# STATISTICAL MACHINE LEARNING TECHNIQUES FOR THE PREDICTION OF LEARNING DISABILITIES IN SCHOOL-AGE CHILDREN

*A Thesis*

*Submitted by*

**JULIE M. DAVID**
**Reg. No. 3763**

*for the award of the degree of*

**DOCTOR OF PHILOSOPHY**
*(Faculty of Technology)*

**DEPARTMENT OF COMPUTER APPLICATIONS**
**COCHIN UNIVERSITY OF SCIENCE AND TECHNOLOGY**
**KOCHI – 682 022**
**INDIA**

*March 2012*

**Dr. B. Kannan**
Associate Professor
Department of Computer Applications
Cochin University of Science and Technology
Cochin – 682 022, India
E-mail: bkannan@cusat.ac.in

# *Certificate*

*Certified that this thesis entitled "**Statistical Machine Learning Techniques for the Prediction of Learning Disabilities in School-Age Children**", submitted to Cochin University of Science and Technology, Cochin – 22, for the award of Ph.D. Degree, is the record of bonafied research carried out by Smt. **Julie M. David**, under my supervision and guidance at Department of Computer Applications, Cochin University of Science and Technology, Cochin - 22. This work did not form part of any dissertation submitted for the award of any degree, diploma, associateship or other similar title or recognition from this or any other institution.*

*Kochi-22*
*24-03-2012*

***Dr. B. Kannan***
*Supervising Guide*

## DECLARATION

I, Julie M. David, hereby declare that the work presented in the thesis entitled "**Statistical Machine Learning Techniques for the Prediction of Learning Disabilities in School-Age Children"**, being submitted to Cochin University of Science and Technology, Cochin – 22, for the award of Doctor of Philosophy under Computer Applications, is the outcome of the original work done by me under the supervision of Dr. B. Kannan, Associate Professor, Department of Computer Applications, Cochin University of Science and Technology, Cochin - 22. This work did not form part of any dissertation submitted for the award of any degree, diploma, associateship or other similar title or recognition from this or any other institution.

Kochi-22 **Julie M. David**

24-03-2012 Reg. No. 3763

*This thesis is the result of last few years of work whereby I have been accompanied and supported by many people. It is a pleasant aspect that I have now the opportunity to express my sincere appreciation to all of them.*

*First and foremost, I would like to thank my guide and mentor, Dr. B. Kannan, Associate Professor, Department of Computer Applications, CUSAT for his overly enthusiasm and integral view on research. It has been the best guidance during my way to earn the degree. Without his encouragement, motivation and insightful advices, I could not have finished this thesis.*

*I express my sincere thanks to Dr. K.V. Pramod, Head, Department of Computer Applications, CUSAT, for his sincere helps, discussions and advise during the various phases of this research work.*

*I am thankful to Dr. G. Raju, Reader, Dept. of Information Technology, Kannur University, Kerala, who had helped me in developing the basic steps of research work.*

*I express my sincere gratitude to Dr. K.C. Sankaranarayanan, Former Director, School of Management Studies, CUSAT, for his valuable information, suggestions and timely guidance given to me during the course of my research work.*

*I am thankful to Dr. P.H. Zainaba Beevi, Pediatrician, ESI Hospital, Ernakulam, Dr. Ajith Murkoth, Director, Vigyan Valley Open School, Cochin, Dr. Anitha Mathew, Child Care Centre, Cochin, for their helps in obtaining relevant data required for the research work and for their valuable guidance and inspirations.*

*I express my deep gratitude to Sri. Tung Kuang Wu, Department of Information Management, National Changhua University of Education,* *Changhua, Taiwan, Republic of China, who had always helped me in getting the datasets for carrying out the research work in a fruitful manner.*

*I also thankful to the authorities of HMT School and Rajagiri School, Kalamassery, Cochin for their helps in conducting interview with students and parents in collecting data.*

*My gratitude also goes to the Management, Principal, Director, Head and other colleagues especially Sri. M.K. Sabu, Senior Lecturer, Department of Computer Applications, of MES College, Marampilly, Aluva, Cochin, other faculties especially Dr. A. Sreekumar and Smt. S. Malathi, non teaching staff, office staff and my co-research scholars at the Department of Computer Applications of CUSAT and to all my friends for all the moral support, helps and encouragements given to me during the course of the research work.*

*I owe my utmost appreciation to my parents. They endowed me not only a material life but the values that really matter in the life. Their understanding and support encouraged me to work hard. Finally, the acknowledgements would not be complete without a heartfelt thank to Mr. M.P. Eldhose, the most significant person in my life. His love, support, continuous encouragement, persistent source of inspiration, courage, patience and confidence helped me to pursue my dreams. To my little beloved kids, my daughters Mintu and Minnu and my son Elias, a special hug and gratitude for their sufferings, tolerance, support and prayers along this journey.*

*Last, but not least, I express my worshipful gratitude to The Almighty who actually guided me in my endeavor.*

# ABSTRACT

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. The learning disabled children are neither slow nor mentally retarded. 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. 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 an average child. 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 and they are life-long. However, children with LD can be high achievers and can be taught ways to get around the learning disability.

The problems of children with specific learning disabilities have been a cause of concern to parents and teachers for some time. With the right help, children with LD can and do learn successfully. Mental retardation, emotional disorders and poor socioeconomic status are not considered learning disabilities. Learning disabilities have been recognized in some countries for much of the 20$^{th}$ century, in other countries only in the latter half of the century and yet not at all in other places. The concept is still new in many developing countries. In India, the research conducted in learning disability has been primarily done over the last two decades and is today comparable with the research carried out in west nearly half a century ago. About 10% children enrolled in schools having LD. 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. Many types of assessment tests are available. Just as there are many different

types of LDs, there are a variety of tests that may be done to pinpoint the problem. Many professionals can be involved in the testing process. The purpose of LD assessment is to determine child's strengths and weaknesses and to understand how he or she best learns and where they have difficulty.

In this research work, data mining using machine learning techniques are used to analyze the symptoms of LD, establish interrelationships between them and evaluate the relative importance of these symptoms. To increase the diagnostic accuracy of learning disability prediction, a knowledge based tool based on statistical machine learning or data mining techniques, with high accuracy, according to the knowledge obtained from the clinical information, is proposed. The basic idea of the developed knowledge based tool is to increase the accuracy of the learning disability assessment and reduce the time used for the same. Different statistical machine learning techniques in data mining are used in the study. Identifying the important parameters of LD prediction using the data mining techniques, identifying the hidden relationship between the symptoms of LD and estimating the relative significance of each symptoms of LD are also the parts of the objectives of this research work. The developed tool has many advantages compared to the traditional methods of using check lists in determination of learning disabilities.

For improving the performance of various classifiers, we developed some pre-processing methods for the LD prediction system. A new system based on fuzzy and rough set models are also developed for LD prediction. Here also the importance of pre-processing is studied. A Graphical User Interface (GUI) is designed for developing an integrated knowledge based tool for prediction of LD as well as its degree. The designed tool stores the details of the children in the student database and retrieves their LD report as and when required.

The developed tool is very user friendly and it not only predicts the LD but also its class like low, minor or major with percentage of LD in each class.

Depending upon the degree of LD, the school authorities/parents can recommend the child for further treatment with councilors/special educators/LD clinics, for proper remedial solutions. Thus the developed tool is helpful in finding the LD at an early stage. With the right help and intervention at proper time, children with LD can succeed in school and go on to be successful later in life, where the research work is found much relevant as early detection of developmental differences is an early signal of a learning disability and thus the problems that are spotted early can be easier to correct.

The present study undoubtedly proves the effectiveness of the tool developed based on various machine learning techniques. It also identifies the important parameters of LD and accurately predicts the learning disability in school age children. This thesis makes several major contributions in technical, general and social areas. The results are found very beneficial to the parents, teachers and the institutions. 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.

# Contents

## *List of Figures*

**Page No.**

# *List of Tables*

Chapter

1

# INTRODUCTION

Contents



## 1.1 Motivation

Learning disability or specific learning disability is a lifelong neuro-developmental disorder which manifest in childhood as persistent difficulties in learning to efficiently read, write or do simple mathematical calculations despite normal intelligence, conventional schooling, intact hearing and vision, adequate motivation and socio-cultural opportunity [1]. It is now known that a learning disability is not connected to mental retardation. The learning disabled frequently have high IQs. It is also not a single disorder, but includes disabilities in any of areas related to reading, language and mathematics.

During the past few decades, the understanding of learning disability has changed. However, it is a tremendous challenge to identify and diagnose and assist children with learning disability. As the concept is still new, in many developing countries including India, the research conducted in learning disability has been primarily done over the last two decades are yet in the infancy stage. In India, the research conducted in learning disability has been primarily done over the last two decades and is today comparable with the research carried out in west nearly half a century ago.

Since no national census of the learning disabled has been taken in India, it is difficult to collect their actual number. In India, the learning disabled children are not identified using reliable tests. However, at least 10% of children in India have a learning disability [2]. We do not have a clear idea about incidence and prevalence of learning disability in India. These facts suggest that the early diagnosis of learning disability in children is critically important to identify and suggest remedial solutions to the parents and children to understand about the learning disability as stumbling blocks such as lack of awareness, indifference and apathy and hamper success.

The problems of LD affected children have been a cause of concern to parents and school authorities for some time. With the right help at right time, right assessment and remediation, children with LD can and do learn successfully and become winners in the society later. Since LD has distinctive symptoms in its early stage, diagnosis approaches have been improved noticeably over the past decades in many counties. Research works done in this area using computer based methods is found very little compared to the magnitude of learning disability affected children.

The present method available to determine LD in children is based on check lists containing the symptoms and signs of LD. This traditional method is time consuming, not accurate and obsolete also. Such LD identification facilities are much less at schools or even in cities. Parents are either unaware or may not willing to take their children to undergo such an evaluation. Even if, teachers are advised, parents may hesitate to such evaluation process because of the unawareness of the society about LD as they might think that the child may be mentally retarded. If the LD determination facility is attached with schools and the check ups are arranged as a routine process, LD can be identified at an early stage.

Under these circumstances, it is felt to design a tool based on machine learning techniques for prediction of learning disability in school-age children. Hence, it is decided to carry out a research work in the topic in a view to increase the diagnostic accuracy of learning disability prediction. Based on the statistical machine learning tool developed, the presence and degree of learning disability in any child can be determined accurately at an early stage.

## 1.2 Problem Statement

The main problem considered in this work for analyzing and solving is the design of a tool based machine learning technique for prediction of learning disability in school-age children. The problem also involves in identifying the important parameters of LD, identifying the hidden relationship between the symptoms of LD and estimating the relative significance of each symptoms of LD using data mining techniques. The drawbacks in the existing classification algorithms have also to be determined first. Then how these algorithms can be effectively modified and used in the prediction and classifications of learning disabilities are studied. Thus the main task of the study is to find out how effectively the different classification algorithms existing in data mining can perform the prediction of learning disabilities. The problem also involves how these drawbacks in the existing algorithms has determined while performing the prediction of LD and how these are eliminated to maximize the accuracy of the optimized solution.

This research work has faced difficulties in collecting data from clinics and schools. Either the doctors or the school authorities are not like to reveal the data related to LD affected children. As there are no such related works in the field, we felt difficulties in acquiring data as well as normalizing the attribute values. However, we convinced the professionals in the field regarding the necessity and effectiveness to be achieved, they agreed to sit along with them during clinical consultations. Then from the experience gained with them for the long term, we could evaluate the deficiencies in their traditional methods of LD evaluation and thereby formed new informal checklists, which have been approved by them and they are now using it.

## 1.3 Data Mining

In recent years the sizes of databases has increased rapidly. The amount of data in the world seems to go on increasing and there is no end. In data mining, the data search is automated by computer and stored electronically. This has lead to the development of tools capable in the automatic extraction of knowledge from data. The term Data Mining or Knowledge Discovery in Databases, has been adopted for a field of research dealing with the automatic discovery of implicit information or knowledge within databases [3]. Data mining is defined as the process of discovering patterns in data. It is the non trivial extraction of implicit previously unknown and potentially useful information about data. It is an analytic process designed to explore data in search of consistent patterns and/or systematic relationships between variables, and then to validate the findings by applying the detected patterns to new subsets of data [4]. The data mining process must be automatic or semiautomatic. The patterns discovered must be meaningful in that they lead to some advantage, usually an economic advantage [5]. Conventionally the construction of model of the semantic structure of the dataset is the information that is mined. The model might be utilized for prediction and categorization of new 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 [6]. A majority of areas related to medical services such as prediction and classification of effectiveness of surgical procedures, medical tests, medication and discovery of relationship among clinical and diagnosis data also make use of data mining methodologies [7]. The kind of learning techniques that do not use the conceptual problems, it is called machine learning.

The mining software is one of the analytical tools for analyzing data. It allows users to analyze data from many different dimensions or angles, categorize it and summarize the relationships identified. Technically, it is the process of finding correlations or patterns among a lot of fields in large relational databases. The patterns, associations, or relationships among all this data can provide information. Data mining provide a link between the transaction and analytical systems. Any one of the following four types of relationships is generally mined for determining the existing patterns in the data base. These relationships are classes, clusters, associations and sequential patterns. Stored data is used to locate data in predetermined groups known as Classes. Data items are grouped according to logical relationships are clusters. Data can be mined to identify associations and sequential patterns are the data mined to anticipate behavior patterns and trends. The three stages consists in the process of data mining are initial exploration, model building and predictions. Different levels of analysis available are artificial neural networks, genetic algorithms, decision trees, rule induction and data visualization.

Data mining has an inherent connection with statistics. Statistics studies the collection, analysis, interpretation or explanation and presentation of data. A statistical model is a set of mathematical functions that describe the behavior of the objects in a target class in terms of random variables and their associated probability distributions. Statistical models are widely used to model data and data classes. Applying statistical methods in data mining is far from trivial [8]. When applying the high complex algorithm, data mining is required to continuously handle fast, real time data stream. In general the statistical methods are categorized in to two, viz. parametric and nonparametric methods.

Classification and prediction are the ultimate goals of data mining. Classification is the process of finding a model that describes and distinguishes data classes or concepts, for the purpose of being able to use the model to predict the class of objects whose class label is unknown. The derived model is based on the analysis of a set of training data.

In this research work, different classifications methods and models such as neural network, decision tree, support vector machine, rough set, fuzzy set and neuro fuzzy are used. The data sets are undergone to classification through these methods or models. Then these models are used for prediction of learning disabilities.

## 1.4 Statistical Machine Learning

Learning is an essential human property. It is the process by which observed data is used for constructing models with the intention to use them for prediction. Machine learning, a branch of artificial intelligence, is a scientific discipline concerned with the design and development of algorithms that allow computers to evolve behaviors based on empirical data, such as from sensor data or databases. The general idea behind the statistical method is the detection of class to learn a generative model fitting the given datasets and then to identify the objects in low probability regions of the model as the class. However in different ways we can learn generative models.

A major area of the machine learning research is to learn automatically in recognizing complex patterns and make intelligent decisions based on data. The difficulty lies in the fact that, the set of all possible behaviors, given all possible inputs is too large, to be covered by the set of observed examples. Hence the learner must generalize from the given datasets, so as to be able to produce a useful output in new cases. Machine learning is concerned with the development of algorithms and technique that allow computers to learn. Machine learning has wide spectrum of applications including natural language processing, search engines, medical diagnosis, detecting credit card fraud, stock market analysis, etc. [6]. Some fundamental types of learning are supervised learning, unsupervised learning, reinforced learning, etc. In the research work different supervised algorithms and unsupervised algorithms are used for predicting the learning disability accurately.

## 1.5 Learning Disability

The term ‘Learning Disabilities (LD)’ is a relatively new one. It was first used by Dr. Samuel Kirk of Chicago, USA in 1963. The Children with Specific Learning Disabilities Act (USA) was passed in 1969. However, it was not until 1990s that the biological basis for LDs found support. Specific learning disability which includes dyslexia, dysgraphia, dyscalculia is commonly referred to as ‘Learning Disability’ or ‘LD’ in India. Now special educators for remediation are available in India [9].

A learning disability is found across all ages and socio-economic classes. It is not a type of mental retardation as sometimes mistakenly thought, in fact, IQ scores could fall in the very high range [10]. LD is a hidden handicap that affects academic achievement, vocational career and social life [11]. Every child born in this world gets the care and comfort from the family in which he or she is born. But the children with disability cannot enjoy and get such care and comfort from the family [11]. If a child grows, he develops into a worthy citizen. These children are able to quantify well and prove their worth. Similarly the society expects the same from the disabled students also, here the children failed. Specific learning disabilities have been recognized in some countries for much of the $20^{th}$ century, in other countries only in the latter half of the century, and yet not at all in other places [12]. These may be detected only after a child begins school and faces difficulties in acquiring basic academic skills.

Learning disability is a general term that describes specific kinds of learning problems. It is a neurological disorder that affects a child's brain and impairs his ability to carry out one or many specific tasks. The LD affected children are neither slow nor mentally retarded [12]. They have either normal or above average intelligence. A child with a learning disability is often wrongly labeled as being smart but lazy. A learning disability can cause a child to have trouble learning and using certain skills. The skills most often affected are: reading, writing, listening, speaking, reasoning and doing math. There is no cure for learning disabilities [13]. 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. They are life-long. However, children with LD can be high achievers. They can be taught ways to get around the learning disability. With the right help, children with LD can and do learn successfully.

Learning disability is a disorder in which a child has difficulty in 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 [14]. 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. 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. 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 labourer. 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.

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. 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 [15].

As many as 1 out of every 10 children in the United States has a learning disability. Almost 3 million children (ages 6 through 21) have some form of a learning disability and receive special education in school [12]. In fact, over half of all children who receive special education have a learning disability [15]. A learning disability often displays a cluster of characteristics over time, in various intensities, which interfere with his/her overall development and achievement. LD affected children can face unique challenges that are often spreading throughout their 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 as well as parents will be a part of the interventions. They can give aid to the children successfully in completing different tasks. School psychologists quite often help to design the intervention and coordinate the execution of the intervention with teachers and parents. With the right support and intervention, LD affected children can succeed in school and go on to be successful later in life. Social support is also a crucial component for these type children in the school system and should not be overlooked in the intervention plan. Parents of LD affected children often find themselves attempting to cope with a bewildering array of problems. Their children appear to be intelligent but they encounter all kinds of obstacles in school [16].

In India the term disability is used synonymously as impairment, and handicap or disability. These terms are different. The impairment means, the loss of physical or sense organs. The child has not able to see, it is disability.

Handicap is the result of impairment and disability. Learning disability is a broad term that covers a wide range of problems, including dyslexia and behavioral problems and the full range of ability. If a child having learning disability, that child requires special education needs. 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 [13].

This lifelong disability can interfere with the students' acquisition of academic and other basic skills necessary for survival as an independent adult. Some of the common signs of learning disabilities and learning disorders in children will be able to catch the problem early and take steps to get help to child. It is very important in paying attention to normal developmental milestones for toddlers and preschoolers. As early detection of developmental differences is an early signal of a learning disability and thus the problems that are spotted early can be easier to correct.

LD is real and a stumbling block for a nation's development process. The problems of children with specific learning disabilities have been a cause of concern to parents and teachers for some time. 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. Many types of assessment tests are available. Child's age and the type of problem determines the tests that child needs. A complete evaluation often begins with a physical examination and testing to rule out any visual or hearing impairment [15]. 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. 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.

Depending on the type and severity of the disability, interventions may be used to help the individual learn strategies that will foster future success. Some interventions can be quite simplistic, while others are intricate and complex. Teachers and parents will be a part of the intervention in terms of how they aid the individual in successfully completing different tasks. School psychologists quite often help to design the intervention and coordinate the execution of the intervention with teachers and parents. Social support can be a crucial component for students with learning disabilities in the school system and should not be overlooked in the intervention plan. With the right support and intervention, children with learning disabilities can become great success in school as well as later in the society.

### 1.5.1 Concept of Learning Disabilities

Learning Disability is a general term that refers to a heterogeneous group of disorders manifested by significant difficulties in the acquisition and use of listening, speaking, reading, writing, reasoning, or mathematical abilities. These disorders are intrinsic to the individual presumed to be due to central nervous dysfunction, and may occur across the life span [10] irrespective of regionwise.

Based on the frequency of occurrence, the important characteristics of learning disability are

(i) Disorders of attention
(ii) Perceptual impairments
(iii) General coordination deficits

(iv) Disorders of memory and thinking

(v) Specific difficulties in the areas of reading, arithmetic, writing and spelling and

(vi) Disorders of speech and hearing

Attention is the ability to concentrate on a task long enough to grasp its essential features [11]. Learning disabled children are having short attention span and lacking in concentration, but this does not mean that all inattenentive children are learning disabled. A child may have an attention disorder for many reasons. A significant characteristic frequently associated with learning disability in a child is hyperactivity, although the relationship between the two is still not clearly substantiated. Some expert refers to hyperactivity as an attention deflects hyperactivity disorder. Hyperactivity is a much misunderstood term and widely misused.

Children with learning disability are generally characterized by inefficient memory systems. There is little doubt that in the area of academic learning, children with learning disability are low achievers. They may have problem in the specific areas of reading, writing and mathematics, although it is not necessary that a child poor in spelling will also be poor in mathematics. Listening is a complex process that requires good attention, discrimination and memory systems. The child may use a lot of fillers such as articulation difficulties. Learning disabled individual could make grammatical errors while talking and use words incorrectly. Research in the area of learning disability has primarily focused on assessment and diagnosis.

There is no one sign that shows a child has a learning disability. Experts look for a noticeable difference between how well a child does in school and how well the child could do, given his or her intelligence or ability. There are certain general symptoms that may mean a child has a learning disability. A child probably won't show all of these signs, or even most of them. However, if a child shows a number of these problems, then parents and the teacher should consider the possibility that the child has a learning disability. Parents need to be especially alert to developmental delays as a child approaches school age [16].

When a child has a learning disability, he or she may show some symptoms. Some of these common symptoms are listed below [17];

(i) may have trouble learning the alphabet, rhyming words or connecting letters to their sounds;

(ii) may make many mistakes when reading aloud and repeat and pause often;

(iii) may not understand what he or she reads;

(iv) may have real trouble with spelling;

(v) may have very messy handwriting or hold a pencil awkwardly;

(vi) may struggle to express ideas in writing;

(vii) may learn language late and have a limited vocabulary;

(viii) may have trouble remembering the sounds that letters make or hearing slight differences between words;

(ix) may have trouble understanding jokes, comic strips and sarcasm;

(x) may have trouble following directions;

(xi) may mispronounce words or use a wrong word that sounds similar;

(xii) may have trouble organizing what he or she wants to say or not be able to think of the word he or she needs for writing or conversation;

(xiii) may not follow the social rules of conversation, such as taking turns, and may stand too close to the listener;

(xiv) may confuse math symbols and misread numbers;

(xv) may not be able to retell a story in order (what happened first, second, third); or

(xvi) may not know where to begin a task or how to go on from there.

### 1.5.2 Basic Types of Learning Disabilities

As pointed out earlier, learning disability is a lifelong disorder that affects the manner in which, individuals with average or above average, intelligence select, retain and express information. It reflects a difficulty in encoding and decoding information as it travels between the senses and the brain. Learning disabilities are also termed as ‘learning differences’, based on the fact that certain individuals learn differently - they are not unable to learn, but respond best to ways of learning that are different from traditional teaching methods.

Many gifted and talented children are often misdiagnosed as having learning disabilities or behavior disorders. This occurs because there are many characteristics of gifted children, both social and emotional, that are mistaken as symptoms of specific learning disorders. It is not uncommon for some gifted children, with IQ scores over 140 to display a significant discrepancy between Verbal IQ and Performance IQ and possess characteristics of a learning disability. Often gifted children have unusual learning styles, and even though they are very intelligent, they may also have learning disorders.

Learning disabilities tend to be diagnosed only when a child reach school age. This is because school focuses on the very things that may be difficult for the child — reading, writing and math, listening, speaking and reasoning. Teachers and parents notice that the child is not learning as expected. The school may ask to evaluate the child to see what is causing the problem. Parents can also ask for their child to be evaluated. With hard work and the proper help, children with LD can learn more easily and successfully. It's important to remember that a child may need help at home as well as in school.

A developmental lag might not be considered as a symptom of a LD until the child is older, but we can intervene early if we recognize it when the child is young. The activity of diagnosing the type of learning disability can be overwhelming and time consuming. Try not to get caught up in trying to determine the label or type of disorder and focus instead on figuring out how best to support the child. The checklist, organized by skill set and age group, can help in evaluating the child's signs and symptoms and indicate whether seek further assistance from a teacher or professional skilled in diagnosing learning disabilities. Types of learning disabilities are often grouped by school area skill set or cognitive weakness. If the child is in school, it will probably be apparent if he or she is struggling with reading, writing, or math, and narrowing down the type will be easier.

The common types of Learning Disabilities are explained below;

i. **Dyslexia-** Difficulty processing language- Problems in reading, writing, speaking.

   It is the most common type of LD affecting in children. It refers to a specific difficulty in the area of reading. The terms generally used instead of dyslexia are severe reading disability, primary reading disability, specific reading disabilities and word blindness. It has been estimated that of the children who attend school, approximately 10% to 15% have some difficulty in reading and 85% to 90% percentage of all learning disabled children have reading problems. Boys with reading problem outnumber girls at the surprising rate of 4 to 1. In fact dyslexia has becomes synonymous with learning disability to such an extent that it has been suggested that dyslexia should be used as an umbrella term for all learning disability in general. All learning disabilities are not dyslexia and the same time dyslexia does not exemplify all learning disabilities. Dyslexia is a broad term describing a reading disability. The word dyslexia implies the meaning *difficulty with words*. It affects a child's ability to read. It is neurological in origin, likely to be present at birth and its effects are life-long. It is characterized by difficulties with accurate and /or fluent word recognition and by poor spelling and decoding abilities. These difficulties typically result from a deficit in the phonological component of language that is often unexpected in relation to other cognitive abilities and the provision of effective classroom instruction [10].

Learning disabilities in writing can involve the physical act of writing or the mental activity of comprehending and synthesizing information. Expressive writing disability indicates a struggle to organize thoughts on paper. Symptoms of a written language learning disability revolve around the act of writing and include. They include problems with neatness and consistency of writing, accurately copying letters and words, spelling consistency, writing organization and coherence. Signs of a language based learning disorder involve problems with verbal language skills, such as the ability to retell a story and the fluency of speech, as well as the ability to understand the meaning of words, parts of speech and directions.

**ii. Dyscalculia** - Difficulty with math- Problems doing math problems, understanding time, using money.

In 1919, Henschen reported that number blindness could occur independently of specific reading disability [10]. Some persons observed that memory and order disorders frequently occurred along with numerical problems. Dyscalculia may result from lesions in widely disparate regions of the brain. Dysfunctions associated within left temporal lobe were characterized by difficulties with complex operations involving a sequence of steps or mental or oral calculation or reasoning. These disorders are termed as secondary arithmetic disturbance. Children with learning disabilities exhibit a variety of deficits in the area of mathematics. Shape discrimination, size discrimination, sets and numbers and counting are the different areas of dyscalculia.

Learning disabilities in math vary greatly depending on the child's other strengths and weaknesses. A child with a math based learning disorder may struggle with memorization and organization of numbers, operation signs, counting principles or have difficulty telling time.

iii. **Dysgraphia -** Difficulty with writing- Problems with handwriting, spelling, organizing ideas.

Disorders of written language are referred to as dysgraphia and this include difficulties in three areas of handwriting, spelling, content [10]. Many learning disabilities authorities have expounded why writing is so important and how it can be taught to learning disabled students [18]. The key characteristics associated with dysgraphia includes the main areas of written language –handwriting, spellings and the content. Usually these are interlinked problems and it is expected that a child having difficulties in any one of these areas may experience a spill over in the others too. Most disabled children hate to write and avoid it wherever it possible. Lack of motivation becomes a real obstacle and a teacher needs to be at her creative best when encouraging learning disabled child to write.

iv. **Dyspraxia -** Difficulty with fine motor skill- Problems with hand–eye coordination, balance, manual dexterity.

Motor difficulty means problems with movement and coordination whether it is with fine motor skills like cutting, writing, etc. or gross motor skills like running, jumping etc. A motor disability relates to the output of information from the brain. In order to run, jump, write or cut something, the brain must be able to communicate with the necessary limbs to complete the action. Signs that the child might have a motor coordination disability include problems with physical abilities that require hand–eye coordination, like holding a pencil or buttoning a shirt.

v. **Auditory Processing Disorder -** Difficulty hearing differences between sounds- Problems with reading, comprehension, language.

The ability to hear things are greatly impacts the ability to read, write and spell, but an inability to distinguish slight differences in sound or hearing sounds at the wrong speed make it hard to explore words and understand the basic concepts of reading and writing. Problems in visual perception include missing slight differences in shapes, skipping lines, skipping words, reversing letters or numbers, misperceiving depth or distance or having problems with eye–hand coordination. In addition to the above, the other disorders that make learning difficulty are anxiety, stressful events, emotion, depression and other conditions affecting concentration make learning more of a challenge.

vi. **Visual Processing Disorder -** Difficulty interpreting visual information- Problems with reading, math, maps, charts, symbols, pictures.

vii. **Attention Deficit Hyperactivity Disorder (ADHD)**

It is a neurobiological disorder that starts early in childhood and can continue into adulthood. The disorder is characterized by a delay or permanent inability to self-regulate behavior or to control behavioral responses.

**viii. Social And Emotional Difficulties**

Sometimes kids have trouble expressing their feelings, calming themselves down and reading nonverbal cues, which can lead to difficulty in the classroom and with their peers. Social and emotional skills are an area where the parent can have a huge impact. For all children, but especially those with learning disabilities, social and emotional skills are the most consistent indicators of success, outweighing everything else, including academic factors. Certain personal qualities and social relationships are very much needed by the learning disabled children to make a satisfactory adjustment to post school life [18].

### 1.5.3 Assessment and Remediation

LD can be a lifetime condition. There may be several apparent overlapping learning disabilities in some children while others may have a single, isolated learning problem that has little impact on their lives. LD is diagnosed by a qualified child psychologist in association with a pediatrician. The process of diagnosing a learning disability can be confusing. It shall be started with the child's school. It involves testing, history taking and observation by a trained specialist. A series of tests may be required to be done to identify the affected areas. Special education brings some solution to the problems the children affected with LD. Finding a reputable referral is important. Since the educational needs of such children are different they will be given special academic sessions in an integrated set up. Other professionals such as speech therapist for children with speech disorder and physiotherapists to those with motor deficits may be required. In some case psychotherapist may also be involved in the treatment process.

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. Periodic evaluation of such children done to check for improvement will help so that they can be integrated into the mainstream society as soon as possible.

### *1.5.3.1 Assessment of LD*

Assessment is the systematic process of collecting information about a child, his past and current levels of performance his strength and weakness, in order to help make better education decisions [10]. Assessment needs to be relevant to the teaching goals and interventions that the child will receive. Assessment is directly linked with how one will go about helping the child. It is linked with intervention methods. The information collected through the assessment must be relevant and of practical help in the class room. Assessment helps parents to better understand their child's problems and adjust their expectations on the basis of the assessment data. It is pertinent to note that, in India the history of LD assessment is still in its infancy.

Many types of assessment tests are available. Child's age and the type of problem determine tests that child needs. Before any formal testing, a conference is usually arranged between the child's parents and representatives from the special education department. A factor that prevents accurate diagnosis of twice exceptional is the prevalent practice of comparing gifted children with the norms for average children. In psychology, as well as in other therapeutic fields, such as audiology, speech pathology, occupational therapy and optometry, the diagnostic question that is usually asked is how this child's performance compares with the norm. If the child scores within the normal range, no disabilities are detected.

The purpose of any evaluation or assessment for LDs is to determine child's strengths and weaknesses and to understand how he or she best learns and where they have difficulty. A major factor that makes it difficult to assess a learning disabled child is the confusing nature of the disability itself. The absence of testing instrument relevant to Indian students is another major drawback. Most tests are designed for native English speakers and have items which lie outside the cultural experience of the average Indian student. Assessment can be expensive too.

Just as there are many different types of LDs, there are a variety of tests that may be done to pinpoint the problem. Examples of the types of testing and evaluations include:

(i) Interviews and direct observation of the child.

(ii) Review of the child's educational and medical history.

(iii) Parent conferences (a parent's input is of great importance because the parent have observed the child since birth and can provide important details about his or her growth and development).

(iv) IQ testing and psychological evaluation usually performed by a psychologist or psychiatrist.

(v) Developmental history or social assessment.

(vi) Speech and language evaluation usually performed by a speech therapist.

(vii) Evaluation of fine motor skills, visual-motor integration and sensory integration usually performed by an occupational therapist.

(viii) Evaluation of gross motor skills, muscle tone and balance usually performed by a physical therapist.

Two approaches are generally used to describe testing or the assessment. They are formal and informal testing. The former uses standardized testing while the latter uses non-standardized testing [10]. The standardized test demands a high degree of uniformity in administration and interpretation. They allow comparison of students of the same age or grade and can be use individually or in groups. Informal tests or assessment are non-standardized procedure used by teachers and other professionals to collect information. The advantage of informal test is that they are simple to construct, administer and score. They targeted at a specific objective and are of greater value for making instructional decisions. The most frequently used informal procedures are observations, interviews, questionnaires and tests. Check lists and rating scales are another way of collecting behavioral data.

Generally, there are about 40 characteristics of LD, any one or more is found in LD affected children. These characteristics in the form of a check lists are used in the traditional methods of assessment of LD. These assessment questions are formed based on discussions made with professionals/doctors engaged in LD assessment fields. These 40 characteristics of LD, expressed in the form of a questionnaire are listed below [17];

1. Do you find words difficult to spell?
2. Are you unsure about the use of full stops capital letters etc.?
3. Do you feel unsure about how to tackle reading unknown words?
4. Do you get very frustrated with your own performance?
5. Do you find it difficult to understand what you have read the first time?
6. Are you unsure how to organize writing a letter, a report or an essay?
7. Do you limit your writing to words you know you can spell?

8. Do you feel you have to read every word on the page?
9. Can you spell a word one day and forget it the next?
10. Do you miss deadlines because you didn't start early enough?
11. Do you find it difficult to listen and write at the same time?
12. Do you find it difficult to take message on the telephone?
13. Do you have difficulty in telling the difference between sounds?
14. Do you read slowly?
15. Do you have to 'see' and 'feel' if a word looks right?
16. Do you mix up dates and appointments so that you miss them or are double booked?
17. Do you get numbers mixed up? e.g. roll numbers, telephone numbers?
18. Do you panic when you get to unknown words?
19. Do you confuse the order of the months of the year?
20. Do you remember words one day and forget them the next?
21. Do you find it difficult to follow written instructions?
22. Do you have days when things go dramatically wrong and equally days when things go really well?
23. Do you find organizing your thoughts difficult?
24. When writing do you think that you end up with really what you wanted to say?
25. Is map reading difficult?
26. Do you find it difficult to remember what you have read?
27. Do you find you have too many ideas to maintain focus?

