fit algorithm is used for imputing missing values. The main idea in developing the LD prediction tool is not only to predict the LD present in children but also to measure its percentage along with its class like low or minor or major. The system is implemented in Mathworks Software MatLab 7.10. The results obtained from this study have illustrated that the designed prediction system or tool is capable of measuring the LD effectively.

Keywords ANN · ANFIS · Closest fit · Data mining · Learning disability · PCA

Communicated by G. Acampora.

J. M. David (✉)
Department of Computer Applications, MES College Marampally, Aluva, Cochin 683 107, India
e-mail: julieeldhosem@yahoo.com

K. Balakrishnan
Department of Computer Applications, Cochin University of Science and Technology, Cochin 682 022, India
e-mail: mullayilkannan@gmail.com

1 Introduction

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 (Julie and Pramod 2008; Sally and Holmes 1999). Data mining is a collection of techniques for efficient automated discovery of previously unknown, valid, novel, useful and understandable patterns in large databases. According to a widely accepted formal definition given subsequently, data mining is the non trivial extraction of implicit previously unknown and potentially useful information about data (Frawley and Piaatetsky 1996). 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. Diverse fields such as marketing, customer relationship management, engineering, medicine, crime analysis, expert prediction, web mining and mobile computing besides others utilize data mining (Hsinchun et al. 2005). 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 (Chapple 1995).

Learning disability (LD) is a general term that describes specific kinds of learning problems. It is a neurological condition that affects a child's brain and impairs his ability to carry out one or many specific tasks (Julie and Balakrishnan 2010a). The LD affected children are neither slow nor mentally retarded (Blackwell Synergy 2007). An affected child can have normal or above average intelligence. They may have difficulty paying attention, with reading or letter recognition, or with mathematics. It does not mean that children who have learning disabilities are less intelligent. In fact, many children who have learning disabilities are more intelligent than the average child. Mental retardation, emotional disorders and poor socioeconomic status are not considered as learning disabilities. The main thing to remember when learning disabilities are discussed is that with proper intervention, teaching and learning techniques, a child with special needs related to one of these disabilities can succeed in school. This is why a child with a LD is often wrongly labeled as being smart but lazy. The LD 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 (Julie and Balakrishnan 2010a). 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 (Chapple 1995; Julie and Balakrishnan 2011a). They are life-long. However, children with LD can be high achievers and can be taught ways to get around the LD. With the right help, children with LD can and do learn successfully.

Neuro fuzzy system combines the learning capabilities of neural networks with the linguistic rule interpretation of fuzzy inference systems (Rahib et al. 2005). The field of neural network was originally kindled by psychologists and neurobiologists who sought to develop and test computational analogous of neurons. A Neural Network is a set of connected input or output units in which each connection has a weight associated with it. During the learning phase, the network learns by adjusting the weights so as to be able to predict the correct class label of input tuples (Han et al. 2011). An advantage of neural network is their high tolerance of noisy data as well as their ability to classify patterns on which they have not been trained. They can be used when we have little knowledge of the relationship between attributes and classes. They are well suited for continuous valued inputs and outputs, unlike most decision tree algorithm. They have been successful on a wide array of real world data, including hand written character recognition, laboratory medicine and pathology. Neural network algorithms are used the parallelization technique to speed up the computation process. In addition, several techniques have recently been developed for the extraction of rule from trained neural network. These factors contribute towards the usefulness of neural network for classification and prediction in data mining (Han et al. 2011).

In this study, we are using the two techniques viz. Artificial Neural Network (ANN) and Adoptive Neuro Fuzzy Inference System (ANFIS) for the prediction of learning disabilities in school age children. Technological analysis tries to model and simulate as accurately as possible the prediction of learning disabilities by different techniques. Works are composed by computing units interconnected so that each neuron can send and receive signals to or from others neural networks are a good answer for the modeling of distributed nonlinear system. As their design is based on the human brain, they were made to acquire knowledge through learning. Neuro fuzzy system is a combination of ANN and fuzzy systems (Han et al. 2011). These systems' most prominent feature is to learning from examples or training sets. Fuzzy logic was found by Lofti A. Zadeh in 1965. It is based on the idea that sets are not crisp but some are fuzzy and these can be modeled in linguistic human terms such as low, minor and major in the case of LD presence. In fuzzy systems rules can be formulated using these linguistic expressions and they are applied to the LD problem. The ANN and ANFIS offer powerful method to predict LD. ANN learns from scratch by adjusting the interconnections between layers. FIS is a popular computing framework based on the concept of fuzzy set theory, fuzzy if-then rules, and fuzzy reasoning. The advantages of a combination of ANN and FIS are obvious (Ajith 2001). Neural networks are one among the widely recognized artificial intelligence machine learning models. A general conviction is that the number of parameters in the network needs to be associated with the number of data points and the expressive power of the network (Shantakumar and Kumaraswamy 2009). Here the ANN is used to predict whether the child has LD or NOT, but ANFIS predicts not only the presence of LD but also its percentage and class viz. low, minor or major. It is very important because we cannot easily say the percentage of LD present in each child. So the developed system is an effective tool in prediction of LD. The underlying principle of fuzzy has the ability to treat non precise and uncertain function, as neural networks may be identified by measuring of input and output signals of the process. The neuro fuzzy networks combine the fuzzy reasoning with neural networks. The basic idea of neuro-fuzzy methods is that the parameters of fuzzy system are computed through learning methods and knowledge data sets. The learning methods are the same as that are used by neural networks.

The purpose of this research paper is to design a prediction tool using ANN and ANFIS methods, with an emphasis in data mining, for prediction and measurement of LD in school age children and also determine the importance of preprocessing in classification. The remaining paper is organized as follows. Section 2 describes elaborately about LD. The proposed methodology is explained in Sect. 3. The design of the proposed tool using ANN and ANFIS methods is explained in Sect. 4. The result analysis and findings and comparison of results are explained in Sects. 5 and 6 respectively. Finally, Sect. 7 deals with conclusion and future research works.

2 Learning disability

Learning disability 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.

The unknown factor is the disorder that affects the brain's ability to receive and process information. This disorder can make it problematic for a child to learn as quickly or in the same way as some child who isn't affected by a LD. 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 (Julie and Balakrishnan 2011b). 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.

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 (Julie and Balakrishnan 2011a). These may be detected only after a child begins school and faces difficulties in acquiring basic academic skills. Even where they have been recognized, the amount of help available varies. 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 LD. A child probably won't show all of these signs, or even most of them (Julie and Balakrishnan 2010b). Almost three million children (ages 6 through 21) have some form of a LD and receive special education in school (Blackwell Synergy 2007). In fact, over half of all children who receive special education have a LD (Chapple 1995).

Even though LD is not indicative of intelligence level, the problems of children with specific learning disabilities have been a cause of concern to parents and teachers for some time. A LD cannot be cured or fixed (Rod 2002). Learning disabilities affect children both academically and socially (Julie and Balakrishnan 2010a). There is no one sign that shows a child has a LD (Julie and Balakrishnan 2011b). 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. However, if a child shows a number of these problems, then parents and the teachers should consider the possibility that the child has a LD. If a child has unexpected problems 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. The child may need to be evaluated to see, if he or she has a learning disability. LD is real and a stumbling block for a nation's development process.

The most frequent clause used in determining whether a child has a LD 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 LD (Hsinchun et al. 2005; Julie and Balakrishnan 2010c). 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 (Julie and Pramod 2008). 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 (Crealock and Kronick 1993). 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 (Julie and Balakrishnan 2011c). Pediatricians are often called on to diagnose specific learning disabilities in school-age children. 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 (Julie and Balakrishnan 2009). 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.

3 Proposed methodology

The aim of the proposed research work is to develop a prediction tool and thereby effectively predict LD and accurately measure its percentage and also determine the importance of preprocessing in classification. The LD prediction tool is designed using the two intelligent methods namely ANN and ANFIS. After applying the data preprocessing, using closest fit algorithm and Principal Component Analysis (PCA), on the data set, the classification methods ANN and ANFIS are applied. A Graphical User Interface (GUI) form is designed in MatLab 7.10. This form contains different text buttons, which performs the functions contained in it while enabling them. The knowledge obtained from the training is used to predict the new data along with the presence of LD and its percentage. In order to make the data appropriate for the mining process, it needs to be preprocessed. In data preprocessing stage, the redundant data or some attributes are unwanted or repetitions of some other attributes are removed, the number of attributes is reduced and the missing values are imputed. In the tool design, we are using closest fit algorithm for missing value imputation and PCA for attribute reduction. The designed tool provides children information collection area which saves the details of the children in the student database. The system flowchart shown at Fig. 1 below represents the proposed methodology adopted in this study. The main stages of information processing adopted in the study include data preprocessing, mining of data and knowledge obtained for the prediction. The information processing stages and variables that affect the prediction of LD are shown in Fig. 2 below.

Even though the methods adopted in the study are not new, it has greater relevance in the present field of application pertaining to the system related to the prediction of LD. Even if, data preprocessing improves the quality of data generally, as we consider the environment of the assessment procedure for LD prediction, the closest fit algorithm is very precise in imputing missing values in the LD data set and hence very suitable in the present context. A general and very recognizable method viz. PCA is the next preprocessing method of attribute reduction adopted in the study. Here, we consider the relevance of PCA for identifying the appropriate symptoms for getting the accurate results. In the case of general assessment of LD, different checklists are used. In view of the fact that, the symptoms of LD are different from person to person, we cannot recognize which check list is more suitable because it contains large number of symptoms related to LD and each one has its own significance. In this study, by applying PCA, we are not deleting any of the symptoms of LD but arranging them in the order of precedence or rank. From the ranked order of the symptoms, we are selecting the high ranked ones as they have very good contributions to the LD prediction. Through the implementation of these concepts with two well known classifiers viz. ANN and ANFIS, a suitable system for prediction especially for medical system can be designed. This concept is exploited in developing the learning disability prediction system in the present study.

