A STUDY ON SOME PREVALENT DISEASES AMONGST NEONATES OF NORTH-EASTERN PARTS OF INDIA AND DEVELOPMENT OF KNOWLEDGE-BASED DECISION SUPPORT SYSTEM FOR TREATMENT PLANNING

A Thesis submitted to the

University of North Bengal

For the Award of

Doctor of Philosophy
in
Computer Science

BY

Dilip Roy Chowdhury

GUIDE

Prof. R. K. Samanta
Principal Supervisor
&
Prof. M. Chatterjee
Co-Supervisor

Department of Computer Science and Application
University of North Bengal

December 2012

Abstract

Title

A STUDY ON SOME PREVALENT DISEASES AMONGST NEONATES OF NORTH-EASTERN PARTS OF INDIA AND DEVELOPMENT OF KNOWLEDGE-BASED DECISION SUPPORT SYSTEM FOR TREATMENT PLANNING

of the

Thesis submitted to the

University of North Bengal

For the Award of

Doctor of Philosophy
in
Computer Science

By

Dilip Roy Chowdhury

Guide

Prof. R. K. Samanta
Principal Supervisor
&
Prof. M. Chatterjee
Co-Supervisor

Department of Computer Science and Application
University of North Bengal
2012

ABSTRACT

A famous quote “Children are the future of the nation” is often used by many people all over the world. This really good listen and appears to be right. But the actual picture is in dark side, that there are a plenty number of sick, filthy, hungry, beggarly kids all around us without having any hope of a secured future. Many children have been deprived of basic needs like food, proper shelter and sanitation, basic education, basic training and access to health care facilities. But every child has the right to protection, basic education, and proper healthcare, a healthy environment and good livelihood opportunities. It is important to look after the children’s health status carefully by the parents or guardians. Particularly new born babies should be given ultimate care to maintain of his/her the future growth.

Neonates, i.e. babies of just few hours/weeks of old after birth, not only constitute a large population group, but also are vulnerable or special risk group. The risk is related with growth, development, disease pattern and survival. Thus by improving the health status of neonates, we contribute to the health of the general population. These considerations have led to the formulation of special health service for children all over the world. The health care system in North Eastern parts of India is not sufficient enough to provide all the necessary medication properly. Mostly, in village areas, there has been acute problem of neonatal prevalent diseases. Some of the identified prevalent diseases and health related problem of neonates are Birth Asphyxia, Neonatal Sepsis, Hypoxic-Ischemic Encephalopathy (HIE), Preterm, Low birth weight, Delayed breastfeeding, Problems in breastfeeding, Diarrhea, Hemorrhage, Conjunctivitis, Skin Infection, Abnormal Jaundice, Meconium Aspiration, Hyaline Membrane Disease (HMD), Pneumonia, Upper Respiratory Infection (URI), Hypothermia, Umbilical Sepsis, Tetanus, Convulsive Disorder, Unexplained fever, Failure to Gain Weight etc.

A Decision Support System (DSS) is actually assists decision making processes based on the available knowledge base of alternatives. It is one of the important and vastly used branches of Artificial Intelligence. DSS supports diagnosis and its probable action that has to be taken on the real time basis. Even a medical decision support system helps to make a diagnosis of diseases and selects an appropriate treatment plan for a sick neonate patient. In some situation, DSS allows for complete automated process of decision making, and provides the mechanism for operational diagnostic intelligence. As we automate more of our decision making process, it increases the speed and consistency in predicting the diseases. This has been a direct impact on productivity, accurateness and time. Knowledge base decision support system helps to filter out inaccurate experiential results and biases around personal judgments. This is particularly important in medical decisions, where a wrong decision has life impacting consequences, particularly for the new born babies. Doctors, medical practitioners and domain related personnel are suffering from high stress levels around decision making. When the concern cases are complex and the outcome of the decision has significant consequences, DSS proves itself to be an important expert hand. In a decision making system, expert involvement is a must. Experts are engaged in their intuitive decision making rather than structured approaches. There are various roles played by Artificial Intelligence in medical informatics and medical scientific research, especially, AI systems have the capacity to learn, leading to the discovery of new phenomena and the creation of medical knowledge.

First part of this work have explored the current status of some prevalent disease amongst neonates of North-Eastern parts of India and the second part have dealt with the development of a Knowledge-Based Decision Support System for Treatment Planning applying several techniques of Artificial Intelligence. Issues like knowledge acquisition (KA), knowledge representation (KR), formalization, tools selection, uncertainty management etc. have been considered in detail in the process. The Object Oriented (OO) approach considered as a suitable candidate for the purpose of knowledge representation and implementation. The validity of the system has tested with live cases of the domain. The issues like Ruled-Based approach, Case Based approach, Learning Mechanisms has also been considered in the process of making decision support system.

Different Data Mining techniques shows significant results to find significant pattern attempted on the collected data. Classification problems have been successfully produced with the use of Decision Tree, ID3, C4.5, and Feed Forward Neural Networks. For human like decisions making, Soft Computing Techniques are being applied in neonatal domain. Several Soft Computing methodologies like Artificial Neural Network, Rough Set Theory have shown great efficiency in decision making and prediction process with higher accuracy rate. The increasing demand of Artificial Neural Network application for predicting the disease shows better performance in the field of medical decision making. The proposed technique involves training a Multi Layer Perceptron (MLP) with a Back Propagation (BP) learning algorithm to recognize a pattern for the diagnosing and prediction of neonatal diseases. Comparative study of using different training algorithm of MLP, Quick Propagation, Conjugate Gradient Descent, shows the higher prediction accuracy.

The Neuro-Genetic fusion process of Artificial Neural Network with Genetic Algorithm has been established by using GA to optimize the parameters for an ANN with specific topology architecture. Back propagation neural network learning done by frequent changing of the weights at the output layer. Neuro-genetic fusion approach shows substantial improvement in predicting the neonatal disease for the development of decision support system. Computational Intelligence (CI) is a sub-branch of Artificial Intelligence. It is the study of adaptive mechanisms to enable or facilitate intelligent behavior in complex and changing environment. In case of neonatal disease diagnosis and management it can also be a great part for decision making and useful tool for domain experts.

Finally, for the fast growing fields in the intelligent system development for any organization, Decision Support System becomes one of the most vital, important, and strategic and most demanded tool. Though the study successfully evident its objectives, still one must know that decision support system is not suppose to replace the domain expert's knowledge, rather it is being helpful for those personnel related with neonatal health care management by giving reliable assistance.

Dilip Roy Chowdhury
Dept. of Computer Science & Application
University of North Bengal

December 2012

DECLARATION

I declare that the thesis entitled "A Study on Some Prevalent Diseases Amongst Neonates of North-Eastern Parts of India and Development of Knowledge-Based Decision Support System for Treatment Planning" has been prepared by me under the guidance of Prof. R. K. Samanta (Principal Supervisor), Professor, Dept. of Computer Science and Application, University of North Bengal and Prof. M Chatterjee (Co-Supervisor), North Bengal Medical College & Hospitals, West Bengal University of Health Sciences. No part of this thesis has formed the basis for the award of any degree or fellowship previously.

Dilip Roy Chowdhury

Department of Computer Science & Application,
University of North Bengal
Raja Rammohunpur, Shivmandir
Siliguri, Darjeeling, Pin: 734 013,
West Bengal, India.

Date : 27th December' 2012

PREFACE

The recent days Artificial Intelligence proves its optimal capability for the successful development of knowledgebase decision support system. The modern history of the computer has significantly elaborates that there is a process to develop an artificial electronic thinking mechanism which may take decision as when required, if the proper information is supplied. In the search of modern technology incorporating medical science and artificial intelligence, the doctors, scientists and researchers were fascinated by the potential of such a technology, and started incorporating in medical decision making. Even it would be able to assist doctors, clinicians in various tasks like diagnosing, managing the disease, giving expert decisions. Mutual efforts by human experts and knowledgebase schemes have resulted in systems which can have the capability of diagnosing neonatal disease.

General people always need the best treatment for the diagnosed disease, especially for neonatal cases. Even we needs better patient management and treatment planning always. Using this proposed system we certainly expect better patient management with ultimate accuracy. Moreover, it would be saving time, huge expenditure and to some extent mental harassment. The growing information technological demands says that proposed system would be of great help not only for the domain experts but for the society also. To mitigate the problems of neonatal disease diagnosis and treatment planning few measures have been taken into consideration on the following chapters.

Chapter 1 provides a historical perspective and some fundamental issues in Decision Support System, Artificial Intelligence in medicine. There is a discussion on how decision support system can be a helpful tool in medical decision making. Some of the useful helps of DSS have been discussed. The objective of the study, that is, development of knowledgebase decision support on prevalent diseases amongst neonates is discussed. Complete knowledge of domain related problems, neonate and their health related problems, health care, examination procedure and clinical terms related to neonates have been reflected. Finally benefits of the proposed research work and chapter summary vividly discussed in this chapter.

The next two chapters, that is, chapter 2 and 3 describe the domain related facts. Some of the statistical overview in global, national and regional respect has been described with specific data. Chapter 2 describes the medical knowledge as well as neonatal domain knowledge and their characteristics. Several statistical views with present status of neonatal health, neonatal disease pattern and their causes have also been discussed. Neonatal mortality and morbidity status and various statistical importances with different related indicator of the said domain have been included in this chapter. Chapter 3 provides a complete study of the status of the new born in study region. This is the domain base of the research, based upon which the complete structural development of the knowledgebase decision support system for neonatal prevalent disease stands.

Chapter 4 describes the making of decision support system using expert system technology. Different issues of knowledge accusation process, inferences and rule based implementation using forward chaining method have explored in chapter 3.

Chapter 5 and 6 provides the detailed knowledge about basics and improved Data Mining process respectively in neonatal disease diagnosis. Knowledge unearthing which is a form of data mining concepts using clustering mechanism has been described in chapter 5. Chapter 6 provides details about data mining through Decision Tree and C4.5 algorithm. Implementation classifier like J48 has been utilized. Different classifiers characteristics like ROC, TP rate, FP Rate, F-measure, Recall has been analyzed with real life data.

Model development based on Soft Computing and Rough Set has described in chapter 7 and 8 respectively. Here in chapter 7 hybridization of rough set theory and C4.5 classification algorithm applied taking consideration of parameters including confidence factor, folding etc. To mange uncertainty in decision making process, Rough set based model designed in chapter 8. Reduct and Core attribute generation using Genetic, Johnson and Exhaustive has been described.

Chapter 9 and 10 describes another combination of Data mining with Rough Set and Rule based reasoning system respectively.

Rest of the chapters, i.e., 11, 12 and 13 are based on the Intelligent System development. Particularly use of Artificial Neural Networks to recognize and train the pattern for the diagnosing and prediction of neonatal diseases is intensely discussed. Neuro-Genetic fusion approach for the development of decision support system has been shown greatly successful in this regard. The fusion process has done with Neural Network with Genetic algorithm. Finally chapter 14 describes the fulfillment of the objectives and future scope of the study.

ACKNOWLEDGEMENTS

The thesis contains the results of the work done during last seven years in the Expert System Laboratory, Department of Computer Science & Application, University of North Bengal. A part of the results has already been published in various journals of computer sciences of reputes. This research was comprehended within the framework of the Artificial Intelligence and Decision Support System incorporating Soft Computing, Data Mining, Rough Set Computing and Artificial Neural Network techniques. Indeed, I would have never reached the point of finishing my thesis without the help and support of others.

The satisfaction that accompanies the successful completion of any task is incomplete without the mention of those people who made it possible. Success is the epitome of hard work, perseverance, undeterred zeal, and determination and, above all, the most encouraging guidance and advice that serve as beacon.

I express my profound gratitude and indebtedness to my Principle Supervisor **Professor (Dr.) Ranjit Kumar Samanta, Professor and Head**, Department of Computer Science & Application, North Bengal University, for his technical guidance, supervision, tireless help and encouragement throughout the research work. He gave me the opportunity to participate in outstanding research project in computer science that have inspired me and essentially shaped my further journey through the scientific world and also uncomplainingly provided a lot of valuable knowledge that helped me to significantly improve my research work and thesis.

I express my sincere thanks to Co-supervisor **Prof. (Dr.) Mridula Chatterjee**, HOD, Dept. of Paediatrics, North Bengal Medical College & Hospitals, Siliguri, Darjeeling, who has been a constant source of encouragement and inspiration guidance. She has been the most instrumental in making the necessary arrangements particularly related to domain. She supported my work as domain expert with her deep knowledge and vast experience on the subjects.

I wish to pay my special thanks to Dr. Indrajit Ghosh, Assistant Professor, Dept. of Computer Science, AC College, Jalpaiguri, Dr. Amal Saha, Expert System Laboratory, Dept. of Computer Science & Application, University of North Bengal and Dr. Sanjay Dutta, Computer Centre, Biswabharati University for their incessant support and help all the time during the work. Their suggestions and advice as senior fellow should duly be acknowledged.

I feel indebted to my parents and elder brother for their moral support which facilitated my studies and without which I could hardly achieved my academic advancement.

My special thanks goes to my wife, Dipanwita Bhattacharjee (Roy Chowdhury) for unwearyingly supporting me throughout the research work. Thank you for never complaining because of my lack of time, for enduring the full continuum of my frame of mind, for helping me wherever you could to keep additional work away from me, and for the thousands of other tiny things unmentioned here for lack of space.

And finally, I express my heartfelt thanks and regards to all teachers, staff members of the Department of Computer Science & Application, University of North Bengal for their support and also every individual who has been associated with my research work including those whom I may inadvertently fail to mention.

(Dilip Roy Chowdhury)
27th December, 2012

Expert System Laboratory
Department of Computer Science & Application
University of North Bengal,
Raja Rammohanpur, Darjeeling,
West Bengal, India.

Table of Contents

Chapter 2

Chapter 3

Chapter 4

Chapter 8

Chapter 9

Chapter 10

EXPERT SYSTEM MODEL DESIGNING IN DIFFERENTIAL DIAGNOSIS OF NEONATAL DISEASE **172**

Chapter 11

ARTIFICIAL NEURAL NETWORK MODEL FOR NEONATAL DISEASE DIAGNOSIS & MANAGEMENT **184**

Chapter 12

NEURO-GENETIC FUSION APPROACH FOR NEONATAL DISEASE DIAGNOSIS: A DECISION SUPPORT SYSTEM 205

Chapter 13

Chapter 14

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

Chapter 1

Introduction

1.1. Introduction

The modern history of the computer has significantly elaborates that there is a process to develop an artificial electronic thinking mechanism which may take decision as when required, if the proper information is supplied. In the search of modern technology incorporating medical science and artificial intelligence, the doctors, scientists and researchers were mostly fascinated by the potential of such a technology. Even they started utilizing this technology in medical decision making. The processes of developing such a technology have been a subject of active research, which would be able to store and process huge medical information. More over it would be able to assist doctors, clinicians in various tasks like diagnosing, managing the disease, giving expert decisions. This actually motivates the scientists and researchers developing a discipline named Artificial Intelligence in medicine (AIM).

Consider a child having various symptoms of disease when comes for the treatment in the medical or any health home. They may likely to be seen by the junior doctor relatively having less clinical knowledge. Whereas the clinical knowledge and wisdom are concentrated among senior doctors, removal of several diagnosing steps may be useful for the patient. This may be one of the causes of medical decision making error. The impact of this kind may play significant role increasing mortality, morbidity and economic loss.

Now a day, clinicians are struggling with information overload. It is estimated that we use nearly 2 million pieces of information in our decision making and that biomedical knowledge is doubling every 20 years [1]**.** Moreover proper and adequate information may not be available all the time. Even the scarcity of domain experts also matters while diagnosing the disease and proper management of the patient.

Health, nutrition and education are the most significant attributes for the development of human resources. The overall progress of any region depends upon these criteria. Unfortunately, the scenario of the studied region covered under the thesis has been found uneven. This might be due to underfunding of heath sector, poor performance by the staff engaged in the health care system, unavailability of the domain expert, improper distribution of facilities in rural areas, poor socio economic condition, and behavioral changes in climates along with covering large population.

The majority of the neonatal deaths may be managed with cost-effective solutions like decision support system (DSS) in this domain. Proper nutrition and hygiene may the important criteria in several cases. For this we need not require highly trained person or sophisticated equipments. Rather any system which has the capability to diagnose the proper disease may be a useful hand of the domain specialists.

In this study, an alarming result regarding the neonatal mortality rate has been found; the details of which are provided in chapter 3. This motivated us towards thinking of such a DSS providing support to decrease mortality rate to some extent using various artificial intelligence (AI) techniques. Modeling techniques and domain knowledge have been used by researchers occasionally for utilizing known information in complex problems. By the use of such modeling techniques in different related practical fields, the system may be highly beneficial and may improve the information retrieval and proper diagnosing performance. This work may bridge the gaps between medical professionals and the patients.

Most of the researchers and scientists are familiar with the statistical approach to data analysis. There have been attempts to discover humanly understandable concepts. Given a particular hypothesis, statistical tests are applied to data to see if any relationships can be found between different parameters. We look at raw data and then attempt to establish relationships between hypotheses within the data; whereas learning systems are able to produce complex characterizations of those relationships. Disease diagnosis for neonate is a specific application domain of AIM. Various expert systems have been developed for different diagnosis. Since only a limited number of such decision support systems have been developed for neonates, it is suggested that each individual can be better served by a decision support system customized to their particular interest.

This motivated us to initiate a study on neonatal diseases of North Eastern parts of India. The result of this study motivates us to develop a DSS for the domain treatment planning.

1.2. Decision Support System in Medical Informatics

1.2.1. Needs

During the research work, we found several reasons for which medical decision support system or the expert system can be applied. The reasoning area may be of the following:

- **Workable and Assisting in Diagnostic Process**

  Decision support system may be helpful if it found workable and capability of assisting the diagnostic process in case of severe and critical case.

- **Assisting Information Retrieval Process**

  This is really useful in the process of medical information retrieval from huge data base of domain related problem. To search any particular disease related problem based on specific criteria this may be immense help for the domain experts in medical informatics.

- **Watchful and Reminding for use of Medicine**

  The system may be requiring sophisticated features for the patients, like in the case of medical test results, if that exceeds from normal range then it may generate signals and reminding the patient as well as doctors. So the system is watchful in nature.

- **Assisting Diagnosing, Planning and Management**

  To get the ultimate result of diagnosis of the disease and thereby curing the patient, a proper planning and management is needed badly. It may be useful on this platform also.

- **Assisting on Interpreting the Medical Image**

  Various forms of medical images like Digital X-Rays, MRI, CT Scan etc. are using different image standards. In this cases interpretation is very much required for the understanding of the proper information related to specific disease. Thus it is also useful in this field.

A Decision Support System (DSS) is actually assists decision making processes based on the available knowledge base of alternatives. DSS supports diagnosis and its probable actions. Even a medical decision support system helps to make a diagnosis of diseases and selects an appropriate treatment plan for a sick neonate patient. In some situation, DSS allows for complete automated process of decision making, and provides the mechanism for operational diagnostic intelligence. As we automate more of our decision making process, it increases the speed and consistency in predicting the diseases. This has been a direct impact on productivity, accurateness and time. Knowledge base decision support system helps to filter out inaccurate experiential results and biases around personal judgments. This is particularly important in medical decisions, where a wrong decision has life impacting consequences, particularly for the new born babies. Doctors, medical practitioners and domain related personnel are suffering from high stress levels around decision making. When the concern cases are complex and the outcome of the decision has significant consequences, DSS proves itself to be an important expert hand. In a decision making system, expert involvement is a must. Experts are engaged in their intuitive decision making rather than structured approaches.

Decision support system may improve quality of neonatal care by increasing clinicians' available time for direct patient care. It also increases applications of clinical pathways and guidelines. DSS facilitates the use of up-to-date clinical evidence, improved clinical documentation and thereby expects patient satisfaction. It also improves efficiency in health care delivery by reducing costs through faster result processing, reductions in test duplication, decreased adverse events, and changed patterns of drug prescribing favouring cheaper but equally effective brand.

1.2.2. Some of the Useful Help of DSS

- It can decrease variation in the quality of care.
- Automatic provision for expert advice, expertise and recommendations sourced from up-to-date, knowledge of best practices.
- It can support medical education and training.
- It can be a helpful tool to overcome problems of inefficient coding of data.
- This may be cost-effective except initial capital and update and maintenance costs.
- It may provide immediate feedback to patients.
- If integrated with a service, which can help in history taking, diagnosis, treatment and encourage more efficient data gathering.
- It also provides an audit trail and support research.
- DSS maintains and improve consistency of care.
- Clinical information can be supplied anytime, anywhere it is needed applying DSS techniques.

In view of the same, it is recommended that, DSS is a not a decision making system, rather it's a decision support system.

1.3. Objectives of Research Study

1.3.1. Objectives

The main objective of our research is to explore conditions under which such a decision support system for neonatal disease diagnosis is feasible, and implement a concept prototype. This research work elaborates that, how an interactive decision support system having capabilities of machine programming with some specific rules enables a medical consultant on finding neonatal abnormalities. For a specific case, consultant works by prompting a doctor for a series of calls about different neonatal blood samples. On every occasion where anomalies are found, the system proceeds to create a chain of data-dependent questions to the user or doctor. For the improvement of diagnostic process, DSS would be beneficial tool for diagnoses of probable illnesses of a neonate and recommends courses of management, also guides for a proper treatment planning.

Intelligent systems that combine user profiling, data mining, and artificial intelligence techniques may be developed to provide many of the services that have traditionally been offered by the domain experts. This is the area of focus where our decision support system(DSS) is different from any others. We have used advanced data-mining (DM) techniques, soft computing (SC) techniques; rough set (RC) based applications, artificial neural networks (ANN), few hybridization techniques and used them to develop knowledgebase decision support system. This would be a perfect diagnosing assistant tool for guiding the domain experts and researchers in their diagnosis part and thereby taking proper neonatal care and management of the neonates. The system will learn about each case history and then be able to recommend items or suggest alternative terms of predicting the disease.