28. Do you say the wrong word in the wrong place or at the wrong time?
29. Do you get the order of the letters wrong when writing?
30. Do you think your spelling weak?
31. Do you find organizing your life difficult?
32. Do you have difficulty telling left from right?
33. Does the print blur or move around as you are reading?
34. Do you find it difficult to skim and scan for information?
35. Is it difficult to see the detailed steps needed to complete a task?
36. Do you forget things quickly?
37. Do you have problems saying long words?
38. Do you lack confidence in yourself and think others are better than you?
39. Do you find it confusing putting sounds together to pronounce words?
40. Do you dislike reading aloud?

In the present research work, the method of informal assessment is adopted for designing the tool for predicting the learning disabilities in children. Even though different types of checklists are generally available for assessing LD characteristics, a check list containing the 16 most frequent and important characteristics (signs & symptoms) of LD collected from the above 40 general characteristics, after eliminating the unwanted and redundant ones, is prepared suiting to the LD conditions generally prevailing in Kerala. This general check list adopted in this research work is shown in Table 1.1 below. As this check list is prepared based on the experience of professionals including outside India, the same is not region wise [24]. This check list is used as first phase of LD assessment. Based on which, new check list is developed incorporating additional symptoms, viz. sub attributes, with different scores for each, as discussed in Chapter 3. The new check list developed given at Table 3.3 is used in the study of determination of LD, which can be used irrespective of region. The increase in number of characteristics, if any, will merely increase the prediction time.

**Table 1.1** General check list

| Sl. No. | Indicators of LD | Yes | No |
|---|---|---|---|
| 1 | Difficulty with Reading | | |
| 2 | Difficulty with Spelling | | |
| 3 | Difficulty with Handwriting | | |
| 4 | Difficulty with Written Expression | | |
| 5 | Difficulty with Basic Arithmetic skills | | |
| 6 | Difficulty with Higher Arithmetic skills | | |
| 7 | Difficulty with Attention | | |
| 8 | Easily Distracted | | |
| 9 | Difficulty with Memory | | |
| 10 | Lack of Motivation | | |
| 11 | Difficulty with Study Skills | | |
| 12 | Does Not like School | | |
| 13 | Difficulty in Learning a Language | | |
| 14 | Difficulty in Learning a Subject | | |
| 15 | Slow To Learn | | |
| 16 | Repeated a Grade | | |

The general check list can be easily used by the parents and teachers. It is used to investigate the presence of learning disability. It includes the general indicators of learning disability and focuses on understanding of the learning disability. The goal is to provide concise and accurate set of diagnostic attributes which can be implemented in a user friendly and automated fashion. If a child has unexpected problems learning to read, write, listen, speak, or do math, then teachers and parents may want to investigate more. The same is true if the child is struggling to do any one of these skills. The child may need to be evaluated to see if he or she has a learning disability.

The expert trained to do psychological testing and result interpretation are clinical psychologists, school psychologists, educational psychologists, developmental psychologists, neuropsychologists, occupational therapists, speech and language therapists. Several professionals coordinate services as a team to obtain an accurate diagnosis, including input from the child's teachers. Recommendations can then be made for special education services or speech language therapy within the public school system. If the public school is not working out, then a nonpublic school provide specializes in treating learning disabilities might be a good alternative. A professional learning disorders specialist might refer to the importance of integration to learning.

#### *1.5.3.2 Remedial solutions*

Although several products are available for the identification and remediation of learning disabilities, most of these are either unable to sustain the progress of a disabled child or not aligned to government standards [19]. Pre-assessment helps separate the regular performers from the children who have special needs. Parents and teachers can and should take an active role in the child’s education.

All children can be both exhilarating and exhausting, but it may seem that the child with a learning disability is especially so. Parents and teachers may experience some frustration trying to work with the child, and it can seem like an uphill battle.  After the parents learn what their specific learning disability is and how it is affecting their behavior, they will be able to start addressing the challenges in school and at home. A child with a learning disability cannot try harder, pay closer attention, or improve motivation on their own; they need help to learn how to do those things [20].

As a parent, discovering that something may stand in the way of the child's success can be unsettling and difficult. Whether or not the child has a learning disability, remember that the way the parents and teachers behave and what they do has the most impact on the child's chances of success. Everyone faces obstacles and the most important thing the parents and teachers can show the child, apart from the consistent love and support and is how to deal with obstacles. A good attitude won't solve the problem, but it can give the child hope and confidence that things can improve. The first task as the parent of a child with a learning disability is to recognize that there are many things the parents can do to help the child.

Parents often teach children to compensate for weaknesses, and gifted children learn compensation strategies more quickly than their less capable peers. What they neglect to tell the children is that compensation can break down under various conditions. It takes more energy to compensate and when one is fatigued, ill, stressed, dieting too strenuously, or adjusting to a new situation, there may not be sufficient energy to support the compensation strategy. So the individual is likely to experience good days when the compensations work well, and bad days when they fail. They need to understand that their high intelligence is revealed on their good days and that there will be bad days, when their compensations, like bad brakes, fail to support them.

Special education brings some solution to the problems of children affected with LD [10]. They have to be given special academic sessions in an integrated set up as their educational needs are different. Periodic evaluations of such individual children have to be done, to check for improvement, which will help them to integrate them into the mainstream society as soon as possible. Nowadays, meeting the challenges in schools, most of the school authorities are compelled to appoint resource persons/counselors in LD to diagnose the same and attain early detection and remedies to improve the school atmosphere and results.

## 1.6 Objectives of the Present Study

The following are the objectives of the research work:

i. Identifying the important parameters of LD using data mining techniques,

ii. Identifying the hidden relationship between the symptoms of LD,

iii. Estimating the relative significance of each symptoms of LD,

iv. Developing a new method for improving the accuracy of classifiers,

v. Developing fuzzy and rough set models in LD prediction; and

vi. Design a tool based on machine learning techniques for prediction of LD.

## 1.7 Scope of Work

The scope of research work includes the following;

i. Collection and preparation of real world data set,

ii. Knowledge extraction from domain experts,

iii. Prediction of LD with the help of machine learning algorithms,

iv. Identifying the frequent symptoms of LD,

v. Development of new approaches for LD prediction,

vi. Identification of the problems related to the classification accuracy,

vii. Development of new algorithms for overcoming the identified problems,

viii. Application of new algorithms developed on existing classifiers,

ix. Developing new models in LD prediction with fuzzy and rough set approaches

x. Development of an integrated knowledge based tool for the prediction of learning disability,

xi. Determination of the performance of the developed tool; and

xii. Applicability and use of the developed tool in LD prediction.

## 1.8 System Framework

The main aim of the proposed research work is to design a tool based on machine learning techniques for accurate prediction of learning disability in school-age children and to effectively measure the percentage of LD present in the child, according to the knowledge obtained from the clinical information. The basic idea of the proposed knowledge based system is to detect the LD at an early stage and to increase the accuracy of the learning disability assessment and reduce the time consumed for the same. For achieving these goals, different statistical machine learning technique are used. Towards the achievement of the full objectives, a learning disability tool is constructed. The proposed tool has many advantages compared to the traditional methods of learning disability determination using check lists. These advantages include; consumption of less manpower and time, the accuracy and efficiency underlying and handling of missing values and redundant data. For the assessment of LD, check lists containing the signs and symptoms of LD (attributes) are used. Obtaining a full fledged information table by interviewing a child will not be a success in all time as the same depends on the mood of the child at the moment of interview. In such occasions, the data may contain missing values. To overcome such a problem, missing value imputing by applying closest fit algorithm and correlation based new methods developed are adopted in this research. Also there are cases of most of the attributes in the data are unwanted and/or redundant. To overcome this problem, dimensionality reduction method using Principal Component Analysis (PCA) is applied.

The learning disability prediction tool is designed using the five classification methods viz. modified neural network, modified decision tree, fuzzy, fuzzy with reduced attributes and ANFIS. After applying the data preprocessing on the data set, using closest fit algorithm or correlation based method, as the case is, and PCA, the classification methods viz. decision tree, fuzzy and ANFIS are applied.

In the developed tool, 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, the redundant data is removed, the number of attributes is reduced and the missing values are imputed. The designed tool is very user friendly. It provides student information record which stores the details of the children in the student database, which can be retrieved as and when required.

## 1.9 Thesis Contributions

This thesis makes several major contributions. Based on the statistical machine learning tool developed, the presence of learning disability in any child with its percentage can be determined. The class of LD like low, minor and major and the percentage of LD in each class can also determined by this tool. The research work provides the new insights into the interrelationships between symptoms of LD, their relative importance and estimating the significance of each symptoms of LD. It contributes in developing new models of LD prediction using fuzzy and rough sets and it succeeds in modifying the data preprocessing with J48 decision tree and neural network for LD prediction. The developed tool classifies LD as well as imputes the missing values in the data set accurately. The work developed new algorithm based on correlation for imputing missing values. The missing value imputing, done by the developed tool contributes strong classification results. The number of attributes is reduced by eliminating the unwanted ones by using PCA, helps in reducing the time. The tool developed gives accurate results in lesser time compared to the traditional assessment methods using check lists. The developed tool is very effective for finding the LD affected child from the large database. This research work has also considered an approach to handle learning disability database to predict frequent symptoms of the learning disabilities in school aged children. Early identification of LD will help the parents and school authorities to recommend the child for early remediation, which will ultimately help them to provide the child with best environment for his success. The study will certainly contribute in the development of the nation as LD is a real stumbling block for a nation’s development process.

Parts of this thesis work have been previously published as peer reviewed journal papers and conference papers as shown in the list of publications below.

### 1.9.1 List of publications

1. Julie M. David, Kannan Balakrishnan: A New Decision Tree Algorithm for Prediction of Learning Disabilities, Journal of Engg. Science and Tech., School of Engg., Taylor's University, Malaysia. (Accepted for publication, 8(3), June 2013).

2. Julie M. David, Kannan Balakrishnan: Attribute Reduction and Missing Value Imputing with ANN: Prediction of Learning Disabilities, Int. J. of Neural Computing and Applications, Springer-Verlag London Limited, DOI: 10.1007/s00521-011-0619, 21(7), Oct.2012, Online First, May 2011. Available at - http:// www.springerlink.com/ content/ evj556771 k561681/

3. 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. Available at - http:// www. ijcsit.com/ ijcsit-v2issue2.php

4. 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 Engg. 3(1), Feb. 2011, pp163-169. Available at-http://www.ijcee.org/abstract/308-JE 351. htm

5. 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, pp 7-14. Available at - http://www.ijcaonline.org/archives/volume9/ number 11 /1432-1931

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. Available at - http://airccse.org/journal/ijaia/currentissue.html

7. Julie M. David, Kannan Balakrishnan: Performance of Classifiers - Significance of New Method in Missing Value Imputation and Dimensionality Reduction: Prediction of Learning Disabilities in School-Age Children, Int. J. of Applied Soft Computing, Elsevier B.V. (communicated).
8. Julie M. David, Kannan Balakrishnan: Learning Disability Prediction Tool using ANN and ANFIS, Int. J. of Transaction on Knowledge and Data Engineering, IEEE. (communicated).
9. Julie M. David, Kannan Balakrishnan: Fuzzy and Rough Set Approaches for Learning Disability Prediction, Int. J. of Artificial Intelligence, Elsevier B.V. (communicated).
10. Julie M. David, Kannan Balakrishnan: Modified Pre-Processing Methods with Machine Learning Algorithms for LD prediction, Int. J. of Human Computer Studies, Elsevier B.V. (communicated).
11. Julie M. David, Kannan Balakrishnan: Prediction of Learning Disabilities in School Age Children using Decision Tree, Proc: of the Int. Conf. on Recent Trends in Network Communications, Lecture Notes in Communication in Computer and Information Science, 90(3), Springer- Verlag Berlin Heidelberg, DOI : 10.1007 / 978-3-642- 14493-6_55, 2010, pp533-542. Available at-http://www.springerlink.com/content/q 66788 rhk 77w7w15//
12. Julie M. David, Kannan Balakrishnan: Prediction of Frequent Signs of Learning Disabilities in School Age Children using Association Rules, Proc: of the Int. Conf. on Advanced Computing, ICAC 2009, MacMillion Publishers Ltd., New York City, ISBN 10:0230-63915-1, ISBN 13:978-0230-63915-7, 2009, pp202–207.

13. Julie M. David, Pramod, K.V.: Prediction of Learning Disabilities in School Age children using Data Mining Techniques, Proc: of AICTE Sponsored National Conf. on Recent Developments and Applications of Probability Theory, Random Process and Random Variables in Computer Science, Thrivikram,T., Nagabhushan, P., Samuel, M.S., (eds), 2008, pp139-146.

## 1.10 Road Map

The rest of this thesis is organized as follows;

A detailed literature survey is given in Chapter 2. Apart from the introduction, the chapter contains detailed literature review on learning disability as well as soft computing methods. The chapter ends with summary and conclusion.

Chapter 3 is detailing about the data collection and implementation of various soft computing methods. It begins with an introduction. The chapter explains about technical background, data collection, data sets, data distribution, knowledge extraction process, data normalization, attributes normalization and entropy of LD attributes. The various implementation methods adopted in this study such as neural network with back propagation algorithm, decision tree with J48 algorithm, support vector machine with sequential minimal optimization algorithm, bagging and fuzzy model are explained well with comparison and results and pointed out about the insights of LD. The chapter ends with summary and conclusion, and contributions.

Chapter 4 deals with improving performance by new pre-processing methods. After the introduction, imputing missing values with closest fit algorithm and with correlation based new algorithm developed, dimensionality reduction with PCA, and modified data pre-processing and performance evaluation with MLP, decision tree, fuzzy model and neuro fuzzy model with reduced attribute are dealt with in detail. The chapter also contains details of study conducted on rough set model, apriori algorithm and clustering with k-means algorithm. This chapter ends with comparisons and results, summary and conclusions, and contributions.

Chapter 5 is development of an integrated knowledge based tool for LD prediction. After the introduction part, the chapter contains system flowchart, architecture of the tool and design of the tool. Tool testing, various screen shots and performance evaluation of the designed tool are also covered in this chapter. The chapter ends with summary and conclusions, and contributions.

The thesis concludes with Chapter 6, containing the summary of the work and contributions of the overall study. This chapter also discusses the future works and final conclusions.

## 1.11 Contemporary Works

There are only little studies available in the area of LD prediction with knowledge based theories, as mentioned below.

Tung-Kuang Wu, et. al. in 2008, studied two well-known artificial intelligence techniques, artificial neural network and support vector machine, to the LD diagnosis problem [24]. This study is based on the formal assessment of LD whereas the present research work relates to informal assessment of LD. 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 [25]. As dyslexia is only a type of LD, the present research on general assessment of LD is entirely different from their study.

## 1.12 Summary and Conclusion

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. However, children with LD can be high achievers and can be taught ways to get around the learning disability.

Individuals with learning disabilities can face unique challenges that are often persistent throughout the lifespan. Depending on the type and severity of the disability, interventions may be used to help the individual learn strategies that will foster future success. Some interventions can be quite simplistic, while others are intricate and complex. Teachers and parents will be a part of the intervention in terms of how they aid the individual in successfully completing different tasks. School psychologists quite often help to design the intervention and coordinate the execution of the intervention with teachers and parents. Social support can be a crucial component for students with learning disabilities in the school system and should not be overlooked in the intervention plan. With the right support and intervention, children with learning disabilities can succeed in school and go on to be successful later in life.

When a LD is suspected based on parent/teacher observations, a formal evaluation of the child is necessary. Parental consent is needed before a child can be tested. Many types of assessment tests are available. Child's age and the type of problem determines the tests that child needs. A complete evaluation often begins with a physical examination and testing to rule out any visual or hearing impairment. Many 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. 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.

At present, the methods available for LD prediction and determination are quite less and not accurate. The identification of such problems are really challenging. In this research work, data mining using machine learning techniques are used to analyze the symptoms of LD, establish interrelationships between them and evaluate the relative importance of these symptoms. The diagnostic accuracy of learning disability prediction is increased by way of a knowledge based tool based on statistical machine learning.  Identifying the important parameters of LD prediction using the data mining techniques, identifying the hidden relationship between the symptoms of LD and estimating the relative significance of each symptoms of LD are also covered. The developed tool has many advantages compared to the traditional methods of using check lists in determination of learning disabilities.

*Chapter*

# 2

# LITERATURE REVIEW

Contents



## 2.1 Introduction

Studies have conducted in the field of Learning Disabilities recently. In this chapter, 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. This survey, through the research work, helps in determination of the characteristics and uses of various classifiers in prediction and classification. It also helps in determining the necessity of development of a knowledge based tool for the assessment of learning disabilities. The detailed literature survey carried out is explained in the ensuing sections.

## 2.2 Literature Survey on Learning Disability

Kenneth A. Kavale, studied about identifying specific learning disability in the year 2005 [21]. 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 [22]. 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 [23] 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.

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 [24]. 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 research work relates to informal assessment of LD.

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 [25]. As dyslexia is only a type of LD, the present research on general assessment of LD is entirely different from their study.

## 2.3 Literature Survey on Soft Computing Methods

Finding frequent item sets is one of the most investigated fields of data mining. The apriori algorithm is the most established algorithm for frequent item sets mining (FIM). Several implementations of the apriori algorithm have been reported and evaluated. One of the implementations optimizing the data structure with a trie by Bodon was studied by Yanbin Yer et.al in 2006 [26]. The effect where input transactions are read by a parallel computer is studied by them.

In another study by Selma and Altay in the year 2001, an algorithm for finding frequent item sets in transaction databases is proposed [27]. The basic idea of the algorithm is inspired from the direct hashing and pruning (DHP) algorithm, which is in fact a variation of the well-known Apriori algorithm. Another paper proposes an efficient SMine (Sorted Mine) Algorithm for finding frequent item sets. This method proposed by Jeba and Victor in 2011, reduces the number of scans in the database [28]. The proposed SMine algorithm works well based on graph construction.

Several studies have applied neural networks in the diagnosis of cardiovascular diseases, primarily in the detection and classification of at-risk people from their ECG waveforms. In the works of Cellar and Chazel [29] in 1998, the application of neural networks to classify normal and abnormal ECG waveforms were studied. The prediction of heart diseases, blood pressure and sugar with the aid of neural networks was proposed by Niti Guru et al. [30] in 2007. In 2008, Xianjun Ni researched in detail, the data mining on neural networks and studied about the combination of data mining method and neural network model for improving the efficiency of data mining methods [31].

Regarding study in decision tree, Hongcui wang and Tatsuya Kawahara, in the year 2008 studied on effective error prediction using decision tree for automatic speech recognition grammar network in call system [32]. Anupama Kumar and Vijayalakshmi, M., studied on Efficiency of decision trees in predicting student's academic performance in 2011 [33].

In a study on SVM, Huilei Xu, Ihor R. Lemischkaand Avi Ma'ayan, conducted a research on SVM classifier to predict genes important for self-renewal and pluripotency of mouse embryonic stem cells in 2010 [34]. In another study, in the year 2011, Betkier, et. al. [35] studied on a prediction of protein phosphorylation sites using classification trees and SVM classifier.

In the area of clustering, Antonia Kyriakopoulou and Theodore Kalamboukis, in 2010, studied on text classification using clustering [36]. Another study on combining clustering with classification for spam detection in social bookmarking systems conducted in the year 2008 by the same authors [37].

Kristina Machova, et. al studied in the year 2006, on bagging methods using decision trees in the role of base classifiers [38]. They describe a set of experiments with bagging, which can improve results of classification algorithms.

Different approaches from rough sets theory are demonstrated on selecting values for the individual interpreted meanings. In 2005, Zhu and Wu introduced solutions on processing missing attribute values by considering the attribute cost [39]. They suggest that it is expensive to predict all the missing attributes, therefore a technique is needed to balance the prediction percentage, the prediction accuracy and the computational cost. Jiye Li and Cercone, N. investigated the effectiveness of assigning missing attribute values from rough sets perspective in the year 2006 [40]. They are of the opinion that, comparing to the closest fit approach proposed by Grzymala-Busse, their new RSFit approach significantly reduces the computation time achieved comparable accuracy. Shichao Zhang, et. al in 2008 proposed an efficient nonparametric missing value imputation method based on clustering, called CMI (Clustering-based Missing value Imputation), for dealing with missing values in target attributes [41].

Apart from the above, many other researchers focused on the topic of imputing missing values. Chen and Chen [42] 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 [43] 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 [44,45,46]. 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 [47] 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 [48], a method for handling inapplicable and unknown missing data [49], different substitution methods for replacement of missing data values [50], robust Bayesian estimator [51], and nonparametric kernel classification rules derived from incomplete or missing data introduced by Pawlak in 1995 [52]. 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 [53]. 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 [54,55]. Chanas and Kutcha in 1992 studied on further remarks on the relation between rough and fuzzy sets, [56] 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 [57]. On a study on inducing fuzzy models for student classification, in the year 2006, Ossi Nykanen, [58] 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 [59].

Fuzzy logic has not been applied to defining ranking function directly; however, fuzzy set model has been used to define fuzzy queries [60], fuzzy relationships between query terms and documents [61.62]. 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 [63].

In the year 2009, Ramasubramanian, P.,et.al [64], analyzed the technique for concept map method of teaching by using rough set theory. The study helps the teacher to find what the lack in concepts of students is and immediately the teacher can supply to them. Roman W. Swiniarski proposed, in the year 2001, [65] a rough sets method and its role in feature selection for pattern recognition. He proposed a sequence of data mining steps, including application of SVD, PCA and rough sets for feature selection.

## 2.4 Summary and Conclusion

The studies conducted in the field of learning disabilities are quite few relative to any other field and that with knowledge based theories are negligible. The available studies on LD are conducted based on formal assessment method while our work is entirely different from these, as we are using the informal method in learning disability assessment. In the best of knowledge, nobody has conducted such a study. The preset literature survey helped us in understanding the characteristics, working methodology and uses of various soft computing methods.

Chapter 3

# DATA COLLECTION AND IMPLEMENTATION OF VARIOUS SOFT COMPUTING METHODS

Contents



## 3.1 Introduction

Data mining is an essential step in the process of knowledge discovery. It consists of an iterative sequence of steps namely data cleaning, data integration, data selection, data transformation, data mining, pattern evaluation and knowledge presentation. Data preprocessing is the pre stage of data mining. The data is ready for the mining process only after completion of data preprocessing. Data preprocessing involves data cleaning, data reduction, data integration and data transformation.

In data cleaning, the data is cleaned by filling in missing values, smoothing noisy data, identifying or removing outliers and resolving inconsistencies. Data reduction is the process to obtain a reduced representation of the data set. It is much smaller in volume but it produces the same analytical results. It consists of dimensionality reduction and numerosity reduction. Data integration is the merging of data from multiple data stores. This will help to reduce and avoid redundancies and inconsistencies in the resulting data sets. It will help to improve the accuracy and speed of the subsequent data mining process. In data transformation, data is transformed or consolidated into forms appropriate for mining. The data mining or knowledge discovery process is shown in Figure 3.1 below [66].

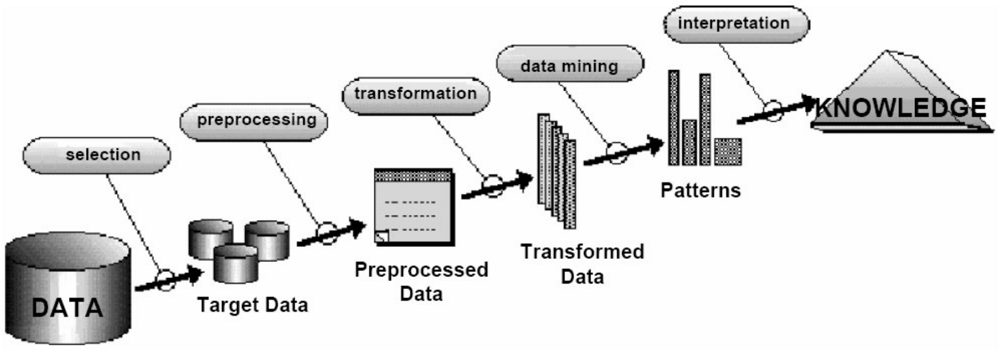


**Figure 3.1** Data mining process

Classification is a form of data analysis that extracts models describing important data classes. Such models called classifiers and predict categorical class labels. Classification consists of two step process. In the first step, a classification model based on previous data is build. The model describes a predetermined set of data classes or concepts [67]. This is the learning step or training phase, where a classification algorithm builds the classifier by analyzing or learning from training sets made up of database tuples and their associated class labels. In the second step, the accuracy of classifier is determined by using the model to classify new data. Using this model, first the predictive accuracy of the classifier is estimated by the help of a test set made up of test tuples and their associated class labels. The accuracy of a classifier on a given test set is the percentage of test set tuples that are correctly classified by the classifier.

The fuzzy classification is a natural extension of the traditional classification. In a sharp classification, each object is assigned to exactly one class, meaning that the membership degree of the object is 1 in this class and 0 in all the others. The belonging of the objects in the classes is therefore mutually exclusive. In contrast, a fuzzy classification allows the objects to belong to several classes at the same time; furthermore, each object has membership degrees which express to what extent this object belongs to the different class [68]. The fuzzy classification approach can be used for instance for diagnosis and for decision making support. In the case of a diagnosis system for ill persons, the classification procedure can derive the illness based on the symptoms of the patient. In a decision making process, the classification is used to derive management decisions based on several characteristics of the objects. A major issue in this field is the complexity of the data. This complexity is a source of uncertainty due to the limited capability of human beings to observe and handle large amounts of data simultaneously. The fuzzy classification, in contrast to the classical one, by allowing objects to belong to several classes at the same time, reduces the complexity of the data and also provides much more precise information about the classified elements. Among the different approaches to intelligent computation, fuzzy logic provides a strong framework for achieving robust and yet simple solutions [69]. Fuzzy logic can be further strengthened by the introduction of learning capabilities, such as those of artificial neural networks.

In this chapter, the technical background and problem statement of the research work along with different sources and methods of data collection, the data sets, data distribution and normalization and attribute normalization are explained. It is also well explained, in this chapter, about the implementation of different classification algorithms, viz, neural network, decision tree and support vector machine, ensembled method-bagging and fuzzy model for the prediction of learning disabilities in children.

Two aspects are considered in the study of prediction of LD. First, in this chapter, the problems or drawbacks existing in the above classification and secondly, in the next chapter, a successful attempt to use these algorithms effectively by modifying the data set at the pre-processing stage. The data mining tool weka, is used for the implementation of these different classification algorithms using 1020 datasets.

## 3.2 Technical Background

The research work in prediction of learning disability is implemented in two phases. In the first phase, the implementation is carried out through various existing algorithms to determine the drawbacks of these algorithms and also to see the insights of LD. In the second phase of implementation, modified data pre-processing with classification is used to recover the draw back of the algorithms that determined in the first phase and also new systems for LD prediction, such as fuzzy, neuro fuzzy and rough set are developed. The second phase also consists of designing a tool in LD prediction for social purpose.

The first phase of the work is implemented using weka, a popular suite of machine learning software written in java, developed at the University of Waikato, New Zealand. Weka is freely available on the World-Wide Web and accompanies a new text on data mining which documents and fully explains all the algorithms it contains [70]. Weka contains a collection of visualization tools and algorithms for data analysis and predictive modeling together with graphical user interfaces for easy access to this functionality. The advantages of weka includes free availability and portability, since it is fully implemented in the java programming language and thus runs on almost any modern computing platform. Weka supports several standard data mining tasks specifically data preprocessing, clustering, classification, regression, visualization and feature selection. All of weka’s techniques are predicted on the assumption that the data is available as a single flat file or relation, where each data point is described by a fixed number of attributes. Weka provides access to SQL databases using java database connectivity and can process the result returned by a database query. Weka's main user interface is the Explorer, but essentially the same functionality can be accessed through the component based Knowledge Flow interface and from the command line there is also the Experimenter, which allows the systematic comparison of the predictive performance of Weka's machine learning algorithms on a collection of datasets.

The second phase of the work is carried out in Mathwork software, Matlab 7.10. It is a numerical computing environment and fourth generation programming language. Dr. Cleve Moler, Chief Scientist at MathWorks, Inc., originally wrote MATLAB in late 1970s to provide easy access to matrix software developed in LINPACK and EISPACK projects [71]. It is an interactive program that helps with numeric computation and data visualization. It is built upon a foundation of sophisticated matrix software for analyzing linear systems of equations. Typical areas of application of MATLAB include math and computation, algorithm development, data analysis, data exploration, application development, etc.

## 3.3 Data Collection

Different checklists containing general signs and symptoms or attributes related to LD are used for the general assessment of learning disability. Based on these checklists, basic information about the child can be ascertained. But these much general information collected cannot be provided as such in the study. So it has been planned and decided to watch various assessment methods generally used in Kerala. Accordingly, a child care center, an open school and two other regular schools in and around Cochin, Kerala, where such LD assessment studies are generally conducting, are selected and with the help of Doctors and other Professionals / Resource Persons engaged there, the various LD assessment methods are studied. With their help, a checklist containing the 40 most frequent signs and symptoms of LD is created for LD assessment. This checklist is then used for further studies there and on subsequent evaluation with the help of these professionals and from the experience gained, another checklist, reducing to 16 prominent attributes were evolved which is used in the present research work. The list of these 16 attributes used in LD prediction is shown in Table 3.1 below. Data sets are collected from the above clinics/schools and also from a well developed research lab in the Department of Information Management, National Changhua University of Education, Changhua, Taiwan, Republic of China, by sort of knowledge sharing. Altogether 1020 data sets collected from the above sources are used in the present study.

**Table 3.1** 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 |

## 3.4 Data Sets

In this research work, 1020 real world datasets collected by informal assessment are used. 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.

## 3.5 Data Distribution

The percentage distribution of the 16 attributes present in the data set is shown in Figure 3.2 below. Among the cases in the dataset, 72% cases (children) are having LD (LD – True) and the remaining 28% cases (children) have not LD (LD – False) as represented in Figure 3.3.

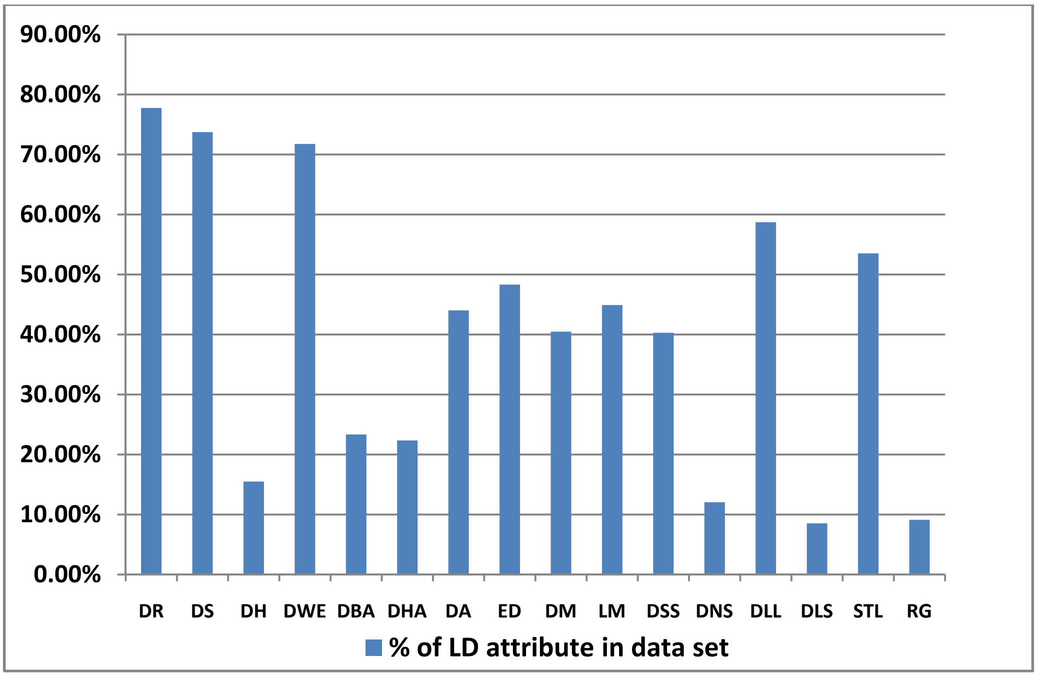


**Figure 3.2** Presence of attributes in data set

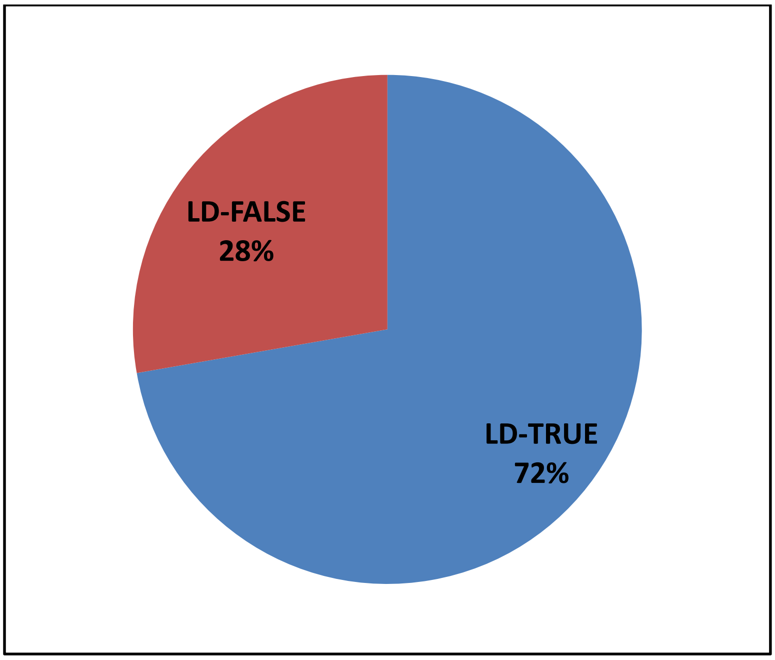


**Figure 3.3** Distribution of LD cases in data set

The significance of attributes in the data set used in the work in LD prediction is shown in Figure 3.4. From the 1020 cases, the number of each attribute dominating in LD True and LD False are indicated therein.