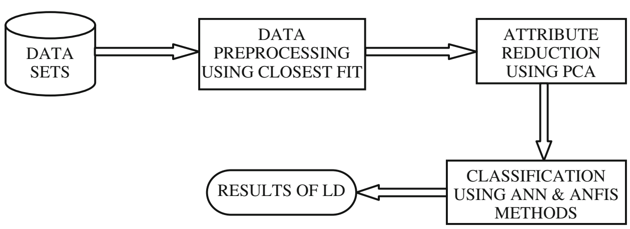

Fig. 1 System flowchart

3.1 Data sets

Data mining techniques are useful for predicting and understanding the frequent signs and symptoms of behavior of LD. There are different types of learning disabilities. If we study the signs and symptoms of LD, which are the attributes in our study, we can easily predict which of the attributes in the data set are more related to the LD. The first task to handle learning disability is the construction of a database consisting of the signs, characteristics and level of difficulties faced by the children. Data mining can be used as a tool for ana-

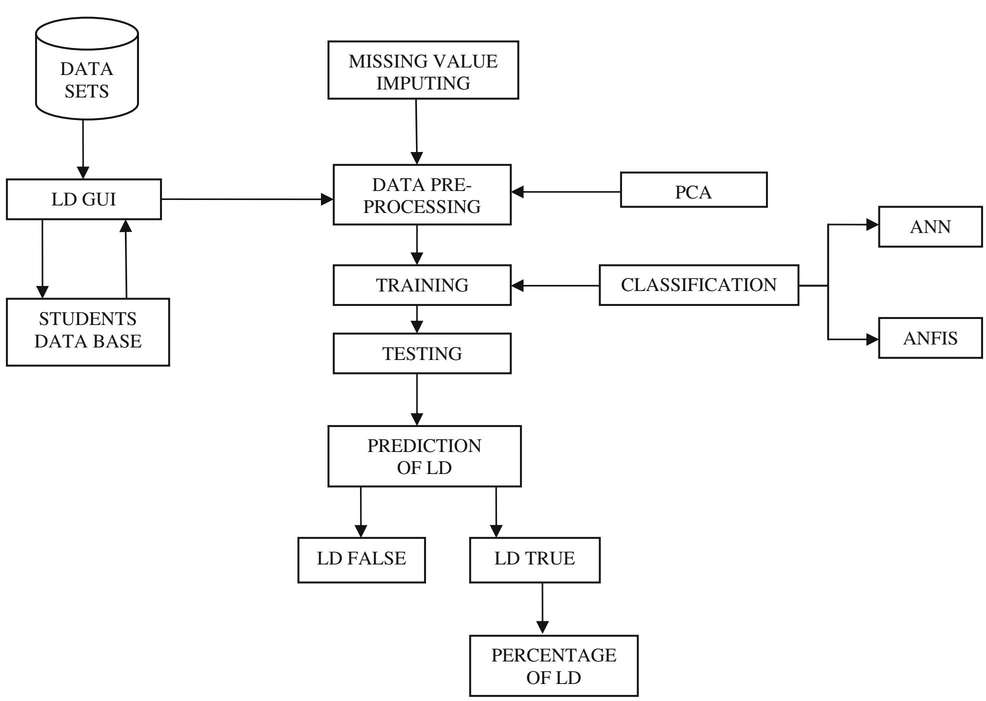

Fig. 2 Information processing stages and variables

Table 1 List of attributes

S. No.	Attribute	Signs and 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 in learning a language
14	DLS	Difficulty in learning a subject
15	STL	Slow to learn
16	RG	Repeated a grade

lyzing 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, classification is performed using ANN and ANFIS classifiers. A checklist is used to investigate the presence of LD. 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 LD. The attributes used in this study are, the same signs and symptoms of learning disabilities used in LD clinics. The attributes used in the study are listed in Table 1 above. In this study, we are used about 1,020 real world datasets. The data is collected from LD clinics and schools in and around Cochin, India. After conducting direct interview with the children, with the help of teachers and/or parents as required, the LD check list is filled, which is ultimately used for preparing the data for conducting the study.

3.2 Data preprocessing

Before the data is analyzed by neural network, it has to be preprocessed in order to increase the accuracy of output and to facilitate the learning process of neural network. Data preprocessing is a broad area and consists of a number of different strategies and techniques that are interrelated in complex ways (Tan et al. 2008). This is a critical operation since neural networks are pattern matches, thus the way data are represented directly influence their behavior. Data pre-processing is a step to be applied to make the data more suitable for data mining. Data preprocessing means the data be pre processed in order to help improve the quality of the data and consequently of the mining results. There are number of data preprocessing techniques which include data cleaning, data integration, data transformation, data reduction, etc. In this system, we are applying two cases (i) imputing missing values and (ii) data reduction for which we are using two techniques, closest fit algorithm and PCA for imputing missing values and reduction of attributes respectively.

The different process exist in the preprocessing stage are dimensionality reductions, feature subset selection, removal of noise from the data, imputing the missing data, etc. In the case of LD datasets, the checklist is the only one assessment for the prediction of LD. As it depends on the mood of child, we cannot obtain a checklist filled in all respects. Incomplete, noisy and inconsistent data are commonplace properties of large real world. Incomplete data can occur for a number of reasons. On assessment of LD, relevant data may not be recorded due to misunderstanding. Our aim is to apply the preprocessing step to make our data more suitable for data mining. Many data mining approaches can be modified to ignore missing values. The main idea of the closest fit, applied to a case with missing attribute values based on searching through the set of all cases, considered as vectors of attributes, for a case that is the most similar to the given case. There are two possible ways to looking for a case that is the closest: we may search in the same class or the entire set of all cases including those with missing attribute values.

3.3 Imputing missing values by closest fit algorithm

Many data mining approaches including ANN are usually ignoring either the case having a attribute with missing values or the attribute having the missing value. In this study, we are applied the closest fit algorithm for imputing the missing values. 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 (Grzymala-Busse et al. 2001). 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 these 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 MatLab. The concept of finding the distance of two vectors cases is given below:

Let e and e′ be the two cases from the training set.

The distance between cases e and e′ is computed as follows;

$$\sum_{i=1}^{n} \text{distance}\,(e_i, e_i'),\ \text{where}$$

$$\text{distance}\,(e_i,\ e_i') = 0 \quad \text{if } e_i = e_i',$$

distance$(e_i,\ e_i') = 1$ if e_i and e_i' are symbolic and $e_i \neq e_i'$ or e_i = ? or e_i' = ? and

distance$(e_i,\ e_i') = 1 - \frac{|e_i - e_i'|}{|a_i - b_i|}$ if e_i and e_i' are numbers and $e_i \neq e_i'$, where, a_i is the maximum of values of A_i, b_i is the minimum of values of A_i and A_i is an attribute.

3.4 Data reduction using principal component analysis

Principal Component Analysis is a quantitatively rigorous method for achieving the simplification of dimensionality reduction (Han et al. 2011). Principal components are new set of variables which are generated by the application of this method. Each principal component is a linear combination of the original variables. There will not be any redundant information as all the principal components are orthogonal to each other. The principal components as a whole form an orthogonal basis for the space of the data. An orthogonal basis for several columns of data can be constructed by an infinite number of ways. The first principal component is a single axis in space. When each observation on that axis will be projected a new variable is formed by the resulting values and the variance of this variable is the maximum among all possible choices of the first axis. The second principal component, which is perpendicular to the first, is another axis in space. Projecting the observations on this axis generates another new variable. The variance of this variable is the maximum among all possible choices of this second axis.

The data to be reduced consists of tuples or data vectors described by n attributes or dimensions are called PCA (Han et al. 2011). The PCA searches for k n-dimensional orthogonal vectors that can be used to represent the data where $k < n$. The original data are thus projected onto a much smaller space, resulting in dimensionality reduction. The basic procedures behind PCA are (i) the inputs data are normalized, so that each attribute falls within the same range. This helps ensure that attributes with large domains will not dominate attributes with smaller domains, (ii) PCA computes principal components; and (iii) the principal components are sorted in order of decreasing strength. Dimensionality reduction is accomplished by choosing enough Eigen vectors to account for some percentage of the variance in the original data. Attribute noise is filtered by transforming to the principal component space, eliminating some of the worst Eigen vectors and then transforming back to the original. In PCA, the eigenvectors are conventionally arranged so that the one with the largest Eigen value is first, which is equivalent the largest variance being first. In our study, we have used the LD data sets having 16 attributes. When we study this dataset, it is seen that some of the attributes are irrelevant. These irreverent attributes are removed by applying the PCA using Mat Lab and the number of attributes is reduced to seven. When we run the PCA algorithm in Mat Lab on 1,020 data set, we got the list of seven ranked reduced attributes of LD as given under:

```
mPca =
  0.4285
  0.4938
  0.1497
  0.3946
  0.2327
  0.2091
  0.3190
  0.3907
  0.2799
  0.3778
  0.3080
  0.2195
  0.3613
  0.0767
  0.3725
  0.1075
eigNum = 16
ans = 'DR' 'DS' 'DWE' 'ED' 'LM' 'DLL'
'STL'
```

4 Design of the proposed tool

The proposed tool for determining LD is designed based on ANN and ANFIS. The design and working of these two intelligent methods in designing the proposed tool are explained below.

4.1 Proposed tool based on artificial neural network

There are several common methods or techniques in data mining, such as statistical analysis, rough sets, covering positive and rejecting inverse cases, fuzzy method, neural network etc., which are used for classification (Xiangun 2008). The neural network can be broadly divided in to three viz. feed forward networks, feedback network and self organization network. At present neural network that commonly used in data mining is back propagation network.

ANN architecture is known to be strong function approximation for prediction and classification problems. It is capable of learning arbitrarily complex nonlinear functions to an arbitrary accuracy level. The back propagation algorithm performs learning on a multilayer feed forward neural network. It iteratively learns a set of weights for prediction of the class label of tuples. A multilayer feed forward neural network consists of an input layer, one or more hidden layers and an output layer. Each layer is made up of units. The input to the network corresponds to the attribute measured for each training tuple. The inputs are fed simultaneously into the units making up the input layer. These input pass through the input layer and are then weighted and fed simultaneously to a second layer known as hidden layer. The output of the hidden layer units can be input of another hidden layer and so on. The weighted outputs of the last hidden layer are input to units making up the output layer. This emits the network's prediction for given tuples. To compute the net input to the unit, each input connected to the unit is multiplied by its corresponding weight, and this is summed. Each unit in the hidden and output layers takes its net input and then applies an activation function to it. The function symbolizes the activation of the neuron represented by the unit (Han et al. 2011). Back propagation learns iteratively processing a data set of training tuples, comparing the network's prediction for each tuple with the actual known targets value known as class labels. Each training tuple, the weights are modified so as to minimize the mean squared error between the network prediction and the actual target value. These modifications are made in the backward direction.