In a point of thinking, it is a matter of concern how the decision is being applied to represent the application domain. In our case, it is neonatal disease diagnosis. It also makes the knowledgebase use for both in theoretical and the practical level. Obviously, an unknown representation of the domain knowledge within the diagnosis task is not viable. Moreover, an unambiguous domain representation also provides the generosity of the diagnosis problem-solving method and facilitates the maintenance and reuse of knowledge as when needed. In the field of medical informatics, Artificial Intelligence (AI) provides a wide range of option to tackle this problem. But the selection of the particular approach implies the novel use of this knowledge representation formalization. In high complex conceptual like neonatal disease diagnosis, any chance of ignorance of knowledge causes serious impact on the neonatal health. Thus development of knowledgebase decision support system would be the primary objective of this research which would be beneficial to the experts and the society as well.

1.3.2. Area of the Study

The Study area where we have mainly worked for is on the North Eastern parts of India. This area actually refers to the North Bengal consisting: Darjeeling, Jalpaiguri, Tarai and Dooars region, Coach Behar. Besides, it includes Seven Sister states and Sikkim also. This area is ethnically different from the other parts of India. The said area has got strong ethnic and cultural ties with East Asia and Southeast Asia. The region has several groups of scheduled caste, tribes and other sub-tribal communities. It is predominantly rural with over 84% of the population living in the countryside. According to 2001 census, the total literacy rate of the population in the region is 68.5%. Female literacy rate is 61.5%, which is higher than the country's average literacy rate of 64.8% and 53.7% respectively.

This area is famous for TTT (Tea, Tourism & Table Tennis) culture, handicrafts, and wooden crafts. The area is full of unique scenic beauty. But there are dark sides as well. This area is suffering from unemployment, infrastructural problems and mostly of illiteracy. Rapid increase of population is one of the important indicators of the dark sides. This leads to various problems in the general people over the region. The health care system is not sufficient enough to provide all the necessary medication properly. Mostly, in village areas, there has been acute problem of neonatal prevalent diseases. We have concentrated mostly on the North Bengal Tarai and Dooars region for knowledge accusation process, which is discussed in chapter 3 vividly.

In this study, it is seen that North Bengal Medical College & Hospital is having tremendous pressure on providing medical health care facilities. This is because of overcrowding situating. People are coming from surrounding areas even from the neighbor countries like Bangladesh, Nepal and Bhutan for the treatment. Particularly Neonatal Intensive Care Unit (NICU) is seriously facing lack of services, neonatologist or the expert and proper infrastructure. Any attempt towards developing decision support system for the neonatal disease diagnosis would defiantly helpful in this region. Though the study has conducted on the said area, still the system might be treated as useful assisting tool for every domain specific users in global respect.

1.3.3. Viability of Decision Support Systems

The popularity of inference and knowledge representation schemes of modern decision support systems make them well appropriate to the development of disease diagnosis systems. Events related to disease, historical profile and sign symptoms directly integrated into knowledgebase and reasoning procedure. Disease diagnosis conditions can be performed concurrently with other reasoning jobs. With the capability of monitoring and evaluating the performances and actions, DSS provides extraordinary supports to the domain experts.

This system for neonatal prevalent disease diagnosis and treatment plan may have been successfully deployed in a number of demanding medical environments where neonatal intensive care is taken care of. Though, the system has generally dealt with monitoring neonatal disease diagnosis, still prediction of prevalent diseases and control management would be performed by the successful use of system.

Viability of the Knowledge-based DSS for neonatal disease diagnosis presented in this study is the evidence for architectural background generation and construction of decision support systems which is able to diagnose and produce the management of neonatal disease. It also describes the way in which this architecture was employed in the adaptation of the existing decision support system to meet the disease diagnosis requirements of neonatal domain. Finally, it analyzes the performance of the system in various platforms of artificial intelligence, data mining, soft computing, rough set computing and artificial neural networks. In viewing all the analytical performance the system measures its significant viability in the field of successful design of decision support system for the neonatal prevalent disease diagnosis and treatment planning.

This work is an attempt to propose a suitable approach to deal with the diagnosis problem and management of the disease in towering complex conceptual domains of prevalent neonatal disease. Initially our research hypothesis was to tackle all the troubles considering several approaches, scopes and their combination which would improve the efficiency of disease diagnose. Later on, our research is concentrated on two complementary approaches; firstly, uses a general point of view of knowledge description about the neonatal problems, and secondly, utilize the particular knowledge to obtain optimal solution and giving an accurate decision.

To solve the problems of a neonate, the basic knowledge regarding neonatal disease pattern have been considered. This proposal describes a theoretical model to represent the evolution of abnormal behaviours like diseases patterns, evolution of diseases and complexity of the said disease. This diagnostic process uses this knowledge to obtain explanations by the use of the disease patterns. This being a deductive approach as we obtain a solution for an actual problem using the general knowledge of neonatal prevalent diseases and their treatment plans. This has been reflected by the theoretical model in this work.

In various research stages, we have utilized the domain knowledge by implementing various theoretical models. The data have been analyzed, pre-processed, classify and then implement in the proposed model. The knowledge needed to solve a problem is based on previous solved problems. As an example, senior neonatologists usually make decisions based on their experience with previous patients. If any automated system assist the experts to diagnose the correct disease, then many problem may solved. Combining various approaches, two advantages could be recognized. Firstly, using various approaches we can represent a wider range of knowledge with prevalent disease specific attributes, and secondly, the diagnostic process would use the advantage of various approaches to obtain an optimal solution towards developing knowledge based decision support system for treatment planning.

The experiments were based on the programs generating data which describe the behavior of the simulated models. Hypothesis about the behavior of the models were tested by statistical analysis with collected data. The program were implemented and after their critical analysis, more complex yet productive model were designed as well as developed using the resulting knowledge then tested the accuracy of predicting the diseases.

1.4. Overview of the Neonates and Neonatal Heath Care

1.4.1. Overview

“Children are the future of the nation”. This famous quote is often used by many people all over the world. This may sounds good and seems to be right; but the reality remains there in dark side, that there are a plenty number of sick, filthy, hungry, beggarly kids all around us without having any hope of a secured future. Many children have been deprived of basic needs like food, proper shelter and sanitation, basic education, basic training and access to health care facilities. But every child has the right to protection, basic education, and proper healthcare, a healthy environment and good livelihood opportunities. This is because; in future they may be of nation builder. There is absolutely no doubt that health is wealth. A wealthy person having ill health cannot enjoy life. On the other hand, a poor man with good health can enjoy his life. Therefore, it is better to have good health. Every nation should provide proper health services to her every citizen. Good health status of every individual is equally important. It also important to looks after the children’s health status carefully by the parents or guardians. Particularly new born babies should be given ultimate care to maintain of his/her the future growth.

Health and education both are the fundamental rights of Indian citizen under the Constitution of India. Health, education, water and sanitation are not commercial services. Hence, the responsibility to ensure that each citizen receives good education and health care is a must. The most efficient of public health systems cannot achieve the best possible results unless they are supported by the welfare organizations having food availability and nutritional status, drinking water supply, housing, transport, education, employment and gender equality. In third world countries these are the most critical inputs without which health systems play a very minor role in preventing diseases. There has been an increased focus on issues that affect children and on improving their health.

Neonates are vital to the nation's present and its future. All neonates needs and deserve a healthy start and has to be properly nourished before birth, in infancy, and during their growing years to receive basic health care both prenatally and during the crucial early years after birth. A proper scientific and public health advancement steps has been proving reduced mortality and morbidity rate of the children in recent days.

1.4.2. Neonates

New born period encompasses the first four weeks of extra uterine life [2]. A neonate is a baby who is 4 weeks of old or younger. A neonate is also called a newborn. The neonatal period is the first 4 weeks of a child's life. This period represents a time when changes occur rapidly, and many critical events can takes place. Newborn or neonate actually comes from Latin word, '*neonatus'*, which refers to an infant in the first 28 days after birth [3]. During the first 28 days, most present from congenital birth defects are discovered.

The total pediatric age group is subdivided as:

i.	Fetal	:	Conception – Birth
ii.	Neonates	:	0 – 4 weeks
iii.	Infants	:	4 weeks – 1year
iv.	Toddler	:	1 year – 3 years
v.	Pre-School	:	3 years – 5 years
vi.	School Going	:	5 years – 10 years and
vii.	Adolescence	:	10 years – 19 years.

First week of life is, less than 7 is or greater than 168 hours is known as early neonatal period. Late neonatal period extends from 7th to less than 28th day.

It should be pointed out that the disease patterns, drug selections, diets and common rearing technologies are different for the different pediatric age groups.

1.4.2.1. Few Terms Related to Neonate [2]

- **Live Born**

  A live born neonate is a product of conception, irrespective of weight or gestational age after separation from mother. It shows any evidence of life such as breathing, heartbeat, pulsation of umbilical cord or definite movement of the voluntary muscle.

- **Still Birth**

  A fetal death is a product of conception that, after separation from the mother, does not show any evidence of life. This is at a gestational age of 20 weeks of more or weighing weighting not more than 500g is designated as still birth.

- **Term Baby**

  Any neonate born between 37 and < 42 weeks of pregnancy irrespective of the birth weight.

- **Pre-term Baby**

  Any neonate born before 37 weeks (<259 days) of pregnancy irrespective of the birth weight.

- **Post-term Baby**

  A neonate born at a gestation age of 42 weeks or more (294 days or more) irrespective of the birth weight.

- **Low Birth Weight(LBW) Baby**

  Any neonate weighing less than 2500g at birth irrespective of the gestational age is the LBW neonate.

- **Very Low Birth Weight(VLBW) Baby**

  Any neonate weighing less than 1500g at birth irrespective of the gestational age is the VLBW neonate.

- **Extremely Low Birth Weight (ELBW) Baby:**

  Any neonate weighing less than 1000g at birth irrespective of the gestational age is the ELBW neonate.

Neonates not only constitute a large population group, but also are vulnerable or special risk group. The risk is related with growth, development, disease pattern and survival. From the commonly accepted indices, it is evident that mortality rates in this age group are higher than adult population especially in developing countries. Thus by improving the health status of neonates, we contribute to the health of the general population. These considerations have led to the formulation of special health service for children all over the world.

1.4.3. Common Conditions in Newborns [13]

Some physical conditions are especially common during the first couple of weeks after birth. They are as follows:

- **Abdominal Distension**

  Most babies' bellies normally stick out, especially after a large feeding. Between feedings, however, they should feel quite soft. Most likely the problem is due to gas or constipation, but it also could signal a more serious intestinal problem.

- **Birth Injuries**

  It is possible for babies to be injured during birth, especially if labor is particularly long or difficult, or when babies are very large. While newborns recover quickly from some of these injuries, others persist longer term. Quite often the injury is a broken collarbone, which will heal quickly if the arm on that side is kept relatively motionless. Incidentally, if after a few weeks a small lump may form at the site of the fracture, this is a positive sign that new bone is forming to mend the injury.

  Muscle weakness is another common birth injury, caused during labor by pressure or stretching of the nerves attached to the muscles. These muscles, usually weakened on one side of the face or one shoulder or arm, generally return to normal after several weeks. At this time one should ask pediatrician to show how to nurse and hold the baby to promote healing.

- **Blue Baby**

  Babies may have mildly blue hands and feet, but this may not be a cause for concern. If their hands and feet turn a bit blue from cold, they should return to pink as soon as they are warm. Occasionally, the face, tongue, and lips may turn a little blue when the newborn is crying hard, but once he becomes calm, his color in these parts of the body should quickly return to normal. However, persistently blue skin coloring, especially with breathing difficulties and feeding difficulties, is a sign that the heart or lungs are not operating properly, and the baby is not getting enough oxygen in the blood. Immediate medical attention is essential.

- **Coughing**

If the baby drinks very fast or tries to drink water for the first time, he may cough and sputter a bit; but this type of coughing should stop as soon as he adjusts to a familiar feeding routine. This may also be related to how strong or fast a breastfeeding mom’s milk comes down. If he coughs persistently or routinely gags during feedings, one should consult the pediatrician. These symptoms could indicate an underlying problem in the lungs or digestive tract.

- **Excessive Crying**

All newborns cry, often for no apparent reason. If anyone made sure that the baby is fed, burped, warm, and dressed in a clean diaper, the best tactic is probably to hold him and talk or sing to him until he stops. One cannot “spoil” a baby this age by giving him too much attention. If this doesn’t work, wrap him snugly in a blanket.

One will become accustomed to baby’s normal pattern of crying. If it ever sounds peculiar—for example, like shrieks of pain—or if it persists for an unusual length of time, it could mean a medical problem. Then one should call the pediatrician and ask for advice.

- **Forceps Marks**

When forceps are used to help during a delivery, they can leave red marks or even superficial scrapes on a newborn’s face and head where the metal pressed against the skin. These generally disappear within a few days. Sometimes a firm, flat lump develops in one of these areas because of minor damage to the tissue under the skin, but this, too, usually will go away within two months.

- **Jaundice**

Many normal, healthy newborns have a yellowish tinge to their skin, which is known as jaundice. It is caused by a buildup of a chemical called bilirubin in the child’s blood. This occurs most often when the immature liver has not yet begun to efficiently do its job of removing bilirubin from the bloodstream (bilirubin is formed from the body’s normal breakdown of red blood cells). While babies often have a mild case of jaundice, which is harmless, it can become a serious condition when bilirubin reaches what the pediatrician considers to be a very high level. Although jaundice is quite treatable, if the bilirubin level is very high and is not treated effectively, it can even lead to nervous system or brain damage in some cases, which is why the condition must be checked for and appropriately treated. Jaundice tends to be more common in newborns who are breastfeeding, most often in those who are not nursing well; breastfeeding mothers should nurse at least eight to twelve times per day, which will help produce enough milk and help keep bilirubin levels low.

Jaundice appears first on the face, then on the chest and abdomen, and finally on the arms and legs in some instances. The whites of the eyes may also be yellow. The pediatrician will examine the baby for jaundice, and if she suspects that it may be present—based not only on the amount of yellow in the skin, but also on the baby's age and other factors—she may order a skin or blood test to definitively diagnose the condition. If jaundice develops before the baby is twenty-four hours old, a bilirubin test is *always* needed to make an accurate diagnosis. At three to five days old, newborns should be checked by a doctor or nurse, since this is the time when the bilirubin level is highest; for that reason, if an infant is discharged before he is seventy-two hours old, he should be seen by the pediatrician within two days of that discharge. Some newborns need to be seen even sooner, including:

- Those with a high bilirubin level before leaving the hospital
- Those born early (more than two weeks before the due date)
- Those whose jaundice is present in the first twenty-four hours after birth
- Those who are not breastfeeding well
- Those with considerable bruising and bleeding under the scalp, associated with labor and delivery
- Those who have a parent or sibling who had high bilirubin levels and underwent treatment for it.

When the doctor determines that jaundice is present and needs to be treated, the bilirubin level can be reduced by placing the infant under special lights when he is undressed, either in the hospital or at home. His/her eyes will be covered to protect them during the light therapy. This kind of treatment can prevent the harmful effects of jaundice. In infants who are breastfed, jaundice may last for more than two to three weeks; in those who are formula-fed, most cases of jaundice go away by two weeks of age.

- ❖ **Lethargy and Sleepiness**

  Every newborn spends most of his time sleeping. As long as he wakes up every few hours, eats well, seems content, and is alert part of the day, it's perfectly normal for him to sleep the rest of the time. But if rarely alert, does not wake up on his/her own for feedings, or seems too tired or uninterested to eat, one should consult your pediatrician. This lethargy, especially if it is a sudden change in his usual pattern, may be a symptom of a serious illness.

- ❖ **Respiratory Distress**

  It may take our baby a few hours after birth to form a normal pattern of breathing, but then he should have no further difficulties. If he/she seems to be breathing in an unusual manner, it is most often from blockage of the nasal passages. The uses of saline nasal drops, followed by the use of a bulb syringe, are what may be needed to fix the problem; both are available over the counter at all pharmacies.

However, if newborn shows any of the following warning signs, notify your pediatrician immediately:

- Fast breathing (more than sixty breaths in one minute), although keep in mind that babies normally breathe more rapidly than adults.
- Retractions (sucking in the muscles between the ribs with each breath, so that her ribs stick out)
- Flaring of his/her nose
- Grunting while breathing
- Persistent blue skin coloring

1.4.4. Neonatal Heath Care

The statistics in developing countries shows that mostly children, adults and particularly neonates are vulnerable to malnutrition because of below standard dietary intakes, infectious and communicable diseases, deficiency in appropriate caring and unbalanced food distribution within the family. Three standard indices of physical growth that describe the nutritional status of children are:

- o *Height-for-age (stunting)*
- o *Weight-for-height (wasting)*
- o *Weight-for-age (underweight*)

As per the Third National Family Health Survey (NFHS-3, 2005-06) [4], almost half of children under five years of age (48%) are stunted and 43% are underweight. The proportion of children who are severely undernourished (more than three standard deviations below the median of the reference population) is also notable, 24% according to height-for-age and 16% according to weight-for-age. Wasting is also quite a serious problem in India, affecting 20% of children under five years of age.

Neonates are generally viewed as healthy when they are assessed by child standards, and there has been a great deal of progress in reducing neonatal death and diseases. But the country should not be unsighted by these facts. Several indicators of neonate's health point to the need for further improvement. A recent improvement in neonates' health needs to be sustained and further efforts are needed to optimize it. To accomplish this, the nation must have an improved understanding of the factors that affect health and effective strategies for measuring, accessing and using information on neonate's health.

The greatest concern in the regionalized specialty care services and special dependence of modern neonatology. Because serious illness is far less prevalent in children than in adults, pediatrics must concentrate patients into networks of regionalized centers with special expertise and resources. Adult care can tolerate far more decentralized services. As a result many cost containment strategies are being directed towards financial contracting networks that create strong disincentives for the use of specialty care facilities. There should be advancement in highly regionalized neonatal care systems may prove increasingly difficult against a rising current of de-regionalization based on adult-focused financial contracting.

The task for the neonatology community, as it is for all those concerned for the health and well-being of children is to create the technical guidance and political voice to ensure that the special requirements of young women, newborns, and children are represented adequately in the often fractious deliberations over the future of health care.

In recent years, there has been an improved focus on issues that affect neonates and on improving their health status. Neonates are forming the largest proportion of immediate workload in primary healthcare. Neonates are vital to the nation's present and its future. All children need and deserve a healthy start - to be properly nourished before birth, in infancy, and during their growing years to receive basic health care both prenatally and during the crucial early years after birth. Scientific and public health advances have improved access to health care and hence, reduced child mortality and morbidity from infectious diseases, accidental causes, even from prevalent diseases. To promote health services today we require considerations of the overall status of neonates, only identification and treatment of specific diseases or deformities is not sufficient to take proper decision.

1.5. Examination Procedure of a Sick Neonate

A neonate must be examined as soon as after delivery, and ideally before discharge from hospital or next day for home births or rapid discharges. The initial examination is usually conducted by midwifery staff in cases of uncomplicated delivery, or by the on-call pediatrician for complicated births.

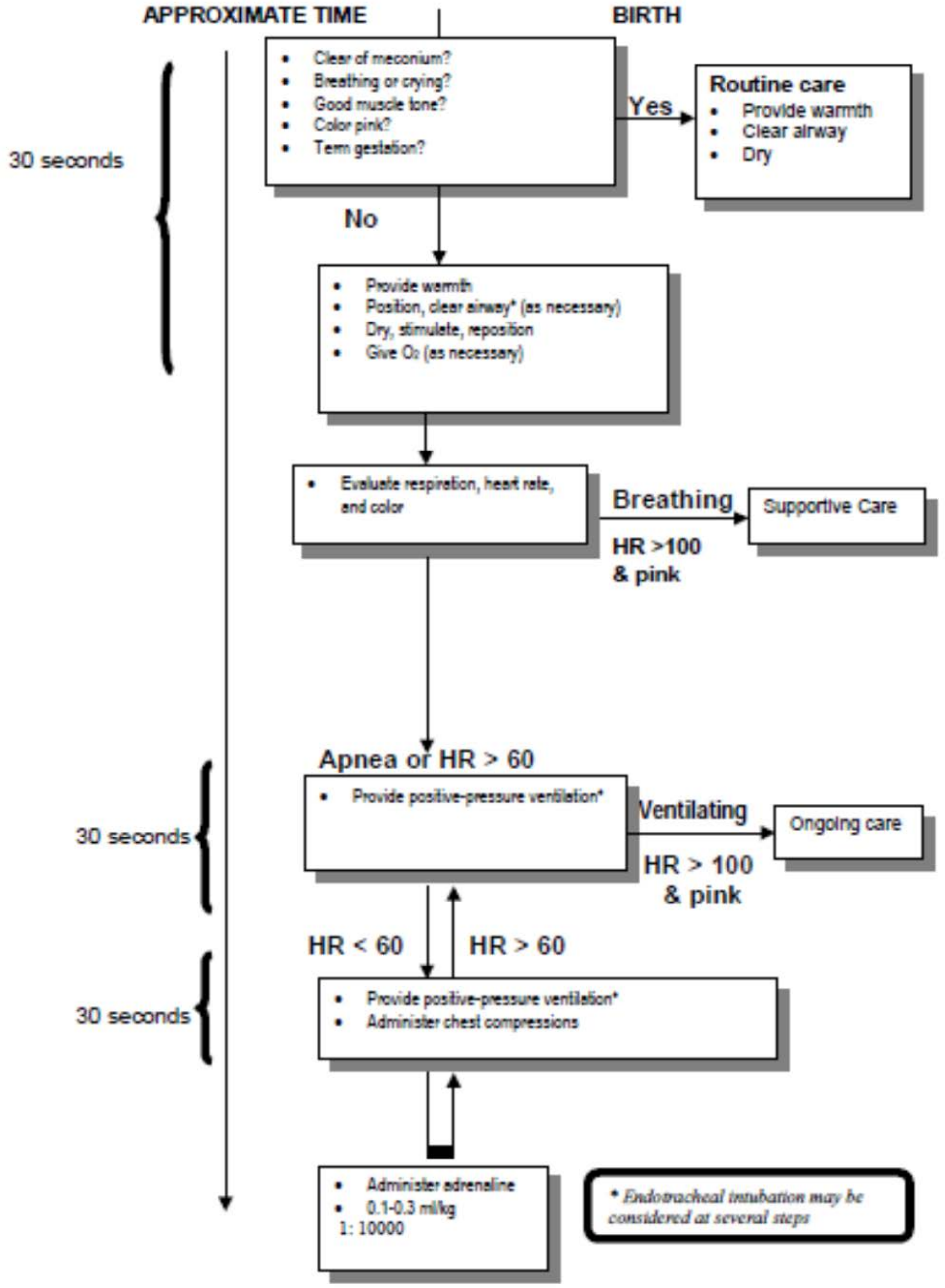

Figure 1.1. Examination Procedure of a Sick Neonate.

