

Performance Improvement of Fuzzy and Neuro Fuzzy Systems: Prediction of Learning Disabilities in School-age Children

Julie M. David

Dept. of Computer Applications, MES College, Marampally, Aluva, Cochin- 683 107, India
E-mail: julieeldhosem@yahoo.com

Kannan Balakrishnan

Dept. of Computer Applications, Cochin University of Science & Technology, Cochin - 682 022, India
E-mail: mullayilkannan@gmail.com

Abstract— Learning Disability (LD) is a classification including several disorders in which a child has difficulty in learning in a typical manner, usually caused by an unknown factor or factors. LD affects about 15% of children enrolled in schools. The prediction of learning disability is a complicated task since the identification of LD from diverse features or signs is a complicated problem. 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. The aim of this paper is to develop a new algorithm for imputing missing values and to determine the significance of the missing value imputation method and dimensionality reduction method in the performance of fuzzy and neuro fuzzy classifiers with specific emphasis on prediction of learning disabilities in school age children. In the basic assessment method for prediction of LD, checklists are generally used and the data cases thus collected fully depends on the mood of children and may have also contain redundant as well as missing values. Therefore, in this study, we are proposing a new algorithm, viz. the correlation based new algorithm for imputing the missing values and Principal Component Analysis (PCA) for reducing the irrelevant attributes. After the study, it is found that, the preprocessing methods applied by us improves the quality of data and thereby increases the accuracy of the classifiers. The system is implemented in Math works Software Mat Lab 7.10. The results obtained from this study have illustrated that the developed missing value imputation method is very good contribution in prediction system and is capable of improving the performance of a classifier.

Index Terms— ANFIS, Data Mining, FIS, Learning Disability, Membership Function

I. 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 [1,2]. 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 [3]. In recent years the sizes of databases has increased rapidly. This has led 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 [4]. 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 [5].

Learning disability 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 [6]. The learning disability affected children are neither slow nor mentally retarded [7]. 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 learning disability is often wrongly labelled as being smart but lazy. The learning disability can cause a child to have trouble in learning and using certain skills. The skills most often affected are: reading, writing, listening, speaking, reasoning and doing math [6]. 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 [5,8]. They are life-long. However, children with LD can be high achievers and can be taught ways to get around the learning disability. With the right help, children with LD can and do learn successfully.

Missing attribute values commonly exist in real world data set. They may occur from very beginning stage viz. data collecting process and during redundant diagnose tests, unknown data, etc. Discarding all data containing the missing attribute values cannot fully preserve the characteristics of the original data [9]. Various approaches on how to cope with the missing attribute values have been proposed in the past years. Values were introduced, such as selecting the most common attribute value, the concept most common attribute value, assigning all possible values of the attribute restricted to the given treating missing attribute values as special values, event covering method and so on [10]. Fuzzy and neuro fuzzy classifiers are two well-known classifiers, both handles missing values in the data set to a certain extent. However, in this paper, we have studied the improving performances of these two classifiers while imputing missing values and removing the redundant attributes by applying the well-known algorithm PCA.

A fuzzy set does not have specific and limited boundaries; the distinction between belonging or not does not exist, but a degree of pertinence. Fuzzy logic is a system of concepts, principals and methods of dealing with modes of reasoning that are approximate rather than exact [11]. It is particularly good at handling uncertainty, vagueness and imprecision and it is especially useful where a problem can be described linguistically. In fuzzy logic the degree of truth of a statement can range between 0 and 1.

Data preprocessing is a broad area and consists of a number of different strategies and techniques that are interrelated in complex ways [12]. We have applied different pre-processing methods in other classification methods viz. decision tree, neural network and SVM and it is seen that the results are improved comparing with the existing algorithms [13]. The fuzzy logic system is handling the missing values in a better way as compared with the other classifiers. This was the motivation in conducting data pre-processing in fuzzy and neuro fuzzy systems. In exploring the possibilities of getting an improvement in accuracy, we have applied the new correlation based algorithm for

imputing missing values and dimensionality reduction using PCA in the pre-processing stage. The pre-processed data, by way of imputing missing values and dimensionality reduction, as explained above, is then applied to the classifiers, fuzzy model and Neuro fuzzy model for evaluating the consistency of data.

The purpose of this paper is to develop a new algorithm for imputing missing values and to determine the significance of the missing value imputation method and dimensionality reduction method (PCA) in the performance of fuzzy and neuro fuzzy classifiers, with an emphasis in data mining, for prediction of LD in school age children and also to determine the importance of preprocessing in classification. The remaining paper is organized as follows. The literature review is explained in Section 2. Section 3 describes elaborately about LD. The design of the proposed tool using ANN and ANFIS methods is explained in Sections 4. The result analysis & findings and comparison of results are explained in Section 5 and Section 6 respectively. Finally, Section 7 deals with conclusion and future research works.

II. Literature Review

Studies in the field of learning disabilities are conducting in 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.

Kenneth A. Kavale, studied about identifying specific learning disability in the year 2005 [14]. In this study he has developed an alternative model for making decision about the presence or absence of special learning disabilities. In 2010, Benjamin J. Lovett conducted a study on extended time testing accommodations for students with disabilities- answers to five fundamental questions [15]. 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. Noona Kiuru et. al., in the year 2011, conducted a study [16] on 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.

Many other researchers focused on the topic of imputing missing values. Chen and Chen [17] 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 [18] 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 [19,20,21]. 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.

Among missing value imputation methods, there are also many existing statistical methods. Statistics based methods include linear regression, replacement under same standard deviation, and mean-mode method. But these methods are not completely satisfactory ways to handle missing value problems. Magnani [22] has reviewed the main missing data techniques (MDTs), and revealed that statistical methods have been mainly developed to manage survey data and proved to be very effective in many situations. However, the main problem of these techniques is the need of strong model assumptions. Other missing data imputation methods include a new family of reconstruction problems for multiple images from minimal data [23], a method for handling inapplicable and unknown missing data [24], different substitution methods for replacement of missing data values [25], robust Bayesian estimator [26], and nonparametric kernel classification rules derived from incomplete or missing data introduced by Pawlak in 1995 [27]. Same as the methods in machine learning, the statistical methods, which handle continuous missing values with missing in class label are very efficient, are not good at handling discrete value with missing in conditional attributes.

There is no literature available on the study of learning disability predictions using fuzzy or rough set models. However, many researchers are conducted studies on various other areas using these models. A comparative study of fuzzy sets and rough sets was conducted by Yao, Y.Y in 1998 [28]. Pawlak, Z. in 1991 and Klir, et. al in 1995 pointed out that there are many formulations and interpretations of theories of fuzzy sets and rough sets [29,30]. Chanas and Kutcha in 1992 studied on further remarks on the relation between rough and fuzzy sets, [31] and opined that the fuzzy set theory deals with the ill-definition of the boundary of a class through a continuous generalization of set characteristic functions. The indiscernibility between objects is not used in fuzzy set theory. Zadeh in 1992 stated that a fuzzy set may be viewed as a class with unsharp boundaries, whereas a

rough set is a crisp set which is coarsely described [32]. On a study on inducing fuzzy models for student classification, in the year 2006, Ossi Nykanen, [33] found that the quality of the fuzzy system is determined by the induced fuzzy model and the rule heuristics. His approach is quite general and suits the needs of various application domains that require inducing fuzzy classifications and decorating archives with intuitive labels. The application of fuzzy logic to the construction of the ranking function of information retrieval systems was studied by Rubens, N.O., in 2006 [34].

Fuzzy logic has not been applied to defining ranking function directly; however, fuzzy set model has been used to define fuzzy queries [35], fuzzy relationships between query terms and documents [36,37]. Each query term defines a fuzzy set and each document has a degree of membership in the corresponding set. The fuzzy set model approach is not popular among the information retrieval community and has been discussed mainly in the literature dedicated to fuzzy theory [38].

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 [39]. 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 [40]. 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 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 learning disability. 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. A learning disability cannot be cured or fixed [41]. 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 [42].

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 [13]. Social support can be a crucial component for children with learning disabilities in the school system and should not be overlooked in the intervention plan. With the right support and intervention, children with learning disabilities can succeed in school and go on to be successful later in life.