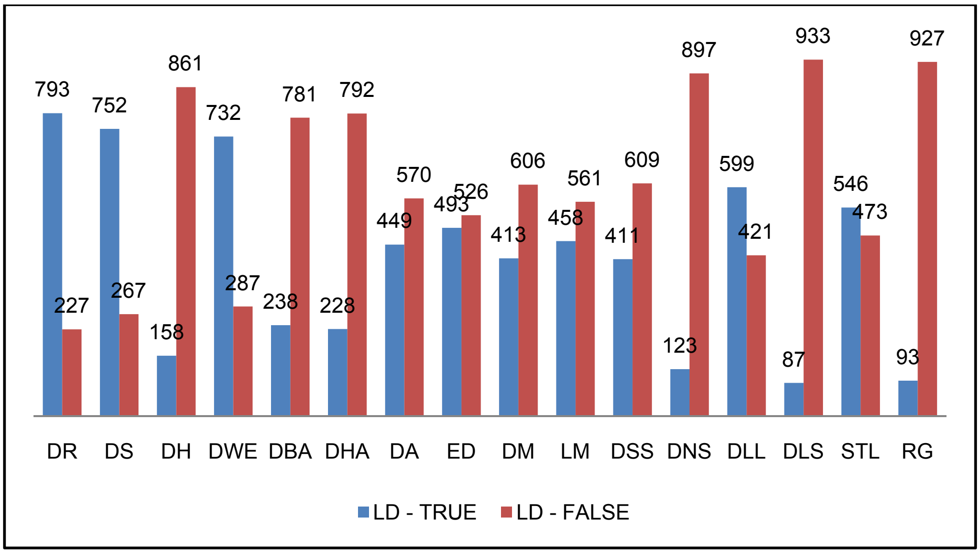


**Figure 3.4** Significance of attributes in LD prediction

1. What is learning disability?
2. How can you assess a student having learning disability?
3. Your professional experience?
4. Tell us your professional experience in identifying learning disability?
5. What are the symptoms of LD?
6. What is the seriousness of the LD in children?
7. What are the different types of LD?
8. How do you find the presence and degree of LD in a child?
9. Describe detail information about the activities of a child having LD?
10. What are the different assessment methods for the determination of degree of LD?
11. Describe your most successful accomplishment at work so far.
12. What are the key challenges of this field of learning disability?
13. What are the benefits of LD affected child if we provide remedial solutions?
14. What is the opinion about the knowledge based integrated tool for informal assessment?

**Figure 3.5** Sample questionnaire - knowledge extraction process

## 3.6 Knowledge Extraction Process

Knowledge extraction process is the creation of knowledge from structured and unstructured sources [72]. After the extraction, the resulting knowledge shall have inference facilitated and be in a machine readable and interpretable format. The check list used for LD knowledge extraction process is fully depends upon the experts in the field. This process of extraction of knowledge from the experts consists of two phases. The first phase is for getting the information about the LD and for finalizing a general assessment format. Various councilors/special educators/LD professionals/doctors, etc. are involved in this stage. A sample questionnaire used in this phase of knowledge extraction process is given in Figure 3.5. In the second phase, a check list for the detailed assessment of LD is formed. It is formed, along with the professionals, after attending different counseling/interviews conducted for the children and parents at the clinics. This new check list, used in the research work for LD prediction with fuzzy models, is the same as that given in Table 3.3.

## 3.7 Data Normalization

Data usually collected from multiple resources and stored in data warehouse. Resources may include multiple databases, data cubes, or flat files. Different issues could arise during integration of data that we wish to have for mining and discovery. These issues include schema integration and redundancy. So, data integration must be done carefully to avoid redundancy and inconsistency that in turn improve the accuracy and speed up the mining process [73].

The careful integration is now acceptable, but it needs to be transformed into forms suitable for mining. In data transformation, the data is transformed or consolidated into forms appropriate for mining. Data transformation involves data smoothing, data aggregation, data generalization, data normalization and data attribute construction or feature construction [8].

Data mining seeks to discover unrecognized associations between data items in an existing database. It is the process of extracting valid, previously unseen or unknown, comprehensible information from large database [74]. The growth of the size of data and number of existing database exceeds the ability of humans to analyze this data, which creates both a need and an opportunity to extract knowledge from databases.

Data transformation may improve the accuracy and efficiency of mining algorithms. The various methods involved in data transformation provide better results, if the data is analyzed. Normalization helps in preventing attributes with initially large range from outweighing attributes with initially smaller range. The main method used is the min-max normalization. In data normalization, the attribute data are scaled so as to fall within a small specified range, such as -1.0 to 1.0 or 0 to 1.0 [75].

The data set containing 1020 cases are collected using the general check list, from various sources by conducting informal assessments of LD determination. The data set thus initially contains non-numeric values, which has then been transformed into boolean values. This data set containing the boolean values is used in the present study for the LD prediction. A sample data set containing both non-numeric and boolean values is shown in Table 3.2.

The general check list used in obtaining the above data set is used as first phase of general assessment in LD determination. Those children identified with LD is subsequently assessed in second and third phases for further confirmation using different check lists. A new checklist is developed using the other LD assessment activities associated in the second and third phases of LD confirmation. This new developed checklist is given in Table 3.3 below. It is the one formed based on the knowledge extraction process. Each attribute is subdivided into sub-attributes with different scores as shown therein.

**Table 3.2** Sample data set - Non-numeric & Boolean values

| CASES | Data type | ATTRIBUTES | | | | | | | | | | | | | | | | LD |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | DR | DS | DH | DWE | DBA | DHA | DA | ED | DM | LM | DSS | DNS | DLL | DLS | STL | RG | |
| 1 | Non-numeric | Y | Y | N | Y | N | Y | Y | Y | N | Y | Y | Y | N | Y | N | Y | Y |
| | Boolean | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 1 |
| 2 | Non-numeric | Y | N | N | N | Y | N | Y | Y | N | Y | Y | N | N | N | Y | N | N |
| | Boolean | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 |
| 3 | Non-numeric | Y | Y | N | N | N | N | N | Y | Y | N | Y | N | Y | N | N | N | N |
| | Boolean | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 |
| 4 | Non-numeric | Y | Y | Y | Y | Y | N | Y | Y | N | N | N | Y | N | Y | Y | N | Y |
| | Boolean | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 1 |
| 5 | Non-numeric | Y | Y | Y | Y | Y | N | N | Y | Y | N | N | N | N | Y | Y | N | Y |
| | Boolean | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 1 |
| 6 | Non-numeric | N | Y | N | Y | N | Y | Y | N | Y | Y | Y | Y | N | N | Y | N | Y |
| | Boolean | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 1 |
| 7 | Non-numeric | N | N | Y | N | N | Y | N | Y | Y | Y | Y | N | N | N | N | Y | N |
| | Boolean | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 0 |
| 8 | Non-numeric | N | N | N | N | Y | N | Y | Y | Y | N | Y | Y | Y | N | N | N | N |
| | Boolean | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 |
| 9 | Non-numeric | Y | Y | Y | N | N | N | N | N | Y | Y | N | Y | Y | Y | Y | Y | Y |
| | Boolean | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 |
| 10 | Non-numeric | N | N | N | Y | N | Y | Y | N | N | Y | Y | Y | Y | Y | Y | Y | Y |
| | Boolean | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

**Table 3.3** New check list

| DR | Score | DS | Score | DH | Score | DWE | Score |
|---|---|---|---|---|---|---|---|
| Poor Reader | 50 | Slow reader | 50 | Messy handwriting | 50 | Copies inaccurately | 50 |
| Do not like reading | 25 | Poor speller | 25 | No space between words | 25 | Tells, but cannot write | 25 |
| Read same line twice | 15 | Similar letters confusion | 15 | Dislike writing | 15 | Frequently reverses letters | 15 |
| Reverse letters when read | 10 | Grammar problem | 10 | Incomplete writing | 10 | Delay in writing | 10 |
| *Total score* | *100* | *Total score* | *100* | *Total score* | *100* | *Total score* | *100* |
| **DBA** | **Score** | **DHA** | **Score** | **DA** | **Score** | **ED** | **Score** |
| Poor basic math skill | 50 | Difficulty in calculations | 50 | Poorly align numbers | 50 | Eyes hurt when read | 50 |
| Makes careless mistakes | 25 | Trouble in formula/rules | 25 | Has trouble in telling time | 25 | Smooth reading skill | 25 |
| Difficulty with word problems | 15 | Trouble in graphs/charts | 15 | Fails to understand ideas | 25 | Inability to locate information | 15 |
| Difficulty in counting principles | 10 | Difficulty in comparisons | 10 | Words tends to blurd | 10 | Undeveloped dictionary skill | 10 |
| *Total score* | *100* | *Total score* | *100* | *Total score* | *100* | *Total score* | *100* |
| **DM** | **Score** | **LM** | **Score** | **DSS** | **Score** | **DNS** | **Score** |
| Trouble in thoughts | 50 | Aimless | 50 | Trouble within lines/spelling/ pronouncing | 50 | Does not likes school | 50 |
| Reverses numbers/symb ols | 25 | Deficiency in ideas | 25 | Forgets quickly | 25 | Is forgetful routine activity | 25 |
| Difficulty in proof reading | 15 | Problem in learning | 15 | Trouble in following directions | 15 | Losses things consistently | 15 |
| Trouble in passage understanding | 10 | Trouble in memorizing | 10 | Difficult to understand what reads | 10 | Difficulty in organizing games | 10 |
| *Total score* | *100* | *Total score* | *100* | *Total score* | *100* | *Total score* | *100* |

| DLL | Score | DLS | Score | STL | Score | RG | Score |
|---|---|---|---|---|---|---|---|
| Unsure about use of full stop | 50 | Has difficulty in listening and taking notes at same time | 50 | Slow to learn new games and puzzles | 50 | Reads reluctantly | 50 |
| Unsure about capital letters | 25 | Has difficulty in generalizing skills from one situation to another | 25 | Does not to follow instructions | 25 | Confuses left and right | 25 |
| Unsure about unknown words | 15 | Fails to pay close attention to details | 15 | Express ideas in a disorganized way | 15 | Makes careless mistakes in school work | 15 |
| Problem in writing | 10 | Trouble reading charts, maps, etc | 10 | Has limited interest in books | 10 | Troubles in achieving goals | 10 |
| *Total score* | *100* | *Total score* | *100* | *Total score* | *100* | *Total score* | *100* |

## 3.8 Attribute Normalization

An attribute is normalized by scaling its values so that they fall within a small specified range, such as 0 to 1.0. Normalization is particularly useful for classification algorithms involving neural networks, or distance measurements such as nearest-neighbor classification and clustering. If using the neural network back propagation algorithm for classification mining, normalizing the input values for each attribute measured in the training samples will help speed up the learning phase. For distance based methods, normalization helps prevent attributes with initially large ranges from outweighing attributes with initially smaller ranges [8].

There are many methods for data normalization such as min-max normalization, z-score normalization, and normalization by decimal scaling

[73]. Min-max normalization performs a linear transformation on the original data. Min-max normalization preserves the relationships among the original data values. In z-score normalization or zero-mean normalization, the values for an attribute A are normalized based on the mean and standard deviation of A. This method of normalization is useful when the actual minimum and maximum of attribute A are unknown, or when there are outliers which dominate the min-max normalization. Normalization by decimal scaling normalizes by moving the decimal point of values of attribute A. The number of decimal points moved depends on the maximum absolute value of A. The attribute data is scaled to fit in a specific range. There are many type of normalization available, we will see one technique called Min Max Normalization here. Min Max Normalization transforms a value A to B which fits in the range[C, D]. It is given by the below formula;

$$B = \left\{ \frac{(A - \text{minimum value of } A)}{(\text{maximum value of } A - \text{minimum value of } A)} \right\} * (D - C) + C$$

A sample set of values obtained after attribute normalization is shown in Table 3.4 below.

**Table 3.4** Sample values of attribute normalization

| Sl. No. | DR | DS | DH | DWE | DBA | DHA | DA | ED | DM | LM | DSS | DNS | DLL | DLS | STL | RG |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0 | 0.965909 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 0.632184 | 0.738636 | 0 | 0.824176 | 0 | 0.909091 | 0 | 0.681818 | 0.543478 | 0.688889 | 0.804348 | 0 | 0.884211 | 0 | 0.647727 | 0 |
| 3 | 0.862069 | 1 | 0 | 0.714286 | 0 | 0.761364 | 0 | 0.625 | 0.684783 | 0.688889 | 0.630435 | 0 | 0.6 | 0 | 0.886364 | 0 |
| 4 | 0.758621 | 0.647727 | 0 | 0 | 0 | 0 | 0.738636 | 0.784091 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 5 | 0 | 0 | 0 | 0.43956 | 0.363636 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 6 | 0 | 0 | 0 | 0 | 0 | 0 | 0.25 | 0.284091 | 0.326087 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 7 | 0.655172 | 0.852273 | 0.903614 | 0.681319 | 0.738636 | 0 | 0.659091 | 0.852273 | 0.597826 | 0.977778 | 0 | 0 | 0 | 0 | 0.818182 | 0 |
| 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.333333 | 0 | 0 | 0.368421 | 0 | 0.340909 | 0 |
| 9 | 0.862069 | 0.738636 | 0 | 0.604396 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.789474 | 0 | 0 | 0 |
| 10 | 0.862069 | 0.670455 | 0 | 0.626374 | 0 | 0.738636 | 0.784091 | 0.784091 | 0.73913 | 0.866667 | 0.619565 | 0 | 0.789474 | 0.943182 | 0.943182 | 0 |
| 11 | 0.862069 | 0.840909 | 0 | 0.714286 | 0 | 0.852273 | 0.647727 | 0.840909 | 0.782609 | 0.788889 | 0.815217 | 0 | 0 | 0 | 0.829545 | 0 |
| 12 | 0 | 0 | 0 | 0.406593 | 0 | 0 | 0.409091 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 13 | 0.965517 | 0.977273 | 0 | 0.725275 | 0 | 0.943182 | 0 | 0.943182 | 0 | 0.922222 | 0.923913 | 0 | 1 | 0 | 0.727273 | 0 |
| 14 | 0 | 0 | 0 | 0 | 0 | 0.375 | 0 | 0.420455 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 15 | 0.747126 | 0.954545 | 0 | 0.010989 | 0 | 0 | 0 | 0.954545 | 0 | 0 | 0.934783 | 0.954545 | 0.894737 | 0 | 0.875 | 0 |
| 16 | 0 | 0.375 | 0 | 0 | 0 | 0 | 0 | 0.363636 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 17 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.272727 | 0.477273 |
| 18 | 0.758621 | 0.761364 | 0.795181 | 0.736264 | 0.852273 | 0 | 1 | 0.761364 | 0.728261 | 0.744444 | 1 | 0.75 | 0.705263 | 0 | 0.852273 | 0.761364 |
| 19 | 1 | 0.761364 | 0.903614 | 0.736264 | 0.943182 | 0 | 0.943182 | 0.761364 | 0.956522 | 0.744444 | 0.728261 | 0.852273 | 0.673684 | 0 | 0.829545 | 0.715909 |
| 20 | 0.862069 | 0.954545 | 1 | 0.967033 | 0.954545 | 0 | 0.840909 | 0.977273 | 0.695652 | 0.822222 | 0.836957 | 0.886364 | 0.768421 | 0 | 0.738636 | 0 |
| 21 | 0 | 0.965909 | 0 | 0 | 0 | 0 | 0 | 0.965909 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 22 | 0.632184 | 0.738636 | 0 | 0.824176 | 0 | 0.909091 | 0.625 | 0.738636 | 0 | 0.833333 | 0 | 0.909091 | 0.884211 | 0 | 0.647727 | 0 |
| 23 | 0.862069 | 1 | 0 | 0.714286 | 0 | 0.761364 | 0.852273 | 1 | 0 | 0.722222 | 0 | 0.761364 | 0.6 | 0 | 0.886364 | 0 |
| 24 | 0.758621 | 0.647727 | 0 | 0 | 0 | 0 | 0.75 | 0.647727 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

## 3.9 Entropy of LD Attributes

Data mining and knowledge discovery has made predominant progress during the past decades [76]. It uses algorithms, and techniques from many disciplines, including statistics, databases, machine learning, pattern recognition, artificial intelligence, data visualization, and optimization [77]. One of the most important data mining tasks is to determine the importance or rating of features. Different algorithms are used to rank the features, such as fisher score and CFS and gain ratio algorithms. But so far decision making algorithms in determining the importance of features are not used.

A measure used from Information Theory in the ID3 algorithm and many others used in decision tree construction is that of Entropy. Informally, the entropy of a dataset can be considered to be how disordered it is. It has been shown that entropy is related to information, in the sense that the higher the entropy, or uncertainty, of some data, then the more information is required in order to completely describe that data. In building a decision tree, we aim to decrease the entropy of the dataset until we reach leaf nodes at which point the subset that we are left with is pure, or has zero entropy and represents instances all of one class, all instances have the same value for the target attribute.

The entropy of a dataset, S, with respect to one attribute, in this case the target attribute is measured with the following calculation:

$$Entropy(s) = -\sum_{i=1}^{C} p_i \log_2 p_i$$

where Pi is the proportion of instances in the dataset that take the $i^{th}$ value of the target attribute, which has C different values [78].

This probability measures give us an indication of how uncertain we are about the data and we use a log2 measure as this represents how many bits we would need to use in order to specify what the class or value of the target attribute is of a random instance. This entropy is used to remove the irrelevant data or attribute and thus we can avoid the anomalies also. Higher entropy attribute will give more information in classification and prediction.

## 3.10 Implementation Methods

This section covers the implementation of the study with various classification methods such as neural network, decision tree, support vector machine, bagging which are performed in weka and fuzzy models performed in matlab for the prediction of learning disabilities in children

### 3.10.1 Neural Network

Neural networks (NN) have emerged as an important tool for classification. The recent vast research activities in neural classification have established that neural networks are a promising alternative to various conventional classification methods [79]. Neural network method is used for classification, clustering, feature mining, prediction and pattern recognition. When the output of the network is continuous, it is performing prediction and when the output has discrete values and then it is doing classification. A simple rearrangement of the neurons and the network becomes adept at detecting clusters.

#### *3.10.1.1 Proposed work*

In this part of the study, it is tried to implement neural network as a classifier for the prediction of learning disability problem. The concept of Multi-Layer Perceptron (MLP) with back propagation is implemented using weka and the drawbacks of NN are also studied. The data for this study has contains 1020 cases having symptoms of learning disability. The various inputs of the system are the various symptoms of LD. The attributes are taken from signs and symptoms that are present in the general checklist for the informal assessment of LD. These inputs are given to the system and the output of the system is LD true or LD false. But the neural network method used for the classification will also classify the other classes present in the system.

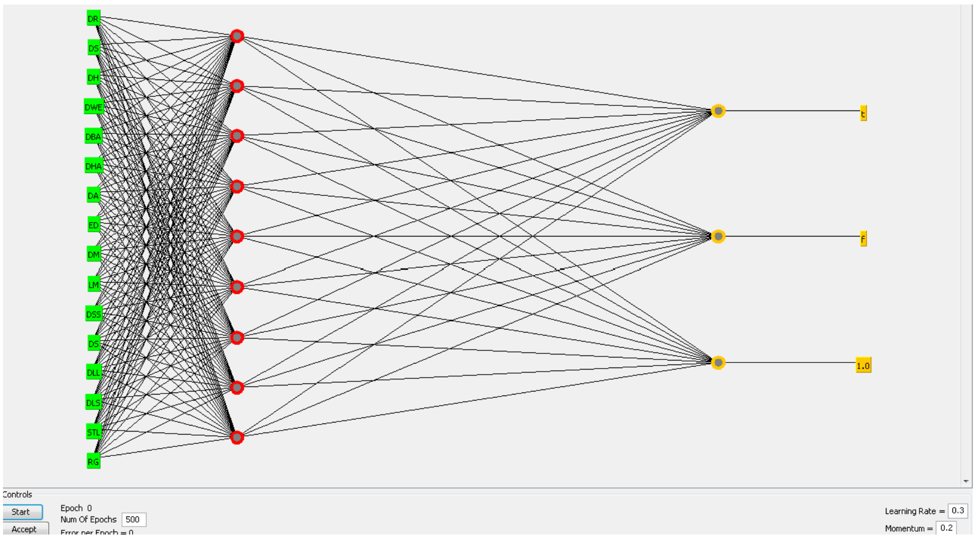


**Figure 3.6** Architecture of MLP network

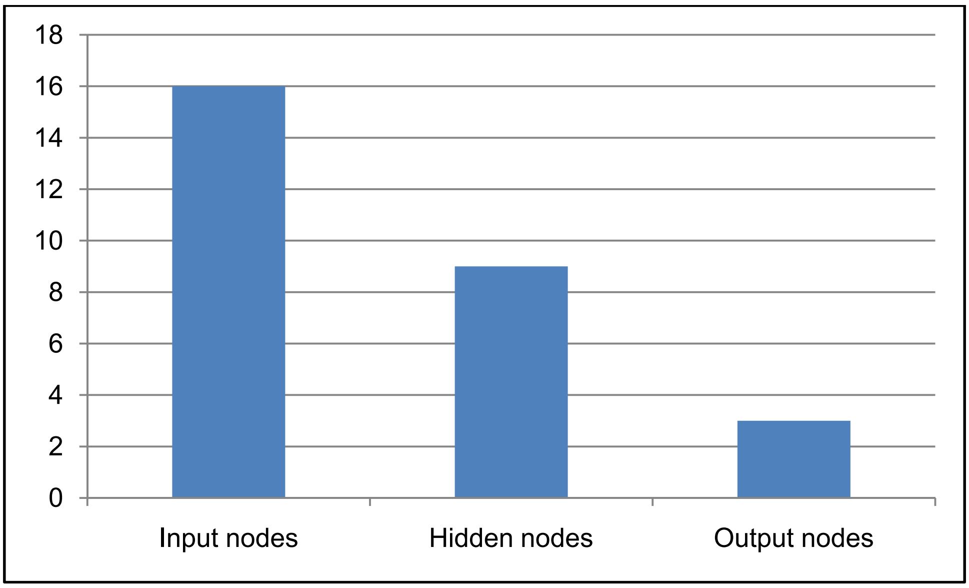


**Figure 3.7** Representation of nodes

As pointed out above, the datasets used in this section contain 1020 cases. Each case has 16 attributes and one class as LD true or false. The architecture of the neural network used in the study is the MLP with back propagation algorithm. MLP is a neural network that trains using back propagation. This network has three layers an input layer with green color and hidden node with red color and output node with yellow color as shown in Figure 3.6. A graphical representation of these nodes is shown in Figure 3.7.

#### *3.10.1.2 Methodology and results*

The classification algorithm MLP with back propagation is implemented through the data mining tool, weka. This obtained system contains 16 attributes and 9 hidden neurons and 3 output nodes. The two output nodes are true and false. The learning rate obtained is 0.3 and epoch is 500 and momentum 0.2, error of epoch=0. The number of nodes is obtained through trial and error. Back propagation is the most widely used learning method. The method of learning involves modifying the weights and biases of the network in order to minimize a cost function. The activation function considered for each node in the network is the binary sigmoid function defined as output. The parameters learning rate and momentum set values for these variables, which can be overridden in the graphical interface. A decay parameters causes to the learning rate. Here one hidden layer is used. The number of hidden neuron is reduced is the main challenge in the training phase and determines the appropriate number of hidden is the experimentation. The classifier model of full training set and the stratified cross-validation summary obtained while implementing MLP in weka on 1020 data set is given in Tables 3.5 and 3.6 respectively.

**Table 3.5** Classifier model full training set

| Sl. No | Sigmoid | Threshold |
|---|---|---|
| 1 | Node 0 | -0.016241901923750777 |
| 2 | Node 1 | -0.598040902392985 |
| 3 | Node 2 | -1.3241745517865613 |
| 4 | Node 3 | -0.5541893062773827 |
| 5 | Node 4 | 0.24625321785824697 |
| 6 | Node 5 | 0.21384450430679258 |
| 7 | Node 6 | 0.5757657748550636 |
| 8 | Node 7 | -0.479722035810239 |
| 9 | Node 8 | 0.44695728552405717 |
| 10 | Node 9 | 0.4199671265900112 |
| 11 | Node 10 | 0.20015351031343415 |
| 12 | Node 11 | -1.1056611347641336 |

**Table 3.6** Stratified cross-validation summary

| Sl. No. | Particulars | Value |
|---|---|---|
| 1 | Correctly Classified instances | 986 (96.6666 %) |
| 2 | Incorrectly Classified instances | 34 (3.3333 %) |
| 3 | Kappa statistic | 0.9329 |
| 4 | Mean absolute error | 0.03164 |
| 5 | Root mean squared error | 0.1056 |
| 6 | Relative absolute error | 6.174 % |
| 7 | Root relative squared error | 29.5363 % |
| 8 | Total Number of Instances | 1020 |
| 9 | Time taken to build model | 29.2 seconds |

**Table 3.7** Detailed accuracy by class

| TP Rate | FP Rate | Precision | Recall | F-Measure | ROC Area | Class |
|---|---|---|---|---|---|---|
| 0.967 | 0.0325 | 0.9624 | 0.967 | 0.966 | 0.962 | t |
| 0.96 | 0.013 | 0.96 | 0.96 | 0.96 | 0.962 | f |

**Table 3.8** Confusion matrix

| a | b | Classified as |
|---|---|---|
| 614 | 17 | a = t |
| 17 | 372 | b = f |

The detailed accuracy by class and confusion matrix are shown in Tables 3.7 and 3.8 above respectively. 96.67% cases are classified correctly and 3.33 % cases are classified incorrectly. The different types of error functions used in this study are mean squared error, mean absolute error, root mean squared error, relative absolute error and root relative squared error, the values of which are also shown in Table 3.6.

### *3.10.1.3 Review of this work*

From the results obtained, it is understood that how effectively the MLP with back propagation algorithm classifies the LD dataset. The major issue studied from this study of prediction of LD in children is failure of the classifier in handling the missing values in the datasets. 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 results obtained shows that 96.67% accuracy with correctly classified instances and 3.33% accuracy in incorrectly classified instances. The findings show that there is no solution in the case of missing values present. Also some attributes are unwanted and hence have no contributions in predicting the LD. But from the output of this classification method, it is understand how easily the learning disability can be predicted in the early stages itself.

### 3.10.2 Decision Tree

Decision trees (DTs) are powerful and popular tool for classification and prediction. It is a classifier in the form of a tree structure where each node is either a leaf node-indicates the value of the target attribute of examples or a decision node–specifies some test to be carried out on a single attribute-with one branch and sub tree for each possible outcome of the test [5]. Classifiers do not require any domain knowledge or parameter setting and therefore is appropriate for exploratory knowledge discovery. Decision tree can handle high dimensional data. The learning and classification step of decision tree are simple and fast. Thus it is a flow chart like structure, where each internal node denotes a test on an attribute, each branch of the tree represents an outcome of the test and each leaf node holds a class label [5]. The topmost node in a tree is the root node.

A decision tree can be used to classify an example by starting at the root of the tree and moving through it until a leaf node, which provides the classification of the instance [5]. There are a variety of algorithms for building the decision tree that share the desirable of interpretability. In this section of the research work, the well known and frequently used algorithm J48 is used for the classification of LD.

The key requirements to do mining with decision trees are: attribute value description, predefined classes, discrete classes and sufficient data. A divide and conquer approach to the problem of learning disability from a set of independent instances leads naturally to a style of representation called a decision tree. To classify an unknown instance, it is routed down the tree according to the values of the attributes tested in successive nodes and when a leaf is reached, the instance is classified according to the class assigned to the leaf [5].

### *3.10.2.1 Classification by decision tree*

Data mining techniques are useful for predicting and understanding the frequent signs and symptoms of behavior of LD. If we study the attributes of LD, we can easily predict which attribute is from the data sets more related to learning disability. The first task to handle learning disability is to construct a database consisting of the signs, characteristics and level of difficulties faced by those children. After identifying the dependencies between these diagnostic attributes, rules are generated and these rules are then be used to predict learning disability.

Based on the information obtained from the checklist, a data set is generated. This set is in the form of an information system containing cases, attributes and class. A complete information system expresses all the knowledge available about objects being studied. Decision tree induction is the learning of decisions from class labeled training tuples. Given a data set D = {t1, t2,…..….., tn} where ti = <ti1,….., tih>. In the study, each tuple is represented by 16 attributes and the class is LD. Then, Decision or Classification Tree is a tree associated with D such that each internal node is labeled with attributes DR, DS, DH, DWE, etc. Each arc is labeled with predicate, which can be applied to the attribute at the parent node. Each leaf node is labeled with a class LD. The basic steps in the decision tree are building the tree by using the training data sets and applying the tree to the new data sets. Decision tree induction is the process of learning about the classification using the inductive approach [8]. During this process, a new decision tree is created from the training data. This decision tree can be used for making classifications. Here the J48 algorithm is used, which is a greedy approach in which decision trees are constructed in a top-down recursive divide and conquer manner.

Most algorithms for decision tree approach are following such a top down approach. It starts with a training set of tuples and their associated class labels. The training set is recursively partitioned into smaller subsets as a tree is being built. This algorithm consists of three parameters – attribute list, attribute selection method and classification. The attribute list is a list of attributes describing the tuples. Attribute selection method specifies a heuristic procedure for selecting the attribute that best discriminate the given tuples according to the class. The procedure employs an attribute selection measure such as information gain that allows a multi-way splits. Attribute selection method determines the splitting criteria. The splitting criteria tells as which attribute to test at a node by determining the best way to separate or partition the tuples into individual classes. Here we are using the data mining tool weka for attribute selection and classification. Classification is a data mining (Machine Learning) technique, used to predict group membership from data instances [7].

### *3.10.2.2 Methodology used for making the decision tree*

J48 algorithm is used for classifying the Learning Disability. The procedure consists of three steps viz. (i) data partition based on cross validation test, (ii) attribute list and (iii) attribute selection method based on information gain. Cross validation approach is used for the sub sampling of datasets. In this approach, each record is used the same number of times for training and exactly once for testing. To illustrate this method, first partition the datasets into two subsets and choose one of the subsets for training and other for testing. Then swap the roles of the subsets so that the previous training set becomes the test set and vice versa. The attribute list, attribute selection method by gain ratio and classification of the study are as given under.

**(i) Attribute List**

The list of 16 attributes used in this part of the study is same as that given in Table 3.1.

**(ii) Attribute Selection Method by Gain Ratio**

The Information Gain Ratio for a test is defined as follows. IGR (Ex, a) = IG / IV, where IG is the Information Gain and IV is the Gain Ratio. Information gain ratio biases the decision tree against considering attributes with a large number of distinct values. So it solves the drawback of information gain.

**(iii) Classification**

As shown in the stratified cross-validation summary at Table 3.9, 98.24% cases are correctly classified and 1.76 % cases are incorrectly classified.

**Table 3.9** Stratified cross-validation summary – decision tree

| Sl. No. | Particulars | Value |
|---|---|---|
| 1 | Correctly classified instances | 1002 Nos., 98.2353 % |
| 2 | Incorrectly classified instances | 18 Nos., 1.7647% |
| 3 | Kappa statistic | 0.9627 |
| 4 | Mean absolute error | 0.0263 |
| 5 | Root mean squared error | 0.1317 |
| 6 | Relative absolute error | 5.5664 % |
| 7 | Root relative squared error | 27.1166 % |
| 8 | Total Number of Instances | 1020 |
| 9 | Time taken to build model | 0.22 seconds |

**Table 3.10** Detailed accuracy by class

| TP Rate | FP Rate | Precision | Recall | F-Measure | ROC Area | Class |
|---|---|---|---|---|---|---|
| 0.984 | 0.021 | 0.987 | 0.984 | 0.986 | 0.975 | t |
| 0.979 | 0.016 | 0.974 | 0.979 | 0.977 | 0.975 | f |

**Table 3.11** Confusion matrix

| a | b | Classified as |
|---|---|---|
| 621 | 9 | a = t |
| 9 | 381 | b = f |

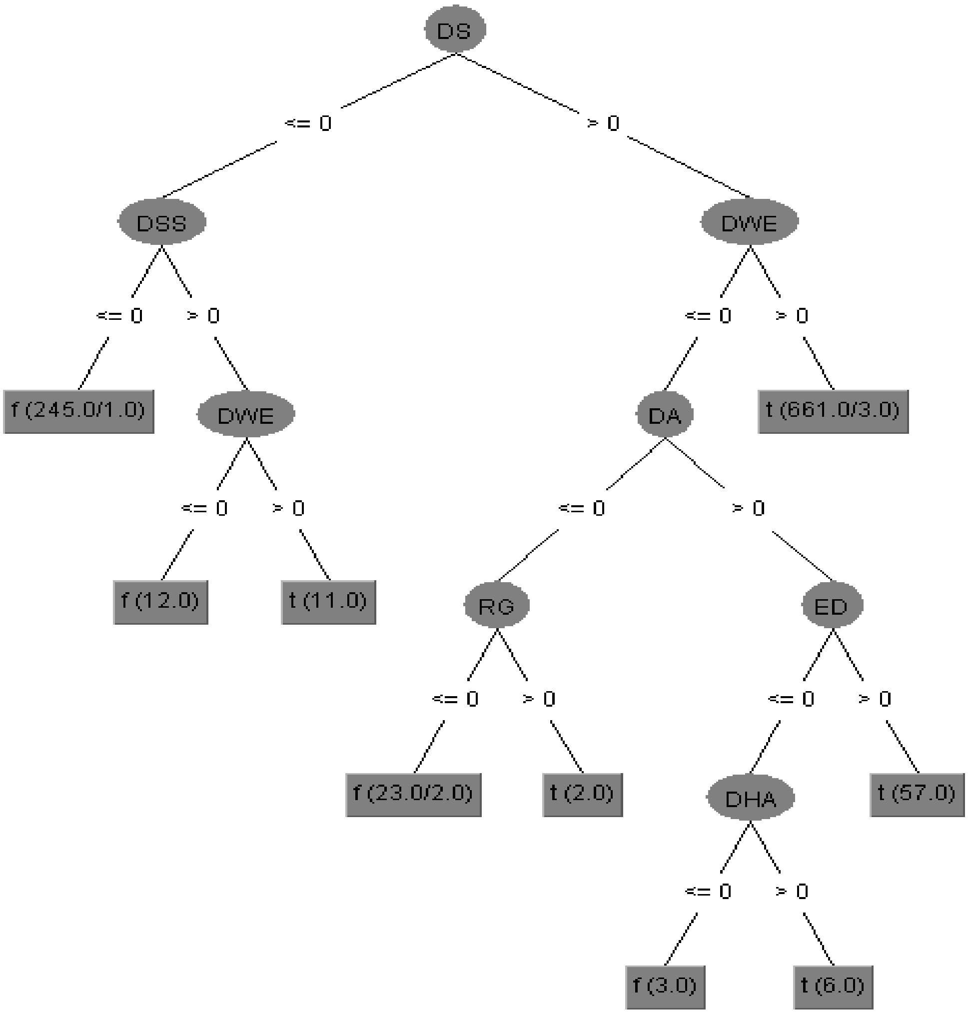


**Figure 3.8** Decision tree

The detailed accuracy by class and confusion matrix of the decision tree are shown in Tables 3.10 and 3.11. The TP Rate (True Positive Rate) and the FP Rate (False Positive Rate) are indicated therein. TP Rate is the ratio of low weight cases predicted correctly cases to the total of positive cases. The FP Rate is then the ratio normal weight cases of incorrectly predicted as low weight cases to the total of normal weight cases. A decision tree formed based on the methodology adopted is shown in Figure 3.8. The visualization of cost curve for LD = true and, LD = false are shown in Figures 3.9 and 3.10 respectively.