Source: Resuscitation algorithm, Essential Pediatrics, 6th edition, CSB Publishers, pp. 139.

Each and every newborn baby is carefully checked at birth for signs of problems or complications. A complete physical assessment will be performed that includes every body system. Throughout the staying period of hospital, physicians, nurses, and other healthcare providers continually assess a baby for changes in health and for signs of problems or illness.

1.5.1. Neonatal Examination Procedure

1.5.1.1. NEST: Neonatal Examination and Screening Trial [5]

It is seen that few hospitals and tertiary centers have been maintain a policy of conducting two pediatric neonatal examinations, Firstly, just shortly after birth to detect any abnormalities that require urgent action, and secondly, before the discharge of the neonate. A trial comparing one with two neonatal examinations found no evidence of a net health gain from the policy of double examination. There were no differences between the two groups with respect to use of community, outpatient, and inpatient resources, or in health care received. It concluded that a single examination policy would save resources both during the postnatal hospital stay, and through fewer outpatient consultations**.**

1.5.1.2. Initial Post-Delivery Examination

Neonates are undergone a brief screening examination. It includes checking the face, eyes, mouth, chest, abdomen, spine and limbs to exclude major abnormalities. Neonates may be well if, a strong cry and a widespread pink blush over the face and body are found. Some children may be born with indiscriminate genitalia. In that case it is important not to guess at the likely gender of the child, but advise that it is uncertain and that further tests will be needed. If you have sufficient clinical experience an orogastric tube should be passed when the neonate's mother has suffered polyhydramnios.

1.5.1.3. The APGAR score

The very first test given to any newborn is the APGAR scoring. This occurs in the delivery or birthing room right after your baby's birth. The test was designed to quickly evaluate a newborn's physical condition and to determine any immediate need for extra medical or emergency care.

The acronym for APGAR is:

A **:** **A**ppearance,

P **:** **P**ulse,

G : **G**rimace,

A : **A**ctivity, and

R : **R**espiration.

The APGAR score gives a reproducible, quantitative, semi-objective assessment of neonatal condition that is useful for assessing a baby's progress or deterioration immediately after delivery [6]. It is important to document it for medical reasons. It is most useful following complicated births or where there are unanticipated problems

Table 1.1. APGAR Scoring

APGAR Scoring			
APGAR Sign	**2**	**1**	**0**
Heart Rate (pulse)	Normal (above 100 beats per minute)	Below 100 beats per minute	Absent (no pulse)
Breathing (rate and effort)	Normal rate and effort, good cry	Slow or irregular breathing, weak cry	Absent (no breathing)
Grimace (responsiveness or "reflex irritability")	Pulls away, sneezes, coughs, or cries with stimulation	Facial movement only (grimace) with stimulation	Absent (no response to stimulation)
Activity (muscle tone)	Active, spontaneous movement	Arms and legs flexed with little movement	No movement, "floppy" tone
Appearance (skin coloration)	Normal color all over (hands and feet are pink)	Normal color (but hands and feet are bluish)	Bluish-gray or pale all over

with the baby after delivery. It should be checked at delivery and 2 and 5 minutes subsequently; these results should be documented in the maternal and neonatal notes.

Five factors are used to evaluate the baby's condition and each factor is scored on a scale of 0 to 2, with 2 being the best score:

1. appearance (skin coloration)
2. pulse (heart rate)
3. grimace response (medically known as "reflex irritability")
4. activity and muscle tone
5. respiration (breathing rte and effort)

There is an excess of mortality and an increased risk of severe neurological morbidity in infants with total APGAR score <7 at 5 minutes [7].

1.5.1.4. Stepwise Routine for New Born Check-Up [8]

First wash hands thoroughly to reduce the risk of cross-infection. After that performs the followings:

- **Listen and Observe**
  - ***Assess overall appearance.*** Note general tone, sleepiness and also observe general condition, proportions and maturity.
  - ***Look* carefully** for evidence of jaundice (preferably in bright, natural light). Are there any birthmarks, rashes or other skin abnormalities?
  - ***Listen* t**o the baby's cry and note its sound.
  - ***Weigh*** the baby and plot this reading on its growth chart.

- **Perform a systematic 'head-to-toes' examination**

This should be done carefully and in good light to detect abnormalities:

- **Head:**
  - Shape, presence of fontanel and whether normal, sunken or bulging.
  - Measure and record head circumference on growth chart.
  - Assess facial appearance and eye position.
  - Look for any asymmetry or abnormality of facial form.
- **Eyes:**
  - Normal shape and appearance?
  - Check for presence of red reflex.
  - Look for obvious cataracts or signs of ophthalmic infection.
- **Ears:**
  - Shape and size.
  - Are they set at the normal level or 'low set'?
  - Check patency of external auditory meatus.
- **Mouth:**
  - Colour of mucous membrane, observe the palate.
  - Check suckling reflex by inserting a **clean** little finger gently inside baby's mouth.
- **Arms and hands:**
  - Are they of normal shape and moving normally?

  - Look for evidence of traction birth injury by checking neck, shoulders and clavicles.
  - Count fingers and observe their shape – is there any evidence of incurving of fingers?
  - Check palmar creases – are they multiple or single? A single palmar crease may be normal, but can be a sign of Down's syndrome.
- **Peripheral pulses:**
  - Check brachial, radial and femoral pulses for rate, rhythm and volume.
  - A hyperdynamic pulse may suggest persistent ductus arteriosus.
  - A weak pulse may occur with a congenital cardiac anomaly (impairing cardiac output and in conjunction with other signs from the examination).
  - Check for radio-femoral delay (aortic coarctation).
- **Heart:**
  - Check cardiac position by palpation and feel for any thrill or heave.
  - Listen to the heart sounds carefully and for any added sounds or murmurs.
  - Suspected abnormalities require further examination (and often more expert opinion and investigation).
- **Lungs:**
  - Watch respiratory pattern, rate and depth for a few seconds.
  - Look for any evidence of intercostals recession.
  - Auscultate lung fields for added sounds.
- **Abdomen:**
  - Look at abdominal girth and shape.
  - Carefully check the umbilical stump for infection or surrounding hernia.
  - Palpate gently for organs, masses or hernia.
  - it is common to be able to feel the liver and/or spleen in healthy newborns.
  - Check the external genitalia carefully (see Ambiguous Genitalia).
  - Palpate for testicles in boys.
  - Inspect the anus (has meconium been passed?).

- **Back:**
    - Look carefully at skin over back and at spinal curvature/symmetry.
    - Is there any evidence of spina-pilonidal sinus hidden by flesh creases or dimples?
    - Palpate the spine gently.
- **Hips:**
    - Specifically test for congenital dislocation of the hip using combination of Barlow and Ortolani manoeuvres.
- **Legs:**
    - Watch movements at each joint.
    - Count toes and check shape.
- **CNS:**
    - Observe tone, behaviour, movements and posture.
    - Elicit newborn reflexes only if there is cause for concern.

Further examination should be conducted as necessary according to any abnormalities that are detected, or suspicions of undetected illness in the baby.

❖ **Record findings**

Always document the findings of the examination in the postnatal care plan and personal child health record. A pro-forma for the examination, kept within the notes, can save time and act as a prompt to ensure that no element of the examination is missed.

1.6. Three Levels of Benefits

Three levels of benefits are proposed by Basden are as follows [9]:

a) Feature benefits;

b) Task benefits;

c) Role benefits.

a) **Feature benefits:**

The advantages which begin from technological features of functionality and user interface are termed as feature benefits. Ease of manipulation can arise, for instance, from graphical user interfaces. An example found in DTI [11] is "critical items are highlighted".

b) **Task Befits:**

The advantages which takes place from the use of any expert system to support a task is task benefits.

c) **Role Benefits:**

Role benefits arise from the effect the expert system has on the roles the user fulfils by carrying out the supported tasks, such as "improved supplier relations", which in DTI [11] is seen to arise from the two task benefits above.

After classifying the benefits into three sets, it is important to have the indication of some form of causal, or at least enabling, link between them. Feature benefits can lead to task level benefits, which can in turn lead to role level benefits.

1.6.1. Usage of Decision Support System

As we have mentioned earlier also that, Decision Support System for neonatal disease diagnosis mainly focused on the point of care. Such a system may be of automated in nature. The pediatricians would interact with the system for diagnosis, analysis, and even management of the data provided to system [11]. Earlier, these types of system were basically used for making decision only. The user had to input the data to be processed and then waits for the outputs generated by the system only. But now the concept has changed. The new system actually forces the user to interact with the system with having user or domain expert's knowledge and also with system knowledgebase to make a utmost analysis and proper diagnosis.

Clinical Decision System may categorize in two parts:

- Knowledge-Based
- Non-Knowledge-Based

1.6.2 Knowledge-Based DSS

In knowledge-base DSS, there are mainly three parts: (i) Knowledge Base, (ii) Inference engine, and (iii) a way to communicate. The data there in the knowledge base are compiled mostly with rules and associations and that are in the form of IF-THEN-ELSE rules. The inference engine combines the rules from the knowledge base with the patient's data and then communication method will allow the system to show the results to the user as well as have input into the system [11]. This has been performed in an environment which is termed as Expert System Cell. We have used this kind of system in our research work using neonatal data.

1.6.3. Non-Knowledge-Based DSS

This kind of decision support system uses a form of mostly machine learning and soft computing techniques, which is a field of Artificial Intelligence. This uses system to learn from the experience they have gathered earlier and finds suitable patterns in the data provided to them. Genetic algorithm and artificial neural network are the most common types of non-knowledge based decision support system. In our research we have used this paradigm efficiently to find the maximum accuracy of disease classification.

1.7. Benefit That We May Expect From Proposed System

For getting ultimate benefit of discussed neonatal domain, an automated knowledge-based decision support system would be beneficial. Rural health centre are lacking of neonatologists or pediatricians or domain experts to serve the medical facility, thus, the proposed system may be used by general people which is related to domain, particularly medical practitioners. This system would also assist the general physicians, working at sub-divisional/district hospitals or engaged in private practices. If there is an infrastructural problem of setting up huge machinery, this system might be useful as a mobile medical unit having one personal computer with required power supply. This system even helpful to medical students also, as if they are not having expert guide on real time, the proposed system might be act as assisting tool, using which they can get the proper decision or may get the second opinion to tally.

As the population increases every day, disease also increases on the same ratio. Thus creating overburden problems for the domain experts, neonatologists or pediatricians. This problem may also be reduced by the use of the proposed system.

Above all, we the general people need the best treatment for the diagnosed disease. Even we needs better patient management and treatment planning. Using this proposed system we certainly expect better patient management with ultimate accuracy. Moreover, it would be saving time, huge expenditure and to some extent mental harassment.

The proposed system does not require any knowledge of Artificial Intelligence (AI) or Decision Support System (DSS) terminology, not even advance training on computer is required to run this system. Any user having little domain knowledge and operational knowledge may handle this efficiently.

Finally, Different funding agencies of Govt. of India are really encourages research projects for the development of such decision support systems using Artificial Intelligence and Expert Systems technology under human resource development schemes[12]. In the growing age of ICT, every medical practitioner is now computer savvy. They are adopting such kinds of system as their assisting tool. Currently, there are good evidences that doctors and bio-medical researchers are actively participating with computer professionals in their projects. Moreover medical professionals are using equipments where the kinds of expert systems are embedded.

Thus taking consideration of growing information technological demands proposed system would be of great help not only for the domain experts but for the society too.

1.8. Summary of the Research Work

The chapter wise scheme of presentation is organized as follows:

i. **Present Medical and Neonatal Domain and their Characteristics with Statistical Importance -- Chapter 2:**

   In this chapter, brief discussions have done on the medical knowledge as well as neonatal domain knowledge and their characteristics. Various neonatal disease pattern and their causes has also been discussed. Neonatal mortality and morbidity status and various statistical importance with different related indicator of the said domain in global, thereafter, Indian perspective of has been included in this chapter.

ii. **Study on the Status of New Born in Terai Region of West Bengal -- Chapter 3 :**

   This chapter has been devoted on the study of the status of New Born in Terai Region of West Bengal. A comprehensive study with different factor of neonatal diseases along with prevalence factor finding has been summarized. We find various neonatal prevalent diseases and their causes. Data have collected, summarized and then analyzed for statistical analysis. Analyzed data are used in the subsequent study mentioned in the following chapters.

iii. A Decision Support System for Prevalent Disease Diagnose & Management for Neonates using ES-Technology -- Chapter 4:

This chapter begins with the representation formalisms which are conventionally used in Expert System. A detail neonatal problem in the study area has been given. A system flow diagram reflects how actually the expert system works in this case. How expert system as a tool or agent helps diagnosing the prevalent disease among the neonates has been discussed. Total information related to the information and knowledgebase development, inference processed is presented here. An outcome of the diagnosing procedure is described. We have explored different issues of knowledge accusation used in the current research. Also relative suitability of the problem discussed in this context in present domain.

iv. A Data Mining and Knowledge Unearthing Concept -- Chapter 5:

The need of knowledge and data mining, knowledge unearthing process in medical disease diagnosis, particularly for the neonatal disease data, has been discussed in this chapter. Several issues related to knowledge base development and data mining using clustering mechanism shows effective in this study, which have been presented here.

v. An Improved Data Mining Concept -- Chapter 6:

This chapter deals with the development of an improved data mining concepts for neonatal disease diagnosis in North Bengal Districts. Data mining through Decision Tree and C4.5 algorithm shows better results in this chapter. Implementation classifier like J48 has been utilized. Different classifiers characteristics like ROC, TP rate, FP Rate, F-measure, Recall has been analyzed with real life data. Finally finds the classifier accuracy depending on those characteristics. For evaluating model quality, there are different popular test metrics for classification models such as confusion matrix, prediction accuracy, receiver operating characteristics and lift have been discussed. Also the chapter presents a confusion matrix generated by the model.

vi. Diagnosis with Soft Computing -- Chapter 7:

In this chapter we present method of designing a hybrid decision support system in Soft Computing paradigm. This includes the application of rough set theory and C4.5 classification algorithm. Generation of rules based on the rough set approach and classification through C4.5 algorithm has done on the data set. The methodology of different classification parameters including confidence factor, folding etc. also presented for the model.

vii. Rough Set Based Model for Differential Diagnosis -- Chapter 8:

We present an effective approach for managing uncertainties using Rough Set based model. How insignificant rules can sidetrack using reduct and core generation has been discussed in this chapter. ID3 Entropy and information gain has been discussed. There is comparison analysis among the different algorithm like Exhaustive, Genetic and Johnson for reduct and core generation using the data set is presented. A detailed comparison with Rough Set and ID3 has been provided.

viii. Coupling with Data Mining and Rough Set Theory -- Chapter 9:

The projected work in this chapter reports the results of a study where some coupling with data mining and rough set theory is proposed for differential neonatal disease diagnosis. Study ensures us that with the use of conventional analysis along with data mining and statistical studies in patient data can improve better disease diagnosing capacity with good accuracy rate. Even using data mining techniques data quality and standard of data, diagnosing plans and treatment procedures and decreases of treatment timings must be improved no doubt. This study in this chapter also suggests that, as this is differential diagnosis, the results might be accepted as first order inference. The next higher order performance is achieved with the results of laboratory tests.

ix. Expert System Model Using Rule Based Reasoning System – Chapter 10:

In this chapter we expressed how the expert system model can be useful for disease diagnosis. This might be helpful as any information that enables individuals to understand their health and make health-related decisions for their family and take care of the little child easily. We have discussed how rule have been generated using Level 5 Object, an AI Shell, and further validated the same using RSES, ROSETA, WEKA, rough set data mining tools. The chapter significantly elaborates that there is absolutely no contribution of three attributes ‘Excessive_Jaundice’, ‘Sclerema’, and ‘GI_Disorder’ in decision making. Overall clarification has been discussed here in this chapter. A comparative study shows that this model may give a useful and better tool for expert opinions and decision making.

x. Artificial Neural Network (ANN) Model for Neonatal Disease Diagnosis – Chapter 11:

This study represents the use of artificial neural networks in predicting neonatal disease diagnosis. How training in a Multi Layer Perceptron (MLP) with a Back Propagation (BP) learning algorithm for recognizing pattern for the diagnosing and prediction of neonatal diseases is vividly discussed. Comparative study of using different training algorithm of MLP, Quick Propagation, and Conjugate Gradient Descent has done for showing better prediction accuracy. There is a discussion on conventional computing method and Neural Network. Development of Neural Network Architecture, hidden layer selection, pre-processing, data set partitioning and training on the neonatal data set has been described in this chapter.

xi. Neuro-Genetic Fusion Approach for Development of DSS -- Chapter 12:

In this chapter, a Neuro-Genetic Fusion approach has been proposed to find and select the best input features for diagnosis of neonatal disease. The fusion process has done with Neural Network with Genetic algorithm. Attributes are selected based on the use of Genetic Algorithm. Total description of genetic algorithm and how it worked in our study has been discussed. Selection for attribute subsets has done by the Attribute Subset Evaluator using Genetic Algorithm. Selected subset then pre-processed and analyzed for choosing best network architecture. We discussed about Batch Back Propagation training algorithm to train the network using various training algorithm parameters. Data set partitioning, pre-processing and training using different Neural Network using batch back propagation algorithm has been provided in this chapter. A comparative analysis of typical ANN with Neuro-Genetic Fusion has focused the better performance and accuracy of predicting the neonatal disease.

xii. Intelligent Diagnostic System Development in Soft Computing Environment using Artificial Neural Network – Chapter 13:

This chapter deals with development of computer aided intelligent diagnostic system. Description of computational intelligence and artificial intelligence is given with specific examples related to our research study. Quick propagation and conjugate gradient descent algorithm has been applied for training the data set because of huge network weight. Feature extraction and reduction is one of the important steps for pattern recognition for that implementation of Correlation-based Feature Reduction has been reflected on our specified area of study.

xiii. Accomplishment of the Objectives and Future Scope – Chapter 14:

This is the final chapter of our research work. It includes the how we accomplish our objectives that mentioned in the chapter I. Even our future extensions of study have been discussed. Here we mentioned how to accomplish the research objectives and reaching the goal after a numbers of research analyses. The chapter also illustrates about research objectives that sets the purpose and focus of our study with the fundamental questions that have already been addressed in different chapters in the thesis. We conclude our discussion after discussing the future scope of this research work.

References

1. Wyatt JC. *"Use and Sources of Medical Knowledge"*, Lancet, 338, pp.1368–1373, 1991.

2. Ghai O.P., Gupta Piyush, Paul V.K., *"Ghai Essential Pediatrics"*, CBS publishers & Distributors, 6th Edition, pp. 136.

3. *"Neonate"*, Merriam-Webster online dictionary, Merriam-Webster.

4. National Family Health Survey, 2005-06, Ministry of Health and Family Welfare.

5. Glazener CM, Ramsay CR, Campbell MK, et al, *"Neonatal Examination and Screening trial (NEST): A Randomised, Controlled, Switchback Trial of Alternative Policies for Low Risk Infants"*, BMJ, 318(7184), pp. 627-631, 1999.

6. *"Postnatal Care: Routine Postnatal Care of Women and Their Babies"*, NICE Clinical Guideline, 2006.

7. Thorngren-Jerneck K, Herbst A, *"Low 5-minute Apgar Score: A Population-Based Register Study of 1 million Term Births"*; Obstet Gynecol, 98(1), pp. 65-70, Jul 2001.

8. *"Current Child Health Screening Subgroup Recommendations For Screening Programmes"*, UK National Screening Committee.

9. Basden, *"Three Levels of Benefits in Expert Systems"*, Expert systems, Vol. 11, no. 2, pp. 99-107, May 1994.

10. DTI, *"Manufacturing Intelligence: A Decision Makers' Briefing. Applying Knowledge Based Systems Technology in the Manufacturing & Process Industries"*, Department of Trade and Industry, London, 1991.

11. Berner, Eta S., ed. *"Clinical Decision Support Systems"*, New York, NY: Springer, 2007.

12. *"Directory of Extramural Research and Development Projects"*. Approved for funding by selected Central Govt. Agencies / Departments during 1994-95.

13. *Caring for Your Baby and Young Child: Birth to Age 5*, American Academy of Pediatrics, 2009, http://www.healtheychildren.org, Available: [Last accessed 19th January, 2012].