LDs affect about 15% of all children enrolled in schools. Almost 3 million children (ages 6 through 21) have some form of a learning disability and receive special education in school [7]. In fact, over half of all children who receive special education have a learning disability [5]. The problems of children with specific learning disabilities have been a cause of concern to parents and teachers for some time. Pediatricians are often called on to diagnose specific learning disabilities in school-age children. Learning disabilities affect children both academically and socially [6]. 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 [8]. 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 from no services to their universal provision. This unevenness in intervention services is tragic since most children with learning disabilities who receive sufficient, knowledgeable remediation can proceed through the school system and attain jobs that range from professor to laborer. Conversely, if they are not helped, the possibility of adjustment of problems arising is considerable. As our world becomes more complex, the knowledge base increases and the concepts more abstract, an increasing number of children will experience difficulty and be assumed to have learning to our collective lives is not forfeited.

There is no one sign that shows a child has a learning disability [43]. 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 learning disability. 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. 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 [43]. The discrepancy clause states there is a significant disparity between aspects of specific functioning and general ability; the exclusion clause states the disparity is not primarily due to intellectual, physical, emotional, or environmental problems; and the etiologic clause speaks to causation involving genetic, biochemical, or neurological factors.

The most frequent clause used in determining whether a child has a learning disability is the difference between areas of functioning. When a 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 [4,44]. 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 [1]. 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 [45]. 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 [46]. 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. Proposed Methodology

The aim of this paper is to determine the significance of new missing value imputation method and dimensionality reduction method through the performance of a two well known classifiers, viz. fuzzy and neuro fuzzy.

He major issue found in the study of prediction of LD in children is the failure of the classifier in handling the missing values in the datasets [13]. The missing values contribution may be some times very

important and significant. The second issue noticed is that some of the attributes in the check list have less contribution in LD prediction. So we have to reduce the number of attributes for improving the performance of the classifier. Reducing the number of attributes is very effective and that will help to reduce the time taken for constructing the model. The findings from that studies shows that there is no solution in the case

of missing values present in the data set. Also some attributes are unwanted and hence have no contributions in predicting the LD. Hence, as a solution to overcome these problems, here we propose a new algorithm based on correlation. The system flow chart of the proposed methodology is shown in Fig. 1.

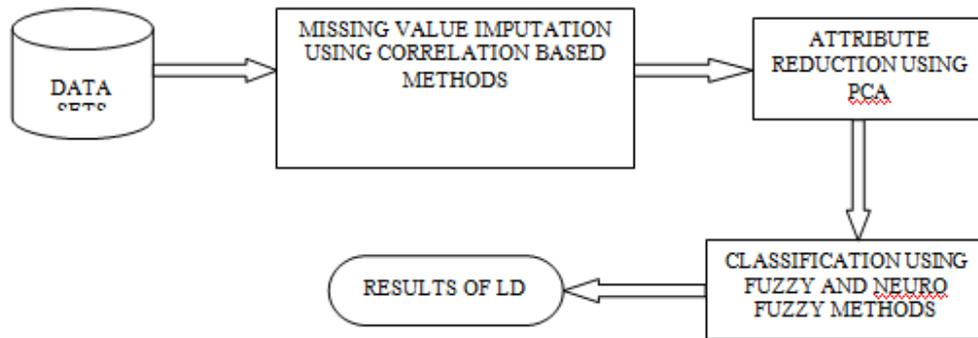


Fig. 1: System Flowchart

4.1 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 learning disability. 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 analyzing complex decision tables associated with the learning disabilities. Our goal is to determine the significance of new missing value imputation method and dimensionality reduction method through the performance of a two well known classifiers, fuzzy and neuro fuzzy. A checklist is used 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. 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 below. In this study, we are used 1020 real world datasets collected by informal assessment methods. The data is mainly collected from learning disability clinics and schools in and around Cochin, India and from the Department of Information Management, National Changhua University of Education, Changhua, Taiwan, Republic of China. For choosing the data, a check list which containing the same signs and symptoms of LD, is used. After conducting direct interview with the children, with the help of teachers and/or parents as required, the check list is filled, which is ultimately used for preparing the data for conducting the study. The

checklist contains numerous amount of attributes related to LD. Various assessment methods are watched in consultation with specialists engaged in the profession and the data set is finalized with their help.

Table 1: List of Attributes

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

4.2 Imputing Missing Values

Missing values imputation is an actual yet challenging issue confronted in machine learning and data mining [47,48]. Missing values may generate bias and affect the quality of the supervised learning process or the performance of classification algorithms [49,50]. However, most learning algorithms are not

well adapted to some application domains due to the difficulty with missing values as most existed algorithms are designed under the assumption that there are no missing values in datasets. That implies that a reliable method for dealing with those missing values is necessary. Generally, dealing with missing values means to find an approach that can fill them and maintain, or approximate as closely as possible, the original distribution of the data.

Missing values may appear either in conditional attributes or in class attribute or target attribute. There are many approaches to deal with missing values described in [51], for instance: (a) ignore objects containing missing values, which usually lost too much useful information; (b) fill the missing value manually, which is time-consuming and expensive in cost, so it is infeasible in many applications; (c) substitute the missing values by a global constant or the mean of the objects, which assumes that all missing values are with the same value, probably leading to considerable distortions in data distribution; and (d) Get the most probable value to fill in the missing values. However, Han, et al. [51] and Zhang, et al. [48] think that the method of imputation is a popular strategy. In comparison to other methods, it uses as more information as possible from the observed data to predict missing values.

Traditional missing value imputation techniques can be roughly classified into parametric imputation (e.g., the linear regression) and non-parametric imputation (e.g., non-parametric kernel-based regression method [52,53,54], Nearest Neighbor (NN) method [55,56]). The parametric regression imputation is superior, if a dataset can be adequately modeled parametrically, or if users can correctly specify the parametric forms for the dataset. For instance, the linear regression methods usually can treat well the continuous target attribute, which is a linear combination of the conditional attributes. However, when we don't know the actual relation between the conditional attributes and the target attribute, the performance of the linear regression for imputing missing values is very poor. In real application, if the model is mis specified, in fact, it is usually impossible for us to know the distribution of the real dataset, the estimations of parametric method may be highly biased and the optimal control factor settings may be miscalculated.

Non-parametric imputation algorithm, which can provide superior fit by capturing structure in the dataset, note that a mis-specified parametric model cannot, offers a nice alternative if users have no idea on the actual distribution of a dataset. The NN method is regarded as one of non-parametric techniques used to compensate for missing values in sample surveys [55]. And it has been successfully used in, for instance, U.S. Census Bureau and Canadian Census Bureau. What's more, using a non-parametric algorithm is beneficial when the form of relationship between the conditional attributes and the target attribute is not known a priori

[56].

While nonparametric imputation method is of low-efficiency, the popular NN method faces two issues: (i) each instance with missing values requires the calculation of the distances from it to all other instances in a dataset; and (ii) there are only a few random chances for selecting the nearest neighbor. Before the data is analyzed, it has to be preprocessed in order to increase the accuracy of the output and to facilitate the learning process. This is a critical operation. Data pre-processing is the step to be applied to make the data more suitable for data mining. It is a broad area and consists of a number of different strategies and techniques that are interrelated in complex ways [57]. The different process exist in the preprocessing stage are dimensionality reduction, feature subset selection, removal of noise from the data and imputing the missing values. In the case of LD prediction, using of checklists is an informal assessment method. The information collected by using check lists solely depends on the mood of child. So we cannot expect to obtain a full filled checklist. Incomplete data can occur for a number of reasons. On assessment of learning disability, relevant data may not be recorded due to misunderstanding. Also, incomplete, noisy and inconsistent data are commonplace properties of large real world databases and data warehouses [51]. Many data mining approaches are usually ignoring either the case having an attribute with missing values or the attribute having the missing value. We have applied the closest fit algorithm and correlation based new algorithm for imputing the missing values.