Artificial Neural Network can be described as a neural network model capable of mapping sets of input data on to a set of appropriate output. It is an alteration of the typical linear perceptron where it employs one or more layers of neurons with non-linear activation functions. The primary task of the neuron in the input layer is the division of the input signal among neurons in the hidden layer. Every neuron in the hidden layer adds up its input signals and weights them with the strength of the respective connections from the input layer and determines its output as a function. The back propagation algorithm is used to train the neural networks. It is widely recognized for applications to layered feed forward networks or multilayer perceptrons (Savcovic 1994). The architecture consists of 16 attributes and two output nodes. The output nodes obtained are LD-false and LD-true. The main advantage of the neural network is its flexibility with multiple data and its main drawback is long time required for training feed forward network with back propagation training algorithm (Tarafdar and Najafi 2005). Neural network consists of number of independent processors or neurons that communicate with each other via weighted connections (Gray and George 1999).

ANN is capable of learning arbitrarily complex nonlinear functions to an arbitrary accuracy level. It iteratively learns a set of weights for prediction of the class label of tuples. A multilayer feed forward neural network consists of an input layer, one or more hidden layers and an output layer. Each layer is made up of units. The inputs are fed simultaneously into the units making up the input layer. The output of the hidden layer units can be input of another hidden layer and so on. The weighted outputs of the last hidden layer are input to units making up the output layer. This emits the network's prediction for given tuples. Back propagation learns iteratively processing a data set of training tuples, comparing the network's prediction for each tuple with the actual known targets value known as class labels. Each training tuple, the weights are modified so as to minimize the mean squared error between the network prediction and the actual target value. These modifications are made in the backward direction. ANN model is capable of mapping sets of input data on to a set of appropriate output. It is an alteration of the typical linear perceptron where it employs one or more layers of neurons with non-linear activation functions.

The gradient descent back propagation with adaptive learning is used for the performance measures using the sum squared error. A sum squared error is a measure of how well the back propagation trained neural network is doing at a particular point during its learning. It is obtained by adding up the sum-squared errors for each output neuron. The error is propagated backward by updating the weights and biases to reflect the error of the network prediction.

For a unit in the output layer, the error *Err* is computed by the equation, $Err = O(1 - O)(T - O)$, where O is the actual output of a unit and T is the non-target value of the given training tuple. $O(1 - O)$ is the derivative of logistic function.

At present neural network that commonly used in data mining is back propagation network. Neural network architecture is known to be strong function approximation for prediction and classification problems. Back propagation Neural Network is one of the most widely used neural network model, with extensive applications in function approximation. It composed of three layers, viz. input, hidden and output (Weixiang et al. 2008). The training process of the ANN used in this study is shown in Fig. 3 below.

The neural network train tool shows the input, output and hidden layers. We are provided the maximum epoch, 5,000 iterations, for yielding more accurate results and best performance, which are executed the training in 1.28 s, setting a performance of 151. One of the best training performances obtained from the NN train tool is shown in Fig. 4 below.

The gradient of 1.14e−07 used in the back propagation search for a set of weights that fits the training data so as to minimize the mean squared distance between the network class prediction and the known target value of the tuple. The

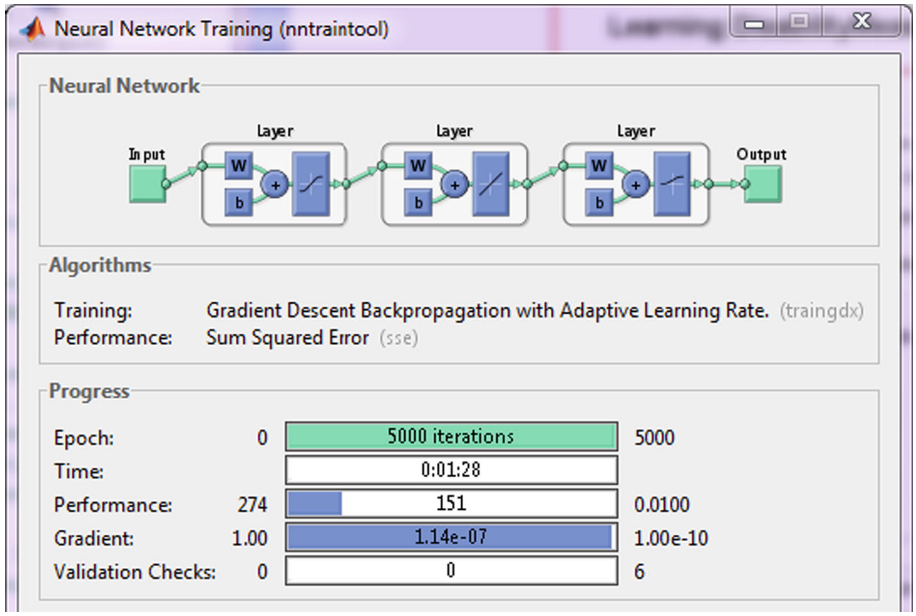

Fig. 3 Training process

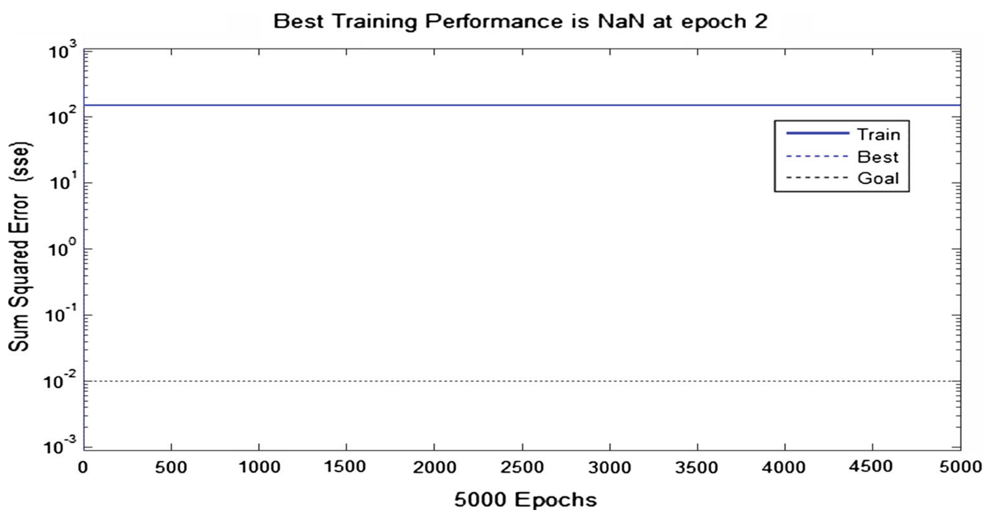

Fig. 4 Best training performance

validation check is 0. The learning rate obtained is 0.14636 which is well between 0.0 and 1.0. The performance of the training, ie. the gradient, validation checks and learning rate obtained are represented in Fig. 5 below. The training stops when the 5,000 epochs has expired. The knowledge so obtained from this training process is used to predict the learning disability of a new client.

The knowledge obtained after training the neural network with 507 cases is used for testing other 513 new cases. The result thus obtained shows 99.03 % accuracy. The result obtained when we run the training program of neural network in MatLab is shown under. The testing results are shown in Table 2 below.

A GUI is designed in MatLab 7.10 for the proposed LD tool in ANN as shown in Fig. 6 below. This form contains different text buttons, which performs the functions contained in it while enabling them. The system is designed such that, the intelligent methods to be used can be selected from the scroll bar. The details about the children are saved in the student database of the designed tool. In the designed GUI, the *Student Data* such as name, DOB, sex, standard, etc. are entered first. When the *FILE* button in the GUI is enabled, the trained data set is loaded into the system. Then the data pre processing and training are performed. Thereafter from the 16 attributes or signs and symptoms of LD, which is/are applicable in the case, is given and testing is performed. Out of 1,020 real world dataset, we have taken 507 dataset for carrying out training process and the remaining for testing purpose. We have obtained the results of LD prediction as *TRUE* or *FALSE* and we have compared it with the actual.

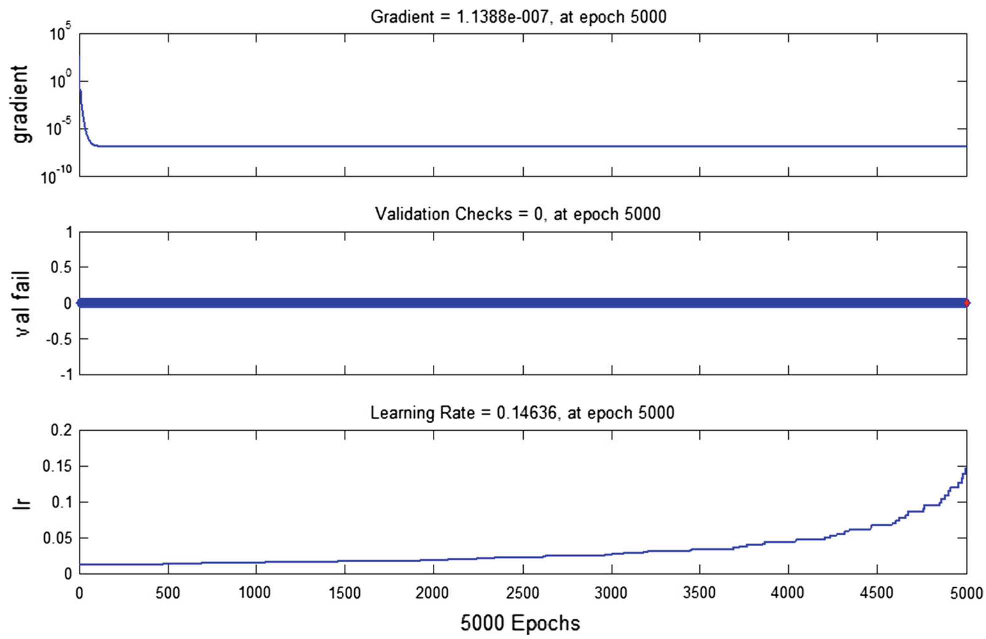

Fig. 5 Representation of gradient, validation check and learning rate

Table 2 Testing results of ANN

S. No.	Particulars	Numbers
1	Data set used for training	507
2	Number of data set used for testing	513
3	Number of instances correctly classified	508
4	Number of instances incorrectly classified	5

From the study, we have obtained 99.03 % accuracy in LD prediction.