CHAPTER 2

Present Medical Scenario: Neonatal Domain

2.1. Introduction

Every day number of baby dies due to neonatal diseases, improper baby care and lack of information relating to the neonates. As we know neonatal period, i.e. first 28 days after birth, is the most crucial period for the survival of a child and that too it is most vulnerable period of their entire life. Thus there is ultimate need of giving attention and focus on the care of the newborn babies for their growth and development. Below, we would discuss the tragic scenario of the neonatal related problems. This chapter is also focusing on the reasons for neonatal mortality. Most of the babies are dying due to infection, prematurity, low-birth weight and some other preventable causes. Newborn health and its contribution to child mortality is one of the important factors. Even there is a clear link between new born health with maternal health which is also an important factor and should be taken care of.

2.2. Maternal Mortality Ratio

When a woman is pregnant or within 42 days of termination of pregnancy and child birth, if death occurs, that death is called maternal death. This maternal death is exclusive of the accidental and incidental cases. Maternal Mortality Ratio (MMR) is the annual number of the maternal deaths by pregnancy or its management, per 100000 live births, for a specified year. According to World Health Organization [1], maternal mortality rate is quite high in global aspect. Every day, nearly about 1000 women die due to complicacy in child birth and from pregnancy in all over the world. Statistics says, in 2008, the number of women died during and following the pregnancy is 358000. Out of that, most of the cases occurred in developing countries. Out of eighth goals of Millennium Development Goals (MDG), maternal mortality is one of the issues. Under MDG5, countries are committed to reduce maternal mortality by three quarters between 1990 and 2015. After 1990, maternal deaths worldwide have decreased by 34%. The national MMR level has come down from 398 per 100,000 live births in 1997-98 to 254 per 100,000 live births in 2004-06, a 36% decline over a span of seven years as compared to a 25% decline in the preceding eight years from 1990-97. However, between 1990 and 2008, the global maternal mortality ratio i.e. the number of maternal deaths per 100000 live births declined by only 2.3% per year. In contrast, to achieve MDG5, annual decline of 5.5% is required.

The Table 2.1. shows the country wise comparison of maternal mortality rate worldwide. In rank wise distribution Afghanistan (1400), Chad (1200) and Somalia (1000) are in the higher rank, where maximum deaths occur. However, we found that the rate of MMR in South East Asia including India (230), Bangladesh (340), Bhutan (200) and Pakistan (260) are low compared to top ranked countries.

Table 2.1. Maternal Mortality Rate [3]

Rank	Country	Maternal Mortality Rate (Deaths/100,000 Live Births)
1	Afghanistan	1,400
2	Chad	1,200
3	Somalia	1,200
4	Guinea-Bissau	1,000
22	Mauritania	550
35	South Africa	410
38	Nepal	380
42	Bangladesh	340
43	Comoros	340
51	Pakistan	260
54	Burma	240
55	Indonesia	240
56	India	230
58	Bhutan	200
94	Brazil	58
95	Vietnam	56
99	Georgia	48

In Indian context, the maternal death, being a rare event, requires large sample size to provide robust estimates, particularly for study area region.

It is seen that Maternal Mortality Rate viz. maternal deaths to women in the ages 15-49 per lakh of women in that age group, and the life time risk have been presented. The life time risk is defined as the probability that at least one women of reproductive age(15-49) will die due to child birth or puerperium assuming that chance of death is uniformly distributed across the entire reproductive span. This has been calculated by the following formula: [2]

$$\mathbf{LifeTimeRisk = 1 - (1 - (MMR / 10000))^{35}} \qquad (1)$$

Though there has been immense reduction in below-five mortality in compared to neonatal mortality. Still the challenges of the medical experts are of both on operational and technical aspects. Spanning health system deficiencies are the issues which actually relate the development and implementation of the right policies and strategies to scale up neonatal development programmes. Thus for a good knowledge-base decision support system, one must requires such a system which will guide the experts as when necessary.

Newborn health continues to be a sensitive indicator of national development. Hence, there is a need of adopting and focusing on the special kinds of strategies. Evidence-based diagnosis has been reflecting best practices for the huge reduction of maternal morbidity and mortality in world. Table 2.2. shows the maternal mortality rate of India. Depending upon this data, figure 2.1. reflects that there have been enormous changes on the number of mortality rate.

Table 2.2. Maternal Mortality Ratio in India [17]

Year	**MMR/1000 Live Births**
1999-2001	327
2001-2003	301
2004-2006	254
2007-2009	212

Source: Special bulletin on maternal mortality, office of the Registrar General, India, June, 2011.

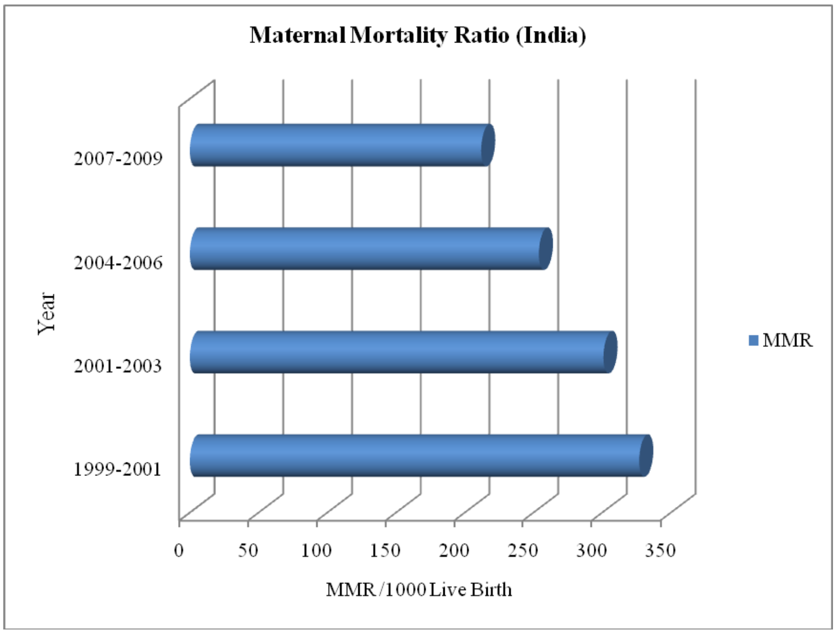

Figure 2.1. Maternal Mortality Rate

Source : Table 2.2.

2.3. Neonatal Mortality: A Point of Discussion

For the growth of population and socio-economic development of any region, neonatal survival is one of the most important indicators. Neonatal mortality is related with the deaths occurring during the first four weeks after birth.

Even we cannot think that newborn is suffering with disease, still information reveals that, a large numbers of children dies soon after birth. Many of them die in the first four weeks of life, i.e. neonatal deaths, and most of those during the first week i.e. early neonatal deaths. For every baby who dies in the first week after birth, is called neonatal death or is born dead. Causes and determinants of neonatal deaths and stillbirths are differs from the causes of post-neonatal and child deaths.

2.3.1. Neonatal Mortality: Global Context

There is slow declining in neonatal mortality than the under five mortality rate in all most all regions. Globally, neonatal mortality rate has declined 28 percent from 32 deaths per 1,000 live births in 1990. Which are 23 percent deaths per 1,000 live births in the year 2010. The declining rate is about an average of 1.7 percent a year, much slower than for under-five mortality (2.2 percent per year) and for maternal mortality (2.3 percent per year). The fastest reduction was in Northern Africa (55 percent), followed by Eastern Asia and Latin America and the Caribbean (52 percent); the slowest reduction was in Oceania and Sub-Saharan Africa (19 percent).

Table 2.3. NMR, Neonatal Deaths and Under Five Deaths by MDG Region

Region	Neonatal mortality rate (deaths per 1,000 live births)			Number of neonatal deaths (thousands)		Neonatal deaths as a share of under-five deaths (percent)		
	1990	2010	Decline (percent) 1990–2010	1990	2010	1990	2010	Relative increase (percent) 1990–2010
Developed regions	7	4	43	106	53	47	53	15
Developing regions	36	25	31	4,319	3,019	37	40	10
Northern Africa	29	13	55	107	46	35	49	37
Sub-Saharan Africa	43	35	19	969	1,123	26	30	17
Latin America and the Caribbean	23	11	52	265	117	42	47	11
Caucasus and Central Asia	30	21	30	58	34	37	44	18
Eastern Asia	23	11	52	589	189	45	57	27
Excluding China	12	9	25	14	8	47	48	1
Southern Asia	48	32	33	1,875	1,256	41	50	20
Excluding India	48	33	31	576	381	40	46	15
South-eastern Asia	28	15	46	335	169	39	48	23
Western Asia	28	16	43	116	79	43	48	12
Oceania	26	21	19	5	5	37	40	7
World	**32**	**23**	**28**	**4,425**	**3,072**	**37**	**40**	**9**

Source: Levels & Trends in Child Mortality", Report 2011 by Unicef.

In Southern Asia, neonatal deaths account for 50 percent of under-five deaths, and almost 30 percent of global neonatal deaths occurred in India. Sub-Saharan Africa, which accounts for more than a third of global neonatal deaths, has the highest neonatal mortality rate (35 deaths per 1,000 live births in 2010) [4].

2.3.2. Neonatal Mortality: Indian Context

The survival of the neonates is one of the important indicators among the population for their growth and socio-economic development. There is stagnancy of the neonatal mortality rate during the year 2003 to 2006, it was 37/1000 live births. This was little came down to 36/1000 live births in 2007, 35/1000 live births in 2008. A significant change noticed during the year 2009, the NMR come down to 34/1000 live births.

Out of all the mortality cases, maximum mortality rate is there in the rural area. It is 38/1000 live births. But in the urban area it is 21/1000 live births. The neonatal mortality rate also varies significantly among all Indian States. Higher neonatal mortality rate as compared to national average, has been recorded, Himachal Pradesh (36), Haryana(35), Gujarat(34), J&K (37), Chhattisgarh(38), Rajasthan (41), Orissa (43),Uttar Pradesh (45), Madhya Pradesh (47). Kerala is having the lowest neonatal mortality rate that is (7). Table 2.4. shows the mortality indicators in major states of India, A significant feature is that, the neonatal mortality rate came down or remained stagnant in 2009 as compared to 2008 except in the case of Haryana, Jharkhand, Himachal Pradesh, and Karnataka[5] .
This neonatal mortality again can be classified into two major types:

- Perinatal Mortality Rate, and
- Post neonatal Mortality Rate

Perinatal Mortality Rate:

Perinatal Mortality Rate actually deals with number of still birth and deaths during first seven days of delivery per 1000 live births. Perinatal Mortality Rate is high in rural areas (39) as compared to urban areas (23) during 2009. This rate varies in the range of 37 to 35 since 2001. In 2009 this mortality rate is found 35 per 1000 live births. Kerala is found to be a best performing state in terms of perinatal mortality ration having only 13. The other states like Madhya Pradesh and Chhattisgarh (45) are least performing States during 2009.

Post neonatal Mortality Rate:

Post neonatal Mortality Rate deals with the number of infant deaths at 28 days to 365 days of age per 1000 live births. The Post Neo natal Mortality Rate is also came down. In 2002 it was 24, compared to 2009 it was 16. The rural area of India is very much affected by this Post neonatal Mortality Rate. It is high and nearly 17 in rural areas (17) as compared to urban areas (13).

Table 2.4. Mortality Indicator, India and Major States, 2004-2009

Sl. No.	India/Major States	Neo-Natal Mortality Rates						Early Neo-Natal Mortality Rates					
		2004	2005	2006	2007	2008	2009	2004	2005	2006	2007	2008	2009
1	2	3	4	5	6	7	8	9	10	11	12	13	14
	INDIA	**37**	**37**	**37**	**36**	**35**	**34**	**26**	**28**	**28**	**29**	**27**	**27**
1	Andhra Pradesh	36	35	33	33	34	33	23	26	26	26	26	27
2	Assam	35	33	35	34	34	33	24	25	26	28	27	29
3	Bihar	33	32	32	31	32	31	23	28	28	27	27	26
4	Chhatisgarh	43	45	43	41	39	38	37	36	36	36	36	33
5	Delhi	20	20	22	20	19	18	16	16	18	16	15	16
6	Gujarat	37	36	38	37	37	34	24	28	27	29	30	26
7	Haryana	31	35	34	34	34	35	17	24	22	23	24	23
8	Himachal Pr.	31	33	30	31	33	36	21	19	20	19	27	31
9	J & K	38	36	39	39	39	37	23	29	30	31	33	32
10	Jharkhand	26	28	29	28	25	28	19	22	22	24	25	24
11	Karnataka	25	28	28	26	24	25	21	23	20	20	20	19
12	Kerala	9	11	10	7	7	7	8	9	8	6	5	5
13	Madhya Pradesh	50	51	51	49	48	47	33	38	40	38	38	37
14	Maharastra	26	25	27	25	24	24	19	20	21	21	19	20
15	Orissa	49	53	52	49	47	43	36	41	38	37	34	35
16	Punjab	30	30	30	29	28	27	20	18	17	20	20	16
17	Rajasthan	42	43	45	44	43	41	32	33	33	34	33	33
18	Tamil Nadu	29	26	24	23	21	18	21	19	18	17	15	13
19	Uttar Pradesh	50	45	46	48	45	45	32	32	35	36	33	35
20	West Bengal	29	30	28	28	26	25	20	23	20	23	21	19

Source: Sample Registration System - Registrar General, India, 2011

2.3.3. Infant Mortality Rate (IMR) – India

The Infant Mortality Rate entry gives the number of deaths of infants under one year old in a given year per 1,000 live births in the same year. This includes the total death rate, and deaths by sex, male and female. Counting IMR rate is used as an indicator of the level of health in a country and also for socio-economic growth and development of the country. Table 2.5. shows the data about Infant Mortality in India during the year 2000 to 2012. There is reduction of the rate of IMR during this tenure, which is positive sign. Still special care should be taken for reducing this rate more.

Table 2.5. Infant Mortality Rate in India

Year	IMR(Deaths/10000 Live Births)
2000	64.9
2001	63.19
2002	61.47
2003	59.59
2004	57.92
2005	56.29
2006	54.63
2007	34.61
2008	32.31
2009	30.15
2010	49.13
2011	47.57
2012	46.07

Source: CIA World Factbook – July, 2012.

According to SRS (Sample Registration System) 2009, the IMR at national level was 50 per 1000 live births in 2009 as compared to 53 in 2008. The IMR is higher in respect of Female (52) as compared to Male (49). The highest infant mortality rate has been reported from Madhya Pradesh (67) and lowest from Kerala (12). Assam (61), Bihar (52), Chhattisgarh (54), Haryana (51), Madhya Pradesh (67), Orissa (65), Rajasthan (59) and Uttar Pradesh (63) recorded higher IMR as compared to the national average [5].

The IMR is very high in rural areas (55 per 1000 live births) as compared to urban areas (34). Rural areas of Madhya Pradesh registered the highest IMR (72) followed by Orissa (68), Uttar Pradesh (66). Rural areas of Kerala State recorded the Lowest IMR (12) in the country. Uttar Pradesh and Chhattisgarh recorded highest IMR in urban areas. Kerala had the lowest IMR (11) in urban areas. Amongst the smaller states, Rural and Urban areas of Goa recorded lowest IMR during 2009. The increase in medical attention to the pregnant women at the time of live births may have resulted

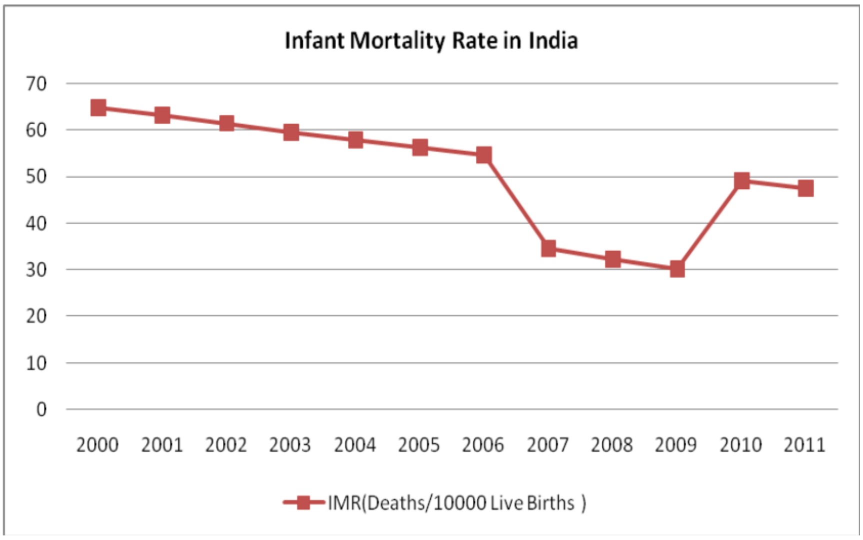

Figure 2.2. Infant Mortality Rate in India (2000-2011).

Source: CIA World Factbook – July, 2012.

in decline in IMR over the period. But in the rural areas, the medical attention is still on the lower side. Table 2.6. shows the Infant Mortality Rate by residence in all states and union territories of India. This comprises a total infant mortality rate along with rural area and urban area mortality rate of infant during the year 2005 to 2009.

2.4. Reasons of Neonatal and Infant Mortality

2.4.1 Statistical Overview

The survival rate of female infants correlates to subsequent population replacement. For these reasons, the issue of neonatal deaths is a serious national health concern, especially in developing countries where 96% of the world’s approximate 5 million annual neonatal deaths occur. It is seen that in India more than one million newborns die before completing their first moth. This is more lass about 30% of the world's neonatal deaths. India’s current neonatal mortality rate of 44 per 1000 live births represents 1.2 million children who die each year. Neonatal mortality is higher in rural areas at 49 per 1000 live births (vs 27/1000 in urban areas). The neonatal mortality rate also varies significantly among Indian states. Orissa and Madhya Pradesh have the highest neonatal mortality rates of 61 (rural 63, urban 42) and 59 (rural 63, urban 40) per 1000 live births, respectively. In Uttar Pradesh the rate is 53/1000 (rural 56, urban 39), and 31/1000 (rural 33, urban 21) in West Bengal. Kerala has the lowest neonatal mortality of 10/1000((rural 10, urban 9), followed by Punjab 29/1000 (rural 32, urban 19). Targeting a reduction in infant mortality, the Government of India is committed to reducing the neonatal mortality rate to approximately 20/1000 live births by 2014. However, this aim requires the consideration of many contributing elements. Maternal factors that contribute to neonatal mortality have their origin long before the baby is born. This necessitates the collation of information on maternal health and women's status throughout women's early (high risk of neonatal mortality among mothers less than 20 years of age) and reproductive life. This is actually difficult to obtain. Two-thirds of infant deaths in India occur in the first month of life out of which one-third of all neonatal deaths associated with low birth weight. Approximately three-quarters of Indian neonatal deaths occur within one week of birth while 90% occur within the first two weeks of life.

A major challenge in Indian rural areas is that most births take place at home, assisted by untrained personnel. Risks are also associated with the use of traditional birth attendants and the environment into which a child is born also influences survival. It is, therefore, important to know about risk factors such as the smoking of parents and unsafe water and sanitation, as well as specific neonatal disease and injury factors, such as diarrhea, neonatal sepsis, pneumonia or neonatal jaundice.

Table 2.6. Infant Mortality Rate by Residence - All States.

Sl. No.	India/Major States	Infant Mortality Rate														
		Total					Rural					Urban				
		2005	2006	2007	2008	2009	2005	2006	2007	2008	2009	2005	2006	2007	2008	2009
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
	INDIA	**58**	**57**	**55**	**53**	**50**	**64**	**62**	**61**	**58**	**55**	**40**	**39**	**37**	**36**	**34**
1.	Andhra Pradesh	57	56	54	52	49	63	62	60	58	54	39	38	37	36	35
2.	Assam	68	67	66	64	61	71	70	68	66	64	39	42	41	39	37
3.	Bihar	61	60	58	56	52	62	62	59	57	53	47	45	44	42	40
4.	Chhatisgarh	63	61	59	57	54	65	62	61	59	55	52	50	49	48	47
5.	Delhi	35	37	36	35	33	44	42	41	40	40	33	36	35	34	31
6.	Gujarat	54	53	52	50	48	63	62	60	58	55	37	37	36	35	33
7.	Haryana	60	57	55	54	51	64	62	60	58	54	45	45	44	43	41
8.	Himachal Pradesh	49	50	47	44	45	50	52	49	45	46	20	26	25	27	28
9.	J & K	50	52	51	49	45	53	54	53	51	48	39	38	38	37	34
10.	Jharkhand	50	49	48	46	44	53	52	51	49	46	33	32	31	32	30
11	Karnataka	50	48	47	45	41	54	53	52	50	47	39	36	35	33	31
17.	Rajasthan	68	67	65	63	59	75	74	72	69	65	43	41	40	38	35
18.	Tamil Nadu	37	37	35	31	28	39	39	38	34	30	34	33	31	28	26
19.	Uttar Pradesh	73	71	69	67	63	77	75	72	70	66	54	53	51	49	47
20.	West Bengal	38	38	37	35	33	40	40	39	37	34	31	29	29	29	27
Smaller States/UTs																
21	Arunachal Pr.	37	40	37	32	32	39	44	41	34	35	17	19	15	19	14
22	Goa	16	15	13	10	11	16	14	11	10	11	15	16	13	11	10
23	Manipur	13	11	12	14	16	12	11	13	16	18	14	11	9	8	11
24	Meghalaya	49	53	56	58	59	50	54	57	60	61	42	43	46	43	40
25	Mizoram	20	25	23	37	36	26	32	27	45	45	10	13	16	24	19
26	Nagaland	18	20	21	26	26	17	18	18	25	27	22	27	29	28	23
27	Sikkim	30	33	34	33	34	31	35	36	35	36	15	16	20	19	21
28	Tripura	31	36	39	34	31	31	37	40	36	33	29	30	32	26	20
29	Uttarakhand	42	43	48	44	41	56	54	52	48	44	19	22	25	24	27
30	A&N Islands	27	31	34	31	27	30	35	38	35	31	18	21	23	23	20
31	Chandigarh	19	23	27	28	25	25	23	25	22	25	18	23	28	29	25
32	D&N Haveli	42	35	34	34	37	45	38	38	38	41	29	24	18	20	24
33	Daman & Diu	28	28	27	31	24	32	33	29	29	21	21	18	23	36	30
34	Lakshadweep	22	25	24	31	25	17	19	25	28	22	27	31	23	35	28
35	Puducherry	28	28	25	25	22	38	35	31	31	28	22	24	22	22	19

Source: Sample Ragistration System, Registrar General, India.