4.3 Development of Correlation Based New Algorithm

In medical diagnosis system, symptoms of diseases are correlated, but we cannot predict which symptoms are more related. In the case of learning disability also, different symptoms are there. If one symptom's value is missing from the data set, we cannot say the contribution of that symptom's value in the prediction of LD. Missing value can be imputed by using different methods. Compared to other methods [13,58], the correlation based new algorithm developed, shows a good performance. In this new method of missing value imputing developed, we proposed to find out all the missing places. Then the missing place's left and right values are determined. Using the correlation coefficient equation, the correlation factor is then determined, by putting the left value of missing attribute in the missing places, and finds the correlation coefficient. Similarly put the right value in the missing place and repeat the process. Compare the right and left correlation coefficient to determine which one is higher. That will be imputed in the missing place. In the case where, the right or left positions of missing place may occur as vacant, the next neighboring position where

an attribute value exists is determined and the value at this neighboring position is taken for further process. The algorithm of this new method is given in Fig. 2 and the detailed computing steps of the new algorithm are given there under.

```

Algorithm:
  Procedure: correlation based missing value
  imputing algorithm
  Input: Missing-valued dataset S,
  Output: Complete dataset S';
  1. Read the data set S;
  2. For each missing -valued instance Ai in data
  set S
  3. Compute the correlation coefficient CRC with
  left and right attribute of missing value CRCL and
  CRCR respectively.
  4. If CRCL > CRCR,
      missing place is filled with left attribute value
  of missing value
  5. Else
      filled with right attribute value of missing
  value
  End.
    
```

Fig. 2: Algorithm of Correlation Based New Missing Value Imputing Method

Computing steps:

- Read the data set
- If $a_i = null$ then
 - $a_{i_r_R} = corr_coeff_R$
 - $a_{i_r_L} = corr_coeff_L$
- If $a_{i_r_R} > a_{i_r_L}$
 - $a_i = a_{i+1}$
- else
 - $a_i = a_{i-1}$
- elseif \ if missing value have only left value \
 - $a_i = a_{i-1}$
- else
 - $a_i = a_{i+1}$
 - \if both sides of the missing value have null
 values\
 - If $a_{i+1} \ \&\& \ a_{i-1} = null$ then
 - $posn_ai = find \ isnan(a_i)$
 - $a_{i-r-R} = corr_coeff \ of \ posn_ai_R$
 - $a_{i-r-L} = corr_coeff \ of \ posn_ai_L$
 - If $a_{i-r-R} > a_{i-r-L}$ then
 - $a_i = posn(a_{i_r_R})$
 - else
 - $a_i = posn(a_{i-r-L})$

4.4 Dimensionality Reduction

Summarization of data with many variables by smaller set of derived variables is called dimensionality reduction. In the case of learning disability data set we aim to reduce the number of attributes or the symptoms

of LD. Among the large number o of symptoms of LD, the unwanted and irrelevant symptoms or attributes of LD are to be removed.

Principal Component Analysis is a method of dimensionality reduction. The data to be reduced consists of tuples or data vectors described by n attributes or dimensions are called PCA [51]. The PCA searches for k n-dimensional orthogonal vectors, that can be used to represent the data where $k \leq n$. 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 k orthonormal vectors that provides a basis for the normalized input data. These are unit vectors that each point in a direction perpendicular to the others. These vectors are referred to as the principal components; and (iii) the principal components are sorted in order of decreasing strength.

In our study, we have used the LD data sets, having 16 attributes. By applying the PCA, the number of attributes is further reduced to seven. After applying on the 1020 datasets, we got the ranked reduced attributes as shown in Table 2. Using conjunction with a ranker search, dimensionality reduction is accomplished by choosing enough eigen vectors to account for some percentage (95%) 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".

Table 2: Ranked Reduced Attributes

Sl. No.	Attribute	Rank
1	DR	0.6846
2	DS	0.5412
3	DH	0.4136
4	DWE	0.3337
5	DBA	0.2683
6	DHA	0.2116
7	DA	0.1657

4.5 Modified Data Pre-Processing and Performance Evaluation

This section consists of two parts. The first part of study is on modified data pre-processing with fuzzy, the second part is on modified data pre-processing with neuro fuzzy. After this study, it is found that these developed systems in data pre-processing shows very accurate results. The performance of these new systems

is compared with that of MLP and decision tree classifiers and models. In the case of percentage of correctly classified instances, these new systems have better results and much lesser time has taken for building the models.

4.5.1 Fuzzy model with data pre-processing

Fuzzy logic was invented by Zadeh [59] in 1965 for handling imprecise and uncertain knowledge in real world applications. It has proved to be a powerful tool for decision-making, and to handle and manipulate imprecise and noisy data. The notion central to fuzzy systems is that truth values or membership values are indicated by a value on the range 0 to 1, with 0 representing absolute Falseness and 1 representing absolute Truth. A fuzzy system is characterized by a set of linguistic statements based on expert knowledge. The expert knowledge is usually in the form of if-then rules. A fuzzy set X in Y is characterized by a membership function which is easily implemented by fuzzy conditional statements. If the antecedent is true to some degree of membership, then the consequent is also true to that same degree [60]. In a fuzzy classification system, a case or an object can be classified by applying a set of fuzzy rules based on the linguistic values of its attributes. Every rule has a weight, which is a number between 0 and 1, and this is applied to the number given by the antecedent. It involves evaluating the antecedent, fuzzifying the input and applying any necessary fuzzy operators. To build a fuzzy classification system, the most difficult task is to find a set of fuzzy rules pertaining to the specific classification problem. A fuzzy inference system is a rule-based system that uses fuzzy logic, rather than Boolean logic, to reason about data. Its basic structure includes four main components (1) a fuzzifier, which translates crisp (real-valued) inputs into fuzzy values; (2) an inference engine that applies a fuzzy reasoning mechanism to obtain a fuzzy output; (3) a defuzzifier, which translates this latter output into a crisp value; and (4) a knowledge base, which contains both an ensemble of fuzzy rules, known as the rule base, and an ensemble of membership functions known as the database. The decision-making process is performed by the inference engine using the rules contained in the rule base. These fuzzy rules define the connection between input and output fuzzy variables.

Fuzzy set concepts are often used to represent quantitative data and membership functions in intelligent systems because of its simplicity and similarity to human reasoning. They have been applied to many fields such as manufacturing, engineering, diagnosis and economics.

4.5.1.1 Fuzzy classification

Fuzzy classification is the process of grouping elements into a fuzzy set whose membership function

defined by the truth value of a fuzzy propositional function [61,62]. It corresponds to a membership function that indicates whether an individual is a member of a class. The steps of fuzzy classification are;

- 1) Define the linguistic variables and terms (initialization)
- 2) Construct the membership functions (initialization)
- 3) Construct the rule base (initialization)
- 4) Convert crisp input data to fuzzy values using the membership functions (fuzzification)
- 5) Evaluate the rules in the rule base (inference)
- 6) Combine the results of each rule (inference)
- 7) Convert the output data to non-fuzzy values (defuzzification)

4.5.1.2 Fuzzification of attributes

Fuzzification is a process of making crisp quantity fuzzy [63]. The form of uncertainty or vagueness is probably fuzzy and can be represented by a membership function. The first step is to take the crisp numerical values of the inputs and determine the degree to which they belong to each of the appropriate fuzzy sets via membership functions [64, 34]. A crisp set of input data are gathered and converted to a fuzzy set using fuzzy linguistic variables, fuzzy linguistic terms and membership functions. This step is known as fuzzification. Afterwards, an inference is made based on a set of rules. Lastly, the resulting fuzzy output is mapped to a crisp output using the membership functions, in the defuzzification step.

For example: $DR = 0.4$, this would be translated into 0.4 degree of membership in fuzzy set "low", 0.6 degree of membership in fuzzy set "minor" and 0.8 degree of membership in fuzzy set "major." Same procedure would be applied to all of the inputs. The system flow chart of fuzzification of attributes is shown in Fig. 3.

4.5.1.3 Linguistic variables

Linguistic variables are the input or output variables of the system whose values are words or sentences from a natural language, instead of numerical values [65,66]. A linguistic variable is generally decomposed into a set of linguistic terms. In this system, the symptoms of learning disability like DR, DHA, DS, etc. and LD probability are the linguistic variables. These attributes can be quantified as low, minor and major, which shows the significance of each attribute in LD prediction.