4.2 Proposed tool based on adaptive neuro-fuzzy inference system

The ANFIS derives its name from adaptive neuro fuzzy inference system. Using a given input or output data set, ANFIS construct a fuzzy inference system whose membership function parameters are tuned using a back propagation algorithm. This adjustment allows the fuzzy system to learn from the data we are modeling. Neural network provide learning capacity and ability for generalization, on the other side fuzzy logic provide a logical reasoning based on inference rules. The combination of neural and fuzzy has the ability to learn linguistic rules or membership functions. To create a linguistic variable or membership functions based on training with a set of data values presented to these models. In order to build a set of fuzzy rules, at least the initial membership function must be defined. There are two very common approaches, one of them consists on the parametric description of the membership functions, and parameters must be optimized during the learning process and the second proposal where a neural network is used to generate membership values according to the input data. The first option is most used. The fuzzy sets have a membership function associated and which defines the distribution of the membership grade for the sets. Fuzzy logic system does not possess any inherent method of learning. It is the expert's knowledge that makes the fuzzy system work to satisfaction. The design and development of neuro fuzzy approach for prediction of LD includes creations of membership function, fuzzy rule system and fuzzy inference system, ANFIS training and prediction as represented in the flowchart given in Fig. 7 below.

4.2.1 Creation of membership function

Three membership functions or linguistic variable such as low, minor and major are defined here. Seven attributes extracted from the check list using dimensionality reduction method PCA and probability LD are the inputs given to the system. Only one output, LD, is there. Each of the inputs has two membership functions viz. MF1 and MF2 and the output has three membership functions viz. MF1, MF2 and MF3. The details of input and output membership functions are shown in Tables 3 and 4 respectively.

The input membership function editor of one of the seven extracted attributes, DS, and that of probability LD are shown below in Figs. 8 and 9 respectively. In our research work, the membership function of output variable in LD prediction is obtained as shown in Fig. 10 below.

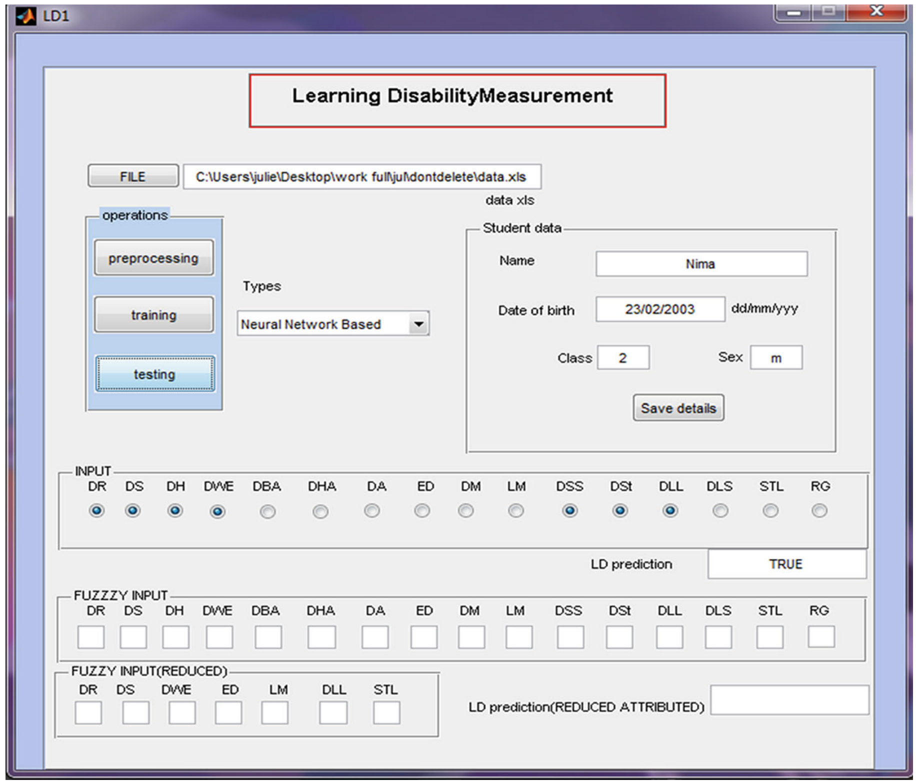

Fig. 6 Designed GUI of LD tool in ANN

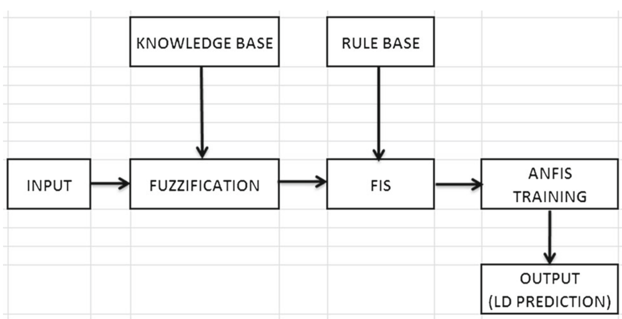

Fig. 7 Neuro fuzzy system flowchart for LD prediction

Table 3 Details of input membership functions

Name of MF	Type of MF	Range of MF
MF1	trapmf	[−0.164 −0.0198 0.4352 0.556]
MF2	trapmf	[0.45 0.597 1.07010582010582 1.17]

4.2.2 Creation of fuzzy rule system

The rule viewer of fuzzy inference system formed based on the relationship existing between input and output of LD is shown in Fig. 11 below. Fuzzy systems are characterized by fuzzy sets and fuzzy IF-THEN rules (Rahib et al. 2005). The fuzzy sets have a membership function associated with it, which defines the distribution of the membership grades for

Table 4 Details of output membership functions

Name of MF	Type of MF	Range of MF
MF1—LOW	trimf	[0.0071 0.0992063492063491 0.234]
MF2—MINOR	trimf	[0.302936507936508 0.507936507936508 0.71031746031746]
MF3—MAJOR	trimf	[0.842544973544973 0.908544973544974 1.00354497354497]

the set. The LD prediction can be determined based on the status of the attribute. The fuzzy system can be optimized by either manually or by using suitable optimization techniques.

In our system, the loaded data set is evaluated by initial fuzzy inference system. This contains 26 rules for defining and tuning the parameters of membership function. After generating the new fuzzy inference system 243 rules are contained in ANFIS. Some of the rules of LD prediction defined by us in this system are shown below.

If (LD is MF3) then (Output1 is MF3)
If (LM) is MF3 and (LD is MF2) then (Output1 is MF2)
If (ED is MF3) and (LD is MF2) then (Output1 is MF2)
If (DWE is MF3) and (LD is MF2) then (Output1 is MF2)

In more complex domains, even an expert may not be able to provide error free data or knowledge. Here