Finally, reports of neonatal deaths in India are often included indices of infant and child mortality, making it difficult to define the scope of the problem and design effective strategies. We should start thinking more sophisticated and useful strategies which can be identify the incidence of, and influences on, neonatal deaths in India, with an explicit focus on defining the problem in rural areas and remote places [6].

As per the online article published by Jagran Josh in the current affair section [7], “Trends in Maternal Mortality” jointly released by WHO, Unicef, UNFPA and World Bank on 15 Sept 2010, “India's maternal mortality rate (MMR) is the highest though the country managed to pull down maternal mortality rate (MMR) by 59% between 1990 and 2008. India’s MMR that stood at 570 in 1990 fell progressively in the following years to 470 per 100,000 live births in 1995,390 in 2000,280 in 2005 and 230 in 2008. India’s MMR currently stands at 63000 maternal deaths a year and since 1990 the country witnessed an annual decrease in MMR by 4.9%”. However, the report pointed out that the annual rate of fall in MMR is less than half needed to achieve the MDG target of reducing the MMR by 75% between 1990 and 2015. To achieve the MDG target an annual decline of 5.5% would be required. According to the report, the number of maternal deaths in Asia showed 52% decline. The adult lifetime risk of maternal death is highest in sub-Saharan Africa (1 in 31), followed by Oceania (1 in 110) and South Asia (1 in 120). The developed regions recorded the smallest lifetime risk (1 in 4,300).

2.4.2 Still Birth in India

According to the Stillbirth series published in the British medical journal, The Lancet, India has topped the list of 10 countries that have the highest number of stillbirths.

As high as 66 per cent (1.8 million) stillbirths in the world occur in just 10 countries namely- India followed by Pakistan, Nigeria, China, Bangladesh, Congo, Ethiopia, Indonesia, Afghanistan and Tanzania.

In India, the stillbirth rates varied from 20 to 66 per 1000 total births in different States. Almost half of all stillbirths, 1.2 million, happen when the woman is in labour. These deaths occur primarily due to lack of skilled care at this critical time for mothers and babies. Two-thirds of stillbirth occurs in rural areas, where skilled birth attendants, midwives and physicians in particular are not always available for essential care during childbirth and for obstetric emergencies, including Caesarean section. About 2.6 million stillbirths occurred worldwide in 2009. As per Lancet more than 7200 babies are stillborn every day. 98 per cent of these occur in low and middle-income countries. These deaths occur mainly during the last trimester of pregnancy (after 28 weeks' gestation). Approximately 1.2 million stillbirths occur during birth (intrapartum) and 1.4 million before birth (antepartum). Most intrapartum stillbirths are associated with obstetric emergencies (childbirth complications).

2.4.3. Reasons of Still Birth

In high-income countries, obesity, smoking, and advanced maternal age are among the big risk factors that tends to result in stillbirths. Childbirth complications, maternal infections in pregnancy, maternal disorders, especially hypertension and diabetes, fetal growth restriction, and congenital abnormalities were named as the biggest reasons for stillbirths. The overall number of stillbirths fell from an estimated 3.03 million in 1995 to 2.64 million in 2009.The global rate was reduced from 22 stillbirths per 1000 total births to 19. Stillbirths had decreased by 1.1 per cent per year since 1995, lower than the 2.3 per cent annual reduction rate in child under-five mortality, and 2.5 per cent annual reduction in maternal mortality.

2.4.4. Reasons

If we look globally, the leading causes of neonatal deaths are infections, asphyxia and prematurity. A large proportion of neonatal mortality is contributed by infections, a largely preventable cause. The incidence of neonatal tetanus, formerly a major cause of mortality, has declined dramatically since 1980s. The risk of infant death at the individual level is determined by considering the survival status of live births in India. Besides this, there are more causes of neonatal deaths occurs, they are. maternal age, age at effective marriage of wife, number of live births, interval between last two births, mother's literacy status, prenatal care taken by mother, type of delivery (full term or premature) and birth weight of the baby are considered which may be having some impact on infant mortality.

About 2·35 million children died in India in 2005. This child death contributes more than 20% of all deaths in children younger than 5 years globally. The main reasons for these deaths were from five causes: prematurity, pneumonia, and extremely low birth weight, and birth trauma, neonatal infections, diarrheal diseases, and birth asphyxia etc.

This causes of neonatal deaths can be prevented or treated and reduce the death rates. For this, skilled worker and domain experts are very much needed. Besides known, highly effective and widely practicable interventions such as skilled attendance during delivery, emergency obstetric care, and simple immediate care for newborn babies [8]. Apart from the above causes, deaths happen due to rural emergencies, selective abortion of girls, especially for pregnancies after a firstborn girl, has increased substantially in India. Figure 2.3 shows the reasons for infant mortality. On the figure 2.4, information regarding various causes for neonatal deaths is presented with the percentile distribution of each category in a chart. Below are the few descriptions or the reasons for neonatal mortality [9]:

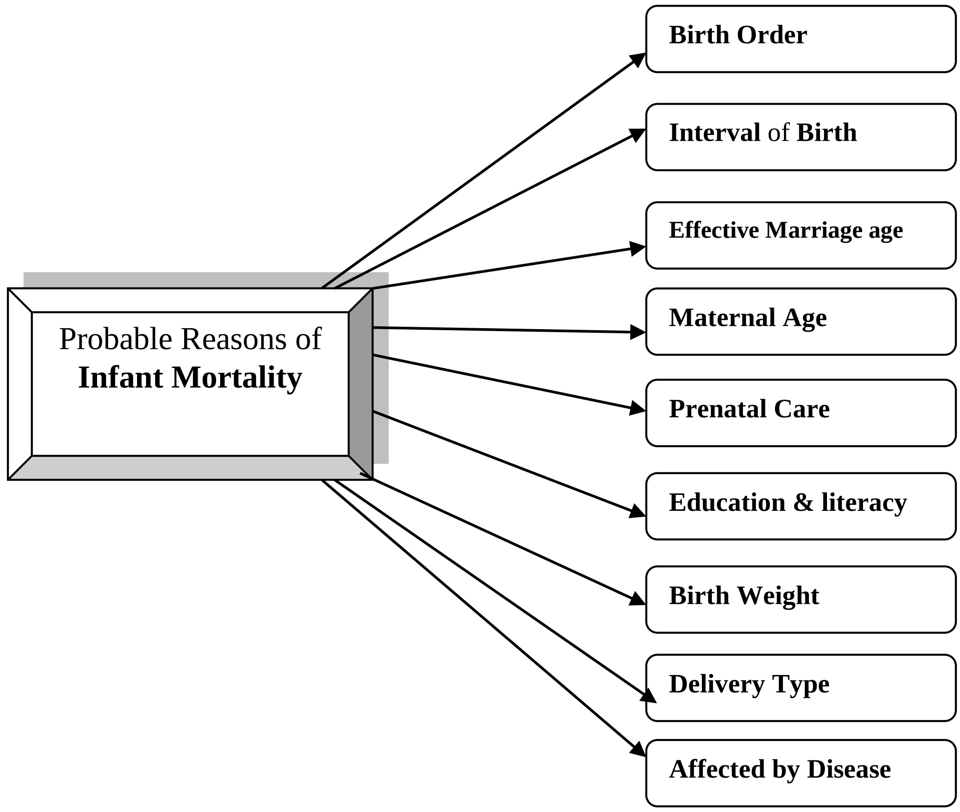

Figure 2.3. Reasons of Infant Mortality.

This causes of neonatal deaths can be prevented or treated and reduce the death rates. For this, skilled worker and domain experts are very much needed. Besides known, highly effective and widely practicable interventions such as skilled attendance during delivery, emergency obstetric care, and simple immediate care for newborn babies**.** Apart from the above causes, deaths happen due to rural emergencies, selective abortion of girls, especially for pregnancies after a firstborn girl, has increased substantially in India. Below are the descriptions or the reasons for neonatal mortality:

- **Birth Order**

The possibility of infant death again rises with the increase in parity. It is seen that the first birth had a much higher risk of dying than 2-3 births. The risk of infant death among the births of order 2-3 was low. This cause seems to have an independent affect on the risk of infant mortality. The percentage turn down in the level of infant mortality from birth order 1 to 2-3 as well percentage boost in the level of infant mortality from birth order 2-3 to 4 and over was much sharpen.

- **Interval of Birth**

Shorter birth interval, likely below eighteen months, between two births of child is also a potential problem of neonatal death. If a gap of more than 18 months is not maintained between the two pregnancies the risk of infant death becomes high.

- **Effective Marriage Age**

Another reason is the effective marriage age. The mother seems to be associated with the risk of infant mortality inversely i.e. lower the age at effective marriage, higher would be the risk of infant mortality. So this lower age marriage tends to neonatal mortality much faster.

- **Maternal Age**

Infants of teen aged mothers had higher risk of dying in India, compared to the infants of mothers in the age group of 20 to 29, specifically this happens in my study area also. Even births to mothers in the older age group, 30 and over, also had a higher risk of death. This is why the possibility of infant death seems to be higher when the mother is either very young or relatively old, more than 30.

- **Prenatal Care**

Prenatal care during pregnancy in terms of taking Tetanus Tosid doses and other prenatal care suggested by the doctor must be taken care of. Inertia of this leads to neonatal mortality. So it is the most important dominant of infant mortality.

- **Education & Literacy**

Less literacy level of any mother leads to poor survival status of neonates. It is seen that less literate or illiterate mother faced relatively high risk than comparatively literate mother during the birth of a child. To maintain the proper way of giving birth of a child, education related to hygiene, care, proper diet and purity are the other indicators, mother must be learned. And all this indicators are related with good education.

- **Birth Weight**

Low birth weight and extremely low birth weight of the babies whose nutritional intake is low from mother, is another causes of infant deaths. It was observed that among those babies who have a birth weight of less than 2400 grams have a much higher risk of death than the babies having birth weight to be more than 2400 grams. Thus, it is an important indicator of infant death.

- **Delivery Type**

Type of delivery is a biological factor and may be difficult to control unless proper medical care is taken by the pregnant mothers during the course of their pregnancies.

Even premature deliveries are likely to have very high risk of infant death as compared to the full-term deliveries. Thus it is recommended always for fulltime deliveries.

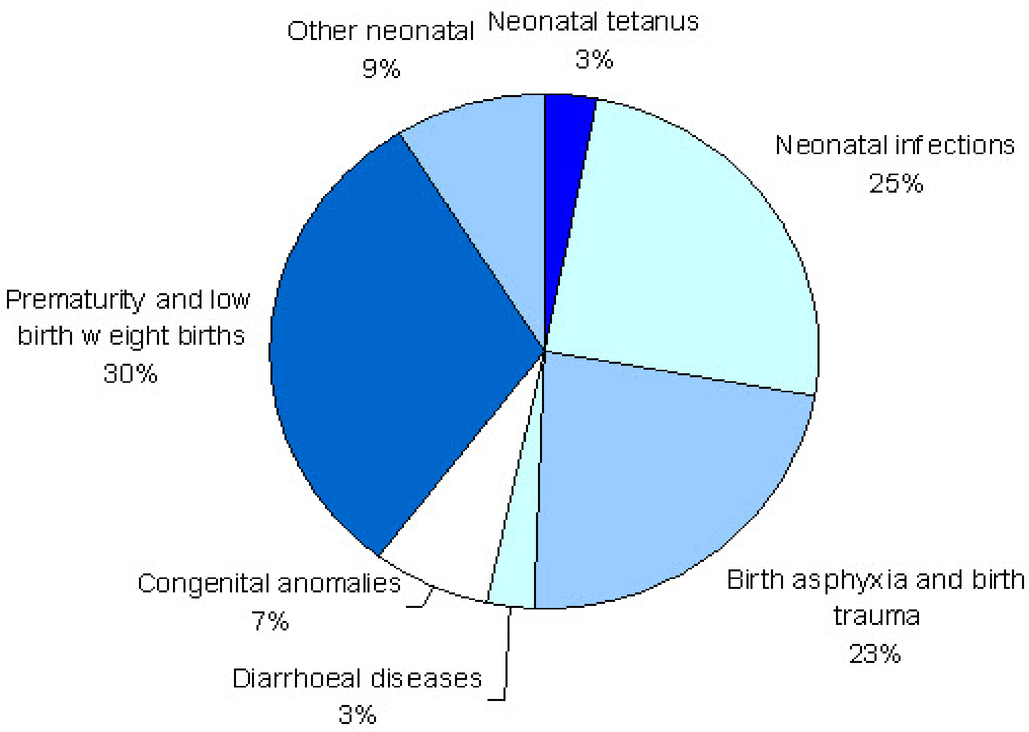

Figure 2.4. Causes of Neonatal Deaths.

Source: WHO, The Global Burden of Disease: 2004 update, WHO, Geneva, 2008.

2.5. Issues Related to Present Health Care System: Overcome Strategies

There are several issues which are related to the present poor health care system. It is high time to mitigate these problems. Some of the issues are described below [10]:

- **Health Care Systems Improvement**

There is still high mortality and morbidity rate in rural and urban areas as per the existing data. Factors that affect causes the same has to be planned with ultimate intervention. In order to plan programs and policies, there is need to collect data on maternal and neonatal health. Improvement on health care system in present scenario has to be taken with high preference, particularly in neonatal domain.

- **Improvement of Literacy of Female**

Due to poor literacy level of mother, survival status is poor of the neonates. It is seen that less literate or illiterate mother faced relatively high risk than comparatively literate mother during the birth of a child. In Kerala, female literacy coupled with woman empowerment on right to health, have been demonstrated to show resounding success in lowering neonatal mortality rate to 10/1000 live births [11].

- **Home Based Care Versus Hospital Care**

Rural and remote areas are facing problems of medical care. For the birth of a child safely, people completely depends on attendants and auxiliary nurse midwives. That is completely home base and does not all the time gets overall medical facilities. Even they use traditional ways of delivery which causes maximum risk. Over two thirds of deaths take place in the first week of life as a result of perinatal asphyxia and sepsis. These problems can be best tackled by technically skilled personnel in hospitals. People are afraid of going to the hospital, because they have the thinking that in hospital there is lack of awareness for antenatal and neonatal care, attitude and behaviour of the staff is worst and huge cost of hospitalization. For increasing the awareness and the issues discussed, need to be tackled if hospital deliveries have to be promoted. Reducing the neonatal deaths one has to promote awareness for institutional deliveries.

- **Referral System Setting**

Adequate utilization of primary and secondary level hospitals and prevent overcrowding in tertiary hospitals. In order to facilitate maximum and effective utilization of health services in rural or urban areas, it is necessary to set up a definite system of referral. There is also a need to create linkages between domiciliary, health centre and hospital level. Protocols for admissions to primary, secondary and tertiary levels must be laid down.

- **Quality of Perinatal Care Improvement**

There is an absolute need of improvements in knowledge and technical skills of the staff that are associated with health care system. Emphasis must also be laid on improving behavior and attitude of the health personnel at all levels. Health institutions must be made "Mother and Child Friendly". This would require training both in technical and communication skills. Mothers and babies should be given free service as per possible resources available.

- **Health Workers Motivation**

Involvement of variety health workers is there, for example workers are form Government, Municipal Corporation, NGOs etc. are involved in neonatal health care.

Early detection of neonatal disease can prevent neonatal mortality, for that workers must be given specialized training. This has shown reduction in neonatal mortality in the rural community.

- **Community Participation**

There is need to understand about the mortality of the neonate and infants, the causes of disease, how to prevent it, care seeking practice. Thus a participation of the community people along with mothers and parents of child has to be ensured. If we understand how families behave and why they do so then we could focus our interventions to improve them. Inability to recognize serious illness has been cited as a main cause for late medical advice. Mother and family are among the key players in reduction of neonatal mortality and improvement in neonatal health status.

- **Community Based Medical Practitioners**

Majority of the peoples living in rural area and remote areas visits general practitioners for medical treatment and advice. Most of the rural practitioners are not qualified and may be of different streams [12]. The practitioner could be of different systems of medicine and include non-qualified practitioners. These care providers are excluded from training under national programs. Training of all practitioners with respect to diagnosis, early management and referral of newborns should be considered.

- **Services Integration for Neonatal Domain**

To get the effective management and treatment of large group of neonates, integration of public, private and NGOs with existing health care system there should be proper distribution of jobs for each group of people to avoid redundancy in job. Importance must be given on community participation, education, and training.

These invisible disasters need to be addressed to ensure India’s progress towards Millennium Development Goal (MDG) 4. Spending on health care in India also remains low (only 4.2 % of GDP), in comparison to many countries, which will also have a direct impact on child mortality [13].

2.6. Rule of Thumb – A Heuristic Concept

In artificial intelligence, Heuristics is the branch that uses the rational rules drawn from experience and derived knowledge to solve problems relating to any means. “Self-learning and get better with experience” is the characteristics of Heuristic programming. Due to these unpredictable characteristics heuristic programming does not give the better output, still normally an optimal result expects from this. The term ‘heuristic’ was initially introduced by the Greeks, its original form was heuriskein. This actually means “discover”.

A heuristic-decision making model may be defined as any 'rule of thumb'. This helps enabling fast decision making in practical domains where lack of information available and also regarding the decision making process which is unstructured and complex in nature. The heuristics can be classified in two broad categories:

(a) Descriptive Heuristics and
(b) Prescriptive Heuristics

(a) Descriptive heuristics describes a situation or context. This is not endowing with any guidance on how to resolve it.

(b) Prescriptive heuristics provide guidance about what to do about a specific situation.

Traditionally, heuristic methods have been used to solve complex problem types, such as frequently occur in finance, other business areas, and particularly to great extent of medicine. In the arena of medical decision-support systems or expert system, the term 'heuristic systems' is often relates with artificial intelligence systems in medicine, which would employ an informal model to solve complex problem in decision making. In the field of medical decision making, most physicians every day deals with various patients having problems in terms of sign and symptoms. Out of all the information provided by patients not always very specific and related. It sometimes contains unspecific and unstructured information. Thus, they always think more than one possible diagnosis, after screening and meeting a patient. They really develop their own guesswork from that unstructured and incomplete little information. Actually most of the physicians work with this heuristics, but they may be biases in nature and leads to uncertainty [14]. To make diagnoses fast, most of domain experts rely on this kind of shortcuts and rules of thumb.

When doctors or any domain expert make errors and when their thinking is excessively influenced by true information; they fail to consider possibilities that contradict their mental pattern of a disease. Thus attribute symptoms to the wrong cause, bias may occurs. In case of availability bias, there is a tendency to judge the likelihood of an event. There is also a tendency to resolve it with similar problems occurred previously. An affective error actually helps the process of decision making in a better way if there is a chance of uncertainty of diagnosing disease.

2.6.1. Uncertain and Incomplete Nature of Medical Knowledge

Uncertainty in medical domain is a big matter of thinking. This can't be ruled out completely. Generally medical knowledge deals with complex and huge domain. Due to this uncertain and incomplete information decision may be vague. To deal with this kind of problem, there were several attempts taken towards developing medical decision making with the help of heuristics. Optimal situational awareness occurs when physicians find the appropriate fit between cognitive level and task complexity- this leads to good decision-making process [15]. Table 2.7 describes the cognitive range of decision-making runs from informal/intuition at one end to calculation/analytical at the other and the nature of tasks runs from simple to complex. Here the system propose a procedure of matching the appropriate cognitive activity to the particular task [16].

In field of medical reasoning, uncertainty is the most critical and central fact of concern. It is seen that many traditional mathematics techniques and algorithm approaches have been giving wrong results and poor output, when it has applied on medical problems. Still medical domain is a methodical and experimental scientific discipline for research and study. Uncertainty caused by the incompleteness of information has been modeled by non-numerical characterizations. Medical Expert Systems developers and researchers have mainly focused on 'uncertainty in narrower sense' applying method of probability and evidence theory to reason under uncertainty. A different approach of uncertainty modeling, similar kinds of identity are matched from the overall evidence and then accumulate them for based on certainty factors (CF). On the following chapters, there has been detail discussion about how CF helps in the perspective of decision making in neonatal medical domain. This certainty factors or the confidence factors actually helps to build the knowledgebase. The method of applying certainty factors have been discuss on chapter 4**.**

Table 2.7. Intuitive and Analytical Approaches in Decision

Characteristics	Intuitive	Analytical
Cognitive style	Heuristic	Systematic
Cognitive awareness	Low	High
Conscious control	Low	High
Automaticity	High	Low
Rate	Fast	Slow
Reliability	Low	High
Errors	Normative distribution	Few but large
Compliance	High for answer Low for method	Low for answer High for method
Effort	Low	High
Predictive power	Low	High
Emotional valence	High	Low
Detail on judgment process	Low	High
Scientific rigidity	Low	High

2.7. Needs of Alternate Thinking

This has been great challenge in rural areas of India, that most of the births are takes place at home, assisted by untrained and unskilled personnel. This includes high risks associated with the use of old traditional birth attendants. The impact of environment where a child is born also influences survival of the neonate. Above all, scarcity of immediate assistance and experience experts in rural areas are the vital cause of infant mortality. It is, therefore, important to know about risk factors such as the smoking of parents, unsafe water and sanitation, specific neonatal disease and injury factors, such as diarrhea, neonatal sepsis, pneumonia or neonatal jaundice etc.