4.5.1.4 Membership functions

Memberships are used in the fuzzification and defuzzification steps of a fuzzy logic system, to map the non-fuzzy input values to fuzzy linguistic terms and vice versa. A membership function is used to quantify a linguistic term. For instance, membership functions for the linguistic variables are plotted in terms of input variable DS, input variable LD and output variable. These are given in Fig. 4, 5 and 6 respectively. The important characteristic of fuzzy logic is that a numerical value does not have to be fuzzified using only one membership function. In other words, a value can belong to multiple sets at the same time. The most common types of membership functions are triangular, trapezoidal, and Gaussian shapes. The type of the membership function can be context dependent and it is generally chosen arbitrarily according to the user experience. The membership function is defined as are trapezoidal.

4.5.1.5 Proposed Methodology

In this part of study, as explained, missing values are imputed in the data set by applying correlation based algorithm. Then the number of attributes is reduced to seven by way of dimensionality reduction using PCA in the pre-processing stage. Then this LD data set containing the seven attributes and LD probability are given for fuzzification as inputs. Apart from the LD probability, the input attributes are DR, DS, DH, DWE, DBA, DHA and DA.

Table 3: Information about FIS

Sl. No.	Particulars	
1	Name	new2
2	Type	mamdani
3	Version	2.0
4	Num Inputs	8
5	Num Outputs	1
6	Num Rules	26
7	And Method	min
8	Or Method	max
9	ImpMethod	min
10	AggMethod	max
8	DefuzzMethod	centroid

The information about FIS is given in Table 3. There are three membership functions associated with each input variable. Some of the input variables of LD prediction and its range along with the number and type of membership functions are shown in the structure of FIS given at Table 4. The visualization of FIS in LD prediction is shown in Fig. 7.

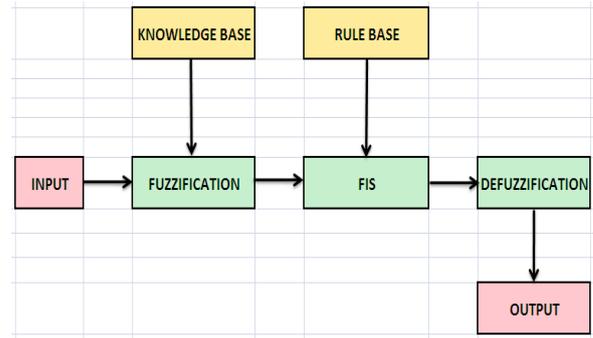


Fig. 3: System Flow Chart of Fuzzification of Attributes

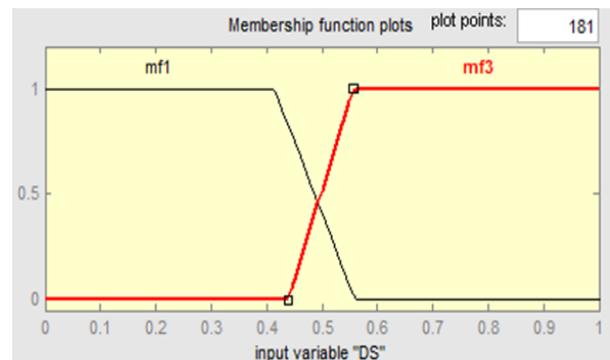


Fig. 4: Membership Function of Input Variable DS

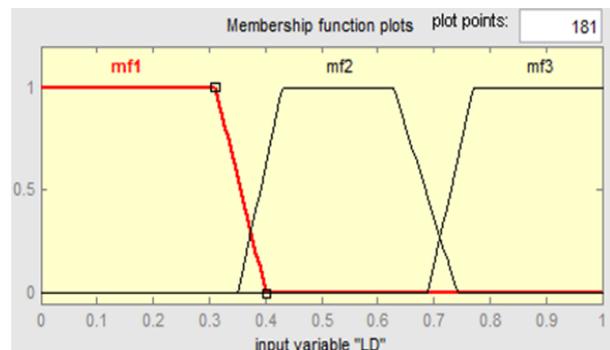


Fig. 5: Membership Function of Input Variable LD

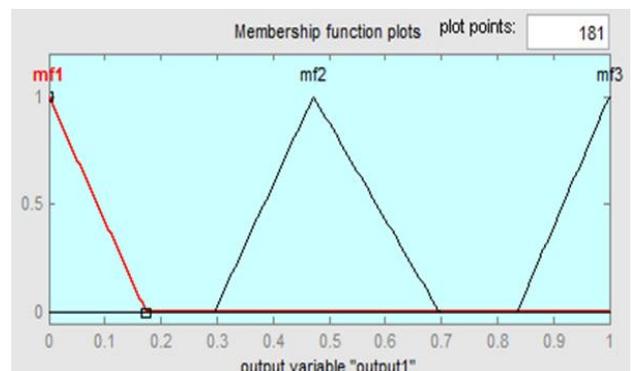


Fig. 6: Membership Function of Output Variable

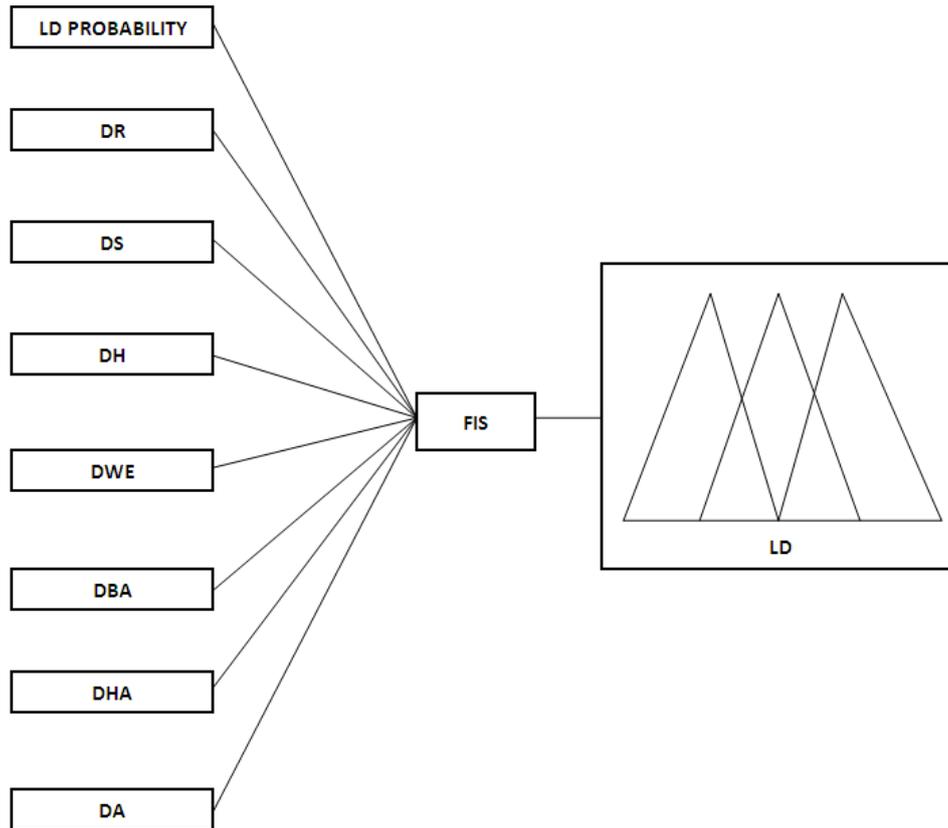


Fig. 7: Visualization of FIS

Table 4: Structure of FIS

Name='LD_PROBLITY'	Range=[0 1]
NumMFs=3	MF1='mf1':'trapmf',[-0.518518518519 -0.198518518519 0.266481481481481 0.366481481481481]
MF2='mf2':'trapmf',[0.377 0.475 0.593915343915344 0.694]	MF3='mf3':'trapmf',[0.702 0.765873015873016 1.06 1.38]
Name='DS'	Range=[0 1]
NumMFs=2	MF1='mf1':'trapmf',[-0.164 -0.0198 0.4352 0.556]
MF2='mf3':'trapmf',[0.444 0.586 1.01455026455026 1.16]	----

The rule viewer and surface viewer of LD fuzzy system with attribute reduction are shown in Fig. 8 and 9 respectively. Sample fuzzy rules for the LD prediction system are listed below;

- R1: If (LD is mf3) then (output1 is mf3)
- R2: If (LM is mf3) and (LD is mf2) then (output1 is mf2)
- R3: If (ED is mf3) and (LD is mf2) then (output1 is mf2)
- R4: If (DWE is mf3) and (LD is mf2) then (output1 is mf2)

We defined an evaluation function and it performs the evaluation of the fuzzy systems. Here also, the

effectiveness of LD prediction data set is evaluated. Out of the 513 testing data set 510 are correctly evaluated and another 507 data set is used for training. Based on the defined fuzzy inference system, we evaluate some data and obtained the results as given in Table 5.