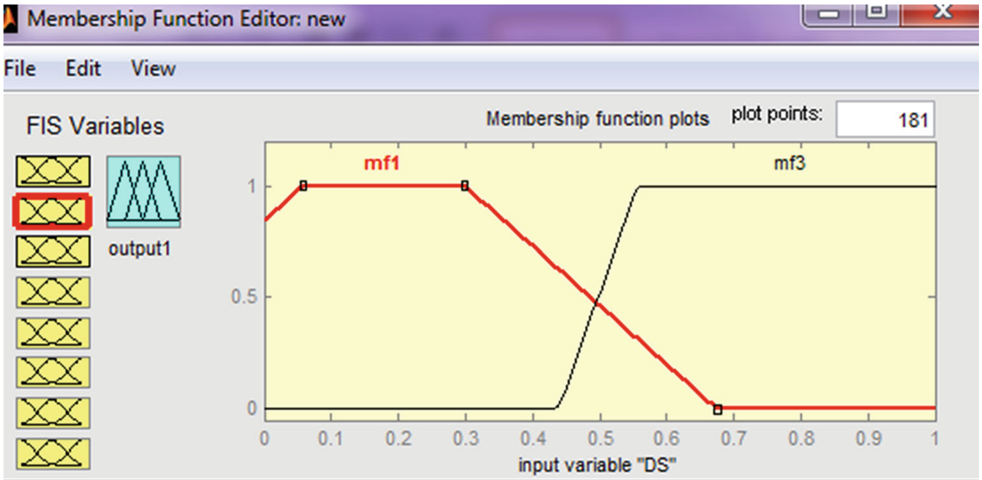

Fig. 8 Input membership function of attribute DS

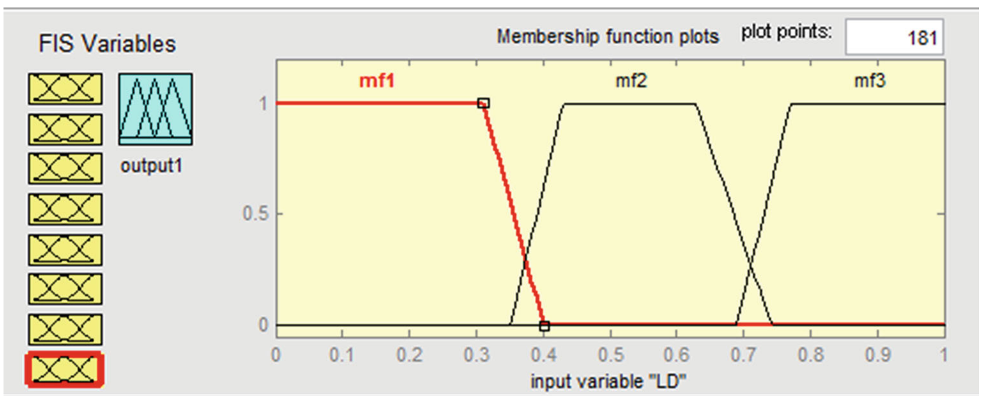

Fig. 9 Input membership function of LD

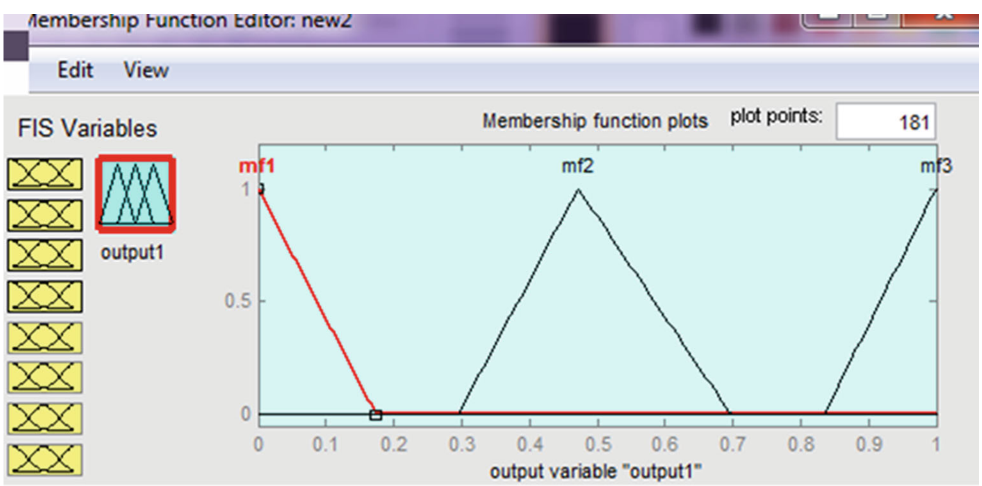

Fig. 10 Membership function of output variable LD

a neuro fuzzy system could be used to learn to tune the system and reject redundant rules. A neuro fuzzy system has layers that embedded in the fuzzy system. The basic idea of neuro fuzzy method is that the parameters of fuzzy system are computed through learning methods, knowing data sets. The learning methods are same as those used by neural networks (Simona et al. 2008).

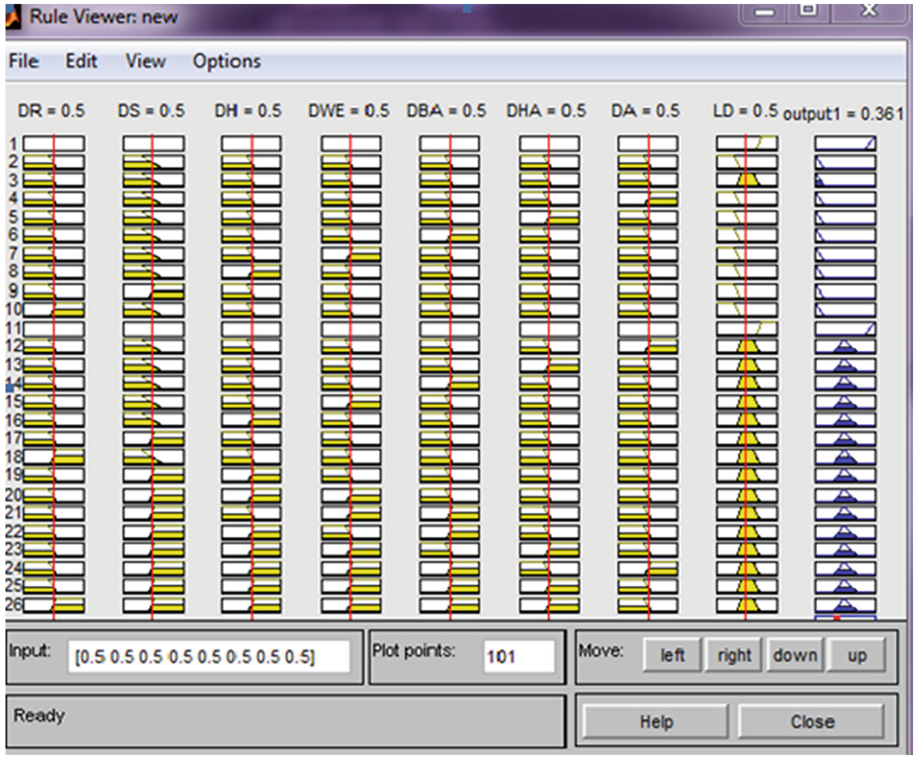

Fig. 11 The rule viewer of the system

4.2.3 Creation of fuzzy inference system

The basic structure of the fuzzy inference system seen thus far is a model that maps input characteristics to input membership functions, input membership function to rules, rules to a set of output characteristics, output characteristics to output membership functions and the output membership function to a single-valued output or a decision associated with the output (Sivarao et al. 2009). Neuro fuzzy systems, consists of sets of rules and inference systems combined with a connectionist structure for optimization and adaptation to given data. The neuro-adaptive learning method works similarly to that of neural networks. It provides a method for the fuzzy modeling procedure to learn information about a data set. Fuzzy logic toolbox software in Matlab computes the membership function parameters that best allow the associated fuzzy inference system to track the given input or output data. The Fuzzy logic toolbox function that accomplishes this membership function parameter adjustment is called ANFIS. The structure of initial fuzzy inference system used for creating the *newFIS* is shown in Table 5 below and the fuzzy inference system viewer is shown in Fig. 12 below.

Table 5 Structure of initial fuzzy inference system

Name=new1	Type = mamdani
Version=2.0	Num Inputs = 8
Num Outputs=1	Number rules = 26
And Method=min	Or Method=max
Imp Method=min	Agg Method=max
Defuzz Method=centroid	—

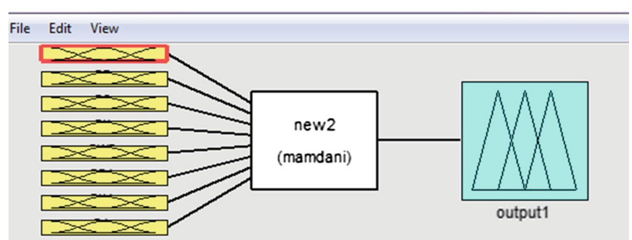

Fig. 12 Fuzzy inference system viewer

4.2.4 Structure of ANFIS

The Adaptive Neuro Fuzzy Inference System is a fuzzy inference system implemented in the frame work of an adaptive neural network by using a hybrid learning procedure. ANFIS can construct an input–output mapping based on both human knowledge as fuzzy IF THEN rules and approximate membership function from the stipulated input-output data pairs for neural network training. This procedure of developing a FIS using the framework of adaptive neural network is called an ANFIS (Sidda et al. 2009). Five network layers are used by ANFIS to perform the following fuzzy inference steps.

(i) Input fuzzification
(ii) Fuzzy set data base construction

(iii) Fuzzy rule base construction
(iv) Decision making and
(v) Output defuzzification.

The ANFIS architecture is shown in Fig. 13 below (Jang et al. 2008). The functions of each layer of ANFIS architecture are explained in Table 6 there under.

The new fuzzy inference system contains 243 rules with three member ship functions contained in ANFIS. Some of the rules of LD prediction defined by us in this system are shown in Sect. 4.2.2. The ANFIS architecture contains different layers as shown in Fig. 13. Each layer is associated with a particular function. The output of each layer is the input of next layer. The input, output and process of each layer are explained, in detail, in Table 7. The layers of neuro fuzzy system has embedded in the fuzzy system and the parameters are computed through the learning methods which are same as those used by neural networks. The mechanism of network of ANFIS is shown in Fig. 14. Details of input membership functions and output membership functions are given at Tables 3 and 4 respectively. The pattern of various membership functions are as shown in Figs. 8, 9 and 10.

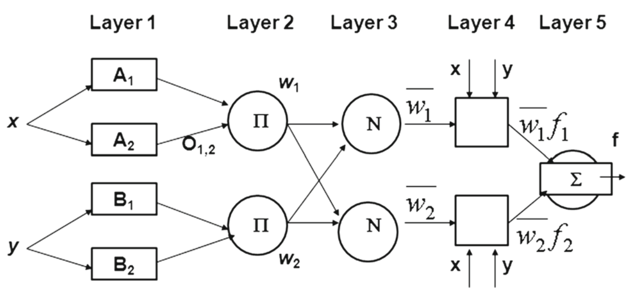

Fig. 13 ANFIS architecture

Table 6 Functions of ANFIS architecture layers

Layer	Functions
Layer 1	This layer accepts the inputs and transition it the input values to layer 2
Layer 2	This is the membership function node. Nodes in this layer correspond to linguistic variables of the input variable in layer 1. The membership value specifying the degree to which an input value belongs to a fuzzy set is calculated in this layer
Layer 3	This is the rule node. The output of each node in this layer is determined by the fuzzy AND operation. Here the product operation is utilized to determine the firing strength of each rule
Layer 4	Nodes in this layer are called consequent nodes. The input to a node in layer 4 is the output delivered from layer 3 and the other inputs are the input variables from layer 1
Layer 5	Each node in this layer corresponds to one output variable. The defuzzification is performed here

In the learning process, the parameters associated with the membership functions change, this change is an optimization essentially facilitated by a gradient vector. Using a combination of back-propagation and with the use of a least squares method, the fuzzy inference system is able to learn from the model data. A system is suited for modeling of non-linear systems by interpolating multiple linear models. This gradient vector provides a measure of how well the fuzzy inference system is modeling the input or output data for a given set of parameters. When the gradient vector is obtained, any of several optimization routines can be applied in order to adjust the parameters to reduce some error measure. This error measure is usually defined by the sum of the squared difference between actual and desired outputs.

The modeling concept used by ANFIS is similar to many system identification techniques. First, hypothesize a parameterized model structure relating inputs to membership functions to rules to outputs to membership functions, and so on. Next, collect input/output data in a form that will be usable by ANFIS for parameters is fully representative of the features of the data that the trained FIS is intended to model. In some cases however, data is collected using noisy measurements, and the training data cannot be representative of all the features of the data that will be presented to the model. In such situations, model validation is helpful.