There is a vital need of domain specialist, who can help in real time. But the scenario is just in darker side. In rural areas no specialist doctors or experts are willing to go for treatment, even there is absolute scarcity of expert's assistance also.

The cause's mentioned above leads to consider for alternate thinking. If there is unavailability of domain experts, an automated thinking termed as Expert System (ES) or Decision Support System (DSS) might be useful. The DSS or ES must be having the capability of selecting proper and accurate information about patient, capable of diagnosing the disease, assist for management and also giving treatment plan as well.

As the increasing pressure of patient on the medical institution, huge data about the various symptoms and disease are being processed every day. Out of that data, few data are irrelevant, vogue and unrelated. Thus there is absolutely necessity of proper classification of data and also a proper way mining the data. On the next few chapters discussion about classification and data mining approaches have been reflected.

2.8. Conclusion

As per the various reports, that we have discussed so far, reducing the neonatal problem, worst health care condition, high mortality rate and other factors related to neonates is not all satisfactory. There are several risk factors related with of neonatal mortality of the study region. Few of them pregnancy phases, birth of child, post natal period and more has to be taken care effectively.

After surveying all the factors related with neonatal health, conclusions reveals that there is economic disparity and various belief and behavior of different community people leads to slowing down the neonatal health related problems. There are few factors incorporates declining neonatal health related problems. Few factors are: Low income group people, diverse Indian cultures-belief-behaviors, unhealthy antenatal care, insecure postnatal care, unhealthy food habits, poor hygienic environments etc. Moreover, there is acute scarcity of domain experts, i.e. neonatologists, pediatricians or specified doctors and nursing staffs, which leads to great problem in Indian neonatal health care context.

There have been numerous awareness programs and campaigns related with the neonatal health care runs by the government and other community leaders. Still for improving this worst condition, family members has to be taken responsibility for neonatal care by eliminating old and wrong perceptions. They should be taken new approaches for neonatal health education in the society.

People of the society should aware of the ground reality of the scarcity of neonatal experts. People from every group and every community should at least contribute their knowledge and merits which may be helpful for those limited numbers of experts. In connection of this scenario, being a researcher of a computer science, our contribution to the society would be the proposed development of decision support system for neonatal disease diagnosis. This may be a useful helping hand to the doctors as well as domain related persons for their early and mostly accurate decision making process.

References

1. World Health Organization, *"Maternal Mortality"*, Media Centre, Fact Sheet No.348, November 2010.

2. Registrar General, India, *"Special Bulletin on Maternal Mortality in India 2004-06"*, Ministry of Home Affairs, Govt. of India, 2009.

3. *"Country Comparison-Maternal Mortality Rate"*, CIA World Fact book - January 1, 2011.

4. Danzhen You, Gareth Jones and Tessa Wardlaw, "*Levels & Trends in Child Mortality*", UNICEF, Report 2011.

5. "*Family Welfare Statistics in India–2011*", Statistics Division Ministry of Health and Family Welfare Government of India.

6. Kumar D, Verma A, Sehgal VK. *"Neonatal Mortality in India. Rural and Remote Health"*, 7: 833. (Online) 2007. Available: http://www.rrh.org.au, [Last accessed on 12 November, 2011].

7. Current Affairs-Jagron Josh, *"Trends in Maternal Mortality- A Report*", Report Survey, News Capsule, September 2010 Current Affairs, Jagron Josh, published on: 10th October, 2010.

8. *"Million Death Study Collaborators. Causes of Neonatal and Child Mortality in India: A Nationally Representative Mortality Survey"*. Lancet. 2010; 376:1853-60.

9. Gandotra M.M., Das N.P.,"*Determinates and causes of infant mortality in Gujrat andMaharashtra*"**,** Available: http://prcs-mohfw.nic.in/writereaddata/research/421.htm [Last accessed on 12th November, 2011]

10. Fernandez Armida, Mondkar Jayshree, Mathai Sheila, "*Urban Slum-Specific Issues in Neonatal Survival*", Indian Pediatrics; Vol. 40, pp. 1161-1166, 2003.

11. Government of India, Ministry of Health and Family Welfare, Child Health Division, New Delhi, 2000.

12. Bhandari N, Bahl R, Taneja S, Martines J, Polian MK."*Pathways to Infant Mortality in Urban Slums of Delhi, India: Implications for Improving the Quality of Community and Hospital-based Programs*", J Health Popul Nutr; Vol. 20, pp. 148-155, 2002.

13. World Health Organization. "*India Statistics 2009*". Available: http://www.who.int/countries/ind/en/ [Last accessed on 11th November, 2011].

14. Jerome Groopman. "*Medical Dispatches: What's the Trouble? How Doctors Think?*"Available:http://www.newyorker.com/reporting/2007/01/29/070129fa_fact_groopman [Last accessed on 19/01/2012].

15. Croskerry Pat. "*The Theory and Practice of Clinical Decision-Making*", Canadian Journal of Anesthesia, Vol. 52, No 6, pp. R1-R8, 2005.

16. Dawson NV. "*Physician Judgment in Clinical Settings: Methodological Influences and Cognitive Performance*", Clin. Chem., Vol. 39, No. 7, pp. 1468–1480, 1993.

17. Special Bulletin on Maternal Mortality, Office of the Registrar General, India, June, 2011.

CHAPTER 3

New Born Status in Terai Region of West Bengal - A Study†

3.1. Introduction

In India the infant mortality rate is 67.6 per thousand live births and neonatal mortality rate is 43.4. Thus it is obvious that almost 2/3 total death during infancy occurs in first four weeks [1]. Many changes have taken place now; this has been discussed in chapter 2. Due to various national programs like Diarrhea Control Program, ARI Control Program, Immunization Program, India can reduce post neonatal death. The aim should be reduction of neonatal death, many of which are preventable if proper care is taken during perinatal period.

The main objective of the present study is to find clinical profile and outcome of neonates in Terai region of West Bengal. The aim is also to highlight the importance and the problems in neonatal care of said region. In this study 5278 neonates were considered. Out of this number 2862 were male and 2416 were female babies. The incidence of LBW was found to be 32.35%. The perinatal asphyxia was found to be higher than other Indian studies. The neonatal mortality rate was found out to be 39.5 per thousand live births. The prime causes of neonatal death were found as Perinatal Asphyxia, Septicemia, and LBW with complications. Most of these mortalities are preventable with proper recitation, rational antimicrobial therapy and aggressive support care.

The newborn period encompasses the first four weeks of extra uterine life and is an important like in the chain of conception to adulthood. The correct management of events in the perinatal period will lead to qualitative outcome without any mental and physical disability. Infant Mortality Rate (IMR) is a sensitive indicator of socio economic development of any country. IMR has got two distinct components:

a) Infant Mortality

b) Post Neonatal Mortality

† This chapter is based on the publication made by the author entitled “A Study on the Status of New Born in Terai Region of West Bengal”, on Advances in Modeling C, Vol. 68, Issue 1, A.M.S.E. France, No 05 225(2c), pp. 44-52, 2007.

a) Infant Mortality

Deaths during first four weeks of life is called infant mortality rate. Basically, Infant mortality rate is the number of infants dying before reaching one year of age, per 1,000 live births in a given year. To find IMR of any given region, we need to get the total number of newborns dying under one year of age and then that has to dived by the total number of live births during the year. Finally result has to multiplied by 1,000 to get the rate of infant mortality.

$$\text{Infant Mortality Rate (IMR)} = \frac{\text{Number of Infant Deaths during time period}}{\text{Number of Live Births during time period}} \times 1{,}000$$

For example, the data has given as below:

Number infant deaths in India 2012 = 60

Number of live births in India 2012 = 8,537

Constant = 1,000

Time period = 2012

So, the IMR is,

$$\text{IMR of India 2012} = \frac{\text{Number of Infant Deaths in India in 2012}}{\text{Number of Live Births in India in 2012}} \times 1{,}000$$

$$= (60 / 8537) \times 1{,}000$$

$$= 7.02 / 1000 \text{ live births}$$

b) Post Neonatal Mortality

Post Natal Mortality is the deaths after four weeks up to 12 months of age. As per World Health Organization definitions postnatal mortality as the "number of stillbirths and deaths in the first week of life per 1,000 live births", though there are other definitions also been used. This has already been discussed in Chapter 2.

In India the average birth weight at term is 2.8 kg [2] which is much less than developed countries and nearly 13 of neonates born in India are low birth weight (LBW) [3]. According to WHO, LBW has been defined as babies born less than 2500gm irrespective of gestational age.

The studies done in other parts of the country have shown that common causes of Neonatal deaths in India are Septicemia(52%), Birth Asphyxia(20%), premature(15%), and congenital malfunction(3-5%) [1]. LBW is the single most important determinant of neonatal deaths. Studies have shown that over 70-90% deaths occur among LBW infants [1].

3.2. Objective of the Study

The objective of this particular study is to survey the clinical profile and outcome of neonates in Terai area of West Bengal and is also to highlight the importance and the problems in neonatal care of the said region.

This area actually refers to the North Bengal consisting: Darjeeling, Jalpaiguri, Tarai and Dooars region, Coach Behar. Besides, it includes Seven Sister states, Sikkim also. This area is ethnically different from the other parts of India. The said area has got strong ethnic and cultural ties with East Asia and Southeast Asia. The region has several Scheduled Caste and Tribes and other sub-tribal communities and groups. It is predominantly rural with over 84 per cent of the population living in the countryside. In spite of having unique scenic beauty, area is known for Tea, Tourism & Table Tennis, culture, handicrafts. But the dark side of this is that, area is suffering from unemployment, be deficient of infrastructures and mostly lacking of education. The health care system is not sufficient enough to provide all the necessary medication properly.

Mostly, in village areas, there has been acute problem of neonatal prevalent diseases. The people come for the treatment from the village areas practically in Govt. hospitals or tertiary centres. This causes a huge overload for the practitioners and related staff members, even for the hospitals also. Study actually focusing the status of new born and statistical analysis.

We have concentrated mostly on the North Bengal Tara and Dooars region while knowledge accusation process started.

3.3. Methodology

This prospective observational study was conducted in a tertiary care hospital based at Siliguri Terai area of West Bengal. All new born babies who were delivered in this institution during the year 2004 (1^{st} January – 31^{st} December), were included in the study. Data regarding birth weight, gestational age, delivery status, detail examination report of new born babies, antenatal care, maternal morbidity and diagnosis, relevant investigations, duration of stay and outcome were recorded on a preset pro-forma. Every day the data has been collected from Pediatric Nursery and Labour Room ward of North Bengal Medical College and Hospital, West Bengal. The babies who had complication at or after birth during neonatal period were also followed up. Proper weights of the newborn were recorded on Detecto type of weigh machine. The length of the baby was measured with infant-meter. Diagnosis, relevant investigation and treatment plan during the staying period were also been recorded on the pre-set form.

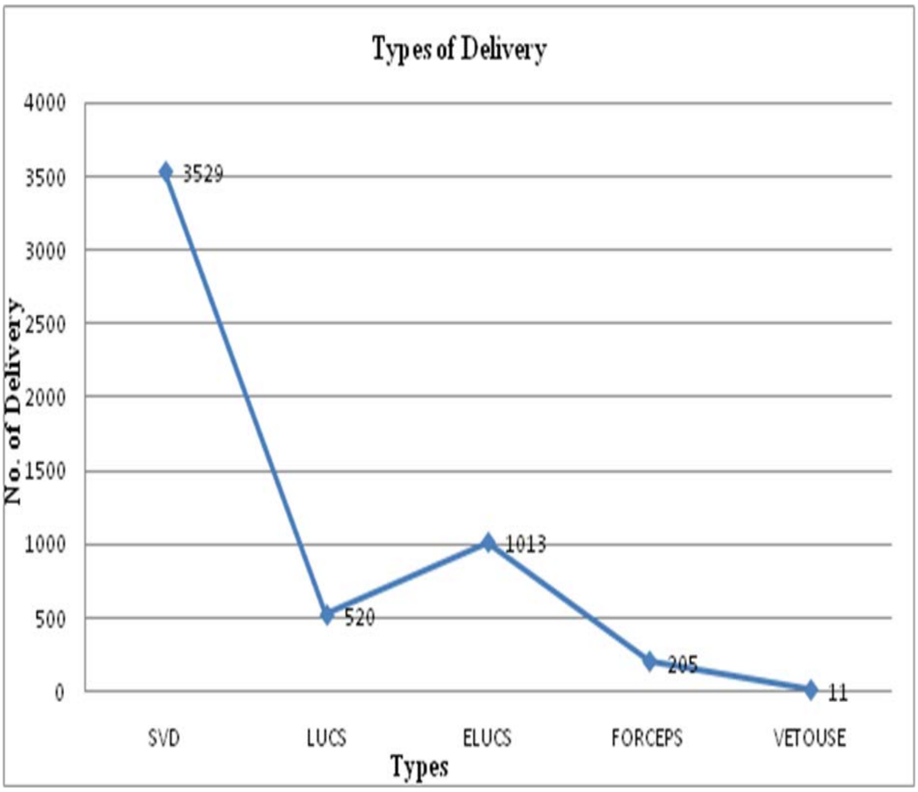

Figure 3.1. Types of Delivery.

After collecting the records of each and every newborn during the study period, analysis was done. The neonatal data was analyzed using appropriate statistical method. Male-Female ratio was calculated from the total data as number of female births per thousand male births [Table 3.1]. Out of total 5278 data, incidence of Low Birth Weight were calculated and categorized among different communities of the region [Table 3.2].

Table 3.1. Birth Rate of Male & Female Babies.

	Total	Male	Female
New born during year 2004	5278	2862	2416
Female births per thousand Male births	844.45		

Table 3.2. Incidence of LBW in Different Community.

Community	Bengali	Rajbangsi	Nepali	Muslim	Tribal
Total New Born	1615	2071	435	575	582
Low Birth Weight	555	609	95	224	223
Percentage of LBW	34.36	29.4	21.84	38.95	38.31

The type of delivery that had been recorded were categorized in five different status like, SVD(spontaneous vaginal delivery), LUCS, ELUCS(Emergency Cessarian Section), Forceps delivery and Ventouse extraction[Table 3.3, Figure 3.1]. The incidence of neonatal death with commonest cause was also found out [Table 3.4].

To find out the incidence of perinatal asphyxia, APGAR score at 1 min. and 5 min. was calculated and was categorized. APGAR score < 4 , between 4 to 7 and above 7. This has been showed on Table 3.5.

Table 3.3. Types of Delivery.

Types of Delivery	SVD	LUCS	ELUCS	FORCEPS	VETOUSE
No. of Delivery	3529	520	1013	205	11

Table 3.4. Causes of Neonatal Death.

Cause of Death	No of Death
Hypoxic Ischemic Encephalopathy / BA	115
Septicemia	29
Preterm with Metabolic Complication	10
Gross Prematurity	7
Congenital Heart Disease	2
Aspiration Pneumonia	6
Hyperbilirubinemia	2
RDS	1
Blood Dyscrasia	1
Meconium Aspiration Syndrome	1

Table 3.5. APGAR Score.

Score	At 1 Min.	At 5 Min.
0	9	1
1	252	107
2	240	74
3	35	11
4	461	270
5	76	18
6	22	80
7	4160	473
8	8	4197

3.4. Results

A total of 5278 babies were delivered during the study period. Out of which female and male babies were 2416 and 2862 respectively. Thus male-female birth ratio is 54% vs. 46%. The overall survival rate was 96.05%. 169(3.95%) neonates died. Birth asphyxia in LBW babies was found to be the commonest cause of neonatal mortality amongst intramural admissions. Death due to Hypoxic Ischemic Encephalopathy / BA is 115/169, Septicemia 21/169, pre term with metabolic complication17/169, Aspiration Pneumonia is 6/169, Congenital Anomaly and Congenital Heart disease is 4/169 and 2/169 respectively. The common morbidities amongst survivors were birth asphyxia, septicemia, hyperbilirubinemia and prematurity. It is found that percentage of Low Birth Weight in this region is 32.35%. The incidence in different community was found to be different and was highest among Muslim (38.95%) and minimum among Nepali (21.83%) population.

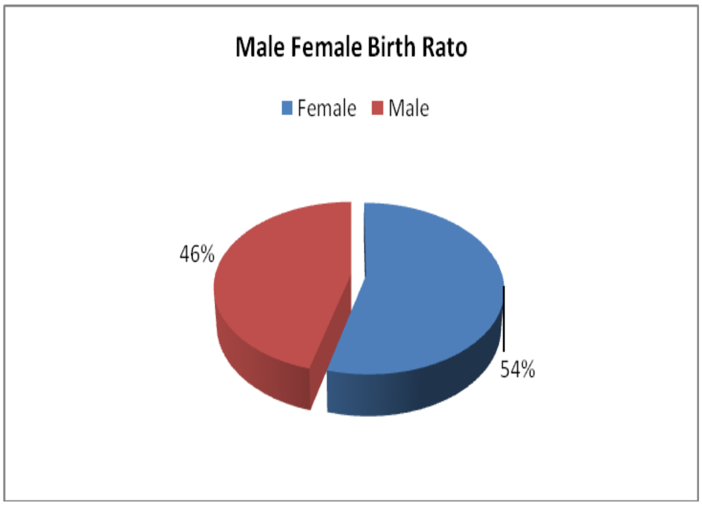

Figure 3.2. Male Female Birth Ratio.

Figure 3.3. shows that the APGAR score below 4 at 1 min. is 10.16% and 5 min. is 3.66% between 4 to 7 at 1 min. and at 5 min is 10.59% and 6.97% respectively. Most of the admitted mothers are found to be Rajbanshi (2071) and Bengali (1615) belonging to poor socioeconomic status.

3.5. Discussion

Current national neonatal mortality rate is 43.4 per thousand live births. There are state to state variations in the country. All most 50% of the neonatal deaths occur within one week and majority within first 24 hours[3]. In our study we find that the incidence is 39.5 per thousand. It is a tertiary care government hospital where usually the complicated pregnancies are referred from the primary and secondary care hospitals.

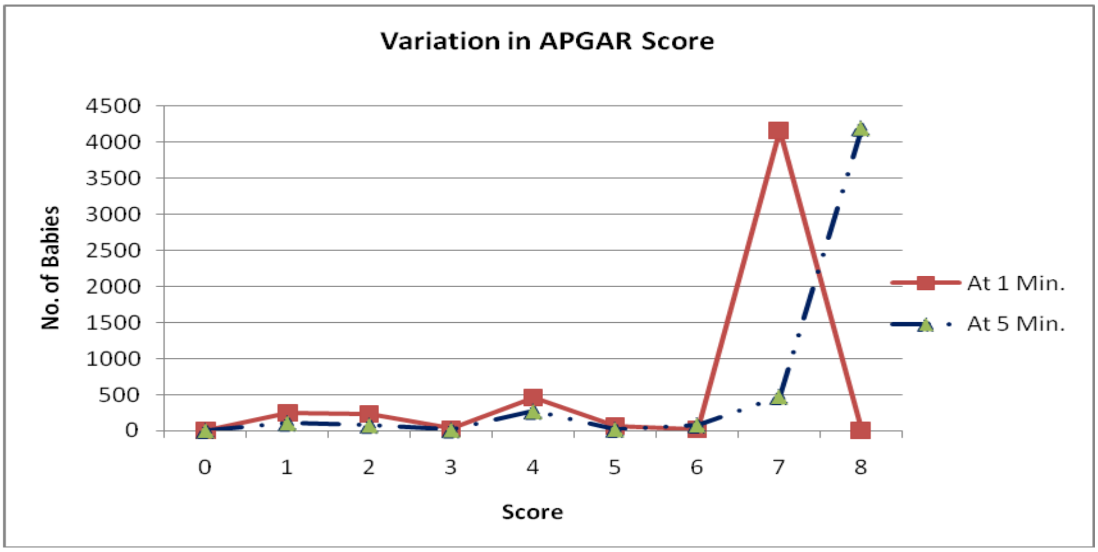

Figure 3.3. Variation in APGAR Score.

The socio economic statuses of majority of the patients are low with poor antenatal care. So the incidence of emergency caesarian section is high (19.19%). Available Indian literature shows sepsis is the commonest cause of mortality and is high probably responsible for 30% to 50 % of total neonatal deaths each year in developing countries [4]. It is estimated that 20% of all neonates develop sepsis and approximately 1% die of sepsis related causes [5]. In our study we find that the most common causes of death is HIE (115/169) and then neonatal septicemia (29/169).

In an another India study [6] the incidence of neonatal sepsis is 38 per thousand intramural live births in tertiary care institution, the incidence of neonatal sepsis is 29 per thousand live births.