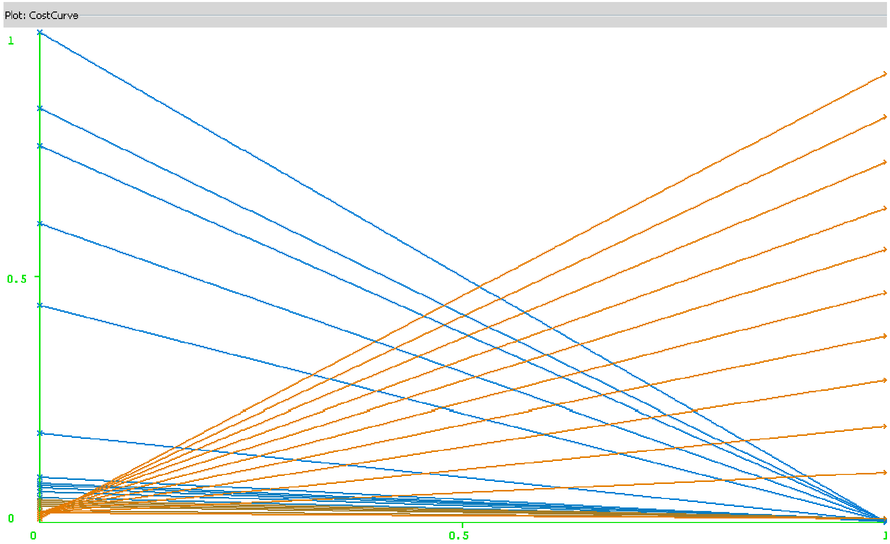


**Figure 3.9** Visualization of cost curve of decision tree for LD = true

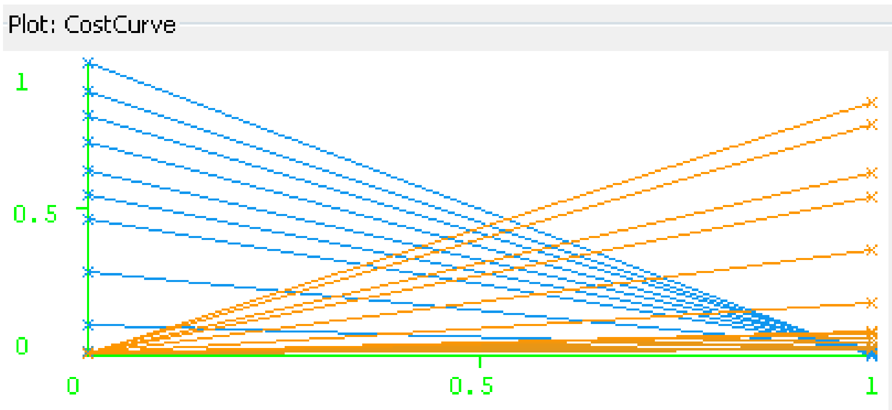


**Figure 3.10** Visualization of cost curve of decision tree for LD=false

### *3.10.2.3 Extraction of rules from tree*

It is easy to read a set of rules directly off a decision tree. One rule is generated for each leaf. The antecedent of the rule includes a condition for every node on the path from the root to that leaf and the consequent of the rule is the class assigned by the leaf [67]. This procedure produces rules that are unambiguous in that the order in which they are executed is irrelevant. However in general, rules that are read directly off a decision tree are far more complex than necessary and rules derived from trees are usually pruned to remove redundant tests. The rules are so popular because each rule represents an independent knowledge. New rule can added to an existing rule sets without disturbing them, whereas to add to a tree structure may require reshaping the whole tree. In this section, we present a method for generating a rule set from a decision tree. In principle, every path from the root node to the leaf node of a decision tree can be expressed as a classification rule. The test conditions encountered along the path form the conjuncts of the rule antecedent, while the class label at the leaf node is assigned to the rule consequent. The rules extracted from the decision tree are given below [80].

R1:IF DS = NO, DSS = NO THEN LD = NO
R2:IF DS = NO, DSS = YES, DWE = NO THEN LD = NO
R3:IF DS = NO, DSS = YES, DWE = YES THEN LD = YES
R4:IF DS = YES, DWE = NO, DA = NO, RG = NO THEN LD = NO
R5:IF DS = YES, DWE = NO, DA = NO, RG = YES THEN LD = YES
R6:IF DS = YES, DWE = NO, DA = YES, ED = NO, DHA = NO THEN LD = NO
R7:IF DS = YES, DWE = NO, DA = YES, ED = NO, DHA = YES THEN LD = YES
R8:IF DS = YES, DWE = NO, DA = YES, ED = YES THEN LD = YES
R9:IF DS = YES, DWE = YES THEN LD = YES

The expressiveness of a rule set is almost equivalent to that of a decision tree because a decision tree can be expressed by a set of mutually exclusive and exhaustive rules.

#### *3.10.2.4 Review of this work*

The nine rules generated from the decision tree data sets are as shown above. If the minimum confidence threshold is, say 90%, then the above rules are output sure, the certainty of the generated rules, confidence of each rule, is calculated by considering the cases given in datasets. The accuracy obtained for these rules are given in Table 3.12. These rules can be used for predicting the learning disability accurately.

In decision tree the main objective of attribute evaluation is based on information gain. Decision trees have pointed at the decision classes, which are not predominant for the given combination of input values like inconsistent data. The result of this study indicates that the rules system represented by the decision trees may be significantly incorrect for inconsistent data as well as for consistent data with large number of variables. The confidence level of the rules of decision trees shows lower accuracy. The inconsistent data may lead to false attribute selection. Here, the input values considered as the symptoms of LD. In the case of decision trees, such values may lead to prediction, which is a good reflection of the general dependencies in training data, and the prediction, which is far from the expectations and impossibility of the prediction.

**Table 3.12** Confidence of rules

| Rules | Confidence |
|---|---|
| R1 | 93% |
| R2 | 92% |
| R3 | 91% |
| R4 | 90% |
| R5 | 91% |
| R6 | 93% |
| R7 | 91% |
| R8 | 90% |
| R9 | 92% |

In construction of decision trees, J48 algorithm is used. The extracted rules are very effective for the prediction. The wrong predictions obtained from decision trees for all inconsistent data sets can be lead to a limited accuracy of decision tree models. Decision trees have pointed at the decision classes, which are not predominant for the given combination of input values like inconsistent data. The result of this study indicates that, the rules system represented by the decision trees may be significantly incorrect for inconsistent data with large number of variables. The computation times of decision tree are generally short and the interpretation of rules obtained from decision tree can be facilitated by the graphical representation of the trees. The main drawback noticed from this study is that the failure in handling inconsistent data. Also the formation of tree and rule generation becomes complex due to the increase of number of attributes. These problems are overcome by the methods developed by us, which are explained in the next chapter.

### 3.10.3 Support Vector Machine

Vladimir Vapnik invented Support Vector Machine in 1979 [81,82]. Support Vector Machine algorithm is based on statistical learning theory, is one of the best classifiers that gives good results. It is a new method for the classification of both linear and non-linear data. The basic idea behind the support vector machine is to map the original data into a feature space with high dimensionality through a non-linear mapping function and construct an optimal hyper plane in new space [83]. In the last few years, there has been a surge of interest in support vector machine [84, 81]. SVM have empirically been shown to give good generalization performance on a wide variety of problems such as hand written character recognition, face detection and pedestrian detection [85].

SVM can be applied to both classification and regression. In the case of classification, an optimal hyper plane is found that separates the data into two classes, whereas in the case of regression a hyper plane is to be constructed that less close to as many points as possible. Separating the classes with a large margin minimizes a bound on the expected generalization error [85].

A minimum generalization error means that when new examples arrive for classification, the chance of making an error in the prediction based on the learned classifier should be minimum. Such a classifier is one, which achieve maximum separation margin between the classes. The two planes parallel to the classifier and which passes through one or more points in the data set are called bounding planes [86]. Support Vector Machines select a small number of critical boundary instances called support vectors from each class and build a linear discriminate function that separates them as widely as possible [5]. The points in the dataset falling on the bounding planes are called support vectors. SVM algorithm transforms the original data in a higher dimension, from where it can find a hyper plane for separation of the data using essential training tuples called support vectors [8]. These points play a crucial role in the theory and hence the name Support Vector Machines. Machine means algorithm. SVMs belong to the class of supervised learning algorithms in which the learning machine is given a set of examples with the associated labels as in the case of decision trees, the examples are in the form of attribute vectors [86].

If the training vectors are separated without errors by an optimal hyper plane, the expected error rate on a test sample is bounded by the ratio of the expectation of the support vectors to the number of training vectors. Since this ratio is independent of the dimension of the problem, if one can find a small set of support vectors, good generalization is guaranteed. In the case, one may simply minimize the number of misclassification while maximizing the margin with respect to the correctly classified instances. In such a case, it is said that the SVM training algorithm allows a training error [86]. There may be another situation; the points are clustered such that the two classes are not linearly separable. It may have to tolerate large training error. In such cases, we prefer nonlinear mapping of data into some higher dimensional space called feature space, where it is linearly separable. In order to distinguish between these two spaces, the original space of data point is called input space [86]. The hyper plane in feature space corresponds to a highly nonlinear separating surface in the original input space. Hence the classifier is called nonlinear classifier [87].

However, the use of SVM is still limited to a small group of researchers. One possible reason is that training algorithm for SVM is slow especially for large problems. Another explanation is that, SVM training algorithm is complex, subtle and difficult for an average engineer to implement [85].

#### *3.10.3.1 Performance of SVM classifiers and results*

In this part of the work, the learning algorithms, Sequential Minimal Optimization (SMO) is used for the prediction of LDs. Due to its analytical foundation the SMO approach is particularly popular and at the moment the widest used, analyzed and still heavily developing algorithm [88]. The performance of this classifier is studied in detail along with the results obtained. SMO is conceptually simple, easy to implement and generally faster.

Sequential minimal optimization algorithm is used for training a support vector classifier. This implementation globally replaces all missing values and transforms nominal attributes into binary ones. It also normalizes all attributes by defaults. In this study, the SMO algorithm which uses the polynomial kernel, correctly classified 98.8235% instances from the real data sets with a complexity parameter of 1. The stratified cross-validation summary, detailed accuracy by class and confusion matrix are shown in Tables 3.13, 3.14 and 3.15 respectively.

**Table 3.13** Stratified cross-validation summary - SVM

| Sl. No. | Particulars | Value |
|---|---|---|
| 1 | Correctly classified instances | 1008 (98.8235 %) |
| 2 | Incorrectly classified instances | 12 (1.1765 %) |
| 3 | Kappa statistic | 0.9706 |
| 4 | Mean absolute error | 0.0193 |
| 5 | Root mean squared error | 0.0932 |
| 6 | Relative absolute error | 4.8162 % |
| 7 | Root relative squared error | 20.8244 % |
| 8 | Total Number of Instances | 1020 |
| 9 | Time taken to build model | 0.61 seconds |

**Table 3.14** Detailed accuracy by class

| TP Rate | FP Rate | Precision | Recall | F-Measure | ROC Area | Class |
|---|---|---|---|---|---|---|
| 0.993 | 0.025 | 0.991 | 0.993 | 0.992 | 0.991 | t |
| 0.975 | 0.007 | 0.982 | 0.975 | 0.979 | 0.991 | f |

**Table 3.15** Confusion matrix

| a | b | Classified as |
|---|---|---|
| 732 | 5 | a = t |
| 7 | 276 | b = f |

### *3.10.3.2 Result Analysis*

We can see that SVM provide algorithm for evaluating conditioning attribute, but the inherent significance is entirely different. The classification is mainly based on the type of kernel choosed. For evaluation of SVM performance, we have used different kernels such as RBF kernel, normalized polynomial kernel and polynomial kernel. The results obtained for these kernels are given in Table 3.16. From this, it can be seen that the performance of polynomial kernel is the best. The cost curves of SVM model is shown in Fig. 3.11.

**Table 3.16** Performance evaluation of different kernels

| Sl. No. | Particulars | Type of kernel | | |
|---|---|---|---|---|
| | | RBF kernel | Normalized polynomial kernel | Polynomial kernel |
| 1 | Total number of instances | 1020 | 1020 | 1020 |
| 2 | Correctly classified instances | 1002 | 1006 | 1008 |
| 3 | Incorrectly classified instances | 18 | 14 | 12 |

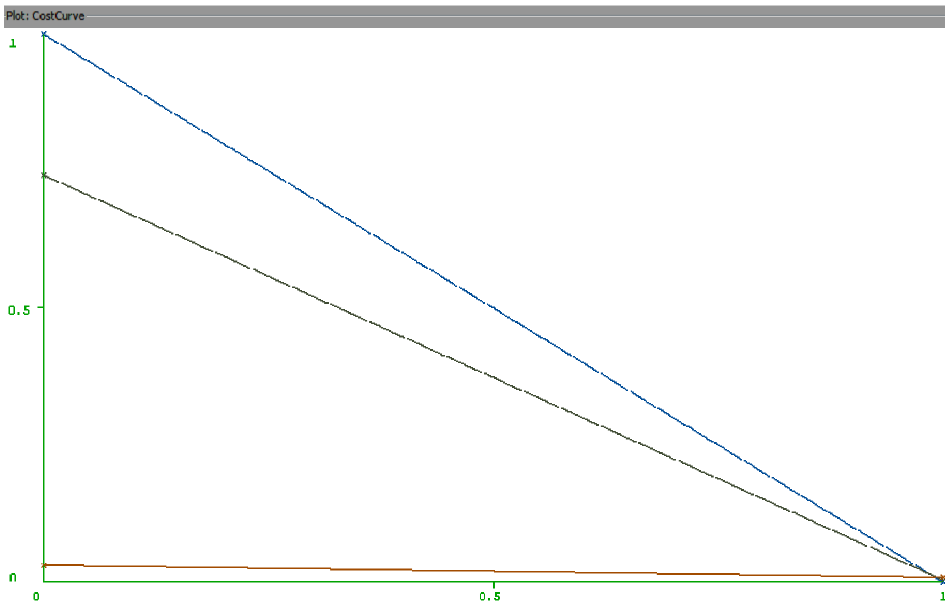


**Figure 3.11** Cost curve of SVM

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

#### *3.10.3.3 Review of this work*

When compared with decision tree, it is seen that the selection of attribute is very important in case of construction of decision tree and the inconsistent data may lead to false attribute selection. But in the case of SVM, alternative methods are available in handling missing data. Information gain is used as the attribute selection method in decision tree. But the inconsistency of the data leads to the false determination of attribute. The input values considered as the symptoms of LD. So the SVM and decision trees consider the inconsistent data in different ways. In the case of decision trees, such values may lead to prediction, which is a good reflection of the general dependencies in training data, and the prediction, which is far from the expectations and impossibility of the prediction. It is seen that, SVM gives more accurate result compared to decision tree, even though the time taken to build the model is much higher. The SVM may require long computational times than decision tree [89].

### 3.10.4 Bagging

Bagging is a method for improving results of machine learning classification algorithms. This method was formulated by Leo Breiman and its name was deduced from the phrase “bootstrap aggregating” [90]. The information had to be transformed into the form of knowledge [91]. This transformation represents a large space for various machine learning algorithms, mainly classification ones. The quality of the transformation heavily depends on the precision of classification algorithms in use. The precision of classification depends on many aspects. Two of most important aspects are the selection of a classification algorithm for a given task and the selection of a training set.

#### *3.10.4.1 Methodology and results*

The approach here we used is based on an idea of making various samples of the training set. A classifier is generated for each of these training set samples by a selected machine learning algorithm. In this way, for k variations of the training set we get k particular classifiers. The result will be given as a combination of individual particular classifiers. This method is called Bagging. In case of classification into two possible classes, a classification algorithm creates a classifier on the base of a training set of example descriptions. The bagging method creates a sequence of classifiers in respect to modifications of the training set. These classifiers are combined into a compound classifier. The prediction of the compound classifier is given as a weighted combination of individual classifier predictions. The theory of classifier voting is well described in [92] and [93]. More precise classifiers have stronger influence on the final prediction than less precise classifiers. The precision of base classifiers can be only a little bit higher than the precision of a random classification. That is why these classifiers are called weak classifiers.

In bagging the original training set is divided into N subsets of the same size. Each subset is used to create one classifier; a particular classifier is learned using this subset. A compound classifier is created as the aggregation of particular classifiers. The most known methods are: disjoint partitions, small bags, no replication small bags and disjoint bags. Bagging method uses multiple versions of a training set to train a different model of classifier. Each version of the training set can be generated by sampling with replacement. The outputs of the models are combined by voting to create a single output.

The stratified cross-validation summary, detailed accuracy by class and confusion matrix are shown in Tables 3.17. 3.18 and 3.19 respectively.

**Table 3.17** Stratified cross-validation summaries - bagging

| Sl. No. | Particulars | Value |
|---|---|---|
| 1 | Correctly Classified Instances | 1013 (99.3137 %) |
| 2 | Incorrectly Classified Instances | 7 (0.6863 %) |
| 3 | Kappa statistic | 0.9829 |
| 4 | Mean absolute error | 0.0069 |
| 5 | Root mean squared error | 0.0828 |
| 6 | Relative absolute error | 1.7107 % |
| 7 | Root relative squared error | 18.5019 % |
| 8 | Total Number of Instances | 1020 |
| 9 | Time taken to build model | 0.53 seconds |

**Table 3.18** Detailed accuracy by class

| TP Rate | FP Rate | Precision | Recall | F-Measure | ROC Area | Class |
|---|---|---|---|---|---|---|
| 0.986 | 0.014 | 0.995 | 0.996 | 0.995 | 0.989 | t |
| 0.986 | 0.004 | 0.989 | 0.986 | 0.988 | 0.991 | f |

**Table 3.19** Confusion matrix

| a | b | Classified as |
|---|---|---|
| 734 | 3 | a = t |
| 4 | 279 | b = f |

### *3.10.4.2 Review of this work*

The accuracy is in the tune of 99.3% shows that the bagged classifier has significantly greater accuracy than a single classifier derived from the original training data. It will not be considerably worse and is more robust to the effects of noisy data. The increased accuracy occurs because the composite model reduces the variance of the individual classifiers.

## 3.10.5 Fuzzy Model

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 [94]. It is particularly good at handling uncertainty, vagueness and imprecision and it is specially useful where a problem can be described linguistically. In fuzzy logic the degree of truth of a statement can range between 0 and 1.

Fuzzy logic was invented by Zadeh [95] 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 [96]. 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.

### *3.10.5.1 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 [97,98]. 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)

### *3.10.5.2 Fuzzification of attributes*

Fuzzification is a process of making crisp quantity fuzzy [99]. 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 [100,59]. 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 Figure 3.12.

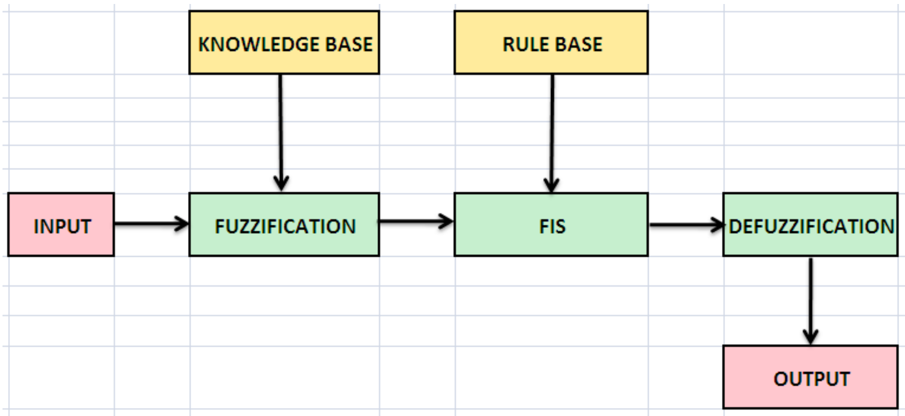


**Figure 3.12** System flow chart of fuzzification of attributes

### *3.10.5.3 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 [101,102]. 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.

### *3.10.5.4 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 Figures 3.13, 3.14 and 3.15 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.

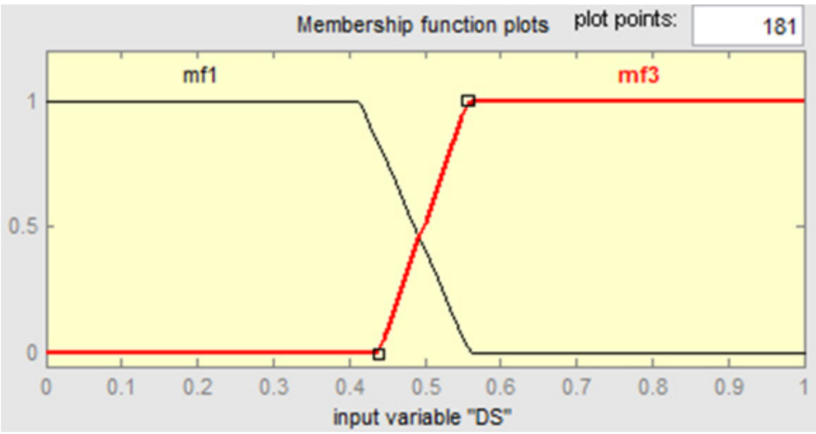


**Figure 3.13** Membership function of input variable DS

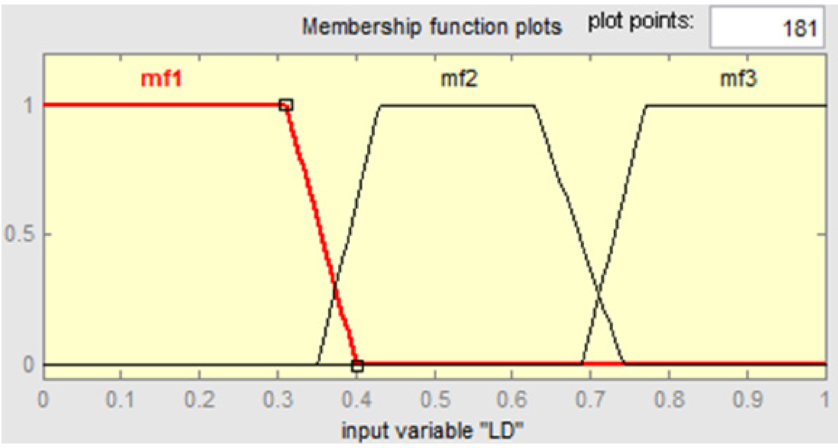


**Figure 3.14** Membership function of input variable LD

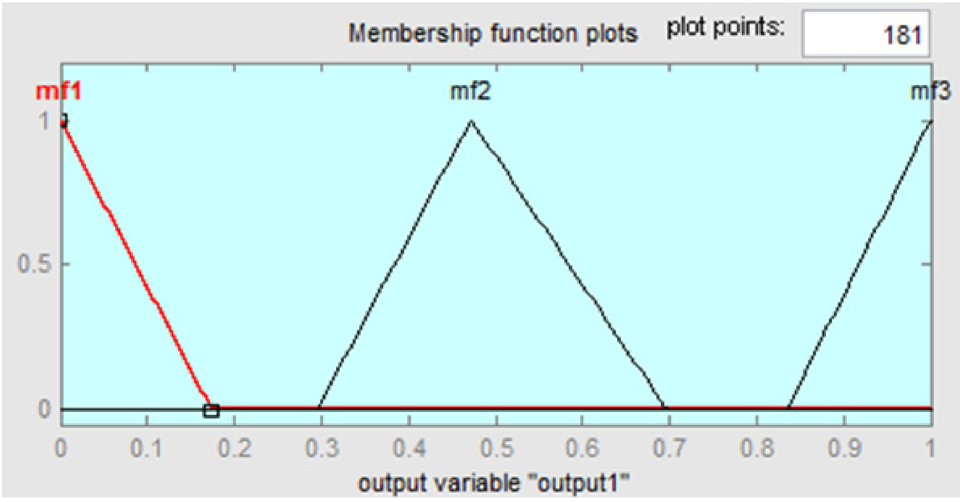


**Figure 3.15** Membership function of output variable

### *3.10.5.5 Fuzzy inference system (FIS)*

FIS have been successfully applied in fields such as automatic control, data classification, decision analysis, expert systems, and computer vision. Because of its multidisciplinary nature, FIS are associated with a number of names, such as fuzzy-rule-based systems, fuzzy expert systems, fuzzy modeling, fuzzy associative memory, fuzzy logic controllers, and simply fuzzy systems [103]. Fuzzy inference is the process of formulating the mapping from a given input to an output using fuzzy logic. The mapping then provides a basis from which decisions can be made, or patterns discerned.

The statement of if-then rules is the main mechanism in the fuzzy inference system. This fuzzy inference system makes the system natural and beneficial to model a complex humanistic in the loop system. A fuzzy inference system is capable of approximating any real continuous function to arbitrary accuracy, and this is a basis from which decisions can be made, or patterns discerned [104]. The components of a fuzzy inference system are the rules, the fuzzifier, the inference engine and the defuzzifier. The inference process is divided into three phases, application of the fuzzy operator in the antecedent, implication from the antecedent to the consequent, and aggregation of the consequents across the rules. The basic function of the inference engine is to compute levels of belief in output fuzzy sets from the levels of belief in the input fuzzy sets. The output is a single belief value for each output fuzzy set. In this stage, the fuzzy operator is applied in order to gain a single number that represents the result of the antecedent for that rule. The inference engine is mainly based on rules. The evaluations of the fuzzy rules and the combination of the results of the individual rules are performed using fuzzy set operations. The operations on fuzzy sets are different than the operations on non-fuzzy sets. The mostly-used operations for OR and AND operators are max and min, respectively [105]. After evaluating the result of each rule, these results should be combined to obtain a final result.

Rules determine the closed-loop behavior of the system. The rules are based on expert opinion, operator experience, and system knowledge. The basic function of the rule base is to represent in a structured way the prediction system of LD is in the form of a set of production rules. The if part of such a rule is called the rule-antecedent and is a description of a process state in terms of a logical combination of fuzzy propositions [106]. Moreover, the then-part of the rule is called the rule-consequent and is again a description of the control output in terms of a logical combination of fuzzy propositions. These propositions state the linguistic values which the control output variables take whenever the current process state matches the process state description in the rule antecedent.

Mamdani's fuzzy inference method is the most commonly seen fuzzy methodology [107] and it expects the output membership functions to be fuzzy sets and we used this inference system in the study for prediction of LD. The information about FIS is given in Table 3.20. The inputs given to the FIS are the symptoms of LD. It consists of 16 attributes of LD and LD probability. The number of rules defined by the system is 154. The output consists of single one, which is the LD prediction. The defuzzification method used in the system is centroid.

**Table 3.20** Information about FIS

| Sl.No. | Particulars | |
|---|---|---|
| 1 | Name | run1 |
| 2 | Type | mamdani |
| 3 | Num Inputs | 17 |
| 4 | Num Outputs | 1 |
| 5 | Num Rules | 154 |
| 6 | And Method | min |
| 7 | Or Method | max |
| 8 | DefuzzMethod | centroid |

The structure of FIS used in LD prediction is shown in Table 3.21 below. It contains the different inputs of the system and its range, type of membership function and number of membership functions related to each input. In FIS the output values from the membership function are linked with several different functions to generate an output set [108]. The visualization of FIS is in LD prediction is shown in Figure 3.16.

**Table 3.21** Structure of FIS

| **[Input1] : Name='LD_PROBLITY'** | |
|---|---|
| Range=[0 1] | NumMFs=3 |
| MF1='mf1':'trapmf',[-0.518518518518519 -0.198518518518519 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] | ---- |
| **[Input2] : Name='DR'** | |
| Range=[0 1] | NumMFs=2 |
| MF1='mf1':'trapmf',[-0.164 -0.0198 0.4352 0.556] | MF2='mf3':'trapmf',[0.443 0.56 1.01984126984127 1.2] |
| **[Input3] : 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] |
| **[Input4] : Name='DH'** | |
| Range=[0 1] | NumMFs=2 |
| MF1='mf1':'trapmf',[-0.164 -0.0198 0.4352 0.556] | MF2='mf3':'trapmf',[0.433883597883598 0.578083597883598 1.0330835978836 1.1538835978836] |
| **[Input5] : Name='DWE'** | |
| Range=[0 1] | NumMFs=2 |
| MF1='mf1':'trapmf',[-0.164 -0.0198 0.4352 0.556] | MF2='mf3':'trapmf',[0.455047619047619 0.599247619047619 1.05424761904762 1.17504761904762] |
| **[Input6] : Name='DBA'** | |
| Range=[0 1] | NumMFs=2 |
| MF1='mf1':'trapmf',[-0.164 -0.0198 0.4352 0.556] | MF2='mf3':'trapmf',[0.4365291005291 0.5807291005291 1.0357291005291 1.1565291005291] |

| **[Input7] : Name='DHA'** | |
|---|---|
| Range=[0 1] | NumMFs=2 |
| MF1='mf1':'trapmf',[-0.164 -0.0198 0.4352 0.556] | MF2='mf3':'trapmf',[0.439 0.578 1.07804232804233 1.16] |
| **[Input8] : Name='DA'** | |
| Range=[0 1] | NumMFs=2 |
| MF1='mf1':'trapmf',[-0.164 -0.0198 0.4352 0.556] | MF2='mf3':'trapmf',[0.447 0.578 1.16005291005291 1.17] |
| **[Input9] : Name='ED'** | |
| Range=[0 1] | NumMFs=2 |
| MF1='mf1':'trapmf',[-0.164 -0.0198 0.4352 0.556] | MF2='mf3':'trapmf',[0.455 0.602 1.16534391534392 1.17] |
| **[Input10] : Name='DM'** | |
| Range=[0 1] | NumMFs=2 |
| MF1='mf1':'trapmf',[-0.164 -0.0198 0.4352 0.556] | MF2='mf3':'trapmf',[0.452 0.581 1.15 1.17] |
| **[Input11] : Name='LM'** | |
| Range=[0 1] | NumMFs=2 |
| MF1='mf1':'trapmf',[-0.164 -0.0198 0.4352 0.556] | MF2='mf3':'trapmf',[0.441645502645503 0.580645502645503 1.0626455026455 1.1626455026455] |
| **[Input12] : Name='DSS'** | |
| Range=[0 1] | NumMFs=2 |
| MF1='mf1':'trapmf',[-0.164 -0.0198 0.4352 0.556] | MF2='mf3':'trapmf',[0.45 0.586 1.16 1.17] |
| **[Input13] : Name='DNS'** | |
| Range=[0 1] | NumMFs=2 |
| MF1='mf1':'trapmf',[-0.164 -0.0198 0.4352 0.556] | MF2='mf3':'trapmf',[0.444 0.597 1.0542328042328 1.16] |
| **[Input14] : Name='DLL'** | |
| Range=[0 1] | NumMFs=2 |
| MF1='mf1':'trapmf',[-0.164 -0.0198 0.4352 0.556] | MF2='mf3':'trapmf',[0.457 0.594 1.01719576719577 1.18] |

| **[Input15] : Name='DLS'** | |
|---|---|
| Range=[0 1] | NumMFs=2 |
| MF1='mf1':'trapmf',[-0.164 -0.0198 0.4352 0.556] | MF2='mf3':'trapmf',[0.45 0.57 1.07010582010582 1.17] |
| **[Input16] : Name='STL'** | |
| Range=[0 1] | NumMFs=2 |
| MF1='mf1':'trapmf',[-0.164 -0.0198 0.4352 0.556] | MF2='mf3':'trapmf',[0.45 0.581 1.09126984126984 1.17] |
| **[Input17] : Name='RG'** | |
| Range=[0 1] | NumMFs=2 |
| MF1='mf1':'trapmf',[-0.164 -0.0198 0.4352 0.556] | MF2='mf3':'trapmf',[0.45 0.597 1.07010582010582 1.17] |
| **[Output1] : Name='LD'** | |
| Range=[0 1] | NumMFs=3 |
| MF1='mf1':'trimf',[0.0071 0.0992063492063491 0.234] | MF2='mf2':'trimf',[0.302936507936508 0.507936507936508 0.71031746031746] |
| MF3='mf3':'trimf',[0.842544973544973 0.908544973544974 1.00354497354497 | ----- |

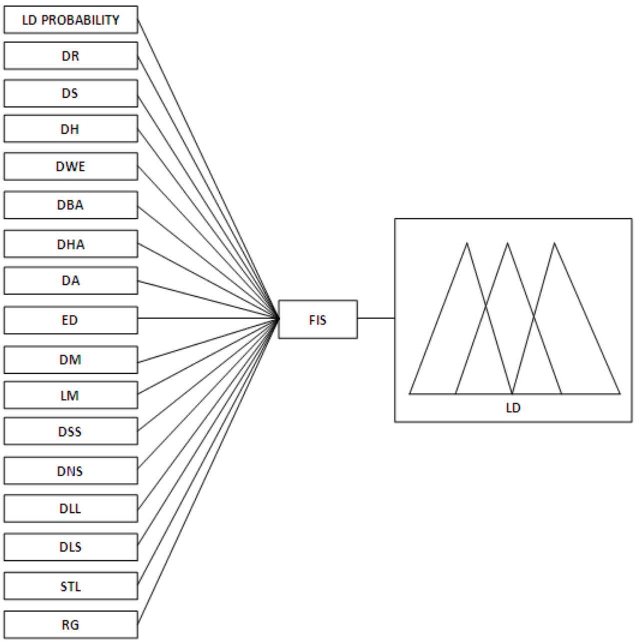


**Figure 3.16** Visualization of FIS

### *3.10.5.6 Fuzzy rules*

A fuzzy rule is a simple If-then rule with a condition and a conclusion. In our Fuzzy Logic System, a rule base is constructed to control the output variable. The rule viewer of LD fuzzy system is shown in Figure 3.16. Sample fuzzy rules for the LD prediction system from the rule viewer are listed below;

R1: If (LD_PROBLITY is mf3) and (DR is mf1) and (DS is mf1) and (DH is mf1) and (DWE is mf1) and (DBA is mf1) and (DHA is mf1) and (DA is mf1) and (ED is mf1) and (DM is mf1) and (LM is mf1) and (DSS is mf1) and (DS is mf1) and (DLL is mf1) and (DLS is mf1) and (STL is mf1) and (RG is mf1) then (LD is mf3)

R2: If (LD_PROBLITY is mf3) and (DR is mf1) and (DS is mf1) and (DH is mf1) and (DWE is mf1) and (DBA is mf1) and (DHA is mf1) and (DA is mf1) and (ED is mf1) and (DM is mf1) and (LM is mf1) and (DSS is mf1) and (DS is mf1) and (DLL is mf1) and (DLS is mf1) and (STL is mf3) and (RG is mf1) then (LD is mf3)

R3: If (LD_PROBLITY is mf1) and (DR is mf1) and (DS is mf3) and (DH is mf1) and (DWE is mf1) and (DBA is mf1) and (DHA is mf1) and (DA is mf1) and (ED is mf1) and (DM is mf1) and (LM is mf1) and (DSS is mf1) and (DS is mf1) and (DLL is mf1) and (DLS is mf1) and (STL is mf1) and (RG is mf1) then (LD is mf1)

R4: If (LD_PROBLITY is mf3) and (DR is mf1) and (DS is mf1) and (DH is mf1) and (DWE is mf1) and (DBA is mf1) and (DHA is mf1) and (DA is mf1) and (ED is mf1) and (DM is mf1) and (LM is mf1) and (DSS is mf1) and (DS is mf1) and (DLL is mf1) and (DLS is mf1) and (STL is mf1) and (RG is mf3) then (LD is mf3)

Once the inputs have been fuzzified, the degree to which each part of the antecedent has been satisfied for each rule is known. If the antecedent of a given rule has more than one part, the fuzzy operator is applied to obtain a number that represents the result of the antecedent for that rule [55]. Let's examine any of the rules. It can be seen that, the input to the fuzzy operator is two membership values from fuzzified input variables.