Learning disability measurement tool using ANFIS method is a nonlinear regression, in which several input attributes such as signs and symptoms of LD are used to predict another continuous attribute as output variable in LD prediction. In the study, we are using 16 attributes which are the signs and symptoms of LD. About 1,020 real data sets, some of which contain missing values are used. In the preprocessing stage, missing values are imputed by closest fit algorithm and the 16 attributes are reduced to seven by PCA. Data scarcity and input space partitioning are the two problems to be taken care while designing the LD tool. For solving the problem of data scarcity, we have divided the data sets into two viz. test data and training data. The test data is used for model evaluation and the training data is used for model building. For solving the problem of input space partitioning, the attributes are reduced using PCA. The percentage of LD to be analyzed is determined by the attribute characteristics. The ANFIS function is used here for training the new system and that result is used to evaluate the system performance.

In our work, we are defining three output variables such as LD-Low, LD-Minor and LD-Major. Here, we are using the Sugano Fuzzy model, where each input is assumed to have two associated membership functions. The eight input space is partitioned into two overlapping fuzzy regions, each of which is governed by a fuzzy IF THEN rule. The surface area of membership function for LD prediction, obtained from the study, is shown in Fig. 15 below. It helps in determining how the inputs are related to output variable LD.

Table 7 Input and output processing of ANFIS architecture layers

Particulars	Layer 1	Layer 2	Layer 3	Layer 4	Layer 5
Functions	Calculate membership value for premise parameter	Firing strength of rule	Normalize firing strength	Consequent parameters	Overall output
Process	Output $O_{1,i}$ for node i = 1, 2 $O_{1,i} = \mu_{A_i}(x_1)$	Use T-norm (min, product, fuzzy AND, ...)	Ratio of ith rule's firing strength vs. all rules' firing strength	Takagi–Sugeno type output	$O_{5,1} = \sum_i \overline{w}_i f_i = \frac{\sum_i w_i f_i}{\sum_i w_i}$
				$O_{4,j} = \overline{w}_i f_i = \overline{w}_i (p_i x_1 + q_i x_2 + r_i)$	Output is linear in consequent parameters p, q, r
	Output $O_{1,i}$ for node i = 3, 4 $O_{1,i} = \mu_{B_{i-2}}(x_2)$ where 'A' is a linguistic label (small, large, ...) $\mu_A(x_1) = \frac{1}{1\mathrm{l}\left\lvert\frac{x_1-c_i}{a_i}\right\rvert^2}$	$O_{2,I} = w_i = \mu_{A_i}(x_1)\mu_{B_i}(x_2)$ (for i = 1, 2)	$O_{3,i} = \overline{w}_i = \frac{w_i}{w_1+w_2}$ (for i = 1,2)	Consequent parameters{p_i, q_i, r_i}	$= \frac{w_1}{w_1+w_2} f_1 + \frac{w_1}{w_1+w_2} f_2 = \overline{w}_1(p_1x_1 + q_1x_2 + r_1) + \overline{w}_2(p_2x_2 + q_2x_2 + r_2) = (\overline{w}_1x_1)p_1 + (\overline{w}_1x_2)q_1 + (\overline{w}_1)r_1 + (\overline{w}_2x_2)p_2 + (\overline{w}_2x_2)q_2 + (\overline{w}_2)r_2$
Node output	Membership value of input	Firing strength of rule	Normalized firing strengths	Evaluation of right hand side polynomials	Weighted evaluation of RHS polynomials

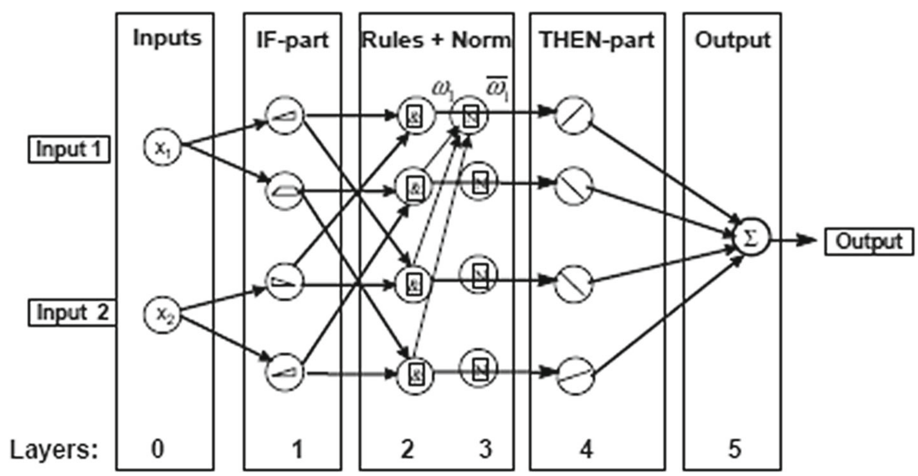

Fig. 14 ANFIS network

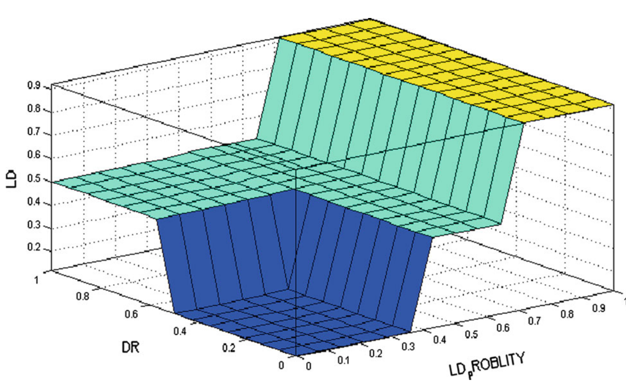

Fig. 15 Surface viewer of LD prediction

Before training a fuzzy inference system, we divided the data set into training and test sets. Training set is used to churn a fuzzy model, while the test set is used to determine when training should be terminated to prevent over fitting. In this study, the fuzzy models trained by the ANFIS command in the fuzzy logic tool box. ANFIS command utilizes iterative optimization technique to fine tune parameters and the training process. An efficient least square method is employed in the inner loop of ANFIS and the performance after the epoch is usually a good index of how well the fuzzy model will perform after further training. Here, we are performing ten fuzzy models each with a single epoch of ANFIS training. The results obtained after ten iterations, using the fuzzy logic tool

Table 8 ANFIS running results

1	1.18109e−005
2	7.70929e−006
3	9.42699e−006
4	6.36613e−006
5	7.41311e−006
6	1.44426e−005
7	7.74266e−006
8	1.35175e−005
9	7.67096e−006
10	9.03833e−006

Table 9 FISMAT information

Fismat 2 =	name: ‘anfis’
type: ‘sugeno’	And Method: ‘prod’
Or Method: ‘max’	defuzzMethod: ‘wtaver’
Imp Method: ‘prod’	Agg Method: ‘max’
input: [1×5 struct]	output: [1×1 struct]
rule: [1×243 struct]	—

Table 10 ANFIS information

Number of nodes: 524	Number of linear parameters: 1,458
Number of nonlinear parameters: 60	Total number of parameters: 1,518
Number of training data pairs: 200	Number of checking data pairs: 0
Number of fuzzy rules: 243	—

box in MatLab, are as given in Table 8 above. The step size decreases to 0.009000 after epoch 9. Designated epoch number reached at epoch 10 and ANFIS training completed there.

The new generated fuzzy inference system is used to evaluate the test cases. After running the program, the information about FISMAT and ANFIS is obtained as given in Tables 9 and 10 respectively above.

4.2.5 LD prediction results

After evaluating the performance of the system based on the new fuzzy inference system, we got the results as shown in Table 11 below.

4.3 Design of LD tool

Learning disability measurement tool is designed based on the MatLab GUI. The facilities present in MatLab GUI are used in this prediction tool. This GUI contains different buttons. These buttons are associated with a function. When the *FILE* button is activated, the type of data used for training is uploaded. The buttons available in the tool for different operations are preprocessing, training and testing. When the *preprocessing* button is activating, it will check the missing value positions and impute the appropriate values there. When the *training* button is activating, based on the selected classification algorithm, it will perform the ANN and ANFIS training. Finally, when the *test* button is activated it will perform the test procedure. The classification algorithm is then selected from the GUI. Then the next process of creation of a report of the student details takes place. In the test procedure based on ANN, there are 16 small buttons from which we can choose the input attributes based on the information collected from the child. After pressing the test button, corresponding information is displayed on the *LD prediction* result box. If we are choosing the ANFIS as the classification algorithm, another concept, viz. PCA, is included in the function of ANFIS program for reducing the number of attributes. Why we prefer to apply the PCA concept is that, some attributes have less contribution in LD prediction. We are using only seven attributes from the total of 16 by eliminating the others using PCA. When the test button is activate, the corresponding result will be displayed on the *LD prediction* (*Reduced Attribute*) result box. Accordingly, we are obtaining a result indicating the class of LD, viz, low, minor or major and its percentage, pertaining to a child. This result is very helpful for the children, parents and teachers to know the seriousness of LD and to recommend for providing remedial solutions. A typical result obtained in the GUI of the tool designed using the neuro fuzzy method is shown in Fig. 16 below. It can be seen from the figure that, the child is having LD, its class is low and percentage of LD is 60 %. The low LD is a normal state but it shows some similarities with that of children having LD. So we have to observe the percentage of LD present in that particular child, whether it is increasing or decreasing. Most probably this may reduce or disappear after some years. On further observations, if it is found increasing, it means that the LD present in the child tends to a minor state, which is at a higher grade than that of low state. As such, the developed tool is very helpful for the teachers as well as the parents for giving more concentrations/remedial measures to the children.