According to estimation by WHO approximately 4 million babies die before they reach one month [7] and 98% of these deaths take place in developing countries [7]. Perinatal asphyxia and birth injuries together contribute to almost 29% of this death [7]. WHO[8] has defined that failure to initiate and sustain breathing immediately after

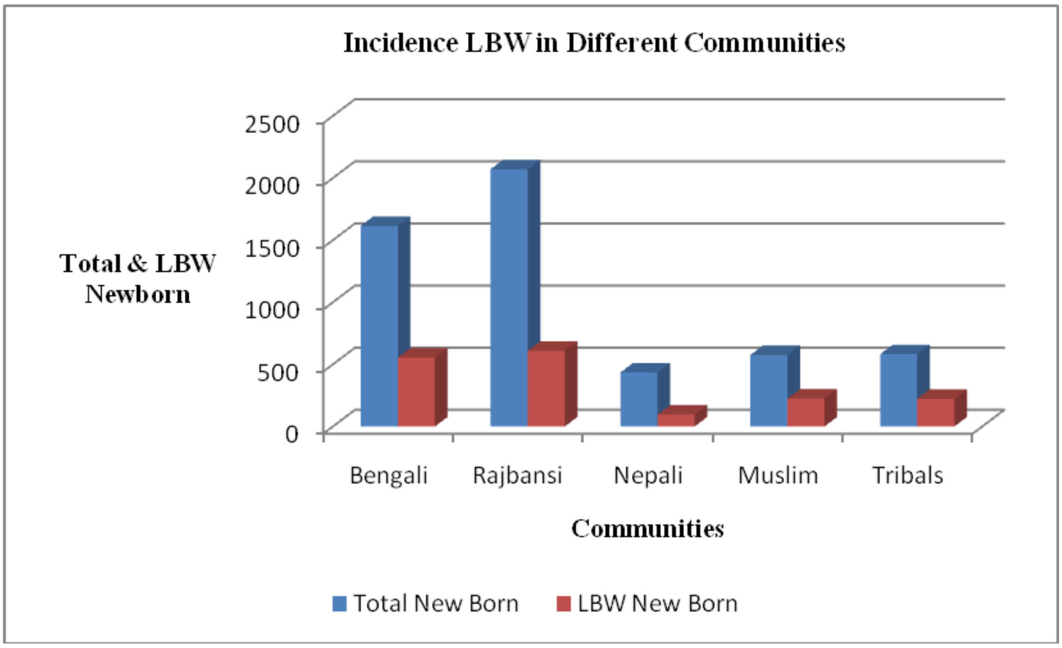

Figure 3.4. Incidence of LBW Babies Among Different Communities.

delivery is perinatal asphyxia and it has been associated with hypoxic ischemic injury to the central nervous system and the clinical manifestations of these injuries is termed as hypoxic ischemic encephalopathy(HIE), which is concern because it can lead to serious long term neuromotor disabilities among survivors.

According to NNPD 2000 data [6] APGAR score below 7 at 1 min. which include moderate and severe asphyxia were documented 9% of all intramural delivery. 2.5% babies continued to have APGAR score less than 7 at 5 min.[8]. In our study we can see that 10.16% of all babies have APGAR score below 4 at 1 min. and 3.66% at 5 min., 10.59% have APGAR between 3 to 6 at 1 min. and 6.97% at 5 min[Table 3.5]. Out of all the babies who died, 68.05% were due to HIE and birth asphyxia. The common reason why there is high incidence of BA and HIE is the late referral, poor antenatal care and low socio economic condition of the patients who attend this medical centre.

The incidence of Low Birth Weight was also found to be high in this region. Figure 3.4. shows the incidence of low birth weight (LBW) among different communities of the region. A study shows nearly1/3 of the newborn in India are low birth weight [weight < 2500 gm. At birth]. The newborn can be low birth weight because of two reasons, firstly the infant may be pre term being delivered before 37 completed weeks, i.e. gestational age, and the second cause is the intrauterine growth restrictions which may be both in term and pre term babies. 2/3 of low birth weight neonates born in India belong to this category [9]. In our study we find that incidence of low birth weights is 32.35%. The incidence is highest among Muslim (38.95%) and local tribes(38%)[Figure. 3.5]. It is relatively less among Nepali population (21.83%). In

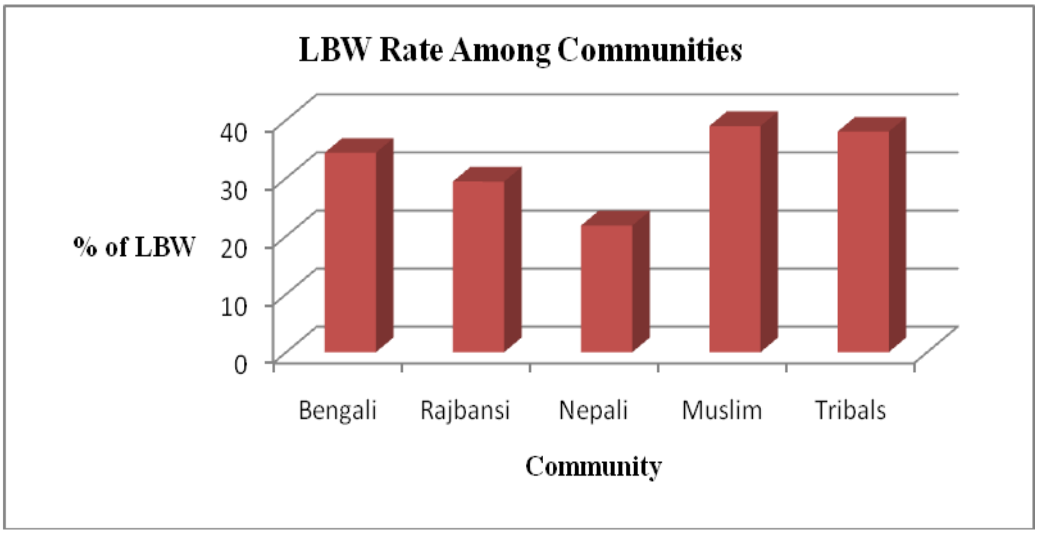

Figure 3.5. LBW Rate Among Different Communities.

Bengali and Rajbanshi the incidence is 34.36% and 29.4% respectively. The reason is low female literacy and health awareness among the Muslims and tribes of this region.

The incidence of postnatal birth asphyxia and septicemia is more among LBW babies. Out of 115 asphyxiated babies and 29 babies with neonatal septicemia, who expired, 72(63.60%) and 21(72.41%) respectively were LBW.

Prenatal and postnatal management of R_h disease has vastly declined the frequency of severe neonatal jaundice and kernicterus since 1960[10]. Higher bilirubin level may be toxic to the developing brain and may cause neurological impairment. Over 60% of newborn develop jaundice by 48-72 hours out of which 5% to 10 % need medical care [11]. The total serum bilirubin level more than 95^{th} percentile in age in hours has been used to describe severe hyper bilirubin for the 120 hours of postnatal age. In our study incidence of death due to hyper bilirubinis is 2 out of 169.

The aspiration pneumonia was cause of death in ELBW babies. Congenital anomalies not compatible with life were heart diseases, neurological malformation and dismorphic babies. A great concern was the birth rate of girl child, which is gradually declining [12] and this is 844.45 female per thousand male births in this area. Further research is warranted to find out causes of this decline.

3.6. Conclusions

This study was conducted to find the status of newborn in Terai Region of West Bengal, India. The alarming result regarding the neonatal mortality rate which come out to be 39.5 per thousand live births. The outcome of this study dictates the need of more neonatal care of the said region. Government of India is taking necessary steps in this regard.

There is message from the government that, it has been decided to have a name-based tracking whereby pregnant women and children can be tracked for their ANCs and immunisation along with a feedback system for the ANM, ASHA etc to ensure that all pregnant women receive their Ante-Natal Care (ANCs) and post-natal care (PNCs) Checkups; and the children receive their full immunization [13][14]. All new pregnancies detected or being registered from 1st April, 2010 at the first point of contact of the pregnant mother are being captured as also all births occurring from 1st December, 2009. A number of States have established the system and other are putting in place systems to capture such information on a regular basis. Mother and Child Tracking System require intense capacity building at various levels primarily at the Block and Sub-Centre levels. The National Informatics Centre (NIC) has developed software application. The roll-out is being monitored centrally [15].

References

1. Gupte Suraj, *"The Short Text Book of Pediatrics"*, 10th Edition, Jaypee Publication, New Delhi, pp. 593, 2003.

2. Ghai O.P, "*Essential Pediatrics*", 6th Edition, CBS Publishers and Distributors, New Delhi, pp. 136, 2004.

3. Deorai K Ashok, Aggarwal Rajiv and Paul K Vinod., "*Management of Infant with Intra-Uterine Growth Restriction*", Protocols of Neonatalogy, Indian J. Pediatrics, Vol. 68, No. 12, pp. 1155-1158, 2001.

4. Bang A.T., Bang R.A., Bactule S.B., Reddy H.N., Deshmukh M.D., "*Effects of Home Based Neonatal Care and Management of Sepsis in Neonatal Mortality: Field trail in Rural India*", Lancet, Vol. 354, pp. 1955-1961, 1999.

5. Stoll B.J, "*The Global Impact of Neonatal Infection*", Clinical Perinatol, Vol. 24 pp.1-21, 1997.

6. Aggarwal Rajiv, Sarkar Nupur, Deorari K Ashok. and Paul K Vinod, "*Sepsis in the New Born*", Protocols of Neonatalogy, India J. Pediatrics, New Delhi, India pp. 52, 2005.

7. Rajiv Aggarwal, Ashok K. Deorari and Vinod K. Paul, "*Post-Resuscitation Management of Asphyxiated Neonates*", Indian J. Pediatrics, Vol. 68 No.12, pp. 1149-1154, 2001.

8. World Health Organization, "*Perinatal Mortality: A Listing of Available Information*", FRH/MSM.96.7.Geneva: WHO, 1996.

9. Central Bureau of Health Intelligence, *Health Information of India*, Ministry of Health and Family Welfare, Government of India, New Delhi, 1998-99.

10. Bhutani Vinod K. and Johnson Lois H., "*Newborn Jaundice and Kernicterus - Health and Social Perspectives*", Indian J. Pediatrics, Vol. 70b No.5, pp. 407-410, 2003.

11. Aggarwal Ramesh, Aggarwal Rajiv, Deorari K Ashok. and Paul K Vinod, "*Jaundice in the Newborn*", Indian J. Pediatrics, Vol. 68 No. 10, pp. 977-980, 2001.

12. Chatterjee M., Samanta, R.K. Saha A.K. and Majumder B, "*Gender Issue in Health Care of Children*" Proc. 23rd International Congress of Pediatrics and 2nd International Congress of Pediatric Nursing, Beijing, China, 2001.

13. "*Family Welfare Statics in India 2011*", Statistics Division Ministry of Health and Family Welfare Government of India.

14. "*Operational Plan for Mother and Child Tracking System*", Ministry of Health & Family Welfare, 2011.

15. "*Annual Report 2010-2012", *Dept. of Health & Family Welfare, 2011. Ministry of Health & Family Welfare, Govt. of India.

C H A P T E R 4

A DSS for Neonatal Prevalent Disease Diagnosis & Management using Expert System Technology†

4.1. Introduction

A child is the potential of a country, especially, the neonates constitutes a large population group, and also vulnerable or special risk group. The risk is related with growth, development, disease pattern and survival. Thus by improving health status of neonates, proper diagnosis of the neonatal diseases is essential. Scarcity of the neonate experts in rural and remote areas causes the higher mortality rate of neonates [1]. To alleviate the lack of human proficiency and assisting the existing experts for their decision making capacity, an expert system for prevalent disease diagnosis and management would be very useful.

Decision Support Systems for disease diagnosis have become most essential component of now days. It is seen that utilization of these kinds of systems are not very satisfactory in every respect. There is still a tendency to avoid taking help of such kinds of decision support system service by the domain related personnel. This leads to less improvement of medical care to the sufferer.

Despite of having applications of knowledge base and rule base kinds of technology for the development of decision support system, many other techniques have also been used rigorously now. The techniques like Artificial Intelligence, Data Mining, Rough Set Computing, Artificial Neural Network, MLP, Genetic Algorithm and many others algorithms have been using efficiently in diagnosis system development. For an example for finding the patterns in the neonatal knowledge base, neural networks may be the very useful tool.

Among the various systems in the field of diagnostic expert system our system is also working efficiently diagnosing neonatal disease and also giving valuable suggestions in terms of management. The system generates the list of diseases by looking at each sign and symptom, and then match with lab findings to determine how strongly each of the physical findings is associated with the disease under consideration of the neonatal disease.

---

† This chapter is based on the publication made by the author entitled "A Decision Support System for Prevalent Disease Diagnose & Management for Neonates using Expert System Technology", Proc. National Conf. on Computing and System 2010, NACCS 2010, Vol. 1, pp 7-12, January 2010.

This chapter presents a rule-based object oriented expert system for prevalent disease diagnosis and management for neonates**.** It shows good performance as apparent from its performance evaluation.

A child is the future of a nation and every nation should take proper care for each and every child. An estimated two-thirds of childhood deaths occur in infancy, and, in turn, two-thirds of infant deaths occur in the first month of life. In addition to these four million neonatal deaths, primarily due to serious bacterial infections, birth asphyxia, and complications of pre-maturity and intrauterine growth restriction, an estimated 3.9 million pregnancies end in stillbirth[2] These deaths, at least half, could have been avoided if proper interventions were taken up. Research suggests that integration of maternal and neonatal health care services is the key to the health benefits for mothers and their newborns. This is particularly happens if a given intervention has potential for direct health benefits to mothers and newborns both. Majority of the neonates in developing countries are born and has been given care in rural homes. The available information is mostly hospital based [3]**.** Neonates not only constitute a large population group, but also vulnerable or special risk group. The risk is related with growth, development, disease pattern and survival. From the commonly accepted indices, it is evident that mortality rates in this age group are higher than adult population especially in developing countries.

Thus by improving health status of neonates, we contribute to the health of the general population. These considerations have led to the formulation of special health services for children all over the world. The study one chapter 3 shows that the incidence of the diseases namely, Neonatal Septicemia, HIE, Metabolic Disturbances, Neonatal jaundice etc. are quite high in the North eastern parts of India specially in Tarai region [4]**.** If neonates suffer from these diseases and proper management is not available in proper time the mortality and morbidity will increase considerably. To overcome the problems mentioned above proper treatment planning is required. And for proper treatment planning the basic requirement is Neonatologists. But the pediatric expert distribution in rural region is not at all satisfactory **[**5**].** To mitigate the scarcity of the domain expertise, an expert system for prevalent disease diagnosis and management for neonates may help considerably.

Expert System technology has been applied in many medical disciplines for new born babies whom we call neonates [6-12]. To the best of our knowledge, no such comprehensive work has been reported for neonates for such prevalent disease diagnosis and management.

4.2. A Probable Architecture of Expert System

4.2.1. Expert System

According to Britannica Concise Encyclopedia, "An expert system is a computer-based system which may act or respond like a human expert in a particular field. Expert systems are built on knowledge gathered from human experts, analogous to a database but containing rules that may be applied to solving a specific problem. An interface allows the user to specify symptoms and to clarify a problem by responding to questions posed by the system. Software tools are exists for helping designers build a special-purpose expert system with minimal effort. An outgrowth of work in artificial intelligence, expert systems show promise for an ever-widening range of applications. There are now widely used expert systems in the fields of medicine, personnel screening, and education."

In another views of McGraw-Hill Science & Technology Encyclopedia, "Expert system is the methods and techniques for constructing human-machine systems with specialized problem-solving expertise. The search of this area of artificial intelligence research has emphasized the knowledge that underlies human expertise and has simultaneously decreased the apparent significance of domain-independent problem-solving theory. In fact, new principles, tools, and techniques have emerged that form the basis of knowledge engineering."

Specifically, Expert system is one of the branches of the Artificial Intelligence. It is an AI program having expertise knowledge level about a particular domain and knows how to use its knowledge to respond properly and more accurately. Here "Domain" is the area or within a range which the task is being performed. Ideally the expert systems process is a substitute process of a human expert. A person without having domain knowledge or having less domain knowledge can solve a problem with the use of expert system. The source of knowledge may come from a human expert or from books, magazines and internet. Expert system sometime called knowledge base system as the in expert system knowledge is having ultimate importance.

4.2.2. Components of Expert System

An Expert system or the knowledge-base system basically contains the following contents and figure 4.2 depicts the Expert System Architecture.

- **Knowledge Base :**

  The knowledge base consists of specific knowledge about some specific domain. Knowledge base includes both implicit knowledge and explicit knowledge. Much of the knowledge in the knowledge base is not stated explicitly, but inferred by the inference engine from explicit statements in the knowledge base. This makes knowledge bases more efficient data storage and gives them the power to exhaustively represent all the knowledge implied by explicit statements of knowledge. Knowledge bases can have many different types of knowledge. Gathering the knowledge for the knowledge-base is called knowledge acquisition.

- **Knowledge-Acquisition Interface**

Knowledge-acquisition is the process of expressing knowledge in the knowledge-base. Knowledge engineer uses the knowledge accusation interface to interact with expert and the procedure of gathering the information for incorporating into knowledge-base. It also assists experts in conveying their knowledge in a form which is suitable for reasoning by the computer and even for the knowledge engineer. Knowledge accusation interface mostly capable of data entry, editing, creating rules, checking the syntax, debugging the errors and also validating the inputs. The interface is giving tremendous benefits to the expert system developers by means of generating specific user interface and on screen helps.

- **Inference Engine**

Inference engine is basically a program which infers new facts from known facts using inference rules. The inference engine uses general rules of inference to reason from the knowledge base and draw conclusions which are not explicitly stated but can be inferred from the knowledge base. Symbolic and mathematical both types of reasoning can be done through inference engines. Inference strategies may be varied from different engines because of varsatality in knowledge representation.

There are basically three stages of inference process which is carried out recursively; they are as follows [24]:

i) ***Match***

ii) ***Select***

iii) ***Execute***

In the ***Match*** stage of inference, working memory contents are compared to the facts and with the rules in knowledgebase. If the match is found then the equivalent rules

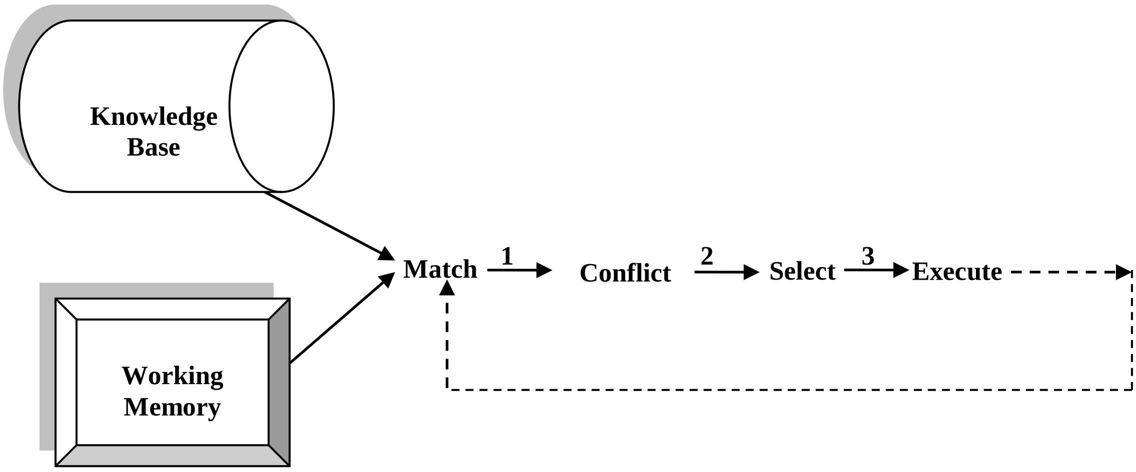

Figure 4.1. System Inference Cycle.

are place in conflict set. When all the matches are found, one of them is selected for the execution. The ***Selection*** procedure is depending upon the number of conjuncts is there in the left. Actually the smallest rule number is selected for execution. The selected rule is then E***xecuted*** with the right hand side rule is then ready for action process. Figure 4.1 shows the Cycle of System Inference.

- **User / Client Interface**

Unlike knowledge accusation interface, user interface also is a part of the program which communicates with the user. Users of the system frequently review a system by its interface. Functionality is not always verified or judged by the user. For avoiding unnecessary errors, user interface is one of the important parts of designing the program. Poor user interface design is the reason why so many software systems are never used. The user interface can be judged by how well it reproduces the kind of interaction one might expect between a human expert and someone consulting that expert. Most users of business systems interact with these systems through graphical user interfaces. The user interface is the part of the program that interacts with the user. It prompts the user for information required to solve a problem, displays conclusions, and explains its reasoning. Thus by designing easy user interface decision making process acts faster.

- **Translator (Rule Base)**

Translator basically used for translating the acquisitioned knowledge and make them prepare for generating rules. In fact, the process frames the rule base.

Converting the information into knowledge is the vital task for an expert system cell. Translator may of various categories depending upon rule specific transformation. For the interpretation process Abstract Data Type (AST) is used. Implementation of this AST makes interpretation easy and efficient by the Inference Engine. This can be used for generating complex and useful rules because of its expressiveness nature. A rule AST is converted into an equivalent form suitable for storage in the Knowledge-base.

- **Object / Class (Rule Base)**

  Expert System Cell may of different paradigm. In an object oriented approach the cell can generate specific Object and Classes using different categories. Parameters involved in producing class and thereby creating instances of the classes, which is called object.

4.2.3. Expert System Shell: Level 5 Object

The Expert System Shell is a component integrating for creating, editing, and executing rules on any problem-independent environment. Afterwards applying inference procedure, make the knowledge-base based on the rules. For developing Expert systems a variety of tools and languages can be applied. Examples of such languages are PROLOG and LISP. Developing the expert system using this kind of languages needs specialized programmer or developer and also it takes longer time to build any knowledge-base. Comparatively, Expert System Cell has a tendency to take less time. Because of having good developmental tools and easy interfaces for development comparatively less expertise person may also successfully use this.

For our expert system development and research purpose we use an Expert System Cell named Level 5 Object. It has capabilities like display and rule editors, debugging aids, and graphical and imaging capabilities or interfaces. LEVEL 5 Object is a product of Microsoft Windows (release 3.0 or higher) by the Information Builders Inc, USA. It has an integrated array of powerful tools like GUI development, Forms and Display builders and also has got capability to chain more than one knowledge base together. Besides this the Level 5 Object is having following facilities:

- Values editor to determine the status of object and attributes.
- Object, display, rule editors.
- Tool for monitoring the session.
- A database interface.
- Windows editor.
- An agenda editor (for specifying expert system goals).

- Knowledge tree editor and navigator.
- Import/export facilities.
- Rule Talk Editor.
- Reasoning Editor.
- Run and Execute facility.