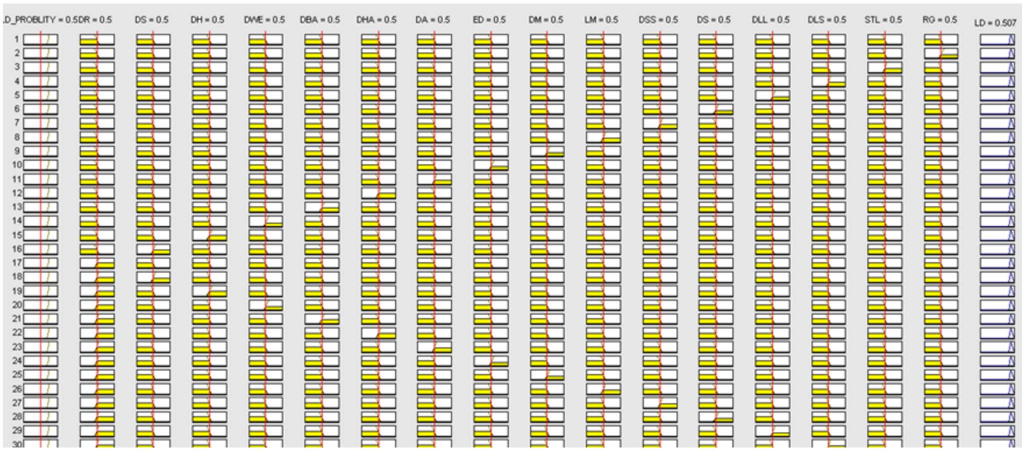


**Figure 3.17** Rule viewer of LD prediction system – fuzzy model

The rule viewer displays a roadmap of the whole fuzzy inference process. It is based on the fuzzy inference diagram. Each rule is a row of plots and each column is a variable. The first two columns of plots show the membership functions referenced by the antecedent, or the if-part of each rule. The third column of plots shows the membership functions referenced by the consequent, or the then-part of each rule. The rule viewer allows interpreting the entire fuzzy inference process at once. It also shows how the shape of certain membership functions influences the overall result. Since it plots every part of every rule, it can become unwieldy for particularly large systems, but, for a relatively small number of inputs and outputs, it performs well.

### *3.10.5.7 Fuzzy surface viewer*

Fuzzy Surface Viewer is presented with a two-dimensional curve that represents the mapping from input to output as shown in Figure 3.18. It has a special capability that is very helpful in cases with two inputs and one output. We can actually grab the axes and reposition them to get a different three-dimensional view on the data.

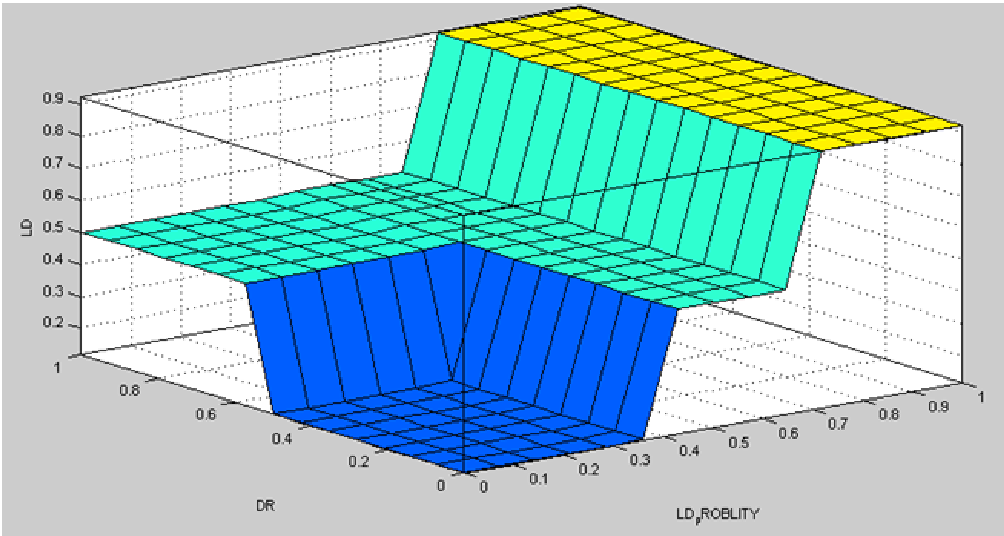


**Figure 3.18** Surface viewer of fuzzy model

### *3.10.5.8 Defuzzification*

After the inference step, the overall result is a fuzzy value. This result should be defuzzified to obtain a final crisp output. This is the purpose of the defuzzifier component of a Fuzzy Logic System. Defuzzification is performed according to the membership function of the output variable. The function of the defuzzifier is to convert the levels of belief in output fuzzy sets to a crisp decision variable of some kind. As described above, the result of fuzzy logic operations with fuzzy sets is invariably a conclusion in the form of a fuzzy set. In practice, the output of the defuzzifier process is a single value from the set.

There are several built-in defuzzifier methods. The centre of gravity method is the most commonly used for extracting a crisp value from a fuzzy set. This method calculates the weighted average of the elements in the support set.

#### *3.10.5.9 Evaluation of data using fuzzy*

We defined an evaluation function and it performs the evaluation of the fuzzy systems. The effectiveness of LD prediction data set is evaluated. Out of the 420 testing data set 417 are correctly evaluated and 600 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 3.22.

**Table 3.22** Evaluation of fuzzy model

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

## 3.11 Comparison and Results

In this chapter, we have performed different machine learning techniques for the prediction of learning disability. In the neural network algorithm, the time taken to build a model is higher, however the accuracy of the result is comparatively good. In the case of decision tree, it is very user friendly, the accuracy is also found good. The extracted rules from the classifier are also performed in a good manner. The wrong predictions obtained from decision trees for inconsistent data sets can be lead to a limited accuracy of decision tree models. Decision trees have pointed at the decision classes, which are not predominant for the given combination of input values like inconsistent data. The SVM algorithm implementation is found complex but accurate results are obtained. SVM is the best classifier because of its characteristics. These studies are pertinent to the supervised learning algorithms and fuzzy models. In the ensembled method of study adopted viz. bagging, higher accuracy is obtained compared to that of single classifiers used. A comparison of the results obtained on the studies, on the existing algorithms, is shown in Table 3.23.

**Table 3.23** Comparison of classification results

| Particulars | Classifier | | | |
|---|---|---|---|---|
| | NN | DT | SVM | Bagging |
| No. of cases | 1020 | 1020 | 1020 | 1020 |
| Correctly classified instances | 986 | 1002 | 1008 | 1013 |
| Incorrectly classified instances | 34 | 18 | 12 | 7 |
| Time taken to build a model (in seconds) | 29.2 | 0.22 | 0.61 | 0.53 |

## 3.12 Insights of LD

Learning disability refers to specific kinds of learning problems found in children. Early invention of LD is very helpful for the affected children to improve their problems and thereby increase their self confidence. Human beings are different from one another. Therefore the symptoms of LD are also different in them. In the present work we use the method of informal assessment for identifying learning disabilities in children. The main LD insights of the study are: development of new check list for LD assessment, significance of attributes in LD prediction, identification of data quality in classification and prediction, the significance of data mining in prediction of LD, how easily informal assessment can be performed with the help of knowledge base, etc. The main drawback found in all these classification algorithms is that, there is no proper solution for handling the inconsistent data in the data base. We have also developed an integrated knowledge based tool for achieving the above insights.

## 3.13 Summary and Conclusions

In this research work, the prediction of learning disability in school age children is implemented through various algorithms. The main problem considered, in the work for analysis and solving, is the design of an LD prediction tool based on machine learning techniques. A checklist containing 16 prominently identified signs and symptoms or attributes related to LD are developed and used for the general assessment of learning disability. Based on the checklists, basic information about the child is ascertained. In this research work, about 1020 real world datasets collected from relevant sources with the help of professionals engaged in the field are used.

A detailed study on the uses of different classification algorithms, viz. neural network, decision tree, support vector machine, bagging and fuzzy model are used for the prediction of learning disabilities in children. The data mining tool weka, is used for implementing the different classification algorithms in 1020 datasets. The main drawback found in all these classification algorithms is that, there is no proper solution for handling the inconsistent or unwanted data in the data base and also the classifier accuracy is low. Hence, the classification accuracy has to be increased by adopting new methods of implementation by proper data preprocessing. Studies, as part of this research work, are conducted to achieve these goals.

## 3.14 Contributions

The main contribution of the study is the development of a new check list, comprising of different attributes and sub attributes with particular score for each and determination of new insights into the relative importance of symptoms of LD. The knowledge obtained from these classification algorithms reveals the need of modification of data preprocessing. Thus new methods for LD prediction, based on machine learning techniques as well as fuzzy models, are developed for enhancing the accuracy of prediction.

*Chapter*

# 4

# IMPROVING PERFORMANCE BY NEW PRE-PROCESSING METHODS

Contents



## 4.1 Introduction

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 [40]. 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 [109]. Experiments on LD data sets were conducted to compare the performances of different classifiers as part of this research work. As explained in Chapter 3, the first draw back that determined in the existing classification algorithms is their problems in handling missing values. The significance of imputing the missing values in the study is proposed by two approaches to compare the vectors of all the attribute pairs from LD data set. These approaches introduced in the data set are closest fit approach, which assigns the value from the most similar pair to the missing value, and correlation based new algorithm, which assigns most correlated attribute value. The second drawback of the existing algorithms, while classification of the data set is the dimensionality reduction. In the checklist used for informal assessment of LD, there are a numerous amount symptoms. From these, we cannot easily predict the symptoms which are having more contribution in LD prediction. The well known algorithm PCA is applied in dimensionality reduction.

The pre-processed data, by way of imputing missing values and dimensionality reduction, as explained above, is then applied in five well known classifiers, viz. neural network (NN), decision tree (DT), fuzzy model, neuro fuzzy model and rough set model for evaluating the consistency of data. Even though, the pre-processed data can be applied in any classifiers, we have used the most developed or optimized classifier NN and the most user friendly classifier DT. We are concentrating only in the significance of pre-processed data in classification accuracy rather than classifier accuracy.

For imputing missing values we applied the closest fit algorithm and a new algorithm based on correlation. For attribute reduction, we are using PCA in fuzzy system and Johnson's reduction algorithm in rough set. In implementing these classifications, we are using most suitable functions and parameters.

Data preprocessing is a broad area and consists of a number of different strategies and techniques that are interrelated in complex ways [101]. The fuzzy logic system is handling the missing values in a better way as compared with the other classifiers. Different pre-processing methods are applied in other classification methods and it is seen that the results are improved comparing with the existing algorithms. 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 correlation based algorithm for imputing missing values and dimensionality reduction using PCA in the pre-processing stage. Subsequently, we have given the data set for fuzzification.

Rough sets theory has many advantages. For instance, it provides efficient algorithms for finding hidden patterns in data, finds minimal sets of data, evaluates significance of data, and generates minimal sets of decision rules from data. It is easy to understand and offer straightforward interpretation of results [110]. Those advantages can make the analysis easy that is why the rough sets approach is applied widely in many researches. The rough sets theory is of fundamental importance in artificial intelligence and cognitive science, especially in the areas of machine learning, knowledge acquisition, decision analysis, knowledge discovery, inductive reasoning and pattern recognition in databases, expert systems and decision support systems.

Using association rule mining based on apriori algorithm, the most frequent symptoms of learning disability are identified. Association rule, one of the most important knowledge representations in data mining, is applied here. The unsupervised classification algorithm k- means is used for clustering the LD data set as LD true and LD false clusters.

## 4.2 Imputing Missing Values

Missing values imputation is an actual yet challenging issue confronted in machine learning and data mining [111, 112]. Missing values may generate bias and affect the quality of the supervised learning process or the performance of classification algorithms [113,114]. 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 [8], 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. [8] and Zhang, et al. [112] 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 [115,116,117], Nearest Neighbor (NN) method [118,119]). 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 [118]. 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 [119].

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 [67]. 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 [8]. 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.

### 4.2.1 Closest fit algorithm

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 [120]. 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 Mathwork software 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:

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

$\text{distance}(e_i, e_i') = 0$ if $e_i = e_i'$,
$\text{distance}(e_i, e_i') = 1$ if $e_i$ and $e_i'$ are symbolic and $e_i \neq e_i'$ or $e_i = ?$ or $e_i' = ?$ and

$$\text{distance}(e_i, e_i') = 1 - \frac{|e_i - e_i'|}{|a_i - b_i|} \text{ if } e_i \text{ and } e_i' \text{ are numbers and } e_i \neq e_i', \text{ where}$$

$a_i$ is the maximum of values of $A_i$, $b_i$ is the minimum of values of $A_i$ and $A_i$ is an attribute.

### 4.2.2 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 including closest fit algorithm as explained in section 4.2.1 above. Compared to these methods, 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 Figure 4.1 and the detailed computing steps of the new algorithm are given in Figure 4.2;

**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.

**Figure 4.1** Algorithm of correlation based new missing value imputing method

## 4.3. 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 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 [8]. 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 using weka, 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 4.1. 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".

```
Computing steps:
  • Read the data set
  • If ai = null then
        ai_r_R = corr_coeff_R
        ai_r_L = corr_coeff_L
  • If ai_r_R > ai_r_L
        ai = ai+1
  • else
        ai = ai-1
  • elseif  \ if missing value have only left value \
        ai = ai-1
  • else
       ai=ai+1
          \if both sides of the missing value have null values\
  • If ai+1 && ai-1 = null then
       posn_ai = find isnan(ai)
       ai-r-R =corr_coeff of posn_ai_R
       ai-r-L = corr_coeff of posn_ai_L
  • If ai-r-R > ai-r-L then
       ai = posn(ai_r_R)
  • else
       ai = posn(ai-r-L)
 End
```

**Figure 4.2** Detailed computing steps of new correlation based algorithm

**Table 4.1** 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 |

## 4.4 Modified Data Pre-Processing and Performance Evaluation

This section consists of four parts. The first part of study is on modified data pre-processing with MLP, the second part is on modified data pre-processing with decision tree, the third part is on modified data pre-processing with fuzzy model and the last part is on modified data pre-processing with neuro fuzzy model. 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 the existing 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.

### 4.4.1 MLP with modified data pre-processing

The MLP with modified data pre processing is carried out in matlab and weka. In data preprocessing, closest fit algorithm is used for missing value imputation in matlab implementation and both closest fit algorithm and PCA are used in weka. These are explained below.

### *4.4.1.1 Implementation in matlab*

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 [121]. The neural network can be broadly divided in to three viz. feed forward networks, feedback network and self organization network. 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 [122].

Artificial Neural 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 [8].

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. 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 [123]. 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 [124]. Neural network consists of number of independent processors or neurons that communicate with each other via weighted connections [125]. For a unit in the output layer, the error Err is computed by the following equation;

$$Err = O(1-O)(T-O) \qquad \ldots\ldots 4.1$$

where, O is the actual output of a unit, T is the non-target value of the given training tuple; and O(1-O) is the derivative of logistic function.

Neural network architecture is known to be strong function approximation for prediction and classification problems. The architecture composed of three layers, viz. input, hidden and output [126]. It consists of 16 attributes and 2 output nodes. Three hidden layers are used. The output nodes obtained are LD-false and LD-true. The training process of the ANN used in this study is shown in Figure 4.3.

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

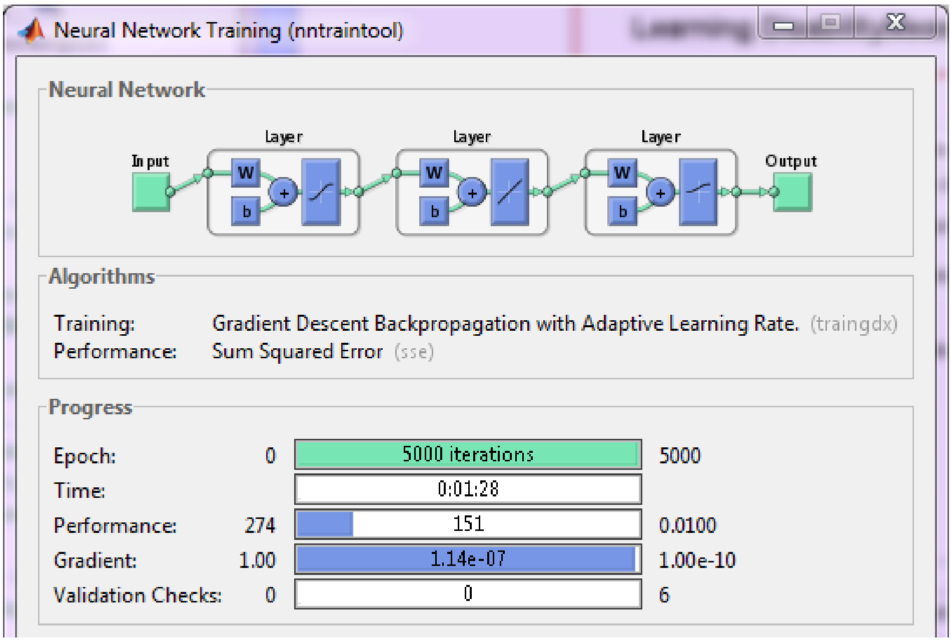


**Figure 4.3** Training process

The neural network train tool shows the input, output and hidden layers. We are provided the maximum epoch, 5000 iterations, for yielding more accurate results and best performance, which are executed the training in 1.28 seconds, setting a performance of 151. The plotting of training error obtained when the epoch is at 2749 from the NN train tool is shown in Figure 4.4.

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 performance of the training, ie. the gradient, validation checks and learning rate obtained at epoch is 2749 are represented in Figure 4.5. The training stops when the 5000 epochs has expired. The knowledge so obtained from this training process is used to predict the learning disability of a new client.

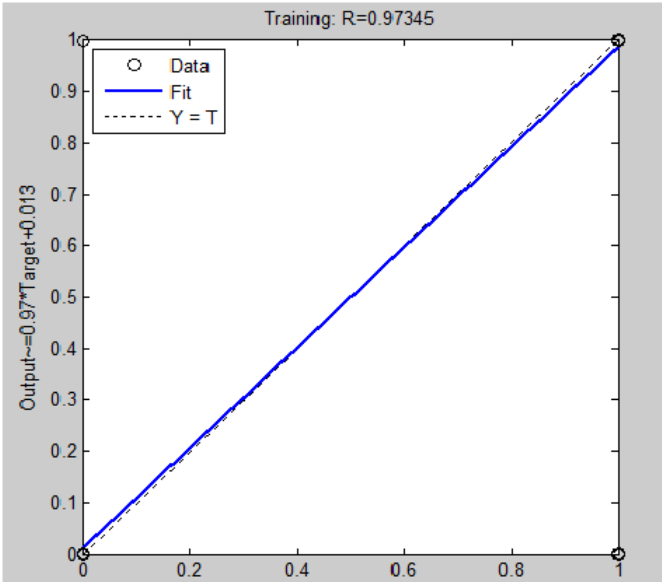


**Figure 4.4** Training error

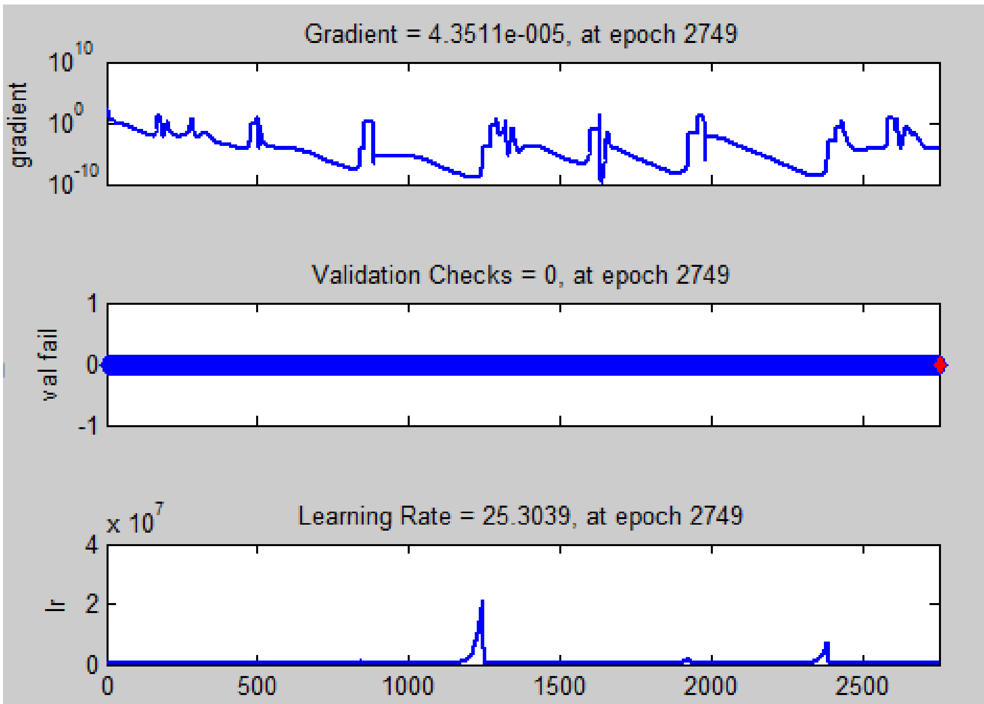


**Figure 4.5** Representation of gradient, validation check and learning rate

The knowledge obtained after training the neural network with 600 cases is used for testing 1020 cases. The result thus obtained shows 99.5% accuracy. The result obtained when we test the new data based on the model constructed in matlab is shown in Table 4.2.

**Table 4.2** Testing results of ANN

| Sl. No. | Particulars | Nos. |
|---|---|---|
| 1. | Data set used for training | 600 |
| 2. | No. of data set used for testing | 1020 |
| 3. | No. of instances correctly classified | 1015 |
| 4. | No. of instances incorrectly classified | 2 |

### *4.4.1.2 Implementation in WEKA*

After performing the classification in neural network, the implementation is carried out in weka, a machine learning workbench. The architecture of the neural network obtained from the study using the multilayer back propagation algorithm have 7 input nodes, 4 hidden nodes and 2 output nodes as shown in Figure 4.6

The output nodes obtained are LD-true and LD-false. The numbers of hidden nodes are determined through trial and error. Learning in a neural network involves modifying the weights and biases of the network in order to minimize cost function. All neural networks are basically trained until the error for each training iteration stopped decreasing. Here from the 1020 data sets we obtained 99.22% accuracy as shown in stratified cross-validation summaries given at Table 4.3 below. The classifier model of full training set, detailed accuracy by class and confusion matrix are shown in Tables 4.4, 4.5 and 4.6 below respectively. This approach of using neural network, MLP with data mining, gives more accurate results in prediction of learning disabilities in children [127].

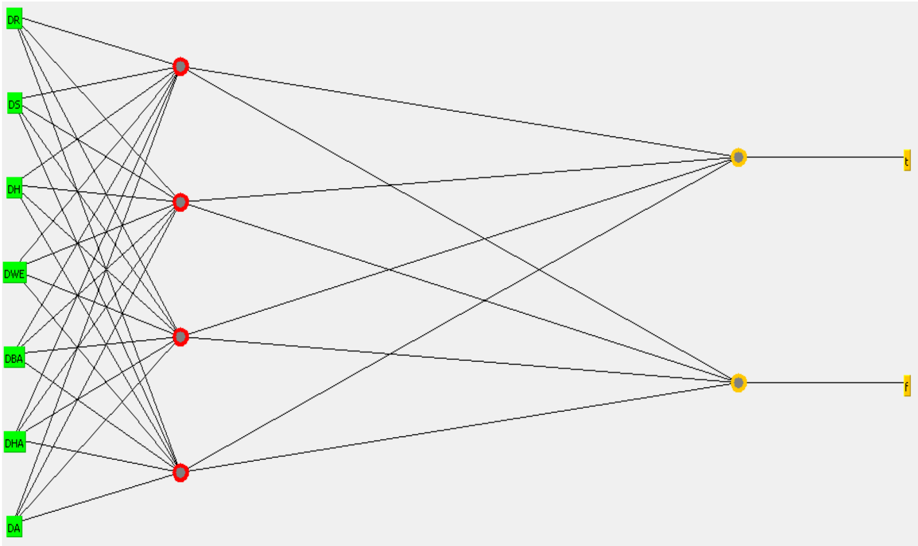


**Figure 4.6** Architecture of MLP using weka

**Table 4.3** Stratified cross-validation summary - ANN

| Sl. No. | Particulars | Value |
|---|---|---|
| 1 | Correctly Classified Instances | 1012 (99.2157%) |
| 2 | Incorrectly Classified Instances | 8 (0.7843 %) |
| 3 | Kappa statistic | 0.9804 |
| 4 | Mean absolute error | 0.01 |
| 5 | Root mean squared error | 0.0844 |
| 6 | Relative absolute error | 2.4849% |
| 7 | Root relative squared error | 18.8474 % |
| 8 | Total Number of Instances | 1020 |
| 9 | Time taken to build model | 15.50 seconds |

**Table 4.4** Classifier model of full training set

| Sl. No | Sigmoid | Threshold |
|---|---|---|
| 1 | Node 0 | -0.9490744090444844 |
| 2 | Node 1 | 0.9490744090444829 |
| 3 | Node 2 | -5.1497410622239945 |
| 4 | Node 3 | -1.7210279833018511 |
| 5 | Node 4 | -3.319582734528363 |
| 6 | Node 5 | 2.515736705210595 |

**Table 4.5** Detailed accuracy by class

| TP Rate | FP Rate | Precision | Recall | F-Measure | ROC Area | Class |
|---|---|---|---|---|---|---|
| 0.996 | 0.018 | 0.993 | 0.996 | 0.995 | 0.987 | t |
| 0.982 | 0.004 | 0.989 | 0.982 | 0.986 | 0.987 | f |

**Table 4.6** Confusion matrix

| a | b | Classified as |
|---|---|---|
| 734 | 3 | a = t |
| 5 | 278 | b = f |

## 4.4.2 Decision Tree with modified data pre-processing

In this part of the study, modified data pre processing with J48 is carried out in matlab as explained below. In the pre processing stage, we have performed correlation based new algorithm for imputing missing values.

### *4.4.2.1 Implementation in matlab*

The purpose of this part of study is to introduce a new supervised learning algorithm in decision tree, which is modified from the existing J48 algorithm, with an emphasis in data mining, for prediction of LD, in such a view to overcome the drawbacks of the existing algorithm.

Incomplete, noisy and inconsistent data are commonplace properties of large real world databases and data warehouses [8]. On assessment of learning disability, incomplete data can occur for a number of reasons including relevant data may not be recorded due to misunderstanding. Our aim is to apply the preprocessing step to make the data more suitable for data mining. For inducing the decision tree, we have developed a new algorithm after incorporating modifications to the existing J48 algorithm. The new algorithm is given below. Here the correlation based new algorithm is used for missing value imputation and information gain ratio is used for attribute selection in decision tree formation. The algorithm developed by us, for this work, is given below.

**Algorithm: Generate_decision_tree**. Generate a decision tree from the training samples of data partition

**Input:**

- Data preprocessing using correlation based new algorithm;
- Data partition, which is a set of training samples and their associated class labels;
- Attribute list, set of candidate attributes;
- Attribute_selection_method.

**Output :** A decision tree.

**Method:**

(1) perform data preprocessing using correlation based algorithm;

(2) create a node *N*;

(3) **if** samples are all of the same class, *C* then

(4) return *N* as a leaf node labeled with the class *C*;

(5) **if** attribute list is empty then

(6) return *N* as a leaf node labeled with the most common class in samples; //majority voting//

(7) select test attribute, the attribute among attribute list with the highest information gain;

(8) label node *N* with test attribute;

(9) **for** each known value $a_i$ of test attribute;

(10) grow a branch from node *N* for the condition test_attribute = $a_i$;

(11) let $s_i$ be the set of samples in samples for which test_attribute = $a_i$;

// a partition//

(12) **if** $s_i$ is empty then

(13) attach a leaf labeled with the most common class in samples;

(14) **else** attach the node returned by Generate_decision_tree to node *N*;

**Endfor**

The modified decision tree is shown in Figure 4.7. From this tree the rules for predicting LD are extracted. After imputing the missing values, which have a good impact in classification and prediction, eight attributes viz. DR, DS, DA, DSS, DWE, ED, RG and DHA are selected from the 16 attributes. The number of leaves and size of the tree are same as that of the existing J48, but the selected attributes are different, hence the rules are also different. The important factor in the proposed approach is that these rules extracted from the tree shows accuracy in the tune of 95%. The test results are shown in Table 4.7

**Table 4.7** Testing results of decision tree

| Sl. No. | Particulars | Nos. |
|---|---|---|
| 1. | Data set used for training | 500 |
| 2. | No. of data set used for testing | 1020 |
| 3. | No. of instances correctly classified | 1017 |
| 4. | No. of instances incorrectly classified | 3 |

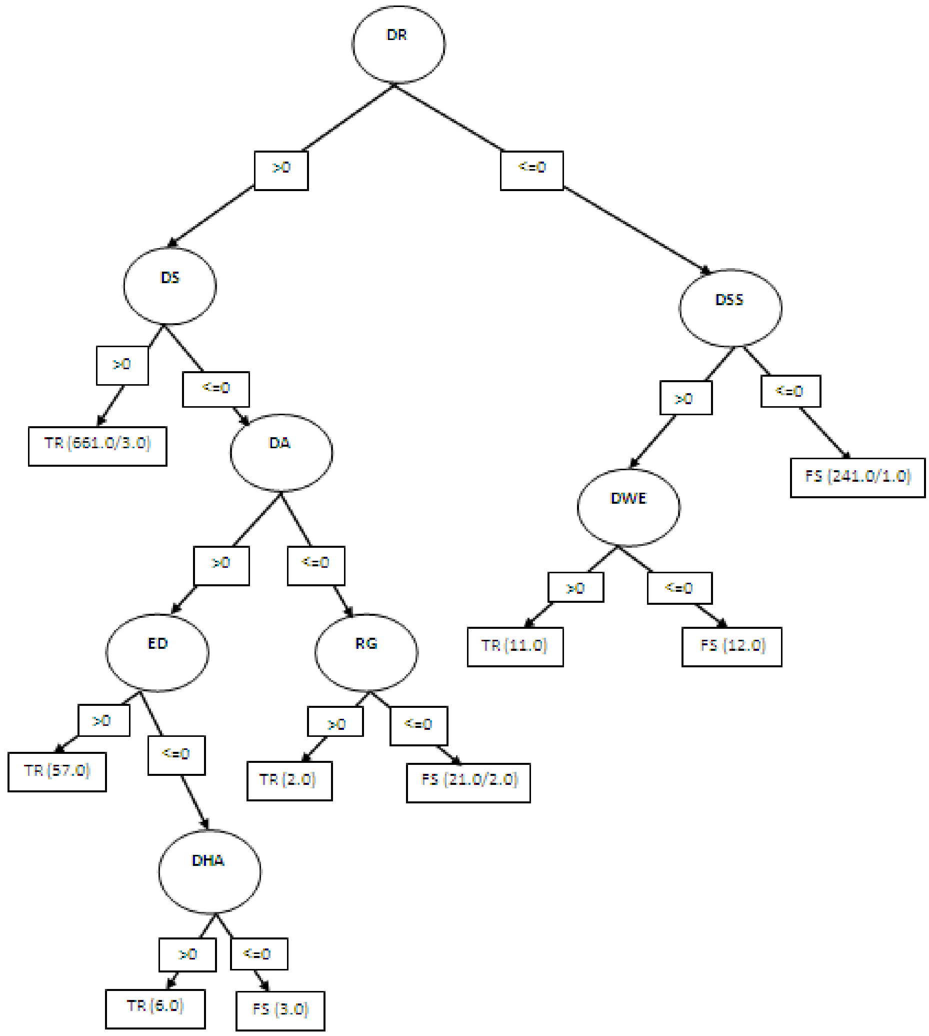


**Figure 4.7** Modified decision tree

The rules extracted from the above modified J48 decision tree are given under:

R1: IF DR=Y, DS=Y THEN LD = Y

R2: IF DR=Y, DS=N, DA=Y, ED=Y THEN LD = Y

R3: IF DR=Y, DS=N, DA=Y, ED=N, DHA=Y THEN LD = Y

R4: IF DR=Y, DS=N, DA=Y, ED=N, DHA=N THEN LD = N

R5: IF DR=Y, DS=N, DA=N, RG=Y THEN LD = Y

R6: IF DR=Y, DS=N, DA=N, RG=N THEN LD = N

R7: IF DR=N, DSS=Y, DWE=Y THEN LD = Y

R8: IF DR=N, DSS=Y, DWE=N THEN LD = N

R9: IF DR=N, DSS=N THEN LD = N

It is easy to read a set of rules directly off a decision tree. One rule is generated for each leaf. The antecedent of the rule includes a condition for every node on the path from the root to that leaf and the consequent of the rule is the class assigned by the leaf [67]. This procedure produces rules that are unambiguous in that the order in which they are executed is irrelevant. However in general, rules that are read directly off a decision tree are far more complex than necessary and rules derived from trees are usually pruned to remove redundant tests. The rules are so popular because each rule represents an independent knowledge. New rule can added to an existing rule sets without disturbing them, whereas to add to a tree structure may require reshaping the whole tree. In this section, we presented a method for generating a rule set from a decision tree [128]. In principle, every path from the root node to the leaf node of a decision tree can be expressed as a classification rule. The test conditions encountered along the path form the conjuncts of the rule antecedent, while the class label at the leaf node is assigned to the rule consequent. The expressiveness of a rule set is almost equivalent to that of a decision tree because a decision tree can be expressed by a set of mutually exclusive and exhaustive rules. We are using the MatLab tool for implementing the data preprocessing and classification. The classification algorithm J48 is now modified. As per the modified algorithm, we got 99.70 % correctly classified instances, and the classifier can be used for classifying and predicting the LD accurately.

### 4.4.3 Fuzzy model with attribute reduction

In this part of study, as explained above, 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.

The information about FIS is given in Table 4.8. 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 4.9. The visualization of FIS in LD prediction is shown in Figure 4.8.