4.5.1.6 LD prediction results

After evaluating the performance of the system based on the new fuzzy inference system, we got the results of LD prediction in fuzzy with data preprocessing as shown in Table 6. These results show accuracy to the tune of 99.50%.

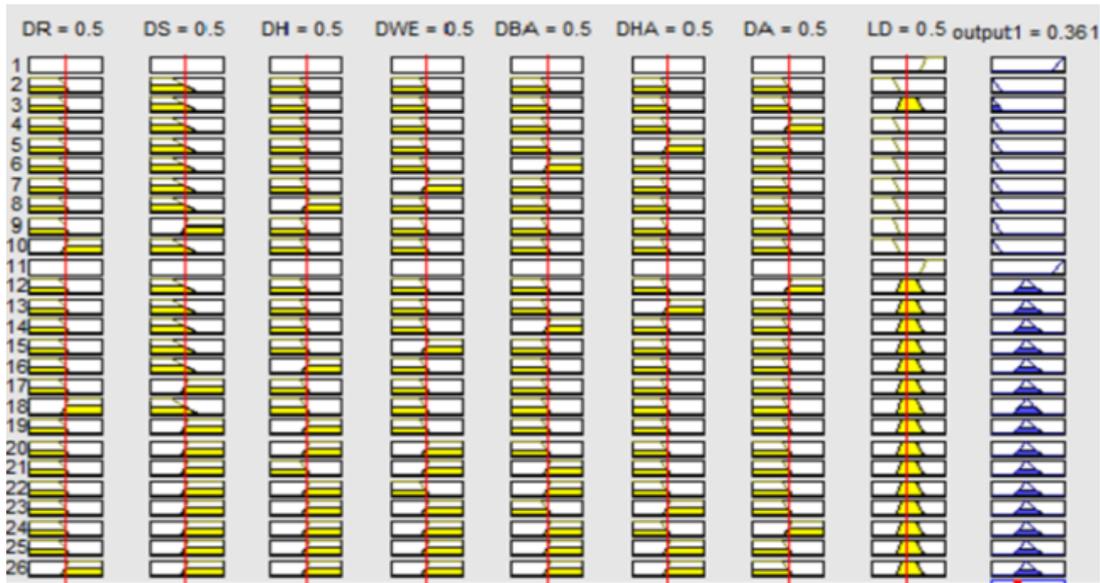


Fig. 8: Rule Viewer of LD Prediction System –Fuzzy with Attribute Reduction

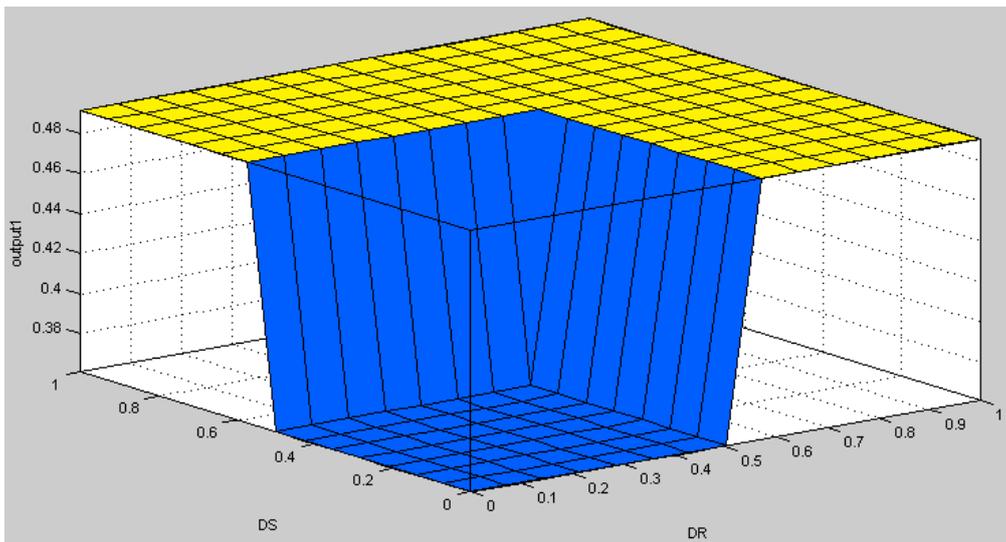


Fig. 9: Surface Viewer of Fuzzy with Attribute Reduction

Table 5: Evaluation of Fuzzy Model with Attribute Reduction

Sl. No.	Evalfis ans
1	0.5013
2	0.0545
3	0.5000
4	0.9447
5	0.9200
6	0.0545

Table 6: Test results of LD Prediction – Fuzzy

Sl. No.	Particulars	
1	Data Set used for training	507
2	No. of data set used for testing	513
3	No. of Instances correctly classified	510
4	No. of Instances incorrectly classified	3

4.5.2 Neuro Fuzzy Model

Neuro fuzzy system is a combination of artificial neural network and fuzzy systems [51]. Neuro fuzzy system combines the learning capabilities of neural networks if the linguistic rule interpretation of fuzzy inference system [67]. 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 [51]. 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. 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. The

advantages of a combination of ANN and FIS are obvious [68].

Technological analysis tries to model and simulate as accurately as possible the prediction of learning disabilities by different techniques. In this part of study, we are using neuro fuzzy techniques for the prediction of learning disabilities. Using fuzzy system, LD prediction rules can be formulated using the linguistic expressions low, minor and major for applying to the learning disability problem. 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 [69]. Here the ANFIS predicts not only the LD but also its percentage and class viz. low, minor or major. It is very important because using the legacy methods of LD assessment; we cannot easily say the percentage of LD present in children. So, the system is very effective in prediction of LD.

4.5.2.1 Adaptive Neuro-Fuzzy Inference System (ANFIS)

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. There are two very common approaches, one of them consists of 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 [70]. The first option is most used. 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. 10.

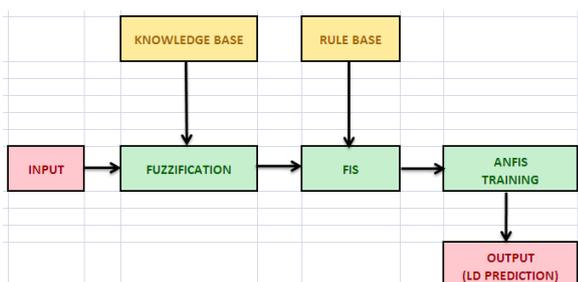


Fig. 10: Neuro Fuzzy System Flowchart for LD Prediction

4.5.2.2 Design issues

Some of the design issues in developing ANFIS system for LD prediction are the following:

(i) Number of input membership functions:

The fuzzy membership functions were set up based on knowledge on LD prediction system. We have determined the two membership functions for each of the input variables of the system and three for the LD probability and output variables.

(ii) Type of input membership functions:

Based on the properties of the LD, we primarily consider trapezoidal membership functions. These are tested as well in the case of input variables and triangular membership functions in the case of output variables.

(iii) Type of output membership functions:

We have used a single output, obtained using weighted average defuzzification. All output membership functions had the same type and are either constant or linear and the number of output membership function is 3.

(iv) The number of rules:

For a well defined fuzzy system we need to define control actions for every possible combination of input membership function values. Based on the initial fis we are created new fis containing 243 rules.

(v) Performance function:

Some of the widely used performance functions in neural networks are sum of squared error, mean squared error, etc. Here we are used sum of squared error for performance function.