Another typical result for a child having LD class minor with LD 52 % is shown in Fig. 17. This minor LD of 52 % is somehow higher than that of low LD of 60 %. From the

Table 11 Test results of LD prediction

S. No.	Particulars	Numbers
1	Data set used for training	507
2	Number of data set used for testing	513
3	Number of instances correctly classified	513
4	Number of instances incorrectly classified	0

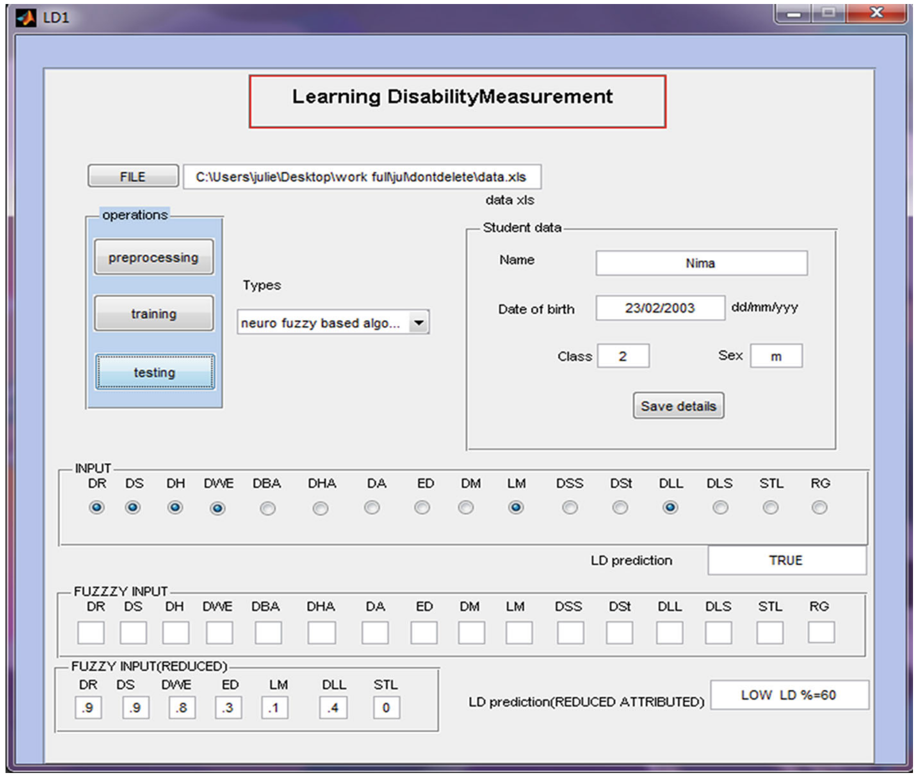

Fig. 16 A typical case of GUI output with class low and LD 60 %

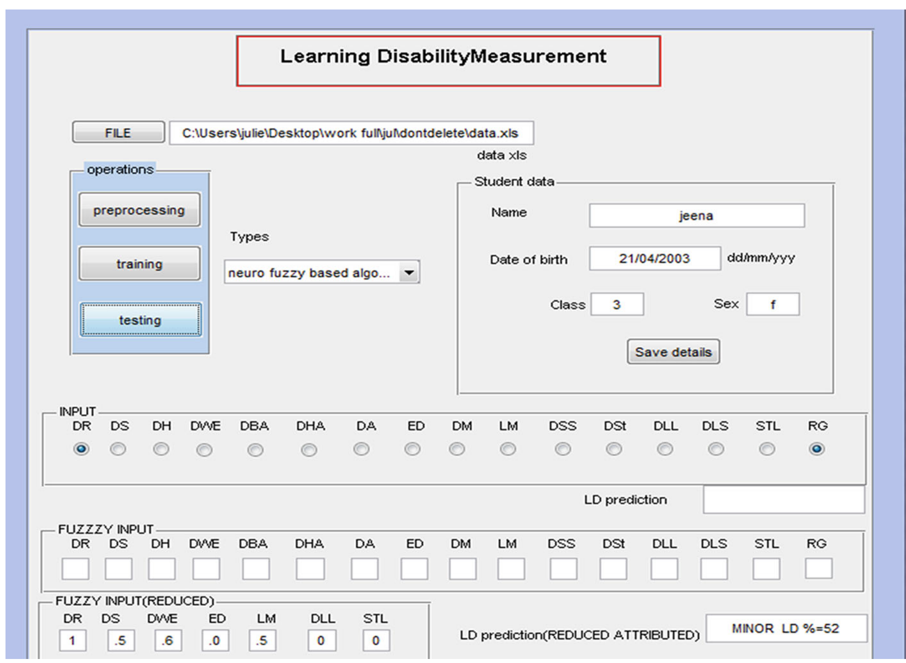

Fig. 17 A typical case of GUI output with class minor and LD 52 %

percentage of LD and its class, we can understand the depth of LD faced by the children. Hence, based on the class and percentage of LD, we can provide the appropriate remedial solution to the child.

The third typical case we have observed is that the class of LD major with LD 88 % as shown in Fig. 18. Even though, it is a very complicated case, if it is identified in an early stage we can provide a very good remedial solution and moral support to the children in early days. This will certainly improve the child's confidence and get a good change in his future. The GUI output obtianed in a typical case for a child having *no LD* is shown at Fig. 19 below.

As such, the designed tool is a very good scaling for the learning disability measurement. In the study, we are

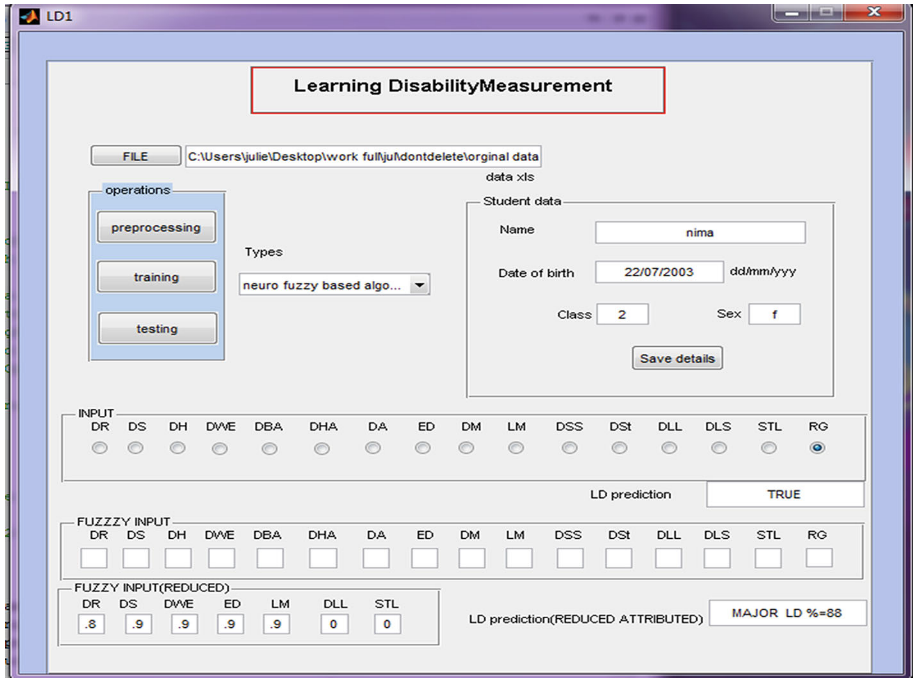

Fig. 18 A typical case of GUI output with class major and LD 88 %

using three input method—one for neural network, second for fuzzy input and third for fuzzy with reduced attribute method. If we are selecting the neuro fuzzy as the classifier then we are choosing the input method as fuzzy wth reduced attribute. From the above, it is seen that both the two basic algorithm ANN and ANFIS used for the prediction of LD in this study are found to be very accurate in classification. These two methods are applied in two ways. The results obtained from the two methods are shown. The results of neural network such as training process, performance, validation, etc. are provided. In the case of ANFIS, the results obtained contain the number of nodes, parameters, number of rules, number of training pairs, etc. The study reveals that, both these functions applied in MatLab GUI for the prediction of learning disabilities in children shows a high accuracy in LD prediction. Ultimately, the tool is found very effective in LD prediction and suggesting remedies for the help of children, parents and teachers.

5 Result analysis and findings

The designed tool for predicting the LD in school age children and determining the class of LD viz. low, minor or major and its percentage measurement using the two data mining methods ANN and ANFIS, is found very accurate. It is very helpful for the children, parents and teachers to identify the learning disabilities at an early stage. ANN with back propagation classification algorithm is used in classification of LD. It is important in this study that we are using data preprocessing for imputing missing attribute values. The missing values have very high impact on the classification algorithm. Obviously almost all the classification algorithms ignore the missing values. In our study, we have imputed the missing values in the data set using closest fit algorithm. After this, we found some attributes are irrelevant. So we have used PCA to reduce the 16 attributes to 7. These seven attributes are applied for ANFIS classification and we obtained very effective results. This approach of using neural network, ANN with data mining and ANFIS gives very accurate and strong results in prediction of learning disabilities in children. The result of our experimental analysis reveals that the significant attributes of LD and predicts the percentage of LD in each child. We have implemented our proposed approach in MatLab. In this study, we have used 1,020 real world dataset which were collected from LD clinics and schools in and around Cochin, India. In this study, based on the PCA method, we have extracted the most important attributes of LD from the 16 attributes. After the data preprocessing, all the predefined classes are obtained or trained based on ANFIS. The 16 attributes used in the form of a check list for the assessment of LD are the very signs and symptoms of LD that are using in LD clinics. In all our previous works on LD prediction (Julie and Balakrishnan 2010a,c, 2011a,c), we could not have to find the strength or percentage of LD and its class. This problem has been well catered by the present tool designed by us and also stores the student details in the student data base which can be retrieved as and when required. It is found that from the 16 attributes, some have a very good role in the prediction of LD. Some of the children have one or two LD attributes present in them. So, in normal case, we are thinking that their case is very mild. But in real situation, the