**Table 4.8** 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 |

**Table 4.9** Structure of FIS

| Name='LD_PROBLITY' | Range=[0 1] |
|---|---|
| NumMFs=3 | MF1='mf1':'trapmf',[-0.518518518518519 -0.198518518518519 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] | ---- |

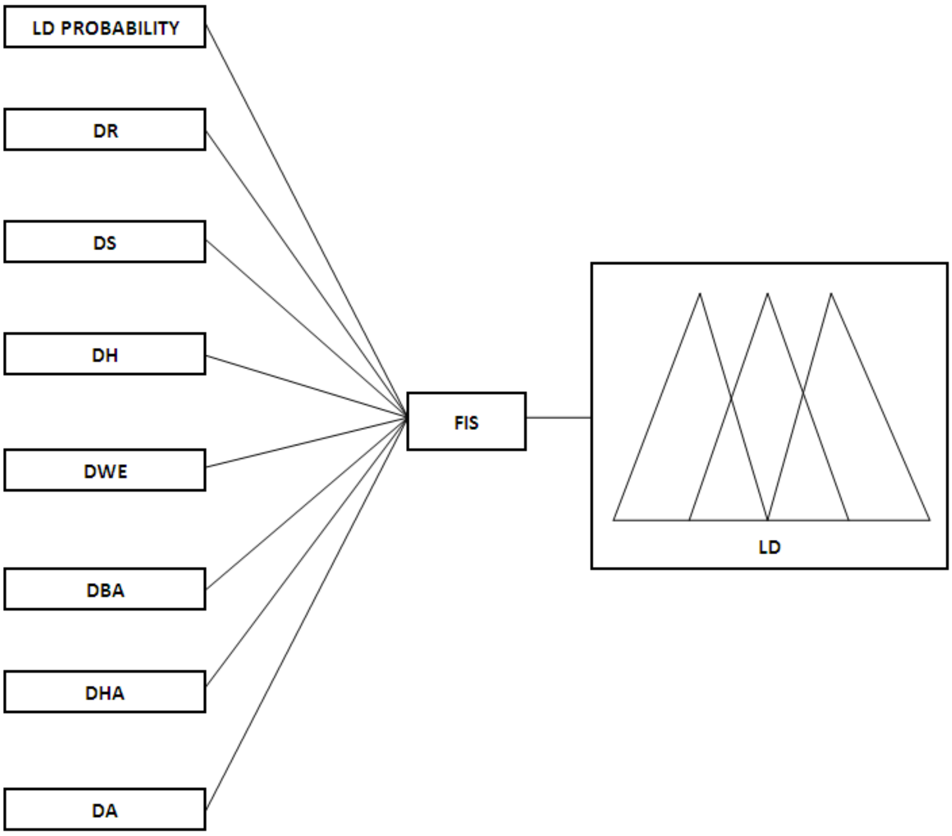


**Figure 4.8** Visualization of FIS

The rule viewer and surface viewer of LD fuzzy system with attribute reduction are shown in Figure 4.9 and 4.10 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)

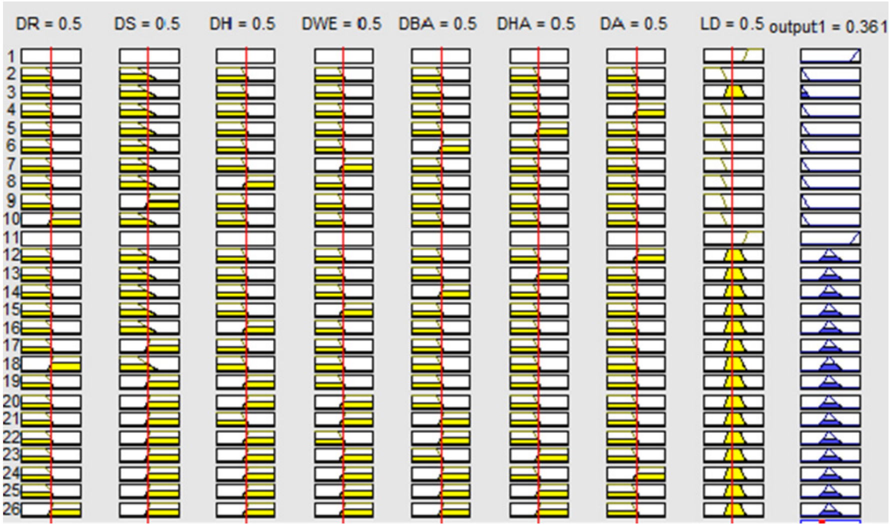


**Figure 4.9** Rule viewer of LD prediction system – fuzzy with attribute reduction

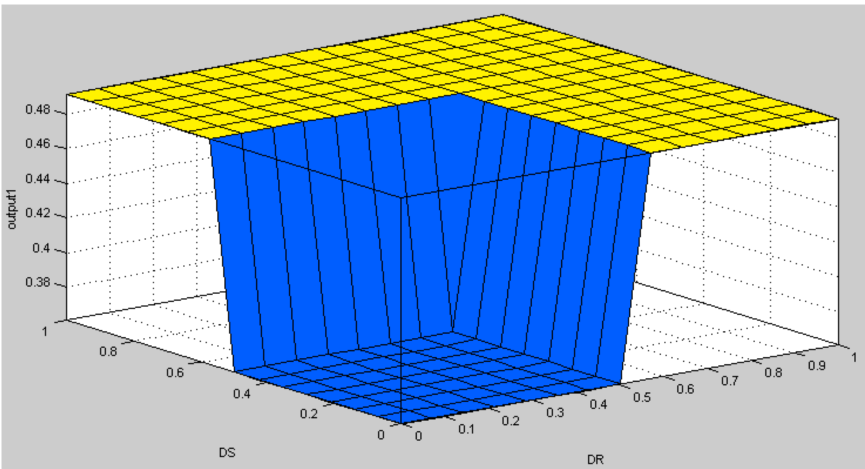


**Figure 4.10** Surface viewer of fuzzy with attribute reduction

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 420 testing data set 419 are correctly evaluated and 600 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 4.10.

**Table 4.10** 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 |

### 4.4.4 Neuro Fuzzy Model

Neuro fuzzy system is a combination of artificial neural network and fuzzy systems [8]. Neuro fuzzy system combines the learning capabilities of neural networks if the linguistic rule interpretation of fuzzy inference system [129]. 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 [8]. 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 [130].

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 [122]. 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.

#### *4.4.4.1 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 [131]. 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 Figure 4.11.

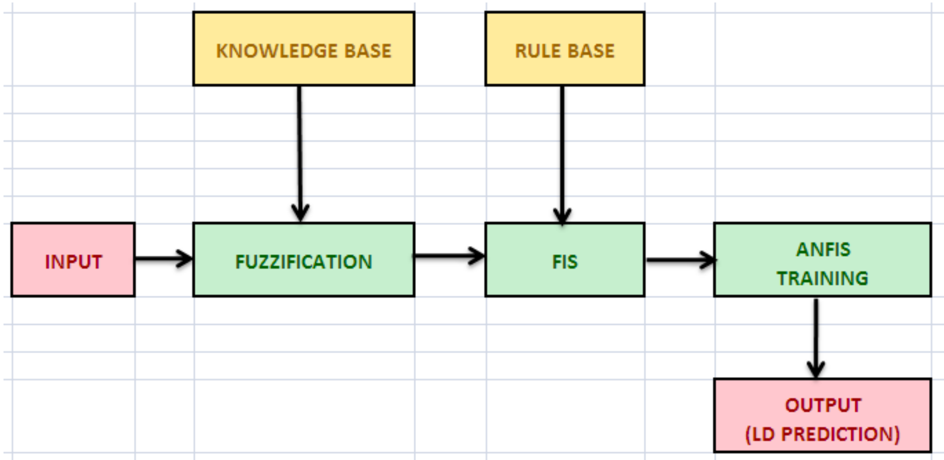


**Figure 4.11** Neuro fuzzy system flowchart for LD prediction

#### 4.4.4.2 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.

#### *4.4.4.3 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 4.11 and 4.12 respectively.

**Table 4.11** 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] |

**Table 4.12** 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 input membership function editor of one of the seven extracted attributes, DS, and that of probability LD are shown in Figures 4.12 and 4.13 respectively. In our research work, the membership function of output variable in LD prediction is obtained as shown in Figure 4.14.

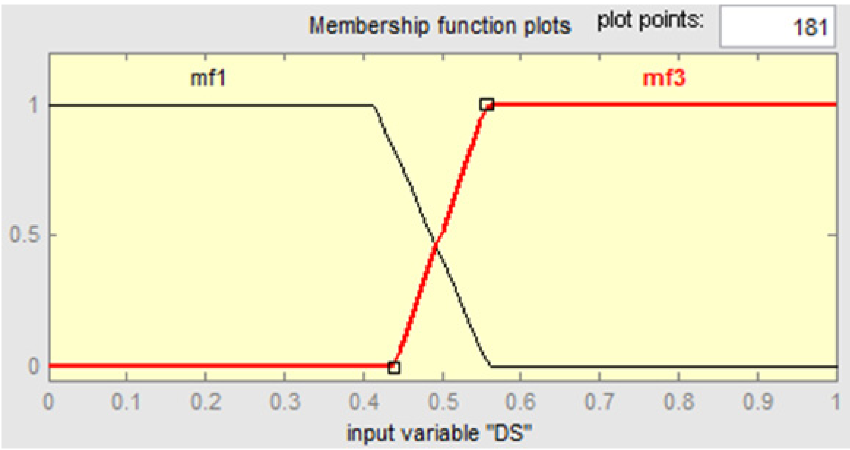


**Figure 4.12** Input membership function of attribute DS

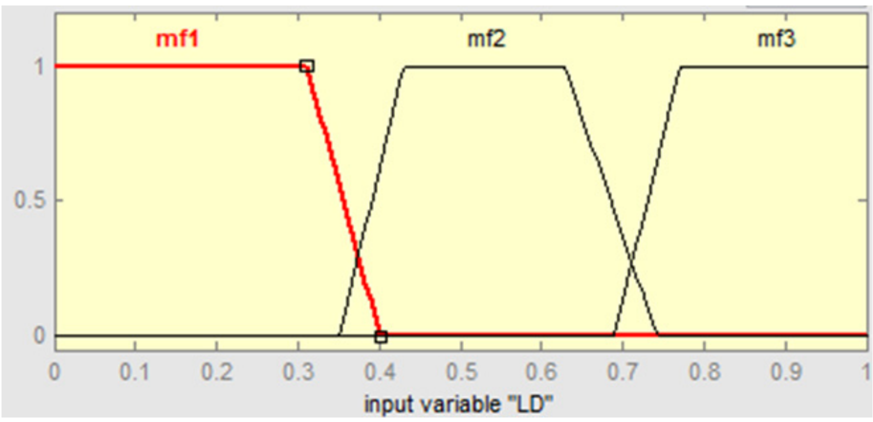


**Figure 4.13** Input membership function of LD

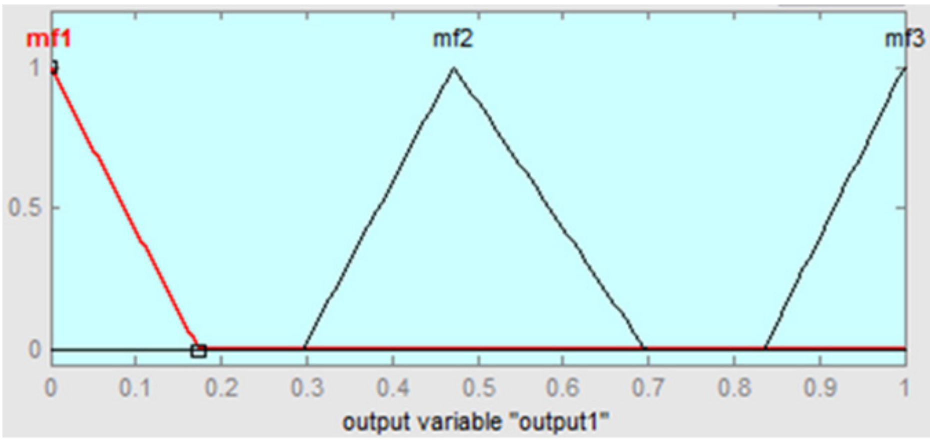


**Figure 4.14** Output membership function of LD

### *4.4.4.4 Fuzzy inference system (FIS)*

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[132]. 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 4.13.

### *4.4.4.5 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 [133]. 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

**Table 4.13** 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 [134]. 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 closest fit algorithm and the 16 attributes are reduced to 5 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.

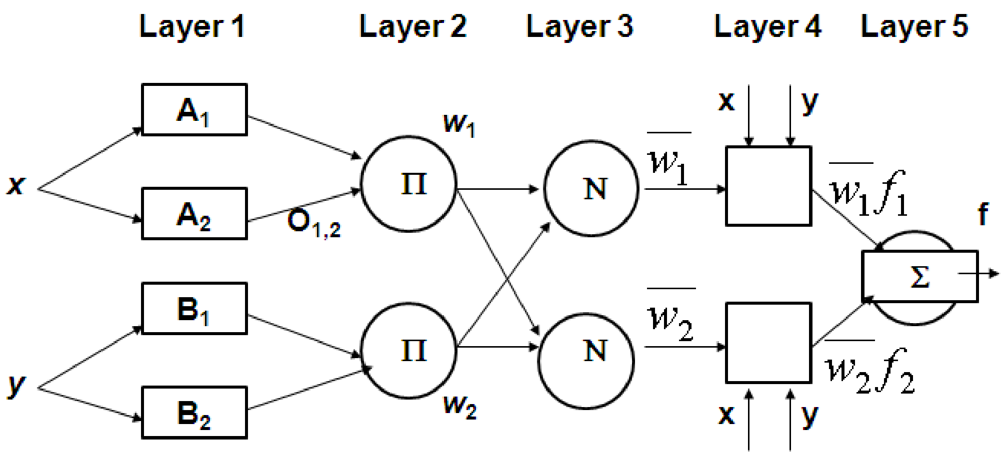


**Figure 4.15** ANFIS architecture

The ANFIS architecture is shown in Figure 4.15 [135]. The functions of each layer of ANFIS architecture are explained in Table 4.14.

**Table 4.14** 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 3and the other inputs are the input variables from layer1. |
| 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 4.15. The step size decreases to 0.009000 after epoch 9. Designated epoch number reached at epoch 10 and ANFIS training completed there.

**Table 4.15** 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 |

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 4.16 and 4.17 respectively.

**Table 4.16** 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 4.17** 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: 600 | Number of checking data pairs: 0 |
| Number of fuzzy rules: 243 | ---- |

#### *4.4.4.6 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 4.18. These results show 100% accuracy.

**Table 4.18** Test results of LD prediction - ANFIS

| Sl. No. | Particulars | Nos. |
|---|---|---|
| 1 | Data Set used for training | 600 |
| 2 | No. of data set used for testing | 420 |
| 3 | No. of Instances correctly classified | 420 |
| 4 | No. of Instances incorrectly classified | 0 |

## 4.5 Rough Set Model

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

Rough set theory (RST) is useful for rule induction from incomplete data sets. Using this approach we can distinguish between three types of missing attribute values: lost values, attribute-concept values and do not care conditions [54].

### 4.5.1 Proposed approach

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

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

$$\underline{P}_X = \{ x \mid [x]_P \subseteq X \}$$

$$P_X = \{ x \mid [x]_P \cap X \neq \emptyset \}$$

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

The lower approximation of a target set is a conservative approximation consisting of only those objects, which can positively be identified as members of the set. The upper approximation is a liberal approximation, which includes all objects that might be members of target set. The accuracy of the rough-set representation of the set X can be given by the following equation [54].

$$\alpha_P(X) = \frac{|PX|}{|\overline{P}X|}$$

In order to consider the features of rough set used to predict the important signs and symptoms of learning disability, we are using the concept of information table, decision table, global covering and data reduct. An information table consists of different variables called attributes and cases called objects. Variables are present in columns and cases in rows. The attributes contained in the information table are the signs and symptoms of learning disabilities.

This part of study consists of two phases. The first phase involves illustration of rough set theory with the aid a small data set consists of 10 cases. In the second phase, the actual implementation using 1020 real cases through rough set tool, Rosetta, is involved.

However, for convenience, the first phase of study, containing only six attributes and five cases in the sample information table given at Table 4.19 are presented for illustration. By using the real time datasets for assessing the LD in children, the type of LD belongs to each child is identified. Since such identification of LD in each child using all the attributes is a very difficult task, certain rules which enable to easily identify different symptoms which are causing LD are used. Based on these, the symptoms of LD in each child are strictly assessed. The mined rules are used for finding the relationship between the symptoms of learning disability.

**Table 4.19** Sample information table

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

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

(x,y) Є IND (B) if and only if ρ(x, a)=ρ(y, a) for all aЄB.................4.2

Obviously, IND(B) is an equivalence relation. Equivalence relation is present through partitions [136]. Partition relation is a family of mutually disjoint nonempty sets of U, called blocks. So the union of all blocks is U. The partition induced by IND (B) will be denoted by B*. Blocks of B* are called elementary set of B.

### 4.5.2 Determination of reduct and core

There is subsets of attributes, which can, by itself, fully characterize the knowledge in the database; such an attribute set is called a reduct [137]. The reduct of an information system is not unique: there may be many subsets of attributes, which preserve the equivalence-class structure expressed in the information system.

The set of attributes which is common to all reducts is called the core: the core is the set of attributes which is possessed by every legitimate reduct, and therefore consists of attributes which cannot be removed from the information system without causing collapse of the equivalence-class structure [137]. It is possible for the core to be empty, which means that there is no indispensable attribute.

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

{DR}*={1,2,3,4,5}; comparing with A* {DR}*≠A*, therefore {DR} is not a reduct.

{DR, DS}*={1,3,4,5},{2}; comparing with A*, {DR, DS}*≠A*, therefore {DR, DS} is not a reduct.

{DR, DS, DH}*={1,3},{2},{4,5}; comparing with A*, {DR,DS,DH}*≠A*, therefore {DR, DS, DH} is not a reduct.

{DR, DS, DH, DWE}*={1},{2},{3}, {4,5}; But here, {DR,DS,DH,DWE}*=A*, therefore {DR, DS,DH,DWE} is a reduct.

Reducts are important subsets of attributes. A subset B of the set A is called a reduct, if and only if (i) B* = A* and (ii) B is minimal with the property (B-{a})* ≠ A* for all a Є B [31]. Based on these properties, only {DR, DS, DH, DWE} is reduct. Similarly, by considering another set of attributes, we are also getting {DH, DWE, DBA, DHA} as reduct.

Computing of all reducts, by this method, is time consuming with respect to the number of attributes considered. In such cases, computation of all the reducts is a complex task. So, it is restricted to compute a single reduct using a heuristic algorithm LEM1 [138]. The first step of this algorithm is elimination of the leftmost attribute from the set and check whether the remaining set is reduct or not. If the set is not reduct, put the attribute back into that set and eliminate the next attribute for similar checking. Like, we are eliminating until the last attribute for reduct checking.

As explained, there are already two properties, {DR, DS, DH, DWE} and {DH, DWE, DBA, DHA}, as reducts from our sample information table. Now, the left most attribute, DR is eliminated from the reducts and check whether the remaining combined set is reduct or not. Then we are getting {DS, DH, DWE, DBA, DHA}* = A*; therefore {DS, DH, DWE, DBA, DHA} is a reduct. Then eliminate the next left most attribute, DS and check whether the remaining set is reduct or not. Now we are getting {DH, DWE, DBA, DHA}*

= A*, therefore {DH, DWE, DBA, DHA} is a reduct. Similarly, after eliminating DH we are getting {DWE, DBA, DHA}* = A*, therefore {DWE, DBA, DHA} is also a reduct. But after eliminating DWE, the remaining set {DBA, DHA}* ≠ A*, therefore, the set {DBA, DHA} is not a reduct. Hence put the attribute DWE into this set and eliminate next attribute DBA, getting the set {DWE, DHA} for reduct checking. Now, {DWE, DHA} ≠ A*, hence {DWE, DHA} is also not a reduct. Finally, after eliminating all other attributes other than the last one, we are getting {DHA}* ≠ A* resulting {DHA} is also not a reduct. From these steps, we are arriving at the conclusion that, the LEM1 algorithm forms the set of attributes {DWE, DBA, DHA} as the core reducts. Core attributes are generated from the preprocessed data set to help evaluating the rules [139].

The determination of reducts from the data set using LEM1 algorithm, as explained above, is tedious, time consuming and complex in nature. Hence, another algorithm viz. Johnson's Reduction Algorithm is used for the entire data set. This algorithm is applied by using the rough set tool kit, Rosetta, for analysis of data, on our 1020 real datasets (cases) with 16 attributes and we are obtaining the set of core attributes (reducts) as {DR, DS, DWE, DA, ED, DM, DLL, RG} with a length of 8 as shown in Table 4.20 below.

**Table 4.20** Reduct results

| Sl. No. | Reduct | Support | Length |
|---|---|---|---|
| 1 | {DR,DS,DWE,DA,ED,DM,DLL,RG} | 100 | 8 |

### 4.5.3 Decision table

One of the important aspects in the analysis of decision tables is the extraction and elimination of redundant attributes. The identification of the most important attribute from the data set is also an equally important aspect.

Redundant attributes are attributes that could be eliminated without affecting the degree of dependency between the remaining attributes and decision [140]. The degree of dependency is a measure used to convey the ability to discern objects from each other. In a decision table, variables are presented in columns. But it contains two categories- attributes and decisions. Decision table has only one decision Y or N, i.e. LD yes or LD no. Rows of decision table, like information tables, are labeled by case names.

A checklist, containing signs and symptoms of LD, ie. attributes, is used for evaluating LD. In the Sample Decision Table given at Table 4.21, there are two elementary sets - {LD}:{1,4,5} for LD has value yes (Y) and {LD}:{2,3} for LD has value no (N). Elementary sets of decisions are called concepts. Decision table contains the cases, which are diagnosed by experts. Decision tables are crucial to data mining. Based on RST, there are two approaches of data mining from complete data sets. They are global covering and local covering [141]. In our study, we are considering only global covering of consistent data in which the entire attributes are used for analysis.

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

**Table 4.21** Sample decision table

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

### 4.5.4 Global covering

A minimal subset of the set of all attribute, such that the substitution partition depends on it, is called global covering. It may be selected on the basis of lower boundaries. In the case of inconsistent data the system computes lower and upper approximations of each concept [141]. In global approach, each concept is represented by the substitution partition.

Relative reducts or rule sets may be induced using global coverings. We start from the definition of a partition being finer than another partition. Let α and β be the partitions of U. α is finer than β, denoted α ≤ β, if and only if, for each block X of α, there exists a block Y of β such that X ≤ Y. Let d be a decision. Then, a subset B of the attribute set A is a global covering if and only if (i) B* ≤ {d}* and (ii) B is minimal with the property (B-{a})* ≤ {d}* is false for any a Є B [136]. Based on these properties, we are checking all subsets of A in the sample decision table, with {LD}* = {1,4,5},{2,3}, with cardinality equal to one.

(i) {DR}* = {1,2,3,4,5}; then {DR}* ≤{LD}* is false.

(ii) {DS}*={1,3,4,5},{2}; then {DS}* ≤{LD}* is false.

(iii) {DH}*={1,2,3,},{4,5}; then {DH}*≤{LD}* is false.

(iv) {DBA}*={1,3},{2,4,5};then{DBA}*≤{LD}* is false.

(v) {DHA}*={1},{2,3,4,5};then {DHA}*≤{LD}* is false.

(vi) {DWE}*={1,4,5}{2,3};then {DWE}*={LD}* is true.

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

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

{DR, DS} = {1,3,4,5},{2}; then {DR, DS}*≤{LD}* is false

{DS, DH} = {1,3},{2,4,5}; then {DS, DH}*≤ {LD}* is false

{DH, DWE} = {1},{2},{3},{4,5}; then {DH, DWE}*≤ {LD}* is false

{DBA, DHA} = {1},{2,4,5},{3}; then{DBA, DHA}*≤{LD}* is false

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

{DR, DS, DH}*={1,3}{2}{4,5}; then {DR, DS, DH}* ≤ {LD}* is false.

{DWE, DBA, DHA}* = {1}{2}{3}{4,5}; then {DWE, DBA, DHA}*≤{LD}* is true.

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

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

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

From the above, we are getting 3 sets of attributes, viz. {DWE}, {DWE, DBA, DHA} and {DR, DS, DH, DWE} as global covering, considering our sample decision table. Obviously, the worst time complexity of the algorithm for computing all global covering is the same as the algorithm for computing all reduct. Thus, we should restrict our attention for computing a single global covering. For this, we are using the same procedure of elimination of left most attribute, one by one, and checking the condition {A}* ≤ {d}*, until the last element is eliminated. A single global covering is used for rule induction [136]. We restrict our attention to attributes from the global covering and check the cases in the decision table. If such a rule condition is not exists in the decision table, it is not consistent and this rule condition can be dropped. From this concept, we can induce certain rules.

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

R1: (DR, Y) (DS, Y) (DH, N) (DWE, Y) = (LD, Y)

R2: (DH, N) (DWE, N) = (LD, N)

R3: (DH, Y) (DWE, Y) = (LD, Y)

R4: (DH, N) (DWE, Y) = (LD, Y)

R5 (DWE, Y) = (LD, Y)

R6: (DWE, N) = (LD, N)

If we are applying the 1020 data set in Rosetta tool of rough set, we obtained 47 rules. Some of these rules having high strength are listed below.

R1: DR(1) AND DS(1) AND DWE(1) AND DA(0) AND ED(0) AND DM(0) AND DLL(1) AND RG(0) => LD(t) [strength – 234]

R2: DR(1) AND DS(1) AND DWE(1) AND DA(1) AND ED(1) AND DM(1) AND DLL(1) AND RG(0) => LD(t) [strength – 112]

R3: DR(1) AND DS(1) AND DWE(1) AND DA(1) AND ED(1) AND DM(1) AND DLL(0) AND RG(0) => LD(t) [strength – 107]

R4: DR(1) AND DS(1) AND DWE(1) AND DA(0) AND ED(1) AND DM(0) AND DLL(1) AND RG(0) => LD(t) [strength – 75]

R5: DR(1) AND DS(1) AND DWE(1) AND DA(0) AND ED(1) AND DM(1) AND DLL(1) AND RG(0) => LD(t) [strength – 60]

R6: DR(0) AND DS(0) AND DWE(0) AND DA(1) AND ED(0) AND DM(0) AND DLL(0) AND RG(0) => LD(f) [strength – 41]

R7: DR(0) AND DS(0) AND DWE(0) AND DA(1) AND ED(1) AND DM(1) AND DLL(0) AND RG(0) => LD(f) [strength – 29]

R8: DR(1) AND DS(1) AND DWE(0) AND DA(1) AND ED(1) AND DM(1) AND DLL(1) AND RG(0) => LD(t) [strength – 28]

R9: DR(1) AND DS(0) AND DWE(0) AND DA(1) AND ED(0) AND DM(0) AND DLL(0) AND RG(0) => LD(f) [strength – 28]

R10: DR(1) AND DS(1) AND DWE(1) AND DA(0) AND ED(0) AND DM(0) AND DLL(0) AND RG(0) => LD(t) [strength – 27]

R11: DR(1) AND DS(1) AND DWE(0) AND DA(1) AND ED(1) AND DM(1) AND DLL(0) AND RG(0) => LD(t) [strength – 27]

R12: DR(0) AND DS(0) AND DWE(1) AND DA(0) AND ED(0) AND DM(0) AND DLL(0) AND RG(0) => LD(f) [strength – 25]

### 4.5.5 Result analysis and findings

The classification results on the 1020 real data sets with 16 attributes are obtained from Rosetta, the rough set tool kit for analysis of data is shown in Table 4.22. In Rosetta tool, Johnson's reduction algorithm is used for obtaining the reduct results and Naive Bayes Batch classifier is used for obtaining the classification results.

**Table 4.22** Classification results – rough set

| | | Predicted | | |
|---|---|---|---|---|
| Actual | | T | F | |
| | T | 574 | 58 | 0.9082 |
| | F | 4 | 384 | 0.9897 |
| | | 0.9930 | 0.8688 | 0.9399 |

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

As a pre-processing before data mining, a subset of original data, which is sufficient to represent the whole data set, is generated from the initial detailed data contained in the information system. This subset contains only minimum number of independent attributes for prediction of LD. This attribute is used to study about the original large data set.

The knowledge is tested against the test set. It is then clearly seen that the patterns found in the training set are valid also for other data. Therefore, if the knowledge gained from the training set is the general knowledge, it is correct for most parts of the test set as well. The learning disability detection process can be considered as a decision making process. The rules generated by considering the original data set give a strong platform for making decisions. We are interested in applying these rules for making decisions

## 4.6 Apriori Algorithm

Data Mining is an analytic process designed to explore data in search of consistent patterns and/or systematic relationships between variables, and then to validate the findings by applying the detected patterns to new subsets of data [4]. In data mining, association rule mining is a well researched method for discovering relationships between variables in large data sets. These powerful exploratory techniques have a wide range of applications in many areas of research [142]. These techniques enable analysts and researchers to uncover hidden patterns in large data sets. The goal of the techniques described in this section is to detect the relationships or associations between specific values of categorical variables in large data sets.

Apriori is the best known algorithm to mine association rules. It uses a breadth-first search strategy to count the support of itemsets. In this section, how the apriori algorithm is used for mining the rules and finding frequent signs and symptoms of learning disability is well explained. Apriori is a seminal algorithm for mining frequent item sets for Boolean association rules.

The name of the algorithm is based on the fact that, it uses prior knowledge of frequent properties. The apriori algorithm is as given below;

**Algorithm**

```
L1 = {large 1-itemsets};
for (k=2; Lk-1 ≠ Ø; k++) do begin
    Ck = apriori-gen (Lk-1);
    for all transactions t Є D do begin
       C1 = subset(Ck, t)
       for all candidates c Є Ct do
          c.count++;
   end
   Lk = {c Є Ck|c.count>minsup}
End
```

### 4.6.1 Extraction of frequent symptoms of LD

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 know the relationship and association of behavior sets or signs and symptoms of LD, then we can easily predict the type of learning disability present and its related symptoms. The first task in handling learning disability is the construction of a database consisting of the signs, characteristics and level of difficulties faced by these children. Data mining can be used as a tool for analyzing the complex decision tables associated with the learning disability. The 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, association rules are generated and these rules are then be used to predict learning disability. The checklist given at Table 4.23 below is used to construct an information table 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, but a checklist to focus our understanding of the learning disability. This check list is only a typical one and it is not the general check list used for the entire study. This part of the work consists of two phases. The first phase is the realization of apriori algorithm through a small set of dataset. In the second phase, we use weka for the implementation of apriori on the 1020 datasets.

**Table 4.23** Check list – apriori algorithm

| Sl. No | Particulars about the Child | Attributes | Yes | No |
|---|---|---|---|---|
| 1 | Difficulty with Reading | DR | | |
| 2 | Difficulty with Spelling | DS | | |
| 3 | Difficulty with Handwriting | DH | | |
| 4 | Difficulty with Written Expression | DWE | | |
| 5 | Difficulty with Basic Arithmetic skills | DBA | | |
| 6 | Difficulty with Higher Arithmetic skills | DHA | | |
| 7 | Difficulty with Attention | DA | | |
| 8 | Easily Distracted | ED | | |
| 9 | Difficulty with Memory | DM | | |
| 10 | Lack of Motivation | LM | | |

Based on the information obtained from the above checklist, a data set is generated. This set is in the form of an information system. A complete information system expresses all the knowledge available about objects being studied. In this section, methods for mining the simplest form of frequent behavior of LD, boolean frequent behavior sets are used. DR, DS, DH, etc. represents the signs and symptoms of LD. As seen in the checklist, only the 10 most common signs of LD are used as a sample set. From these signs, the frequent ones are determined. For finding the accuracy of the rule, a small database is used. From this database, 10 cases or symptoms are considered. These ten symptoms of learning disabilities are considering in the small database, and ten cases in the database, ie. |d| = 10. The information table given at Table 4.24 contains these cases and attributes.

**Table 4.24:** Information table

| CASES | ATTRIBUTES | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | DR | DS | DH | DWE | DBA | DHA | DA | ED | DM | LM |
| 1 | N | Y | N | N | N | N | Y | N | N | N |
| 2 | Y | Y | Y | Y | N | N | Y | N | Y | Y |
| 3 | N | Y | Y | Y | N | N | Y | Y | Y | Y |
| 4 | N | Y | Y | N | N | N | Y | Y | Y | Y |
| 5 | Y | N | Y | Y | N | N | Y | Y | Y | Y |
| 6 | N | N | N | N | Y | N | Y | Y | Y | Y |
| 7 | N | Y | Y | N | Y | N | Y | N | Y | Y |
| 8 | N | Y | N | Y | N | N | Y | Y | Y | N |
| 9 | N | N | Y | Y | Y | N | N | Y | N | Y |
| 10 | Y | Y | Y | N | Y | N | Y | N | N | N |

Apriori employs an iterative approach, known as level wise search, where n symptoms are used to explore (n+1) symptoms sets. First, the set of frequent behavior sets, say B1, is found by scanning the database to accumulate the count for each behavior set 1 and collecting the behaviors that satisfy minimum support. The resulting set is denoted by L1. Next, L1 is used to find L2, the set of frequent behavior set 2, which is used to find L3 and so on until no more frequent n behaviors sets can be found out. Finding of each $L_k$ requires one full scan of the database. Apriori property is all non empty subsets of frequent behaviors sets must also be frequent. By definition, if behaviors sets S do not satisfy the minimum support threshold, min-support, then S is not frequent, i.e. P(S) < min-support. If the behavior A added to the behavior sets S, then the resulting symptom sets cannot occur more frequently than S. Therefore (S U A) is not frequent either; that is, P(S U A) < min-support. This property belongs to a special category of properties called antimonotone [8], in the sense that, if a set cannot pass a test, all of its supersets will fail the same test as well. It is called antimonotone, because the property is monotonic in the context of failing a test. Here we consider two step process, join and prune. The sample training decision tables are given in Table 4.25 [143].

**Table 4.25** Training decision tables

| Behavior Sets | Support Count |
|---|---|
| DR | 3 |
| DS | 6 |
| DH | 7 |
| DWE | 5 |
| DBA | 4 |
| DHA | 0 |
| DA | 9 |
| ED | 6 |
| DM | 7 |
| LM | 7 |

**B1**

| Behavior sets | Support count |
|---|---|
| DS | 6 |
| DH | 7 |
| DWE | 5 |
| DA | 9 |
| ED | 6 |
| DM | 7 |
| LM | 7 |

**L1**

| Behavior Sets | Support Count |
|---|---|
| DS, DH | 6 |
| DS, DA | 5 |
| DS, ED | 6 |
| DH, DM | 5 |
| DH, LM | 5 |
| DA, DM | 6 |
| DA, LM | 5 |
| ED, DM | 5 |
| DM, LM | 6 |

**L2**

| Behavior Sets | Support Count |
|---|---|
| DH, DA, DM, LM | 5 |

**L3**

In the first iteration of algorithm, each behavior is a member of the set of candidate-1 behavior sets B1. The algorithm simply scans all of the cases in order to count the number of occurrence of each behavior. Suppose that, minimum support count required is 5, i.e min-support = 5, we are referring to absolute support because we are using support count. The set of frequent-1 behavior set, L1, can then be determined. It consists of the candidate 1 behaviors sets satisfying minimum support to discover the set of frequent - 2 behaviors set L2. The algorithm uses the join step L1 JOIN L2 to generate the candidate of set-2 behaviors sets. Next find the third frequent behaviors sets from the join step L2 JOIN L3 and determine the L3 sets. Then the algorithm uses L3 JOIN L4 to generate candidate sets of fourth frequent behaviors sets and thus B4 is null set.