4.5.2.3 Membership function

For implementation of ANFIS system, we have used Matlab 7.1 environment with fuzzy logic toolbox. Various types of input membership functions have been tried and as we expected, the trapezoidal membership functions performed best in the case of input variables. Five 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 7 and 8 respectively.

The input membership function editor of one of the seven extracted attributes, DS, and that of probability LD are same as shown in Fig. 4 and 5 respectively. In this study, the membership function of output variable

in LD prediction is obtained is also as same as shown in Fig. 6.

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[71]. 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. 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 new fismat is shown in Table 9.

4.5.2.4 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 [72]. 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
- (iv) Output defuzzification

Table 7: Details o 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]

Table 8: 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]

Table 9: Structure of Initial Fuzzy Inference System

Sl.No	Particulars	
1	Name	New1
2	Type	mamdani
3	Version	2.0
4	Num Inputs	8
5	Num Outputs	1
6	Num Rules	26
7	And Method	min
8	Or Method	max
9	ImpMethod	min
10	AggMethod	max
8	DefuzzMethod	centroid

The modeling concept used by ANFIS for LD prediction 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 [73]. However, as we have adopted data preprocessing into our data set, occurrence of such situations are far away.

Learning disability measurement 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 most important and frequent signs and symptoms of LD. 1020 real cases, some of which contain missing values are used. In the

preprocessing stage, missing values are imputed by correlation based new algorithm and the 16 attributes are reduced to 7 by PCA, which is ultimately used by the ANFIS LD prediction system. 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. The ANFIS architecture is shown in Fig. 11 [74]. The functions of each layer of ANFIS architecture are explained in Table 10.

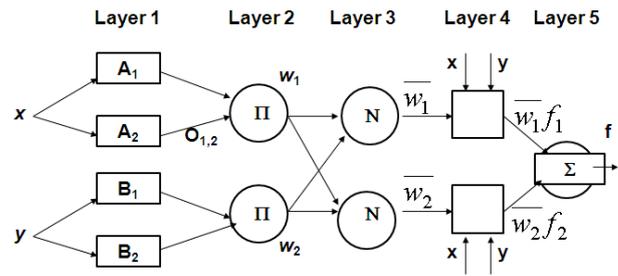


Fig. 11: ANFIS Architecture

Table 10: 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.

Once the ANFIS is trained it would be useful to extract a small number of rules, which can reliably predict LD to be offered based on fuzzy membership function values. Of course extracting rules cannot further improve performance, but it can increase speed and efficiency for further training.

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 10 fuzzy models each with a single epoch of ANFIS training. The results obtained after 10 iterations are as

given in Table 11. 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 12 and 13 respectively.

Table 11: 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 12: FISMAT Information

Fismat 2 =	Name: 'anfis'
Type: 'sugeno'	And Method: 'prod'
Or Method: 'max'	DfuzzMethod: 'wtaver'
Imp Method: 'prod'	Agg Method: 'max'
Input: [1x5 struct]	Output: [1x1 struct]
Rule: [1x243 struct]	---

Table 13: ANFIS Information

Number of nodes: 524	Number of linear parameters: 1458
Number of nonlinear parameters: 60	Total number of parameters: 1518
Number of training data pairs: 253	Number of checking data pairs: 0
Number of fuzzy rules: 243	----

4.5.2.5 LD prediction results

After evaluating the performance of the system based on the new fuzzy inference system created by the ANFIS, we got the results as shown in Table 14. These results show 100% accuracy.

Table 14: Test Results of LD Prediction – ANFIS

Sl. No.	Particulars	Nos.
1	Data Set used for training	507
2	No. of data set used for testing	513
3	No. of Instances correctly classified	513
4	No. of Instances incorrectly classified	0

V. Result Analysis and Findings

The proposed method aims to develop a new algorithm for imputing missing values and to determine the relevance of data preprocessing in fuzzy and neuro fuzzy systems. The two well known classifiers, fuzzy and neuro fuzzy are used for the implementation. From this study, we have identified that the correlation based new algorithm for missing value imputation and dimensionality reduction PCA have effect in the performance of classifier. The missing values have very high impact on the classifier. Obviously almost all the classifiers ignore the missing values. After imputing the missing values in the data set using correlation based new algorithm, developed by us, we found some attributes are irrelevant. So we have used PCA and the number of attributes is reduced to 7 from 16. These 7 attributes are applied for fuzzy and neuro fuzzy classification and we have obtained very effective results. The results are very beneficial to the children, parents and teachers to identify the

learning disabilities at an early stage. If they are able to diagnose the child’s problem at an early stage they can go for the proper treatments/counseling at the correct time so as to avoid the academic and social losses.

The main contributions from this study are the development of new methods for LD prediction based on machine learning techniques, new insights into the interrelationships between symptoms of LD and their relative importance, development of new algorithm based on correlation based missing value imputing and modification of data preprocessing with fuzzy and neuro fuzzy system for LD prediction. The number of attributes is reduced by eliminating the unwanted and redundant ones by using Principal Component Analysis, which helps in reducing the time of classification and the contribution of the study ultimately improves the confidence of children through the early invention of LD and helps in getting the social support to them.

VI. Comparison of Results

The results obtained from this study are compared with the results of our other similar studies conducted based on J48, Naive Bayes, Support Vector Machines (SVM) and Multi Layer Perceptron classifiers [42,13]. The comparison of the results is shown in Table 15 below. From this comparison, we can see that the New Fuzzy and New ANFIS classifiers are better in terms of classification and accuracy. In the data preprocessing stage, we applied some methods such as correlation based new algorithm for imputing missing values and PCA for attribute reduction in these classifiers. From the comparison of results, it can be seen that, these preprocessing methods are effectively improves the accuracy of fuzzy and neuro fuzzy classifiers.

Table 15: Comparison of Classification Results

Parameters	Classifiers						
	Existing Classifiers				Modified Classifiers with Data Pre-Processing		
	J48	Naive Bayes	SVM	ANN	New ANN	New Fuzzy	New ANFIS
No. of instances used for classification	513	513	513	513	513	513	513
Correctly Classified Instances (in Nos. & %)	500 97.47%	426 83.04%	502 97.86%	502 97.86%	508 99.03%	510 99.42%	513 100%
Incorrectly Classified Instances (in Nos. & %)	13 2.53%	87 16.96%	11 2.14%	11 2.14%	5 0.97%	3 0.58%	0 0%
Time taken to build a model (in seconds)	0.08	0.06	3.92	18.03	1.28	1.28	1.28

VII. Conclusion and Future Works

In this paper, we have developed a new approach in two data mining methods, fuzzy and neuro fuzzy, to effectively and accurately predict the learning disability in school age children. This study mainly focuses on developing a new algorithm for imputing missing values and evaluation of performance of classifiers with new preprocessing methods. The accuracy of decision-making is found 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 1020 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 classifiers with fuzzy and neuro fuzzy models 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. 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