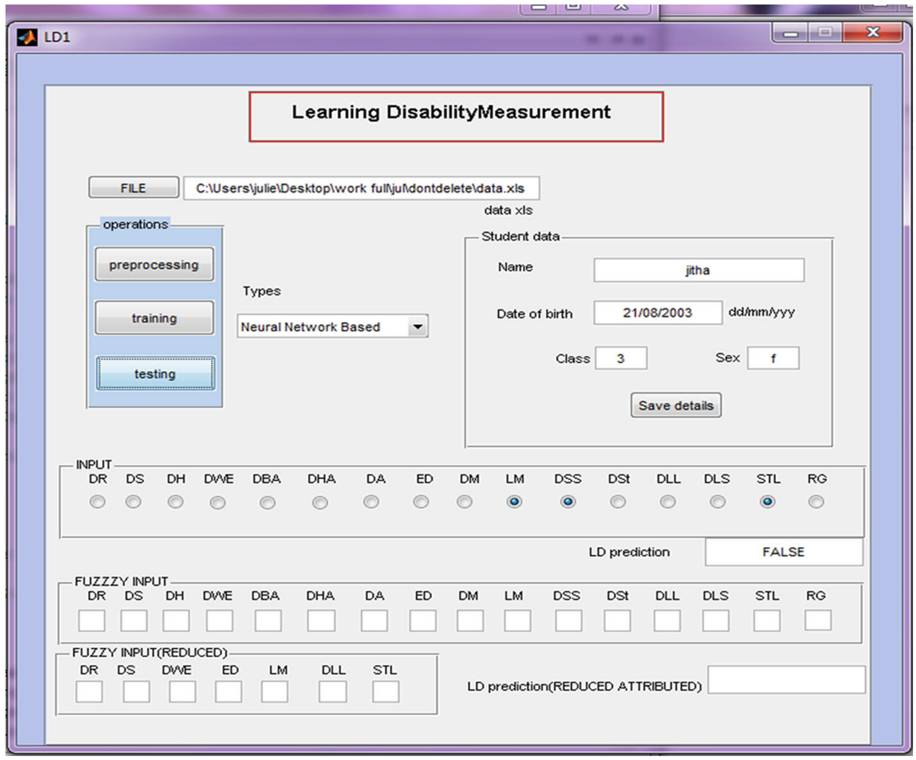

Fig. 19 A typical case of GUI output with no LD

Table 12 Comparison of classification results

Parameters	Classifiers					
	J48	Naive Bayes	SVM	ANN	New ANN	New ANFIS
Number of instances used for classification	513	513	513	513	513	513
Correctly classified instances, nos. (%)	500 (97.47)	426 (83.04)	502 (97.86)	502 (97.86)	508 (99.03)	513 (100)
Incorrectly classified instances, nos. (%)	13 (2.53)	87 (16.96)	11 (2.14)	11 (2.14)	5 (0.97)	0 (0)
Time taken to build a model (in s)	0.08	0.06	3.92	18.03	1.28	1.28

contribution of these one or two attributes may be high in occurring LD. These things, which are not at all least in anyway, can be easily determined by this tool rather manually. With the help of the designed prediction system, we can also predict the different risk level of LD. The results obtained from the study can be shown in the GUI designed by us. Some researchers are doing the identification of dyslexia, a type of LD, in children (Maitrei and Prasad 2010). Compared to their works, our work is entirely different and our results shows more effectiveness in prediction of LD and its percentage. The importance in our work is that, we have precisely performed the data preprocessing and data reduction in LD prediction. The results are very beneficial to the parents, teachers and the institutions. Because they are able to diagnose the child's problem at an early stage and can go for the proper treatments/counseling at the correct time so as to avoid the academic and social losses.

6 Comparison of results

The results obtained from this study are compared with the results of our other similar studies conducted in the data mining tool weka based on J48, Naive Bayes, Support Vector Machines (SVM) and Muliti Layer Perceptron classifiers (Julie and Balakrishnan 2010b, 2011c). The comparison of the results is shown in Table 12 above. From this comparison, we can see that the tool designed by us based on New ANN and New ANFIS is better in terms of classification and accuracy. In the data preprocessing stage, we applied some methods such as closest fit algorithm for imputing missing values and PCA for attribute reduction in New ANN and New ANFIS. From the comparison of results, it can be seen that, these preprocessing methods are effectively improves the accuracy of ANN and ANFIS in prediction of LD.

7 Conclusion and future works

In this paper, we have developed a new approach in two data mining methods, ANN and ANFIS, to effectively and accurately predict the LD in school age children, its class like low, minor or major and LD percentage. This study mainly focuses on designing a tool for the measurement of LD based on data mining concepts. Accuracy of decision-making can be improved by using our good methods of missing value imputing, attributes reduction and applying the classification. While applying on the same dataset, it shows an accuracy of about 100 %. This study has been carried out on 1,020 real world data sets. 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. The work is a study of the proposed approach by applying it to large datasets and analyzing the completeness and effectiveness of the LD tool designed based on ANN and ANFIS methods in data mining.

The designed tool in data mining with ANN and ANFIS methods for LD measurement shows that, it is better than other classifiers such as J48, Naïve Bayes, SVM and MLP in terms of efficiency and accuracy. Imputing missing values and dimensionality reduction using PCA are having good roles in predicting LD. The results from the experiments on these dataset suggest that the designed LD tool gives more precise results for classification and prediction of LD. Our future research work will focus on fuzzy-decision tree for exploring the possibilities of getting much more insights in prediction of learning disabilities.

References

Ajith A (2001) Neuro-fuzzy systems: state-of-the-art modeling techniques, connectionist models of neurons, learning processes, and artificial intelligence. In: Mira J, Prieto A (eds) Lecture notes in computer science, vol. 2084. Springer-Verlag, Germany, pp 269–276 (ISBN 3540422358, Spain)

Blackwell Synergy (2007) Learning disabilities and research practice, vol 22

Chapple M (1995) About.com guide. http://databases.about.Com/od/datamining/g/classification.htm. Accessed 2 Feb 2013

Crealock C, Kronick D (1993) Children and young people with specific learning disabilities. In: Guides for special education, vol 9, UNESCO

Frawley, Piaatetsky (1996) Shaping knowledge discovery in database; an overview. The AAAI/MIT press, Menlo Park

Gray R, George PE, Frank C (1999) Research paper on application of neuro fuzzy system to behavioral representation in computer generated forces. http://citeseer.ist.psu.edu/george99appilcation.html. Accessed 2 Feb 2013

Grzymala-Busse JW, Grzymala-Busse WJ, Goodwin LK (2001) Coping with missing attribute values based on closest fit in pre-term birth date: a rough set approach. Int J Comput Intell 17(3):425–434

Han J, Kamber M, Pei J (2011) Data mining-concepts and techniques, 3rd edn. Elsevier Publishers, Morgan Kaufmann (ISBN: 978-93-80931-91-3)

Hsinchun C, Fuller SS, Carol F, William H (2005) Chapter 1: Knowledge discovery in data mining and text mining in medical informatics, pp 3–34

Jang JSR, Sun CT, Mizutani E (2008) Neuro-fuzzy and soft computing. In: A computational approach to learning and machine intelligence. PHI Learning Pvt. Ltd., New Delhi

Julie MD, Balakrishnan K (2009) 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, pp 202–207 (ISBN 10:0230-63915-1, ISBN 13:978-0230-63915-7)

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

Julie MD, Balakrishnan K (2010b) Machine learning approach for prediction of learning disabilities in school age children. Int J Comput Appl 9(10):7–14 (ISSN-0975-8887)

Julie MD, Balakrishnan K (2010c) Paper on prediction of learning disabilities in school age children using decision tree. In: Meghanathan N, Boumerdassi S, Chaki N, Nagamalai D (eds) Proceedings of the international conference on recent trends in network communications—CCIS, vol 90(3), Springer-Verlag, Berlin/Heidelberg, pp 533–542. doi:10.1007/978-3-642-14493-6_55 [ISSN:1865-0929 (print) 1865-0937 (online), ISBN 978-3-642-14492-9 (print) 978-3-642-14493-6 (online)]

Julie MD, Balakrishnan K (2011a) Prediction of key symptoms of learning disabilities in school-age children using rough sets. Int J Comput Electr Eng 3(1):163–169

Julie MD, Balakrishnan K (2011b) Prediction of learning disabilities in school-age children using SVM and decision tree. Int J Comput Sci Inform Technol 2(2):829–835 (ISSN 0975-9646)

Julie MD, Balakrishnan K (2011c) Attribute reduction and missing value imputing with ANN: prediction of learning disabilities. Int J Neural Comput (Springer-Verlag, London). doi:10.1007/s00521-011-0619-1

Julie MD, Pramod KV (2008) Paper on prediction of learning disabilities in school age children using data mining techniques. In: Thrivikram T, Nagabhushan P, Samuel MS (eds) Proceedings of AICTE sponsored national conference on recent developments and applications of probability theory, random process and random variables in computer science, pp 139–146

Maitrei K, Prasad TV (2010) Paper on identifying dyslexic students by using artificial neural networks. In: Proceedings of the world engineering congress on engineering, vol 1 [ISSN : 2078-0966 (online)]

Rahib A, Vasif H et al (2005) Electricity consumption prediction model using neuro-fuzzy system. World Acad Sci Eng Technol, vol 8

Rod P (Secretary) (2002) US Department of Education. In: 24th 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

Sally JC, Holmes G (1999) Developing innovative applications in agricultural using data mining. In: Proceedings of the Southeast Asia regional computer confederation conference

Savcovic S (1994) Neural networks for process analysis and optimization: modeling and applications. Comput Chem Eng 18(11-12):1149–1155

Shantakumar BP, Kumaraswamy YS (2009) Intelligent and effective heart attack prediction system using data mining and ANN. Eur J Sci Res 31(4):642–656 (ISSN 1450-216X)

Sidda RB, Suresh Kumar J, Vijaya Kumar Reddy K (2009) Prediction of roughness in turning using adaptive neuro-fuzzy inference system. Jordan J Mech Ind Eng 3(4):252–258

Simona D, Ioan F et al (2008) An application of neuro fuzzy modeling to prediction of some incidents in an electrical energy distribution centre. Int J Comput Commun Control (ISSN 1841-9836, III)
Sivarao et al (2009) GUI based Mamdani fuzzy inference system modeling to predict surface roughness in laser machining. Int J Electr Comput Sci 9(9):281–288
Tan PN, Steinbach M, Kumar V (2008) Introduction to data mining, low price edition. Pearson Education Inc (ISBN 978-81-317-1472-0)
Tarafdar HM, Najafi S (2005) Application of neuro fuzzy dynamic programming to improve the reactive power and voltage profile of a distribution substation. World Acad Sci EngTechnol 6:126–129
Weixiang Z, Abhinav B et al (2008) Machine learning: a crusial tool for sensor design, open access algorithms. doi:10.3390/a1020130 (ISSN 1999-4893)
Xiangun N (2008) Research of data mining based on neural network. World Acad Sci Eng Technol, vol 39