It is common to divide the database into two parts to create training and a test set. One of these parts, for instance 70% of the data, is used as a training set and examined by the data mining system. The rest of the original database is used as a test set to see if the knowledge acquired from the training set was general or not. By examining the data in the database, the system tries to create general rules and descriptions of the patterns and relations in the database. The goal is to gain knowledge which is valid not only in the specific database considered but also for other similar data. The knowledge may be tested against the test set. It will then be clear, if the patterns found in the training set are valid also for other data. If the knowledge gained from the training set is the general knowledge, it will be correct for most parts of the test set as well.

### 4.6.2 Generating rules

While we implementing this algorithm in weka on 1020 dataset, the apriori full training set associate model obtained is given in Table 4.26 below.

The association rules that generated from the data sets are given below;

R1 : {LD=YES} => {DR=YES}

R2 : {DS = YES} =>{DR = YES}

R3 : {DHA= NO}=> {DLS=NO}

R4 : {DBA = NO, RG = NO} => {DH = NO}

R5 : {DR=YES} => {DS=YES}

R6 : {DS = NO} => {DLS = NO}

R7 : {DBA = NO, DS = NO} => {DH = NO}

R8 : {DR = YES} => {LD = YES}

R9 : {DBA = NO} => {DH = NO}

R10 : {DBA = NO, DLS = NO} => {DH = NO}

**Table 4.26** Apriori full training set associator model

| Sl. No. | Particulars | Value |
|---|---|---|
| 1 | Minimum support | 0.6 (613 instances) |
| 2 | Minimum metric<confidence> | 0.9 |
| 3 | Number of cycles performed | 8 |
| 4 | Size of set of large item sets L(1) | 10 |
| 5 | Size of set of large item sets L(2) | 13 |
| 6 | Size of set of large item sets L(3) | 9 |

**Table 4.27**: Confidences of the rules

| Rule | Confidence |
|---|---|
| R1 | 99% |
| R2 | 97% |
| R3 | 96% |
| R4 | 94% |
| R5 | 93% |
| R6 | 93% |
| R7 | 93% |
| R8 | 92% |
| R9 | 92% |
| R10 | 92% |

If the minimum confidence threshold is, say 90%, then the above rules are output sure. The certainty of the generated rules or confidence of each rule is calculated by considering the cases given in Table 4.24. The confidences of these rules are given in Table 4.27 above. Based on these studies, we find that DR and DS are significant symptoms in the prediction of LD. This was certified by professionals in the field.

### 4.6.3 Review of this work

In this section, an approach to handle learning disability database to predict frequent symptoms of the learning disability and the interrelationship among them in school aged children are considered. This study mainly focuses on association rule mining, because accuracy of decision making can be improved by applying these rules. The rules generated by considering the original data set give a strong platform for making decisions. It is interested in applying these rules for making decisions.

## 4.7 Clustering

This study consists of using K-means algorithm in creating the clusters of LD. Clustering is a tool for data analysis, which solves classification problem [144]. Its object is to distribute cases into groups, so that the degree of association to be strong between members of same clusters and weak between members of different clusters. This way each cluster describes in terms of data collected, the class to which its members belong. Clustering is a discovery tool. It may reveal associations and structure in data which though not previously evident. The results of cluster analysis may contribute to the definition of a formal classification scheme. Clustering helps us to find natural groups of components based on some similarity. Clustering is the assignment of a set of observations into subsets so that observations in the same cluster are similar in some sense. Clustering is a method of unsupervised learning, and a common technique for statistical data analysis used in many fields, including machine learning, data mining, pattern recognition, image analysis and bioinformatics.

### 4.7.1 Classification by clustering

In clustering, using the k-means algorithm for classifying LD, the clustered instances are classified into two, viz. LD - Yes and LD - No. The K-means algorithms iterates over the whole dataset until convergence is reached.

### 4.7.2 Methodology and results

The K-means algorithm is a most well-known and commonly used partitioning method. It takes the input parameter, K, and partitions a set of N objects into K clusters so that the resulting intra-cluster similarity is high but the inter cluster similarity is low [145]. When clustering a dataset, the right number k of clusters to use is often not obvious, and choosing k automatically is a hard algorithmic problem [146]. Cluster similarity is measured in regard to the mean value of the objects in a cluster [8]. The working of algorithm is like it randomly selects the K objects, each of which initially represents cluster mean or center. For each of the remaining objects, an object is assigned to the cluster to which it is the most similar, based on the distance between the objects and the cluster mean. It then computes the new mean for each cluster. This process iterates until the criterion function converges.

An important step in most clustering is to select a distance measure, which will determine how the similarity of the two elements is calculated. This will influenced the shape of the clusters, as some elements may be close to one another according to one distance and farther away according to one another. Another important distinction is whether the clustering uses symmetric or asymmetric distances [8]. Many of the distance function have the property that distances are symmetric. Here, we are using the binary variables. A binary variable has two states 0 or 1, where 0 means that variable is absent and 1 means that is present. In this study, we use the partitioning method K- means algorithm, where each cluster is represented by the mean value of the objects in the cluster. In this partitioning method, the database has N objects or data tuples, it constructs K partitions of the data, where each partition represents a cluster and it classifies the data into K groups. Each group contains at least one object and each object must belong to exactly one group. The clustering results obtained by us are shown in Table 4.28.

**Table 4.28** Clustering results

| Particulars | LD=0 (No) | LD=1 (Yes) |
|---|---|---|
| Clustered instances | 525 Nos.<br>51.47% | 495 Nos.<br>48.53% |

The clustering history indicating LD = No and LD = Yes is shown in Table 4.29 and the cluster visualizer is shown in Figure 4.16 [147].

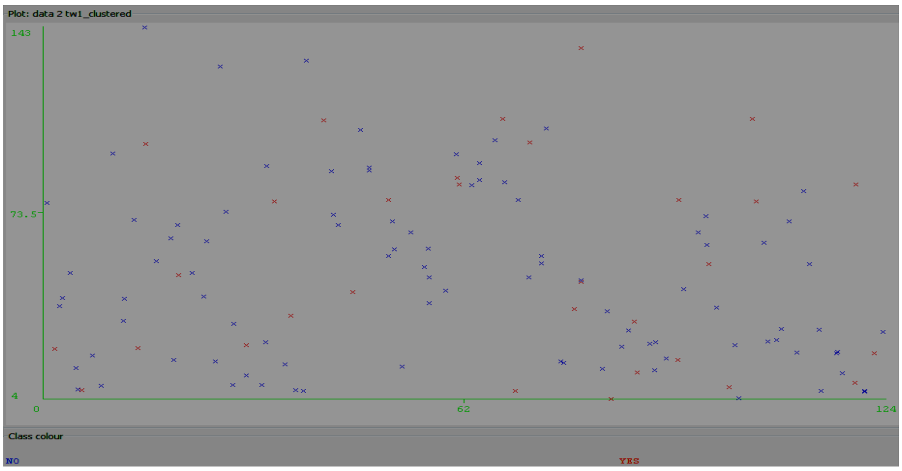


**Figure 4.16** Cluster visulizer

**Table 4.29** Clustering history

| Sl. No | Attributes | Full Data (1020) | LD = 0 (No) (525) | LD = 1 (Yes) (495) |
|---|---|---|---|---|
| 1 | DR | 0.7775 | 0.5676 | 0.9996 |
| 2 | DS | 0.7373 | 0.5105 | 0.9773 |
| 3 | DH | 0.1549 | 0.0210 | 0.2967 |
| 4 | DWE | 0.7176 | 0.5771 | 0.8664 |
| 5 | DBA | 0.2333 | 0.0667 | 0.4097 |
| 6 | DHA | 0.2235 | 0.0533 | 0.4037 |
| 7 | DA | 0.4402 | 0.2800 | 0.6098 |
| 8 | ED | 0.4833 | 0.1162 | 0.8719 |
| 9 | DM | 0.4049 | 0.0629 | 0.7669 |
| 10 | LM | 0.4490 | 0.0571 | 0.8638 |
| 11 | DSS | 0.4029 | 0.0229 | 0.8052 |
| 12 | DNS | 0.1206 | 0.0038 | 0.2442 |
| 13 | DLL | 0.5873 | 0.5162 | 0.6625 |
| 14 | DLS | 0.0853 | 0.0648 | 0.1070 |
| 15 | STL | 0.5353 | 0.1867 | 0.9043 |
| 16 | RG | 0.0912 | 0.0762 | 0.1070 |
| No. of iterations | | | | 3 |
| Within cluster sum of squared errors | | | | 2367.217653 |
| Missing values globally replaced with mean/mode | | | | |

### 4.7.3 Review of this work

In this study, we are used 1020 data sets with 16 attributes. The k means algorithm in clustering classifies the data set into LD - Yes and LD - No. The main drawback of this algorithm is that the noisy or unwanted data are present in each cluster. So it is found that by removing the unwanted data at the pre processing stage, we can improve the accuracy of the unsupervised classification k-means clustering.

## 4.8. Comparisons and Results

The rules generated from the data sets by applying J48 and modified J48 two algorithms are different. The rules derived from the tree have been compared. The certainty of the generated rules, ie. confidence of each rule, is calculated by considering the cases given in datasets. The comparison or quality parameters of the rules R1, R2, R3, etc. for both existing (inconsistent) and modified (consistent) algorithms are graphically represented in Figure 4.17.

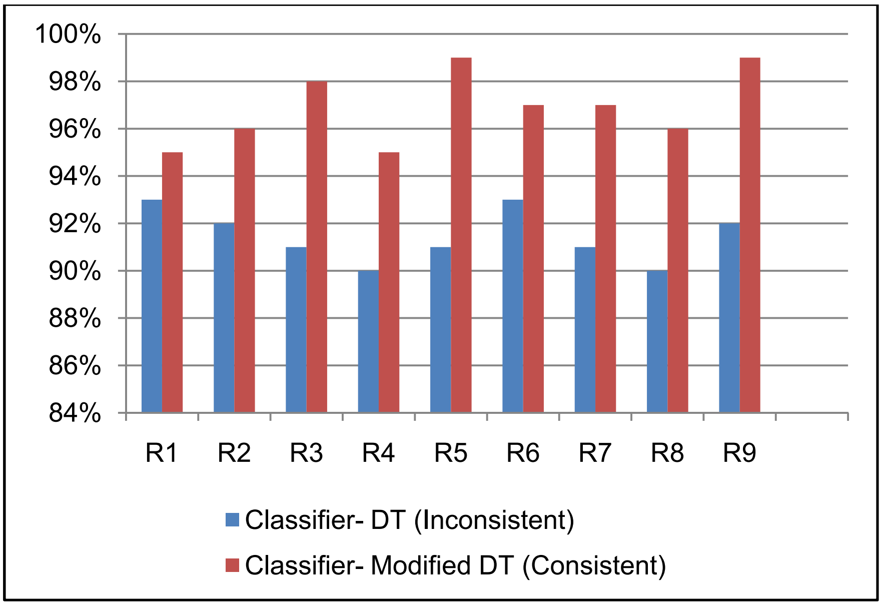


**Figure 4.17** Quality parameters of rules

A comparison of the results of the above studies along with that of the existing algorithms studied in the previous chapter is shown in Table 4.30. From these results, it can be concluded that, our new algorithm for imputing missing values, viz. correlation based, used in J48 decision tree is best. So it can be seen that, performance of a classifier is fully depends on the quality of the data used. Based on this new algorithm, the impact of missing values in the data set can be identified and seen that its contribution is very much affects the ultimate results. Such situations can be avoided by using our new algorithm.

**Table 4.30** Comparison of classification results - classifiers

| Particulars | Classifiers | | | | | | |
|---|---|---|---|---|---|---|---|
| | NN | DT | SVM | Bagging | NN (Modified) | | DT (Modified) |
| | Weka | | | | Matlab | Weka | Matlab |
| Correctly classified instances ( in number and percentage) | 986<br>96.67% | 1002<br>98.24% | 1008<br>98.82% | 1013<br>99.31% | 1015<br>99.51% | 1012<br>99.22% | 1017<br>99.70% |
| Incorrectly classified instances ( in number and percentage) | 34<br>3.33% | 18<br>1.76% | 12<br>1.18% | 7<br>0.79% | 5<br>0.49% | 8<br>0.78% | 3<br>0.30% |
| Time taken to build a model (in seconds) | 29.2 | 0.22 | 0.61 | 0.53 | 1.28 | 15.50 | 0.11 |

From this, it has seen that, the rules extracted from the modified J48 algorithm are more accurate and precise. By using the various rules developed by us using this modified approaches, we can easily and precisely predict the learning disability of any child. This study will be helpful for the parents, teachers and school authorities in diagnosing the child's problem at an early stage. Hence these results will be helpful and beneficial for the educational as well as medical communities.

The output of different models undergone in our study with modified data pre-processing are also compared. A comparison of these results is shown in Table 4.31below. From the comparison, it can be seen that, fuzzy system performs good results in classification compared to rough set approach. However in case of attribute reduction or feature selection, rough set is found performing in a good manner.

**Table 4.31** Comparison of classification results-models

| Particulars | Classifiers | | | |
|---|---|---|---|---|
| | Fuzzy | Fuzzy with data pre-processing | Neuro fuzzy | Rough set |
| No. of cases used | 420 | 420 | 420 | 1020 |
| Correctly classified instances ( in number and percentage) | 417<br>99.29% | 419<br>99.76% | 420<br>100% | 958<br>93.92% |
| Incorrectly classified instances ( in number and percentage) | 3<br>0.71% | 1<br>0.24% | 0<br>0% | 62<br>6.08% |

## 4.9 Summary and Conclusion

In this chapter, we have studied the performances of the algorithms developed in classifiers and models. The new approaches in data pre processing improve the accuracy and performance of classifiers. The accuracy of decision-making can be improved by using our good method of missing value imputing thereby improving the quality of data. This study has been carried out on 1020 data sets. Modified MLP and J48 decision tree application on discrete data shows that these are better than existing MLP and J48 in terms of efficiency and complexity.

The importance of this study is that, using the modified approaches we can easily and accurately predict whether a child has LD or not. This model would be highly beneficial to the parents of students to diagnose their child's problem and then teachers/school authorities to identifying the problems of children at an early stage and provide a good solution. The study is beneficial to the young students to know their problems at an early stage and thereby avoid their academic losses.

In this chapter, we are proposed different models for predicting learning disabilities. For handling vague information, fuzzy and rough set approaches are most suitable. After the study, it is seen that, these approaches are efficient in handling data without pre-processing. In the case of dimensionality reduction, the fuzzy system is complex since rule formation is difficult. In the case of rough set, the attribute reduction is good and classifier results are also comparably good as it is done without pre processing.

The extracted rules by the application of rough set theory in LERS data mining system and use of Rosetta tool in rough set data analysis in particular emphasis to classification, in prediction of the learning disabilities in school age children, are very effective for the prediction [148].

## 4.10 Contribution

The main contribution of this chapter is the development of new methods to improve the performance of different classifiers in LD prediction, development of our new algorithm is based on correlation for missing value imputing. The number of attributes is reduced by eliminating the unwanted and redundant ones by using principal component analysis. This helps in reducing the time of classification. The accurate results from the classifiers help the children in improving their confidence. Adopting closest fit algorithm with classifier is a very good contribution in the determination of accuracy and performance improvement in classifiers.

Occurrence of vague information is quite happen in LD prediction. The present approaches are contributed good results in handling this vague information. These models give good results even without any data preprocessing as done for the existing classifiers. The main contribution in this study is the development of new fuzzy and rough set approaches in LD prediction.

The apriori algorithm is used to determine the frequent signs and symptoms of LD. We got accurate results from this study. The main drawback found is that the dependency of the pruning process on the threshold value, whereas the study performed with clustering algorithm is a type of unsupervised algorithm. This algorithm easily classifies the instances into two clusters, LD - Yes and LD - No. The main drawback found is that, in each instance, if it a noisy data, which includes in any one of the clusters.

Based on these knowledge gained, we are developing different models and classifiers for LD prediction and also a knowledge based tool, viz. Knowledge Based Learning Disability Prediction (KBLDP) tool.

Chapter 5

# DEVELOPMENT OF AN INTEGRATED KNOWLEDGE BASED TOOL FOR LD PREDICTOION

Contents



## 5.1 Introduction

In this chapter, the integration of all the techniques described earlier into a single expert system tool is presented and explained. A Graphical User Interface (GUI) is designed. The developed tool, which stores the details of the children in the student database and retrieves their LD report as and when required, helps in classifying LD as well as imputes the missing values in the data set in an accurate way. The main stages of information processing adopted in the tool design are data preprocessing, mining of data, getting knowledge and prediction of LD.

## 5.2 System Flowchart

The system flowchart and a layout of the system component of the integrated knowledge based tool for LD prediction are given below at Figure 5.1 and Figure 5.2 respectively.

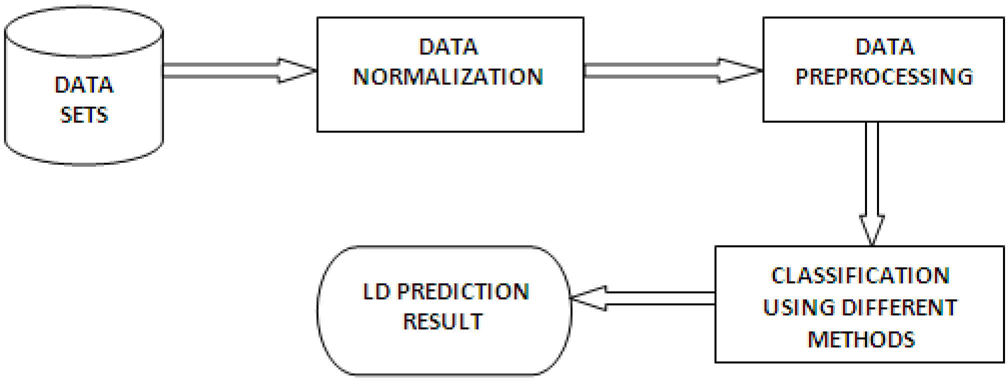


**Figure 5.1** System flowchart

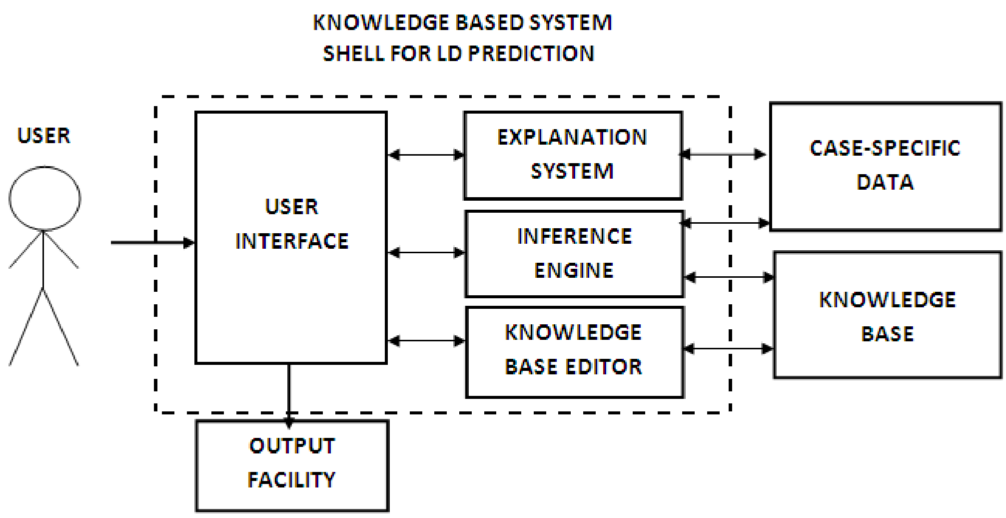


**Figure 5.2** Knowledge based system layout

## 5.3 Architecture of the Tool

The architecture of the developed tool is as shown in Figure 5.3 given below. The main tool components are explained under section 5.4 below.

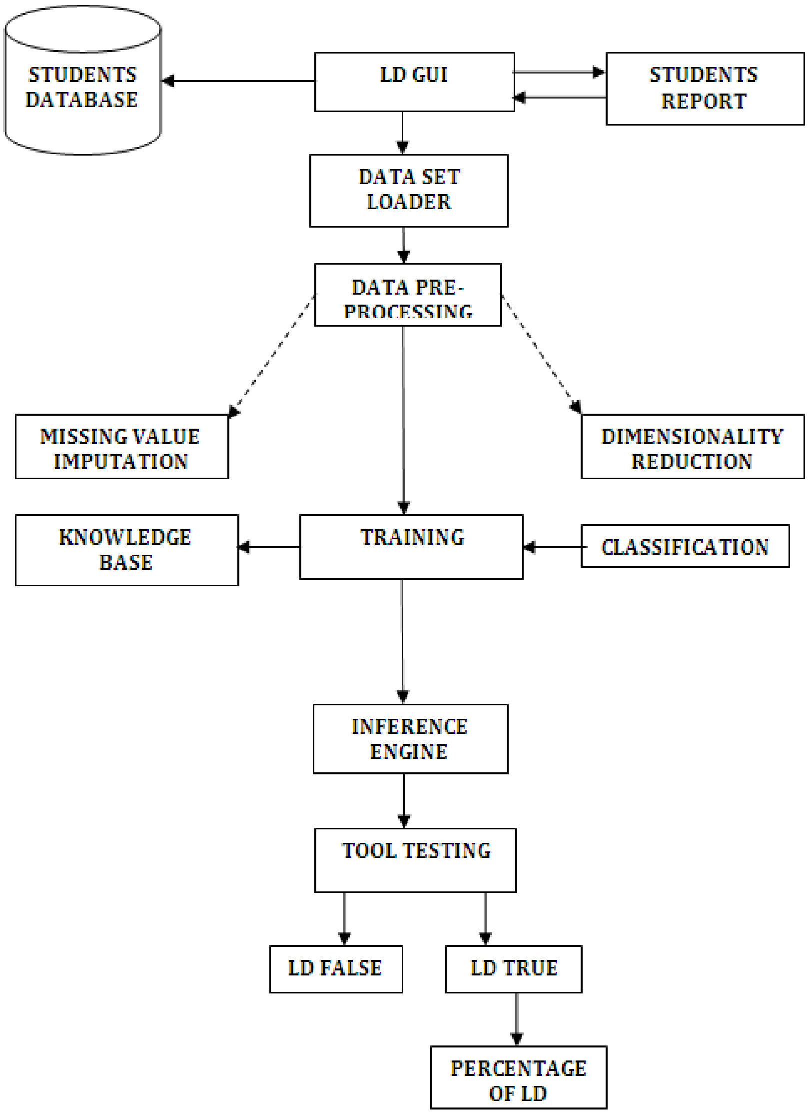


**Figure 5.3** Architecture of the tool

## 5.4 Design of the Tool

Design and structure of the tool are given under. The main components of the tool are students data base, LD GUI, students report, data set loader, data preprocessing, classification, training, knowledge base and inference engine.

### 5.4.1 Students data base

In the tool design architecture, two types of data sets are used. The first type is the data set obtained by means of informal assessment using checklists containing the 16 attributes which are the symptoms of LD.

The second type of data set is prepared for the fuzzy concepts. Here, instead of one attribute, four questions, sub attributes, related to that attribute are used. Each questions having a particular ranking. The rank is between 0 and 100.By summing up the score (rank) of these four sub attributes, the total rank of each attribute is calculated. By using this method very accurate results can be collected and hence this type of data set is very effective and hence this leads to a good classification results. Each case values are calculated automatically by a special microprogramming excel sheet.

### 5.4.2 LD GUI

A GUI is designed in MatLab 7.10 software environment, for the proposed LD tool, is as shown in Figure 5.4. The purpose of this developed tool is to help the child, parents and teachers to assess the child's LD status effectively. Here knowledge of the machine to perform the classification and prediction is used. The rules and facts built in the knowledge base can be viewed and modified, expanded or updated through this component. This interface plays the role of translator where the information displayed to the user is in English-like format. It acts as the user interface. It shows 100% accuracy.

The GUI collects the details about the child. 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. The GUI is very user friendly and the time consuming for the testing is much less. Any type of informal assessment checklist can be loaded in the GUI for testing. After the testing, apart from the degree of LD its percentage can also be assessed by the user from this GUI.

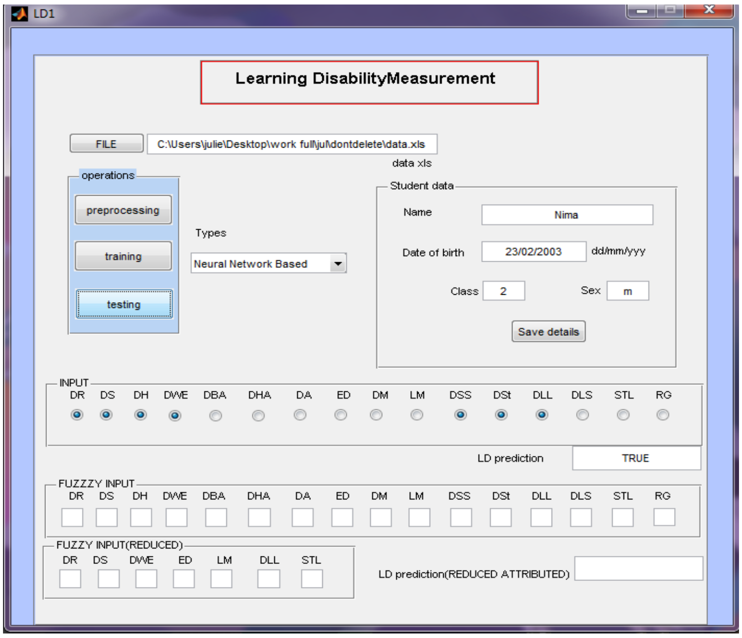


**Figure 5.4** Designed GUI of LD tool

### 5.4.3 Students report

The system is designed such that, the different 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. After the all processing, we can take the detailed d report of the student. It contains details about the child and details about the LD.

### 5.4.4 Data set loader

The purpose of data loader given in the GUI is for loading the data for training. After training, the knowledge is stored in the system.

### 5.4.5 Data preprocessing

Learning disability databases are highly susceptible to noisy, missing and inconsistent data due to their childish nature. They may not respond to the question or the teacher couldn't identify their problems. Sometimes the assessment may not correct. In addition to these problems the data sets should have same characteristics of usual database. The purpose of this data preprocessing is to improve the quality of the data and, consequently of the mining results. The collected data has to be preprocessed so as to improve the efficiently and ease of the mining process.

Here, the different methods are implementing for improving the quality of the data sets are closest fit algorithm and correlation based algorithm. Also, attribute reduction is done using PCA. These methods improve the quality of the datasets and it will help to classify the LD children effectively.

### 5.4.6 Classification

In this system, we are using different classification models such as decision tree, neural network, fuzzy and neuro fuzzy systems. The classifier results fully depend on the quality of the data. So we provides different methods for improving the quality of data and hence these classifiers shows a better results.

### 5.4.7 Training

Different training methods are performed here viz. training for neural network, decision tree and fuzzy. After performing the training, testing will be performed. Based on our data set, in this tool, 50% data is used for testing and the knowledge gained from here is used for testing.

### 5.4.8 Knowledge base

The acquired knowledge of the components is organized in the form of facts and rules in this knowledge base component [149]. The knowledge base provides a means for information collected on LD and how it is organized, shared, searched and utilized. The machine knowledge bases stores knowledge in a computer readable form. They contain a set of data, often in the form of rules, decision tree, neural net, etc. that describe the knowledge in a logically consistent manner. Knowledge base of our tool contains 243 rules from ANFIS, 153 rules from fuzzy model, 26 rules from fuzzy with attribute reduction and 9 rules from decision tree.

### 5.4.9 Inference Engine

The inference engine is the mechanism by which the search for conclusions or reasoning is conducted using a search strategy of the knowledge built in the knowledge base [1]. An inference engine is a program that derives answers from the knowledge base. It is also called testing. It is the brain of the system that uses to reason about the information in the knowledge base for the ultimate purpose of formulating new decisions.

## 5.5 Tool Testing

Testing is carried out on the whole data. For the performance evaluation of the tool, in presence of the professionals, details of new candidates are entered and we obtained 100% accurate LD results. The developed tool, viz. Knowledge Based Learning Disability Prediction (KBLDP) tool is tested for its accuracy and found highly performed for the cases tested. Different types of test results are obtained. These are LD – True and LD – False. Apart from this, the class of LD such as low, minor and major along with the percentage of LD in each class are also obtained in an accurate way while testing.

## 5.6 Screen Shots

A typical result obtained in the GUI of the tool designed using the neuro fuzzy method is shown in Figure 5.5. 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.

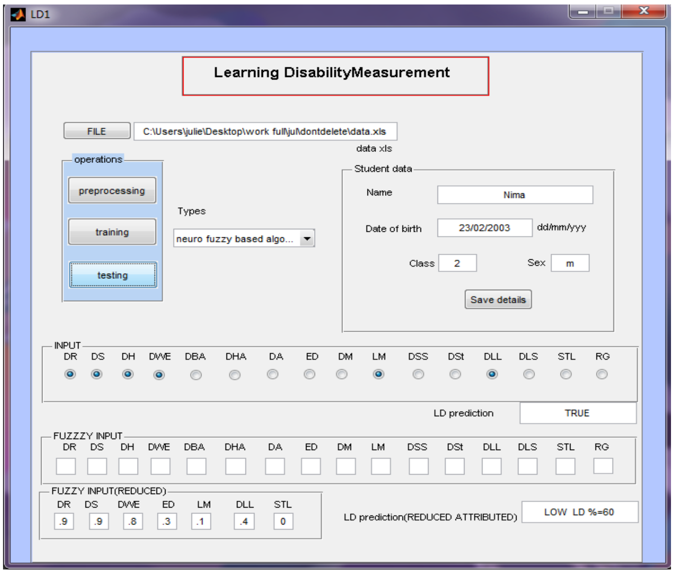


**Figure 5.5** A typical case of GUI output with class low and LD 60%

Another typical result for a child having LD class minor with LD 52% is shown in Figure 5.6. This minor LD of 52% is somehow higher than that of low LD of 60%. From the percentage of LD and its class, we can understand the depth of learning disability faced by the children. Hence, based on the class and percentage of LD, we can provide the appropriate remedial solution to the child.

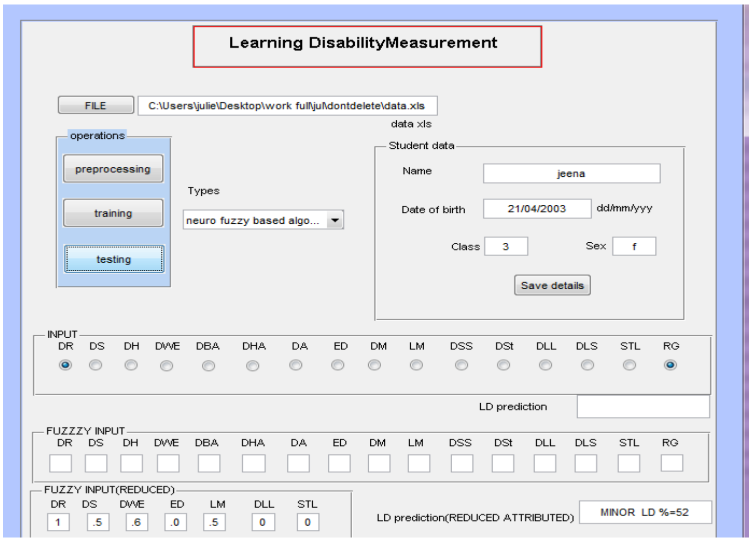


**Figure 5.6** A typical case of GUI output with class minor and LD 52%

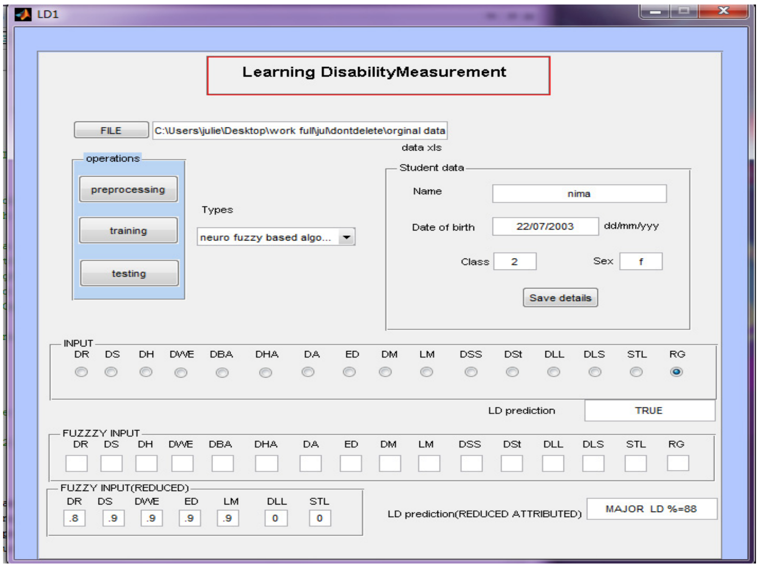


**Figure 5.7** A typical case of GUI output with class major and LD 88%

The third typical case we have observed is that the class of LD major with LD 88% as shown in Figure 5.7. 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 Figure 5.8.

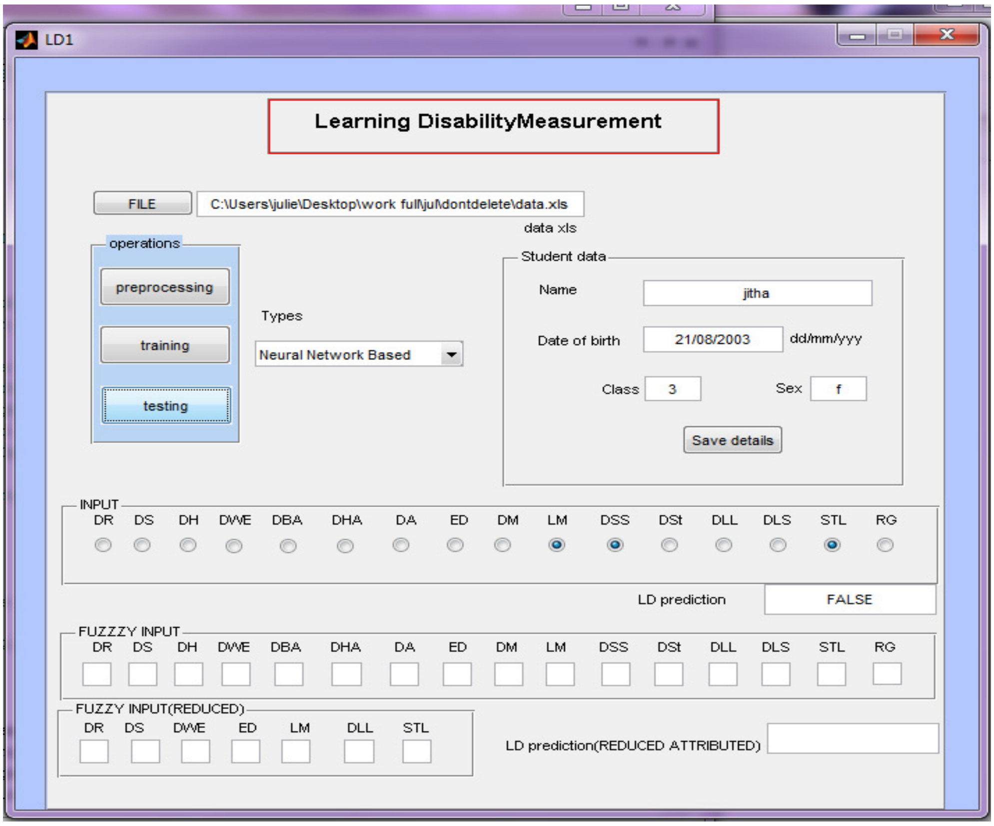


**Figure 5.8** A typical case of GUI output with no LD

As such, the designed tool is a very good scaling for the learning disability measurement. In the study, we are using three input method - one for neural network and decision tree, second for fuzzy input and third for fuzzy with reduced attribute and neuro fuzzy methods. 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.