- [1] Julie M. David, Pramod K.V., "Paper on Prediction of Learning Disabilities in School Age children using Data Mining Techniques", In: Proceedings of AICTE Sponsored National Conf. on Recent Developments and Applications of Probability Theory, Random Process and Random Variables in Computer Science, T. Thirvikram, P. Nagabhushan, M.S. Samuel (eds), 2008, pp139-146.
- [2] Sally Jo Cunningham, Geoffrey Holmes, Developing innovative applications in agricultural using data mining-In the Proceedings of the Southeast Asia Regional Computer Confederation Conference, 1999
- [3] Frawley and Piatetsky, Shaping Knowledge Discovery in Database; an Overview, The AAAI/MIT press, Menlo Park, 1996.
- [4] Hsinchun Chen, Sherrilynne S. Fuller, Carol Friedman and William Hersh. Knowledge Discovery in Data Mining and Text Mining in Medical Informatics, Chapter 1, 2005, pp3-34.
- [5] Chapple Mike, About.com Guide, <http://databases.about.Com/od/datamining/g/classification.htm>
- [6] Julie M. David, Kannan Balakrishnan, Significance of Classification Techniques in Prediction of Learning Disabilities in School Age Children, *Int. J. of Artificial Intelligence & Applications*, 1(4), DOI:10.5121/ijaia.2010.1409, Oct.2010, pp111-120.
- [7] Blackwell Synergy, *Learning Disabilities & Research Practice*, Volume 22, 2007.
- [8] Julie M. David, Kannan Balakrishnan, Prediction of Key Symptoms of Learning Disabilities in School-Age Children using Rough Sets, *Int. J. of Computer and Electrical Engineering*, 3(1), Feb. 2011, pp163-169
- [9] Jiye Li, Nick Cercone: Assigning Missing Attribute Values Based on Rough Sets Theory, *IEEE International Conference on Granular Computing*, May 2006, pp607-610.
- [10] Grzymala-Busse, J.W., Hu, M.: A Comparison of Several Approaches to Missing Attribute Values in Data Mining. Ziarko, W., Yao. Y., (Eds.): *RSCTC 2000, LNAI 2005 (2001)* pp378-385
- [11] Fatma Khcherem, Abdelfettah Bouri: Fuzzy Logic and Investment Strategy Global Economy & Finance Journal, 2(2), 2009, pp 22-37.
- [12] Zadeh, L.A: The Concept of Linguistic Variable and its Application to Approximate Reasoning, *Inf. Sciences*, 8, 1975, pp199-249
- [13] Julie M. David, Kannan Balakrishnan, Attribute Reduction and Missing Value Imputing with ANN: Prediction of Learning Disabilities, *Int. J. of Neural Computing*, Springer-Verlag London Limited, DOI: 10.1007/s00521-011-0619-1, 21 (7), Oct.2012, pp 1757-1763.
- [14] Kenneth A. Kavale: Identifying Specific Learning Disability - Is Responsiveness to Intervention the Answer?, *J. of LDs*, 38, 2005, pp553-562.
- [15] Benjamin J. Lovett: Extended Time Testing Accommodations for Students with Disabilities: Answers to 5 Fundamental Questions, *Review of Edu. Research*, *J. of LDs*, 80, 2010, pp611-638
- [16] Noona Kiuru, et. al.: Students with Reading and Spelling Disabilities-Peer Groups and Educational Attainment in Secondary Education, *Journal of Learning Disabilities*, 44, 2011, pp556-569.
- [17] Chen, S.M., Chen, H.H.: Estimating Null Values in the Distributed Relational Databases Environments, *International Journal on Cybernetics and Systems*, 31, 2000, pp851-871.
- [18] Chen, S.M., Huang, C.M.: Generating Weighted Fuzzy Rules from Relational Database Systems for Estimating Null Values Using Genetic Algorithms. *IEEE Transactions on Fuzzy Systems*. 11, 2003, pp 495-506.

- [19] Quinlan, J.R.: *C4.5 - Programs for Machine Learning*. Morgan Kaufmann, San Mateo, USA, 1993.
- [20] Friedman, J., et al.: Lazy Decision Trees. Proceedings of the 13th National Conference on Artificial Intelligence, 1996, pp717-724.
- [21] Lakshminarayan, K., et al.: Imputation of Missing Data Using Machine Learning Techniques. *KDD-1996*, pp140-145.
- [22] Magnani, M.: Techniques for Dealing with Missing Data in Knowledge Discovery Tasks. Available at: <http://magnanim.web.cs.unibo.it/data/pdf/missingdata.pdf>, Version of June 2004.
- [23] Kahl, F., et al.: Minimal Projective Reconstruction Including Missing Data. *IEEE Trans. Pattern Anal. Mach. Intell.*, 23(4), 2001, pp418-424.
- [24] Gessert, G.: Handling Missing Data by Using Stored Truth Values. *SIGMOD Record*, 20(3), 2001, pp30-42.
- [25] Pesonen, E., et al.: Treatment of Missing Data Values in a Neural Network Based Decision Support System for Acute Abdominal Pain. *Artificial Intelligence in Medicine*, 13(3), 1998, pp139-146.
- [26] Ramoni, M., Sebastiani, P.: Robust Learning with Missing Data. *Machine Learning*, 45(2), 2001, pp147-170.
- [27] Pawlak, M., Kernel: Classification Rules from Missing Data. *IEEE Transactions on Information Theory*, 39(3), 1993, pp979-988
- [28] Yao, Y.Y.: A Comparative Study of Fuzzy Sets and Rough Sets, *Information Sciences*, 109 (1-4), 1998, pp. 227-242
- [29] Pawlak, Z.: *Rough Sets: Theoretical Aspects of Reasoning about Data*, Kluwer Academic Publishers, Boston, 1991
- [30] Klir, G.J., Yuan, B.: *Fuzzy Sets and Fuzzy Logic, Theory and Applications*, Prentice Hall, New Jersey, 1995
- [31] Chanas, S., Kuchta, D.: Further remarks on the relation between rough and fuzzy sets, *Fuzzy Sets and Systems*, 47, 1992, pp391-394.
- [32] Zadeh, L.A.: Forward, in: Orłowska, E. (Ed.): *Incomplete Information: Rough Set Analysis*, Physica-Verlag, Heidelberg, 1998, pp. v-vi.
- [33] Ossi Nykanen: Inducing Fuzzy Model for Students Classification, *Educational Technology & Society*, 9 (2), 2006, pp223-234.
- [34] Rubens, N.O.: The Application of Fuzzy Logic to the Construction of the Ranking Function of Information Retrieval System, *Computer Modelling and New Technologies*, 10 (1), 2006, pp20-27
- [35] Bordogna, G, Pasi, G.: Handling vagueness in information retrieval systems, Proceedings of the Second New Zealand Int. Two-Stream Conf. on Artificial Neural Networks and Expert Systems, 1995, pp110-114.
- [36] Miyamoto, S.: *Fuzzy sets in Information Retrieval and Cluster Analysis*, Kluwer Academic Publishers, 1990.
- [37] Ogawa, Y., Morita, T., Kobayashi, K.: A Fuzzy Document Retrieval System Using the Key Word Connection Matrix and A Learning Method, *Fuzzy Sets and Systems* 39, 1991, pp163-179.
- [38] Baeza Yates, R., Ribeiro Neto, B.: *Modern Information Retrieval*, ACM Press, 1999.
- [39] Tung-Kuang Wu, Shian Chang Huang, Ying Ru: Evaluation of ANN and SVM Classifiers as Predictors to the Diagnosis of Students with LDs, *J. of Expert Systems with Applications*, 34 (3), 2008, pp1846-1856.
- [40] Maitrei Kohli, Prasad, T.V.: Identifying Dyslexic Students by Using Artificial Neural Networks, *Proc. of the World Congress on Engg. I*, 2010,
- [41] Rod Paige (Secretary), US Department of Education, Twenty-fourth Annual Report to Congress on the Implementation of the Individuals with Disabilities Education Act-To Assure the Free Appropriate Public Education of all Children with Disabilities, 2002.
- [42] Julie M. David, Kannan Balakrishnan, Machine Learning Approach for Prediction of Learning Disabilities in School Age Children, *Int. J. of Computer Applications*, ISSN-0975-8887, 9(10), Nov. 2010, pp7-14.
- [43] Julie M. David, Kannan Balakrishnan, Prediction of Learning Disabilities in School-Age Children using SVM and Decision Tree, *Int.J. of Computer Science and Information Technology*, ISSN 0975-9646, 2(2), Mar-Apr. 2011, pp829-835
- [44] Julie M. David, Kannan Balakrishnan, "Paper on Prediction of Learning Disabilities in School Age Children using Decision Tree". In: Proceedings of the Int. Conf. on Recent Trends in Network Communications- CCIS, 90(3), N. Meghanathan, Selma Boumerdassi, Nabendu Chaki, Dhinaharan Nagamalai (eds), Springer-Verlag Berlin Heidelberg, ISSN:1865-0929(print) 1865-0937(online), ISBN 978-3-642-14492-9(print) 978-3-642-14493-6(online), DOI : 10.1007/978-3-642-14493-6_55, 2010, pp533-542
- [45] Crealock Carol, Kronick Doreen. Children and Young People with Specific Learning Disabilities, Guides for Special Education, No. 9, UNESCO, 1993.
- [46] Julie M. David, Kannan Balakrishnan, "Paper on Prediction of Frequent Signs of Learning