## 5.7 Performance Evaluation of the Tool

The designed tool for predicting the learning disability in school age children and determining the class of LD viz. low, minor or major and its percentage measurement has been tested and evaluated in presence of professionals in the field. The tool shows 100% accuracy and the experts/professionals have agreed that the developed tool is very effective in the case of informal assessment and can be used for easy and speedy identification of LD. They opined that the tool can be used without their presence. As in the Indian scenario, parents/children may not be willing to undergo LD assessment and evaluation carries out in LD clinics. In such situations, the tool is found very effective, as it is user friendly and it can be used by teachers at the school level itself or even by the parents.

## 5.8 Summary and Conclusions

An integrated knowledge based tool designed with the help of a GUI for prediction of LD in school children is explained. The developed tool helps in classifying LD after performing different data pre-processing methods also. The developed tool performs in good manner. In future, the tool has to be enhanced to incorporate other assessment methods of LD, including formal assessment. Also, in future, other classifying methods including hybrid computing approaches and advanced features of fuzzy models have also to be studied for these assessment methods.

## 5.9 Contributions

The developed tool is helpful in finding the LD at an early stage. Depending upon the degree of LD present, as determined by the school authorities/parents, they can recommend the child for further treatment with councilors/special educators/LD clinics, etc., for proper remedial solutions. With the right help and intervention at proper time, children with LD can succeed in school and go on to be successful later in life, where the designed tool is found much relevant as early detection of spotting out of signals of learning disability can be easier to correct.

*Chapter*
# 6
# CONCLUSION

Contents



## 6.1 Summary

Specific learning disabilities (LD) have been recognized in some countries for much of the 20$^{th}$ century, in other countries only in the latter half of the century, and yet not at all in other places. 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 a learning disability. In India, about 10% of children enrolled in schools having LD. Learning disabilities are formally defined in many ways in many countries. It is a neurological condition that affects a child's brain and impairs his ability to carry out one or many specific tasks. Learning disabilities affect children both academically and socially. These may be detected only after a child begins school and faces difficulties in acquiring basic academic skills. An affected child can have normal or above average intelligence. Our challenge will be to alter our parenting and teaching approaches so that their potential contribution to our collective lives is not forfeited. The concept is still new in many developing countries. Since no national census of the learning disabled has been taken in India, it is difficult to collect their actual number. In India, the learning disabled children are not identified using reliable tests and the research conducted in learning disability has been primarily done over the last two decades only.

In this thesis, various machine learning techniques are used to analyze the symptoms of LD, establish interrelationships between them and evaluate the relative importance of these symptoms. To increase the diagnostic accuracy of learning disability prediction, a knowledge based system based on statistical machine learning or data mining techniques according to the knowledge obtained from the clinical information is developed. The basic idea of the developed tool is to increase the accuracy of the learning disability assessment and reduce the time used for LD assessment. The tool has many advantages compared to the traditional methods to determine learning disabilities using check lists.

For improving the performance of various classifiers, we developed some pre-processing methods for the LD prediction system. No others have done this type of work and it is very relevant in medical diagnosis system. A new system based on fuzzy and rough set models are also developed for LD prediction. Here also the importance of pre-processing is studied. A Graphical User Interface (GUI) is designed for developing an integrated knowledge based tool for prediction of LD as well as its degree. The designed tool stores the details of the children in the student database and retrieves their LD report as and when required.

## 6.2 Contributions

This thesis makes several major contributions in technical, general and social areas as discussed below.

### 6.2.1 Technical contributions

i. New methods for LD prediction, based on machine learning techniques, are developed,

ii. New insights into the interrelationships between symptoms of LD, their relative importance and estimating the significance of each symptoms of LD

iii. Identification of the problems related to the classification accuracy of different classifiers,

iv. New algorithm based on correlation is developed for imputing missing values,

v. New models of LD prediction using fuzzy and rough sets,

vi. Modification of data preprocessing with J48 decision tree and neural network for LD prediction; and

vii. Developing of an integrated knowledge based tool for LD prediction,

### 6.2.3 General contributions

i. The research works done in the area of prediction of learning disabilities using knowledge based methods is very little compared to the magnitude of LD affected children,

ii. Based on the machine learning tool developed, the presence of learning disability in any child with its percentage can be determined,

iii. The class of LD like low, minor and major and the percentage of LD in each class can also be determined by this tool,

iv. 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,

v. The tool developed gives more accurate results in lesser time compared to the traditional assessment methods using check lists,

vi. The developed tool is very effective for finding the LD affected children from the large database; and

vii. This research work has also considered an approach to handle learning disability database to predict frequent symptoms of the learning disabilities in school age children.

### 6.2.4 Social contributions

i. The study will certainly contribute in the development of the nation as LD is a real stumbling block for a nation's development process,

ii. The contribution of the study in early diagnosis of LD in children is critically important to identify and suggest remedial solutions to children/parents /teachers, which will ultimately help them to provide the child with best environment for them and they can learn successfully and become winners,

iii. The contribution of the study ultimately improves the confidence of children and helps in getting the social support to them; and

iv. As the developed tool is very user friendly, it can be used for LD identification by the parent/teacher/friends of the children.

## 6.3 Future Works

My future work focuses on Hybrid Computing approach and advanced features of fuzzy models for enhancing the tool to incorporate other assessment methods of LD, including formal assessments. If we enhanced the tool by incorporating these, all kinds of assessments can be done. The prediction will then be more accurate and the tool can be used by the teachers at school level itself or even by the parents or friends of the affected children. This work can be enhanced to other areas of medical diagnosis also. As this work is particularly for general assessment of LD, in future more work can be done to categorize the LD like dyslexia, dyscalculia, etc.

The developed tool is found very user friendly and it can be used by teachers at the school level itself or even by the parents. The results are found very beneficial to the parents, teachers and the institutions. 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.

# REFERENCES


[1] Editorial on Current Challenges in Managing Specific Learning Disability in Indian Children, Journal of Postgraduate Medicine, 54(2), April-June 2008, pp75-77

[2] http://articles.timesofindia.indiatimes.com/2012-01-27/chennai/30669918_1_inclusive -education-disability-autism

[3] Sally Jo Cunningham, Geoffrey Holmes: Developing Innovative Applications in Agricultural using Data Mining, Proceedings of the Southeast Asia Regional Computer Confederation Conference, 1999

[4] Frawley, Piaatetsky: *Shaping Knowledge Discovery in Database; An Overview*, The AAAI/MIT press, Menlo Park, 1996

[5] Witten Ian, H., Frank Eibe: *Data Mining–Practical Machine Learning Tools and Techniques,* Morgan Kaufmann - Elsevier Publishers, ISBN: 13: 978-81-312-0050-6, 2005

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

[7] Chapple Mike: About.com Guide, http://databases.about.Com/od/datamining/g / classification.htm

[8] Han Jiawei, Kamber Micheline, Pei Jian: *Data Mining-Concepts and Techniques,* Third Edition, Morgan Kaufmann - Elsevier Publishers, ISBN: 978-93-80931-91-3, 2011

[9] Rawal Swaroop: *Learning Disability in a Nutshell*, B. Jain Publishers Private Limited, New Delhi, 2010

[10] Nakra Onita: *Children and Learning Difficulties*, Allied Publishers Private Limited, New Delhi, 2007

[11] Santhanam, T., Prasad Babu, B., Sugandhi, S., Rao, D.B.: *Children with Learning Disabilities*, Sonali Publications, New Delhi, 2011

[12] Blackwell Synergy: Learning Disabilities & Research Practice, 22, 2007

[13] 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

[14] Elizabeth Broughton Sorby: A paper on Addressing Questions about Learning Difficulties, Jan. 2012, pp1-7

[15] Crealock Carol, Kronick Doree: Children and Young People with Specific Learning Disabilities, Guides for Special Education, No. 9, UNESCO, 1993.

[16] Smith Corinne, Strick Lisa: *Learning Disabilities: A to Z*, Free Press, A Division of Simon & Schuster, Inc, New York, 2010

[17] Learning Disabilities, National Dissemination Center for Children with Disabilities (NICHCY) Fact Sheet No. 7 (FS7), January 2011, Available at: http://nichcy.org/wp-content/uploads/docs/fs7.pdf

[18] Reddy, G.L., Ramar, R., Kusuma, A., *Learning Disabilities, A Practical Guide to Practitioners*, Discovery Publishing House, New Delhi, 2011

[19] Helping Overcome Learning Disabilities, Tata Interactive Systems Learning Disability Forum, 2010, *Available at : http:// www.tatalearningforum.com/tldf/Remediation.htm*

[20] Contributor eHow, Harwell Tabitha, Resources for Children with Learning Disabilities, *Available at : http:// www.ehow.com/ info_ 8028161_resources-children-learning-disabilities.html*

[21] Kenneth A. Kavale: Identifying Specific Learning Disability - Is

Responsiveness to Intervention the Answer?, Journal of Learning Disabilities, 38, 2005, pp553-562.

[22] Benjamin J. Lovett: Extended Time Testing Accommodations for Students with Disabilities: Answers to Five Fundamental Questions, Review of Educational Research, Journal of Learning Disabilities, 80, 2010, pp611-638

[23] 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.

[24] Tung-Kuang Wu, Shian Chang Huang,Ying Ru: Evaluation of ANN and SVM Classifiers as Predictors to the Diagnosis of Students with Learning Disabilities, Journal of Expert Systems with Applications, 34 (3), 2008, pp1846-1856

[25] Maitrei Kohli, Prasad, T.V.: Identifying Dyslexic Students by Using Artificial Neural Networks, Proceedings of the World Congress on Engineering, I, 2010,

[26] Yanbin Yer, Chia-Chu Chiang, Acxiom Corp, Little Rock: Parallel Apriori Algorithm for Frequent Item sets Mining, Proceedings of Fourth International Conference on Software Engineering Research, Management and Applications, 2006, pp87-94

[27] Selma and Altay, H.: An Algorithm for Mining Association Rules using Perfect Hashing and Database Pruning, Proceedings of the Tenth Turkish Symposium on Artificial Intelligence and Neural Networks, Acan, A., Aybay, I., Salamah, M. (Eds.) 2001, pp 257-264

[28] Jeba, J.R., Victor, S.P: A Novel Approach for finding Frequent Item Sets with Hybrid Strategies, International Journal of Computer Applications, 17(5), 2011, pp30-34

[29] Celler, B.G., Chazal, P.: Low Computational Cost Classifiers for ECG Diagnosis using Neural Networks, Proceedings of the International Conference of Engineering in Medicine & Biology Society, 1998, pp1337–1340.

[30] Niti Guru, Anil Dahiya, Navin Rajpal: Decision Support System for Heart Disease Diagnosis using Neural Networks, Delhi Business Review, 8(1), 2007

[31] Xianjun Ni: Research of Data Mining Based on Neural Networks, World Academy of Science, Engineering and Technology, 39, 2008

[32] Hongcui wang, Tatsuya Kawahara: Effective Error Prediction Using Decision Tree for Automatic Speech Recognition Grammar Network in Call Systems, IEEE, ICASSP 2008, pp5069-5072

[33] Anupama Kumar, Vijayalakshmi, M: Efficiency of decision trees in predicting student's academic performance, Wyld, D.C., et al. (Eds), CCSEA 2011, DOI: 10.5121/csit.2011.1230, 2011, pp 335–343

[34] Huilei Xu, Ihor R Lemischkaand Avi Ma'ayan: SVM classifier to predict genes important for self-renewal and pluripotency of mouse embryonic stem cells, BMC Systems Biology 2010, 4: 173 doi: 10. 1186 / 1752 – 0509 – 4 - 173, 2010

[35] Betkier, Piotr, et.al: Prediction of Protein Phosphorylation Sites using Classification Trees and SVM classifier, Romaniuk, Ryszard, S.(eds.): Proceedings of the SPIE, 2011, pp80081N-80081N-8

[36] **A**ntonia Kyriakopoulou, Theodore Kalamboukis: Text Classification Using Clustering, CLEF Working Notes, 2010

[37] Antonia Kyriakopoulou, Theodore Kalamboukis: Combining Clustering with Classification for Spam Detection in Social Bookmarking Systems, ECML/PKDD, Discovery Challenge, 2008

[38] Kristina Machova, Frantisek Barcak, Peter Bednar: A Bagging Method using Decision Trees in the Role of Base Classifiers, Acta Polytechnica Hungarica, 3(2), 2006, pp121-132

[39] Zhu, X., Wu, X.: Cost-Constrained Data Acquisition for Intelligent Data Preparation. IEEE Transactions on Knowledge and Data Engineering, 17(11), 2005, pp1542–1556

[40] Jiye Li, Nick Cercone: Assigning Missing Attribute Values Based on Rough Sets Theory, IEEE International Conference on Granular Computing, May 2006, pp607-610

[41] Shichao Zhang, Jilian Zhang, Xiaofeng Zhu, Yongsong Qin, Chengqi Zhang: Missing Value Imputation Based on Data Clustering, Transactions on Computational Science, Springer-Verlag Berlin, Heidelberg, I, 2008

[42] 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.

[43] 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.

[44] Quinlan, J.R.: *C4.5 - Programs for Machine Learning.* Morgan Kaufmann, San Mateo, USA, 1993.

[45] Friedman, J., et al.: Lazy Decision Trees. Proceedings of the 13th National Conference on Artificial Intelligence, 1996, pp717-724.

[46] Lakshminarayan, K., et al.: Imputation of Missing Data Using Machine Learning Techniques. *KDD-1996, pp*140-145.

[47] 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.

[48] Kahl, F., et al.: Minimal Projective Reconstruction Including Missing Data. IEEE Trans. Pattern Anal. Mach. Intell., 23(4), 2001, pp418-424.

[49] Gessert, G.: Handling Missing Data by Using Stored Truth Values. SIGMOD Record, 20(3), 2001, pp30-42.

[50] 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.

[51] Ramoni, M., Sebastiani, P.: Robust Learning with Missing Data. Machine Learning, 45(2), 2001, pp147-170.

[52] Pawlak, M., Kernel: Classification Rules from Missing Data. IEEE Transactions on Information Theory, 39(3), 1993, pp979-988

[53] Yao, Y.Y.: A Comparative Study of Fuzzy Sets and Rough Sets, Information Sciences, 109 (1-4), 1998, pp. 227-242

[54] Pawlak, Z.: *Rough Sets*: *Theoretical Aspects of Reasoning about Data*, Kluwer Academic Publishers, Boston, 1991

[55] Klir, G.J., Yuan, B.: *Fuzzy Sets and Fuzzy Logic, Theory and Applications*, Prentice Hall, New Jersey, 1995

[56] Chanas, S., Kuchta, D.: Further remarks on the relation between rough and fuzzy sets, Fuzzy Sets and Systems, 47, 1992, pp391-394.

[57] Zadeh, L.A.: Forward, in: Orlowska, E. (Ed.): Incomplete Information: Rough Set Analysis, Physica-Verlag, Heidelberg, 1998, pp. v-vi.

[58] Ossi Nykanen: Inducing Fuzzy Model for Students Classification, Educational Technology & Society, 9 (2), 2006, pp223-234.

[59] 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

[60] Bordogna, G, Pasi, G.: Handling vagueness in information retrieval systems, Proceedings of the Second New Zealand International Two-Stream Conference on Artificial Neural Networks and Expert Systems, 1995, pp110-114

[61] Miyamoto, S.:*Fuzzy sets in Information Retrieval and Cluster Analysis,* Kluver Academic Publishers, 1990

[62] 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.

[63] Baeza Yates, R., Ribeiro Neto, B.: *Modern Information Retrieval*, ACM Press, 1999.

[64] Ramasubramanian, P., Iyakutti, K., Thangavelu, P., Joy Winston, J.: Teaching Result Analysis using Rough Sets and Data Mining, Journal of Computing, 1(1), 2009, pp168-174

[65] Roman W. Swiniarski: Rough Sets Methods in Feature Reduction and Classification, International Journal of Appl. Math. Comput. Sci., 11(3), 2001, pp565-582

[66] Fayyad Usama, Piatetsky Shapiro Gregory, Smyth Padhraic: From Data Mining to Knowledge Discovery in Databases, American Association for Artificial Intelligence, 1996, pp37-54.

[67] Pang-Ning, T., Michael, S., Vipin, K: *Introduction to Data Mining*, Low price edn. Pearson Education, Inc., London, 2008

[68] Del Amo, A., Montero, J., Fernandez, A., Lopez, M., Tordesillas, J.M., Biging, G.: Spectral Fuzzy Classification: An Application, IEEE Transactions on Systems, Man and Cybernetics, Part-C, Application and Reviews, 32(1), 2002, pp42-48

[69] Emmanouilidis, C., MacIntyre, J., Cox, C.: Neurofuzzy Computing Aided Machine Fault Diagnosis, Proceedingds of JCIS'98, The Fourth Joint Conference on Information Sciences. Research Triangle Park, North Carolina, USA, 1998, 1, pp. 207-210.

[70] Witten Ian H., Eibe Frank, Len Trigg, Mark Hall, Geoffrey Holmes, Sally Jo Cunningham: Weka - Practical Machine Learning Tools and Techniques with Java Implementations, 1999

[71] Sivanandam, S.N., Sumathi, S., Deepa, S.N.: *Introduction to Neural Networks using MATLAB 6.0,* Tata McGraw Hill Education Pvt. Ltd., New Delhi, 2011.

[72] Tapati Bandopadhyay, Pradeep Kumar, Anil Kumar Saini: Knowledge Pre-Processing: A Generic Approach based on compiler Function concepts, IRMA International Conference, 2007, pp1510-1513

[73] Shalabi Al Luai, Shaaban Zyad, Kasasbeh Basel: Data Mining - A Preprocessing Engine, Journal of Computer Science, 2(9), 2006, pp735-739

[74] Frawley, Piaatetsky: *Shaping Knowledge Discovery in Database; An Overview,* The AAAI/MIT press, Menlo Park, 1996

[75] Eduardo Ogasawara, Leonardo C. Martinez, Daniel de Oliveira, Geraldo Zimbrao, Gisele L. Pappa, Marta Mattoso: Adaptive Normalization: A Novel Data Normalization Approach for Non-Stationary Time Series, International Joint Conference on Neural Networks, DOI: 10.1109/IJCNN.2012.5596746, IEEE, 2010, pp1-8

[76] Peng, Y., Kou, G., Shi, Y., Chen, Z.: A Descriptive Framework for the Field of Data Mining and Knowledge Discovery, International Journal of Information Technology and Decision Making, 7 (4), 2008, pp639–682.

[77] Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P.: From Data Dining to Knowledge Discovery: An Overview, Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P., Uthurusamy, R. (Eds.): Advances in Knowledge Discovery and Data Mining, AAAI Press, 1996, pp1–34.

[78] Decision Trees & Data Mining, Tutorial (4): ID3 Decision Trees tutorial > ID3 & Entropy, Available at : http://decisiontrees.net/?q=node/27

[79] Guoqiang Peter Zhang: Neural Networks for Classification: A Survey, IEEE Transactions on Systems, Man and Cybernetics – Part C: Applications and Reviews, 30(4), Nov. 2000, pp451-462

[80] Julie M. David, Kannan Balakrishnan: Prediction of Learning Disabilities in School Age Children using Decision Tree, Proceedings of the International Conference on Recent Trends in Network Communications, Lecture Notes in Communication in Computer and Information Science (CCIS), 90(3), Springer- Verlag Berlin Heidelberg, 2010, pp 533-542.

[81] Vapnik,V.: Estimation of Dependences Based on Empirical Data, Springer-Verlag, 1982

[82] Haykin, S.: *Neural Networks-A Comprehensive Foundation,* Prentice Hall, 1999

[83] Radhika, Y., Shashi, M.: Atmospheric Temperature Prediction using Support Vector Machines, International Journal of Computer Theory and Engineering, 1(1), 1793-8201, April 2009, pp55-58.

[84] Vapnik,V.,: The Nature of Statistical Learning theory, Springer-Verlag, 1995

[85] John C. Platt: Sequential Minimal Optimisation – A Fast Algorithm for Training Support Vector Machines, Technical Report MSR-TR-98-14, April 21, 1998

[86] Soman, K.P., Loganathan, R., Ajay, V.: *Machine Learning with SVM and other Kernel Methods*, PHI Learning Pvt. Ltd, New Delhi, ISBN-978-81-203-3435-9, 2009

[87] Boser, B.E., Guyon, I.M., Vapnik V.: A Training Algorithm for optimal Margin Classifiers, Fifth Annual Workshop on Computational Learning Theory, ACM, 1992

[88] Kecman, V., Huang, T.M., Vogt, M.: *Iterative Single Data Algorithm for Training Kernel Machines from Huge Data Sets: Theory and Performance*, StudFuzz, Springer-Verlag Berlin Heidelberg, 177, 2005, pp255–274

[89] Julie M. David, Kannan Balakrishnan: Prediction of Learning Disabilities in School-Age Children using SVM and Decision Tree, International Journal of Computer Science and Information Technology, 2(2), Mar-Apr. 2011, pp 829-835.

[90] Breiman, L.: Bagging Predictors, Technical Report 421, Department of Statistics, University of California at Berkeley, 1994

[91] Kristina Machova, Frantisek Barcak, Peter Bednar: A Bagging Method using Decision Trees in the Role of Base Classifiers, Acta Polytechnica Hungarica, 3(2), 2006, pp121-132

[92] Breiman, L.: Arcing the Edge - Technical Report 486, Department of Statistics, University of California at Berkeley, 1996

[93] Schapire, R., Freund, Y.: Boosting the Margin: A New Explanation for the Effectiveness of Voting Methods. The Annals of Statistics, 26(5), 1998, pp1651-1686

[94] Fatma Khcherem, Abdelfettah Bouri: Fuzzy Logic and Investment Strategy Global Economy & Finanace Journal, 2(2), 2009, pp 22-37.

[95] Zadeh, L.A.,: Fuzzy sets, Information and control, 8, 1965, pp 338-353

[96] 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

[97] 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

[98] Zadeh, L.A.: Fuzzy sets. Infor. Control, 8, 1965, pp338-353

[99] Ross, T.J.: *Fuzzy Logic with Engineering Applications*, Second Edition, John Wiley & Sons Ltd., 2004

[100] Fuzzy Logic Toolbox User's Guide , The MathWorks Inc., 2004

[101] Zadeh, L.A: The Concept of Linguistic Variable and its Application to Approximate Reasoning, Inf. Sciences, 8, 1975, pp199-249

[102] Zadeh, L.A: The Concept of Linguistic Variable and its Application to Approximate Reasoning, Inf. Sciences, 9, 1975, pp43-80

[103] Milindkumar V. Sarode, Ladhake, S.A., Prashant R. Deshmukh: Fuzzy system for color image enhancement, World Academy of Science, Engineering and Technology, 48, 2008, pp311-316

[104] Chi-Cheng Jou: On the Mapping Capabilities of Fuzzy Inference Systems, International Joint Conference on Neural Networks, 2, 1992, pp703-713

[105] Mendel, J.:Fuzzy Logic Systems for Engineering: A Tutorial, Proceedings of the IEEE, 83(3), 1995, pp345-377

[106] Driankov, D., Hellendoorn, H., Reinfrank, M.: *An Introduction to Fuzzy Control*, Second Edn., Springer-Verlag Berlin Heidelberg, 1996

[107] Nageswara Rao, G., Chandra Sekhar, K., Sangameswara Raju, P.: Design and Implementation of Hybrid Active Power Filter, International Journal of Computer Applications,8(10), 2010, pp10-15

[108] Gervin Thomas, Ben Juurlink, Dietmar Tutsch: Traffic Prediction for NoCs using Fuzzy Logic, Proceedings of second International Workshop on New Frontiers in High Performance and Hardware Aware Computing, San Antonio, Texas, USA, 2011

[109] Grzymala-Busse, J.W., Hu, M.: A Comparison of Serveral Approaches to Missing Attribute Values in Data Mining. Ziarko, W., Yao. Y., (Eds.): RSCTC 2000, LNAI 2005 (2001) pp378–385

[110] Pawlak, Z.: Rough set and data analysis, Proceedings of the Asian11-14 Dec.1996, pp1-6.

[111] Zhang, S.C., et al.: Information Enhancement for Data Mining. IEEE Intelligent Systems, 19(2), 2004, pp12-13.

[112] 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.

[113] Qin, Y.S., et al.: Semi-Parametric Optimization for Missing Data Imputation. Applied Intelligence, 27(1), 2007, pp79-88

[114] Zhang, C.Q., et al.: An Imputation Method for Missing Values, PAKDD, LNAI, 4426, 2007, pp1080–1087.

[115] Zhang, S.C., et al.: Optimized Parameters for Missing Data Imputation, *PRICAI06*, 2006, pp1010-1016.

[116] Wang, Q., Rao, J.,: Empirical Likelihood-based Inference in Linear Models with Missing Data, Scand. Journal of Statist., 29, 2002, 563-576.

[117] Wang, Q., Rao, J. N. K.: Empirical likelihood-based inference under imputation for missing response data, Ann. Statist., 30, 2002, pp896-924.

[118] Chen, J., Shao, J., Jackknife: Variance Estimation For Nearest-Neighbor Imputation, Journal of Amer. Statist. Assoc., 96, 2001, pp260-269.

[119] Lall, U., Sharma, A.: A Nearest-Neighbor Bootstrap For Resampling Hydrologic Time Series, Water Resource, Res. 2001, 32, 1996, pp679–693.

[120] Grzymala-Busse, J.W., Grzymala-Busse, W.J., Goodwin, L.K.: Coping with Missing Attribute Values Based on Closest Fit in Pre-term Birth Date : A Rough Set Approach, International Journal on Computational Intelligence, 17(3), 2001, pp425-434.

[121] Xiangun Ni: Research of Data Mining based on Neural Network, World Academy of Science, Engineering & Technology, 39, 2008

[122] Shantakumar B. Patil, Kumaraswamy, Y.S.: Intelligent and Effective Heart Attack Prediction System using Data Mining and ANN, Europian Journal of Scientific Research, 31(4), ISSN 1450-216X, 2009, pp642-656.

[123] Savcovic-Stevanovic: Neural networks for process analysis and optimization: Modeling and Applications, Computer and Chemical Engineering, 18(11-12), 1994, pp1149-1155.

[124] Tarafdar Haque, M., Najafi, S.: Application of Neuro Fuzzy Dynamic Programming to Improve the Reactive Power and Voltage Profile of a Distribution Substation, World Academy of Science Engineering and Technology, 6, 2005, pp126-129

[125] Gray, R., George, P.E., Frank Cardullo: Application of Neuro Fuzzy System to Behavioral Representation in Computer Generated Forces, 1999. Available at http:// citeseer.ist. psu.edu/ George 99 application. html

[126] Weixiang Zhao, Abhinav Bhushan, et. al.: Machine Learning: A Crusial Tool for Sensor Design, Open Access Algorithms, ISSN 1999-4893, DOI:10.3390/a1020130, 2008.

[127] Julie M. David, Kannan Balakrishnan; Attribute Reduction and Missing Value Imputing with ANN: Prediction of Learning Disabilities, International Journal of Neural Computing, Springer-Verlag London Limited, DOI: 10.1007/s00521-011-0619, May 2011.

[128] 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. (Accepted for publication)

[129] Rahib Abiyeb, Vasif, H., et. al.:Elecricity Consumtion Prediction Model using Neuro-Fuzzy System, World Academy of Science, Engineering and Technology, 8, 2005.

[130] 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.

[131] 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

[132] 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

[133] 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

[134] 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

[135] 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

[136] Grzymala-Busse, J.W.: Rough Set Theory with Applications to Data Mining, 2005

[137] Pawlak, Z.: Rough Sets, International Journal Computers and Information Sci., 11, 1982, pp341-356

[138] Grzymala-Busse, J. W.: A New Version of the Rule Induction System LERS, Fundamental Informaticae, 31, 1997, pp27-39

[139] Jiye Li, Nick Cercone:A Rough Set Based Model to Rank the Importance of Association Rules, Proceedings of the 10th international conference on Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing , Springer-Verlag Berlin, Heidelberg, II, 2005

[140] Hameed Al-Qaheri, Aboul Ella Hassanien, Ajith Abraham: Discovering Stock Price Prediction Rules using Rough Sets, 2008, pp1-19

[141] Grzymala-Busse, J.W.: Knowledge Acquisition under Uncertainty-A Rough Set Approach. Journal of Intelligent & Robotic Systems, 1, 1988, pp3-16

[142] Kargupta, Joshi, Sivakumar, Yesha: *Data Mining Next Generation Challenges and Future Directions*, PHI Learning Pvt. Ltd, 2011

[143] Julie M. David, Kannan Balakrishnan: Prediction of Frequent Signs of Learning Disabilities in School Age Children using Association

Rules, Proceedings of the International Conference on Advanced Computing, ICAC 2009, MacMillion Publishers Ltd., New York City, 2009, pp 202–207.

[144] Kasthurirangan Gopalakrishnan, Siddhartha Khaitan, Anshu Manik: Enhanced Clustering Analysis and Visualization using Kohonen's Self-Organizing Feature Map Networks, World Academy of Science, Engineering and Technology, 42, 2008, pp798-805

[145] Forgy, E.: Cluster analysis of multivariate data: Efficiency vs. Interpretability of Classifications, Biometrics, 21, 1965, pp768

[146] Hamerly, H., Elkan, C.: Learning the k in k-means. Proceedings of the 17th International Conference of Neural Information Processing System, 2003.

[147] Julie M. David, Kannan Balakrishnan: Significance of Classification Techniques in Prediction of Learning Disabilities in School Age Children, International Journal of Artificial Intelligence & Applications, 1(4), Oct.2010, pp 111-120

[148] Julie M. David, Kannan Balakrishnan: Prediction of Key Symptoms of Learning Disabilities in School-Age Children using Rough Sets, International Journal of Computer and Electrical Engineering, 3(1), Feb. 2011, pp 163-169

[149] Ibrahim S. Moghram, Faisel A. Ibrahim: An Expert System Aided Tool for Teaching the Design of PID Based Controllers, International Journal of Electrical Engineering Education, 37(4), 2000, pp357-367

# LIST OF PUBLICATIONS

Parts of this thesis work have been previously published as peer reviewed journal papers and conference papers, as listed below;

1. 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. (Accepted for publication, 8(3), June 2013).
2. Julie M. David, Kannan Balakrishnan: Attribute Reduction and Missing Value Imputing with ANN: Prediction of Learning Disabilities, International Journal of Neural Computing and Applications, Springer-Verlag London Limited, DOI: 10.1007/s00521-011-0619, 21 (7), Oct. 2012, Online First May 2011.

   Available at - http://www.springerlink.com/content/evj556771k561681/
3. Julie M. David, Kannan Balakrishnan: Prediction of Learning Disabilities in School-Age Children using SVM and Decision Tree, International Journal of Computer Science and Information Technology, ISSN 0975-9646, 2(2), Mar-Apr. 2011, pp829-835.

   Available at - http://www.ijcsit.com/ijcsit-v2issue2.php
4. Julie M. David, Kannan Balakrishnan: Prediction of Key Symptoms of Learning Disabilities in School-Age Children using Rough Sets, International Journal of Computer and Electrical Engineering, 3(1), Feb. 2011, pp163-169

   Available at - http://www.ijcee.org/abstract/308-JE351.htm

5. Julie M. David, Kannan Balakrishnan: Machine Learning Approach for Prediction of Learning Disabilities in School Age Children, International Journal of Computer Applications, ISSN-0975-8887, 9(10), Nov. 2010, pp 7-14.

   Available at - http://www.ijcaonline.org/archives/volume9/ number 11 /1432-1931

6. Julie M. David, Kannan Balakrishnan: Significance of Classification Techniques in Prediction of Learning Disabilities in School Age Children, International Journal of Artificial Intelligence & Applications, 1(4), DOI:10.5121/ ijaia.2010.1409, Oct.2010, pp111-120.

   Available at - http://airccse.org/journal/ijaia/currentissue.html

7. Julie M. David, Kannan Balakrishnan: Performance of Classifiers - Significance of New Method in Missing Value Imputation and Dimensionality Reduction: Prediction of Learning Disabilities in School-Age Children, International Journal of Applied Soft Computing, Elsevier B.V. (communicated).

8. Julie M. David, Kannan Balakrishnan: Learning Disability Prediction Tool using ANN and ANFIS, International Journal of Transaction on Knowledge and Data Engineering, IEEE. (communicated).

9. Julie M. David, Kannan Balakrishnan: Fuzzy and Rough Set Approaches for Learning Disability Prediction, International Journal of Artificial Intelligence, Elsevier B.V. (communicated).

10. Julie M. David, Kannan Balakrishnan: Modified Pre-Processing Methods with Machine Learning Algorithms for LD prediction, International Journal of Human Computer Studies, Elsevier B.V. (communicated).

11. Julie M. David, Kannan Balakrishnan: Prediction of Learning Disabilities in School Age Children using Decision Tree, Proceedings of the International Conference on Recent Trends in Network Communications, Lecture Notes in Communication in Computer and Information Science (CCIS), 90(3), 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.

    Available at- http://www.springerlink.com/content/q66788rhk77w7w15//

12. Julie M. David, Kannan Balakrishnan: Prediction of Frequent Signs of Learning Disabilities in School Age Children using Association Rules, Proceedings of the International Conference on Advanced Computing, ICAC 2009, MacMillion Publishers Ltd., New York City, ISBN 10:0230-63915-1, ISBN 13:978-0230-63915-7, 2009, pp202–207.

13. Julie M. David, Pramod, K.V.: Prediction of Learning Disabilities in School Age children using Data Mining Techniques, Proceedings of AICTE Sponsored National Conference on Recent Developments and Applications of Probability Theory, Random Process and Random Variables in Computer Science, Thrivikram,T., Nagabhushan, P., Samuel, M.S., (eds), 2008, pp139-146.