- Disabilities in School Age Children using Association Rules”. In: Proceedings of the Int. Conf. on Advanced Computing, ICAC 2009, MacMillan Publishers India Ltd., NYC, ISBN 10:0230-63915-1, ISBN 13:978-0230-63915-7, 2009, pp202–207.
- [47] Zhang, S.C., et al.: Information Enhancement for Data Mining. IEEE Intelligent Systems, 19(2), 2004, pp12-13.
- [48] Zhang, S.C., et al.: Missing is useful - Missing Values in Cost-Sensitive Decision Trees, IEEE Transactions on Knowledge and Data Engineering, 17(12), 2005, pp1689-1693.
- [49] Qin, Y.S., et al.: Semi-Parametric Optimization for Missing Data Imputation. Applied Intelligence, 27(1), 2007, pp79-88.
- [50] Zhang, C.Q., et al.: An Imputation Method for Missing Values, PAKDD, LNAI, 4426, 2007, pp1080–1087.
- [51] Han Jiawei, Kamber Micheline, Pei Jian: *Data Mining-Concepts and Techniques*, Third Edition, Morgan Kaufmann - Elsevier Publishers, ISBN : 978-93-80931-91-3, 2011.
- [52] Zhang, S.C., et al.: Optimized Parameters for Missing Data Imputation, PRICAI06, 2006, pp1010-1016.
- [53] Wang, Q., Rao, J.: Empirical Likelihood-based Inference in Linear Models with Missing Data, Scand. Journal of Statist., 29, 2002, 563-576.
- [54] Wang, Q., Rao, J. N. K.: Empirical likelihood-based inference under imputation for missing response data, Ann. Statist., 30, 2002, pp896-924
- [55] Chen, J., Shao, J., Jackknife: Variance Estimation For Nearest-Neighbor Imputation, Journal of Amer. Statist. Assoc., 96, 2001, pp260-269.
- [56] Lall, U., Sharma, A.: A Nearest-Neighbor Bootstrap For Resampling Hydrologic Time Series, Water Resource, Res. 2001, 32, 1996, pp679–693.
- [57] Tan Pang-Ning, Steinbach Michael, Kumar Vipin : Introduction to Data Mining, Low Price Edition, Pearson Education, Inc., ISBN 978-81-317-1472-0, 2008
- [58] Julie M. David, Kannan Balakrishnan: A New Decision Tree Algorithm for Prediction of Learning Disabilities, Journal of Engineering Science and Technology, School of Engineering, Taylor's University, Malaysia. 8 (2), April 2013, pp120-132.
- [59] Zadeh, L.A.: Fuzzy sets, Information and control, 8, 1965, pp 338-353
- [60] Urathal U alias Swathiga Sri, Chandrasekar, C.: An Efficient Fuzzy based Congestion Control Technique for Wireless Sensor Networks, International Journal of Computer Applications, 40(14), Feb 2012, pp47-55.
- [61] Maged Marghany, Mazlan Hashim, Farideh Moradi: Object recognitions in RADARSAT-1 SAR Data using Fuzzy Classification, International Journal of the Physical Sciences, Academic Journals, 6(16), 2011, pp3933-3938.
- [62] Zadeh, L.A.: Fuzzy sets. Infor. Control, 8, 1965, pp338-353.
- [63] Ross, T.J.: *Fuzzy Logic with Engineering Applications*, Second Edition, John Wiley & Sons Ltd., 2004.
- [64] Fuzzy Logic Toolbox User's Guide , The MathWorks Inc., 2004.
- [65] Zadeh, L.A: The Concept of Linguistic Variable and its Application to Approximate Reasoning, Inf. Sciences, 8, 1975, pp199-249.
- [66] Zadeh, L.A: The Concept of Linguistic Variable and its Application to Approximate Reasoning, Inf. Sciences, 9, 1975, pp43-80.
- [67] Rahib Abiyeb, Vasif, H., et. al.:Electricity Consumption Prediction Model using Neuro-Fuzzy System, World Academy of Science, Engineering and Technology, 8, 2005.
- [68] Ajith Abraham: Neuro-Fuzzy Systems: State-of-the-Art Modeling Techniques, Connectionist Models of Neurons, Learning Processes, and Artificial Intelligence, Lecture Notes in Computer Science, 2084, Springer-Verlag Germany, Jose Mira, Alberto Prieto (Eds.), ISBN 3540422358, Spain, 2001, pp269-276.
- [69] Shantakumar B. Patil, Kumaraswamy, Y.S.: Intelligent and Effective Heart Attack Prediction System using Data Mining and ANN, European Journal of Scientific Research, 31(4), ISSN 1450-216X, 2009, pp642-656.
- [70] Corona-Nakamura, M. A., Ruelas, R., Ojeda Magana, B., Andina, D.: Classification of Domestic Water Consumption Using an ANFIS Model, World Automation Congress, Waikoloa, Hawaii, EEUU, IEEE Computer Society, 2008.
- [71] Sivarao, et. al.: GUI Based Mamdani Fuzzy Inference System Modeling to Predict Surface Roughness in Laser Machining, International Journal of Electrical & Computer Sciences, 9(9), 2009, pp281-288.
- [72] Sidda Reddy, B., Suresh Kumar, J., Vijaya Kumar Reddy, K.: Prediction of Roughness in Turning using Adaptive Neuro-Fuzzy Inference System, Jordan Journal of Mechanical and Industrial Engineering, 3(4), Dec. 2009, pp252-258.
- [73] Erdem Buyukbingol, Arzu Sisman, Murat Akyildiz, Ferda Nur Alparslan, Adeboye Adeja:

Adaptive neuro-fuzzy inference system (ANFIS): A New Approach to Predictive Modeling in QSAR Applications: A Study of Neuro-Fuzzy modeling of PCP-based NMDA Receptor Antagonists, *Bioorganic & Medicinal Chemistry*, 15(12), 2007, pp4265-4282.

- [74] Jang, J.S.R., Sun, C.T., Mizutani, E.: *Neuro – Fuzzy and Soft Computing, A Computational Approach to Learning and Machine Intelligence*, PHI Learning Pvt. Ltd, New Delhi, 2008

Authors' Profiles



Dr. Julie M. David born in 1976 received her MCA degree from Bharathiyar University, Coimbatore, India in 2000, the M.Phil degree in Computer Science from Vinayaka Missions University, Salem, India in 2009 and the Ph. D. degree in the research area of Artificial Intelligence from

Cochin University of Science and Technology, Cochin, India in 2013. During 2000-2007, she was with Mahatma Gandhi University, Kottayam, India, as Lecturer in the Department of Computer Applications. Currently she is working as Assistant Professor in the Department of Computer Applications with MES College, Aluva, Cochin, India. She has published several papers in International Journals and International and National Conference Proceedings. Her research interests include Artificial Intelligence, Data Mining and Machine Learning. She is a life member of International Association of Engineers and Computer Scientists, IAENG Societies of Artificial Intelligence & Data Mining and a Reviewer of Elsevier International Journal of Knowledge Based Systems. Also, she is a reviewer and an Editorial Board Member of various other International Journals. She has coordinated various International and National Conferences.



Dr. Kannan Balakrishnan born in 1960 received his M.Sc and M. Phil degrees in Mathematics from University of Kerala, India, M. Tech degree in Computer and Information Science from Cochin University of Science & Technology, Cochin, India and Ph. D in Futures Studies from

University of Kerala, India in 1982, 1983, 1988 and 2006 respectively. He is currently working with Cochin University of Science & Technology, Cochin, India, as Associate Professor (Reader) in the Department of

Computer Applications. He has visited Netherlands as part of a MHRD project on Computer Networks. Also he visited Slovenia as the co-investigator of Indo-Slovenian joint research project by Department of Science and Technology, Government of India. He has published several papers in International Journals and International and National Conference Proceedings. His present areas of interest are Graph Algorithms, Intelligent Systems, Image Processing, CBIR and Machine Translation. He is a reviewer of American Mathematical Reviews. He is a recognized Research Guide in the Faculties of Technology and Science in the Cochin University of Science and Technology, Cochin, India. He has served in many academic bodies of various universities in Kerala, India. Also currently he is a member of the Board of Studies of Cochin, Calicut and Kannur Universities in India. He is also a member of MIR labs India.