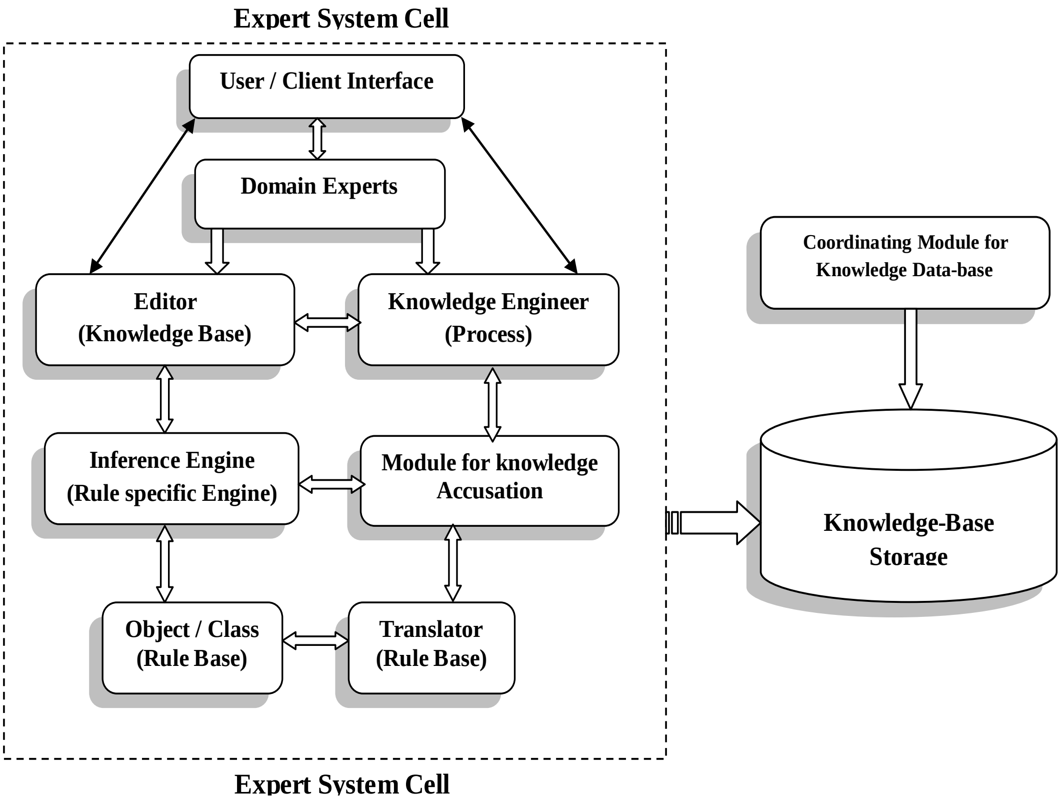

Figure 4.2. Knowledge Base Architecture.

As the name suggests, Level 5 Object, is the development tool based on an object-oriented design and implementation. Object oriented concept is now an emerging trend in developing expert system now a day. Features related to Object-oriented structure in Level 5 Object Cell are as follows:

- Objects can inherit the properties of other objects.
- Object attributes can have attachment that characterize their use and definition (e.g., search method, initial values, displays, queries, "when changed" procedures).

- User displays can be developed that contain objects of many different types and can allow the user to select, modify, or determine their value and status.
- Systems that are object oriented can be easily modified because it is easier to determine where and how knowledge is represented.
- Interfaces to different forms of knowledge are available.

There is huge difference between writing a program in OOPs concept and in conventional programming process. OOPs concept actually increase the productivity of the developer or the programmer.

One of the major advantages of using Level 5 Object was the ability to use different inference procedures. Both forward chaining and backward chaining inference can be possible here.

4.2.4. Forward-Chaining Inference

The forward-chaining inference engine starts with known facts or data and infers new facts about the situation based on the information contained in the application. The process continues until no further conclusions can be deduced from the initial data. Because the process is triggered by initial events, it is also referred to as data-driven or event-driven reasoning. It Involves the use of "when changed" or "when needed" procedures and demons.

Procedures and demons are algorithmic statements that direct the computer to process the knowledge in a previously specified manner. The use of "when changed" procedures involves a change in the value of an attribute which is triggered by the user selecting a response to a question and then clicking the mouse on the "enter" push button. The push button is attached to an attribute that is linked to a "when changed" procedure that contains a set of IF-THEN-ELSE procedural statements used to evaluate the user's response. Another use of forward chaining involves the attachment of displays to a push button. This in effect "advances" the execution of the expert system in a predetermined logical manner.

4.2.5. Backward-Chaining Inference

The backward-chaining inference engine starts with a desired goal or objective and proceeds backwards along a chain of reasoning in an attempt to gather the information needed to verify the goal. In LEVEL5 OBJECT, a chain of reasoning can consist of a series of IF - THEN statements called rules, procedures called methods, default values, and end-user queries.

The goal's search order defines the specific combination of these items and determines how and in what order the backward-chaining inference engine will attempt to verify the goal. It involves the execution of rules in search of a goal. The goal in this case is a conclusion or recommendation. The user supplies information about the structure or procedure and the system attempts to find a suitable goal.

This chapter presents a rule-based, object-oriented expert system for prevalent disease diagnosis and management of neonates. In section 4.3, the common causes of neonatal deaths have been pointed out. Even this has already been discussed in chapter 2. Section 4.4 explains the knowledge engineering process. Section 4.5 presents the system design and implementation. Section 4.6 describes the evaluation and performance of the expert system. In the last section that is section 4.7, we present our conclusion of the study.

4.3. Common Causes of Neonatal Deaths

In previous chapter, we have discussed about the common causes of neonatal death in global and national respect as well. Here again the causes that have been discussed in the research study. The neonatal health problems are mentioned in Table 4.1. The probable health problems of a sick neonate are summarized below.

Table 4.1. Neonatal Health Problems.

Neonatal Health Problems
Birth asphyxia, Neonatal Sepsis, Hypoxic-Ischemic Encephalopathy (HIE), Preterm, Low birth weight, Delayed breastfeeding, Problems in breastfeeding, Diarrhea, Hemorrhage, Conjunctivitis, Skin Infection, Abnormal Jaundice, Me conium Aspiration, Hyaline Membrane Disease (HMD), Pneumonia, Upper Respiratory Infection (URI), Hypothermia, Umbilical Sepsis, Tetanus, Convulsive Disorder, Unexplained fever, Failure to Gain Weight.

4.4. Knowledge Engineering

Knowledge Engineering is the part of expert system development. Basically, Knowledge Engineering refers to a process of building, maintain and developing of knowledge based expert system [13]**.** Integrating the domain knowledge in computer system for solving complex problem is the main task of Knowledge Engineering. To engineer the knowledge a high quality of human expertise is absolutely required [14].

4.4.1 Process of Knowledge Engineering

There are three processes that basically used in knowledge engineering process for the development of any expert system. Fig. 4.1 Shows the KE activities or processes:

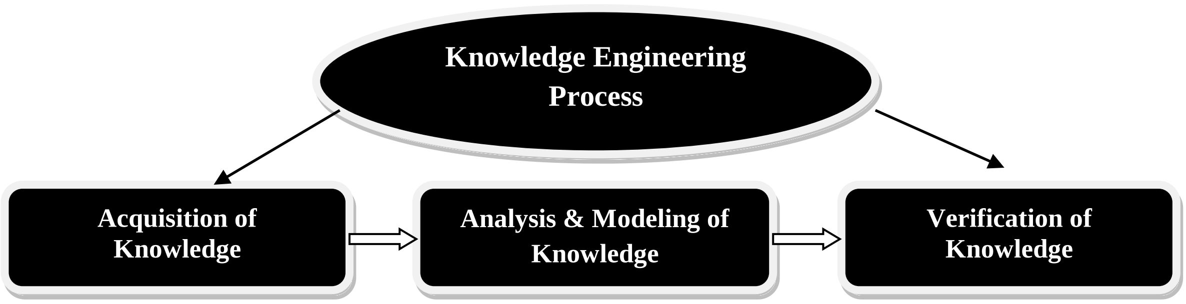

Figure 4.3. Knowledge Engineering Process.

The process of Knowledge Engineering is broadly categorized in three types:

- ***i.*** ***Acquisition of Knowledge***
- ***ii.*** ***Analysis & Modeling of Knowledge***
- ***iii.*** ***Verification of Knowledge*.**

4.4.2. Acquisition of Knowledge

Accruing expert knowledge is a crucial part of knowledge engineering. It is a subsystem that helps the expert for building up the knowledge bases. To develop an actual expert system and to have the highest performance, knowledge acquisition plays one of the most vital roles during the development of the system. This process actually gathers information which is relevant about the domain from domain experts primarily. Here the system's knowledge had been acquired through three main sources namely, (a) Medical Experts, (b) Published literature, (c) Real observation at the hospitals and (d) Text Books [16, 21].

Three Experts in Pediatrics, each of them having 12-33 years of experience were the first source of the main knowledge. The main experts are from Dept. of Pediatrics, North Bengal Medical College and Hospital and others from said department and other hospitals were consulted through structured interviews. To record the knowledge extracted from the domain experts, forms were prepared in consultation with them. The experts were asked to give their judgments for different sets of possible real observation and cases. In this phase, it was decided to involve multiple experts in the knowledge acquisition process. Thereafter, knowledge was gathered from various research publications including books and workshop reports published by Indian Pediatrics, some other journals and text books [16, 21].

For real observed cases, frequent visits were conducted at neonatal unit of Dept of Pediatric Medicine of North Bengal Medical College & Hospitals, Siliguri, West Bengal, India. All real life physical observations were recorded and incorporated in the knowledge base of the system.

4.4.3. Analysis & Modeling of Knowledge

The very next process after knowledge accusation is analyzing and modeling of the knowledgebase. The unstructured knowledge as acquired from the above three sources of knowledge was then structured by the knowledge engineer. The knowledge was then represented in an Object Oriented form for later implementation. An Object Oriented Approach to KR(Knowledge Representation) schemes is more structured than other well known schemes and improves consistency, understandability, maintainability and modifiability of the knowledge base[16]. The knowledge in the system is stored in as group of objects. Each group is represented by a class with its attributes. A class defines the general properties of structure of a group of objects. Attributes describe the object's important characteristics. The knowledge library class serves as a database. Various pictures were collected from North Bengal Medical College during field surveys and from experts. Both the experts and medical persons are accustomed with various medical terms like "APGAR", "HIE" etc. which have been incorporated as attribute values.

We have applied an approach called "Select and Modify" [17]. As per the needs of the system the model is selected and modify whenever there is a need. This process is again subdivided into four activities. They are:

i. Selection of an interpretation model according to a set of selection criteria.
ii. Evaluation of interpretational model and investigate it for checking suitability for the application.
iii. Modification of interpretational model for making suitable applicability. and lastly,
iv. Domain knowledge accusation for modified interpretation model.

4.4.4. Verification of Knowledge

Knowledge verification is the phase whereby making quality assertion of the acquired knowledge. There are two points of concern: Establishment of review procedure and Conflicts Resolving for Multiple Expert procedure.

❖ Establishment of Review Procedure

There are three phases of establishing the review procedure. They are knowledge elicitation, knowledge analysis and modeling, and implementation.

- At the elicitation stage, reviewing is conducted by letting the domain expert's opinion on the results of the knowledge elicitation sessions.
- At the analysis and modeling stage, the domain experts review the filled forms describing the domain knowledge. The knowledge engineer performs this activity by walking through them with the presence of the domain experts.
- Lastly at the implementation stage reviewing is conducted by letting the domain experts review any early prototype and finalize the knowledgebase for implementation.

❖ Conflict Resolution for Multiple Experts

Because when two experts give different knowledge for the same thing, then trying to resolve this conflict yields more reliable knowledge, hopefully, agreed upon by both of them. Thus multiple experts' conflict resolution is considered as a way of verifying the acquired knowledge. If no consensus is reached the expert who is recognized to be more specialized in the area of disagreement is considered for knowledge management.

4.5. System Design and Implementation

The system is designed to aid the decision-making process for identification of common diseases in neonates. It also selects the controls measure taken/ management for confirmation of disease. A system flow diagram of the system is shown on figure 4.4. There are basically three phases in the expert system. During the first phase, preliminary identification of the sick neonate in the field can be done on the basis of neonatal information observed in the field. This identification is related with complain/ sign-symptoms of the sick baby. This identification is further confirmed with the knowledge related to characteristics such as Birth Weight, APGAR Score, Gestational Age etc. The system uses domain knowledge in this phase. The system asks for the inputs related to the sick neonate history and sign & symptoms for which he/she is suffering. Firstly the system asks for history related with mother and child, like Mothers age, LMP, DOD (for calculating term/preterm/post term baby), Parity(no of children), Mothers Blood Group, History of hereditary diseases, weight of the baby(to find out LBW / VLBW/ ELBW / NORMAL or finding out weather the bay is AGA(appropriate for gestational age/ SGA(small for gestational age)/LGA(large for gestational age) etc.

After that it asks for the primary signs such as lethargy, refusal to suck poor cry, poor weight gain, hypothermia, sclerema, excessive jaundice, bleeding, GI disorder, seizure, sluggish neonatal reflex etc. The production of rules of inference are used a good level of accuracy is achieved in the resulting identification. The forward chaining of reasoning is performed with the rule based knowledge.

For inferences in the system “IF {I} AND {C} THEN {D} CF” rules have been used, where {I} is the information of the sick baby, {C} is complaint or symptoms and {D} is the differential diagnosis. Here CF is the certainty factor which often call confidence factor associated with the rule. The confidence factor CF attached to the most of the rule represents actual confidence of that rule. The value assigned to CF ranges from 0 to 100. This knowledge is taken from expert doctors after few interactions. Multiple experts in the pediatric domain actually have suggested the confidence factor (CF) at the time of knowledge acquisition.

The format chosen for the definition of rules allows flexibility in structuring the knowledge [18, 20]. A predecessor of any rule may be a composite of number of clause connected through logical operation OR and AND. In addition to the “if-then” structures, common to all rule-based systems, knowledge representation supports a new type of concept –the *criterion* – that enables a more compact and natural way of expressing rules of the form: “if at least *n* conditions are satisfied then...”[22].

Knowledgebase structure of the system having classes, instances, attributes and their types and other details have been attached in Appendix A.

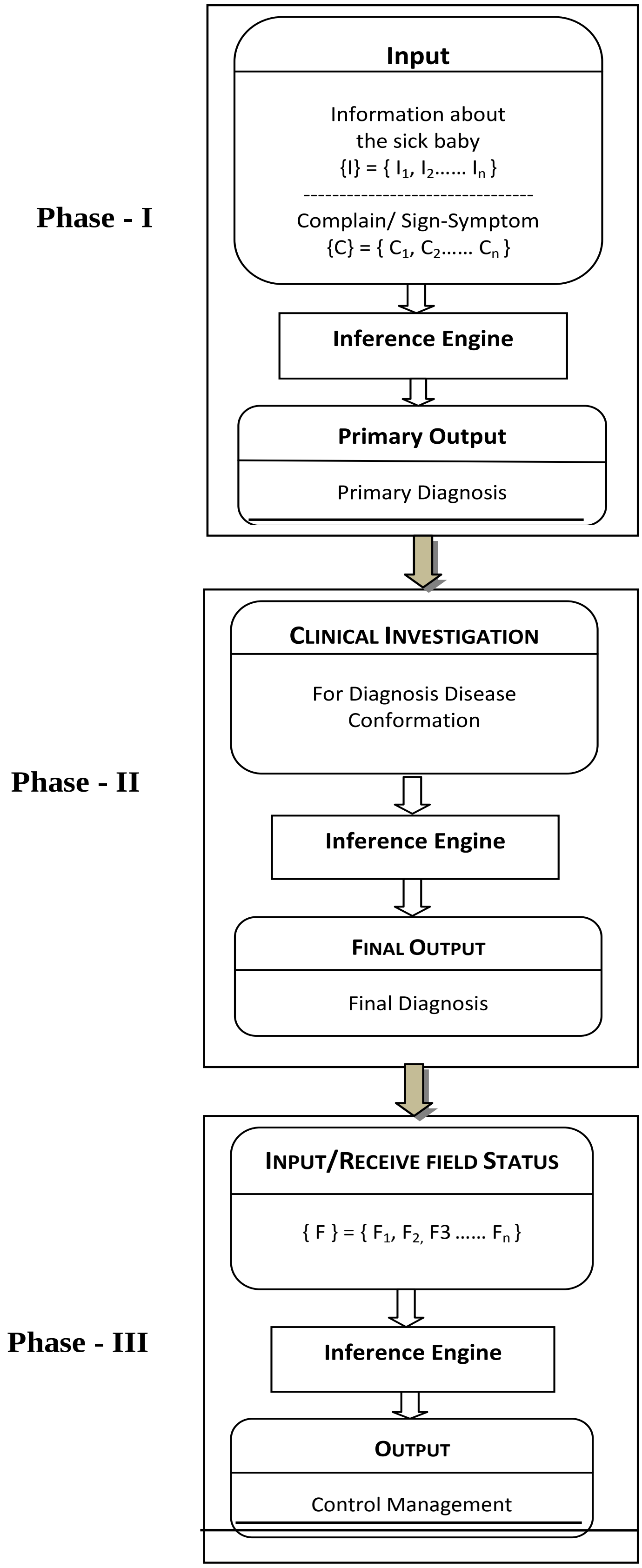

Figure 4.4. System Flow Diagram.

4.6. Evaluation of the System

The system has been analyzed for 68 real field cases in different neonatal disease category. The beginning of using the expert system causes no additional clinical work. Only minor changes in working practice required for the auxiliaries, who normally perform the sampling and for noting the mother's name and hospital identification number before analysis.

The system has been designed for the simplification in the use. Users' quick acceptance of the system is also endorsed with this. The multiple sticky labels for patient notes eliminated the time consuming and error prone process of transcription, and this helped significantly with user acceptability. For trail run, some randomized cases taken from the set of real field case used for evaluation. The study has found that 90% accuracy or certainty of the system. This has been verified by the expert pediatricians. Two case studies are presented in Appendix B as representative examples.

4.7. Conclusion

This chapter presents an integrated environment for the development of knowledge based decision support systems applied to diagnose and manage the common neonatal diseases. There are few researches also used this kinds of system but in various other domains [7,12,23]. This system provides graphical support for every phase of the development cycle, from knowledge acquisition and addition to reasoning explanation and knowledge-base validation. It is an interactive Object Oriented Rule-Based Expert System developed to meet the needs of such places where is scarcity of the medical expert for diagnose of neonatal disease and give proper clinical control management, especially in hospital, and primary health centres in remote villages.

Appropriate Graphical User Interface components incorporated in the system a facility to select and deselect multiple options from the easy to use menu. It needs almost no training for its user. Easy and commonly used linguistic variable used in inputs and output which makes this system more effective. Even less-trained person seeking expert system guidance for proper neonatal disease diagnosis and treatment plan practices also gets help out of using this system. The extensive system testing and validation process in conjunction with the overall structured project methodology, has greatly contributed to the expert system reliability and user-acceptance. It proves itself as a useful and beneficial tool and supporting hand for the domain experts.

References

1. Kumar D. Varma A, Sehgal VK, "*Neonatal Mortality in India*", Rural and Remote Health 7: 833. (Online) 2007. Available: http://www.rrh.org.au [Last accessed on 8th October, 2009].

2. *"Government of India. Annual Report 2006-2007"*, Ministry of Health and Family Welfare, Government of India,. New Delhi, 2007.

3. Bang Abhay T., Bang Rani A., Baitule Sanjay, Deshmukh Mahesh and Reddy M. Hanimi, "*Burden of Morbidities and the Unmet Need for Health Care in Rural Neonates - A Prospective Observational Study in Gadchiroli, India*", Indian Pediatrics, Vol. 38, pp. 952-965, 2001.

4. Roy Chowdhury Dilip, Samanta R.K., Chatterjee M., "*A Study of the Status of New Born in Terai Region of West Bengal*", A.M.S.E. France, No. 05 225 (2C), February 2007.

5. "*Government of India. National Population Policy–India(2000):National Socio-Demographic Goals for 2010*", (Online) 2000, Available: http://populationcommission.nic.in/npp_obj.htm#box2 [Last accessed on 1st November 2009].

6. James A. Taylor, Lori A. Loan, Judy Kamara, et al, "*Medication Administration Variances Before and After Implementation of Computerized Physician Order Entry in a Neonatal Intensive Care Unit*", Pediatrics, Vol. 121, No. 1, 123-128, 2008.

7. Dollberg Gil, Mimouni Michael, Dollberg Shaul, "*Computerized Decision-Making Assistance for Managing Neonatal Hyperbilirubinemia*", Pediatrics, Vol. 117, No. 1, January, 262-263, 2006.

8. Srinivasan S., Mital D. P. and Haque S., "*A Point of care Clinical Decision Support System for the Diagnosis of Neonatal Jaundice by Medical Field Personnel*", Journal of Applied Sciences, Vol. 6, No. 5, pp. 1003-1008, 2006.

9. Karayiannis NB, Tao G, Xiong Y, Sami A, Varughese B, Frost JD Jr, Wise MS, Mizrahi EM. Epilepsia "*Computerized Motion Analysis of Videotaped Neonatal Seizures of Epileptic Origin*", NCBI-PubMed, Vol. 46, No. 6, pp. 901-917, June 2005.

10. Singh A. K., Kohli M., Trell E., Wigertz O., Kohli S., Int. J., "*Bhorugram (India): revisited. A 4 Year Follow-up of a Computer-Based Information System for Distributed MCH Services*", Med. Informatics, Vol. 44, No. 2, 117-125, 1997.

11. Shimomura K., Shono H., Kohara M., Uchiyama A., Ito Y., Sugimori H. "*Neonatal Assessment Using the Apgar Fuzzy Expert System*", Comput. Biol. Med., Vol. 24, No. 3, pp. 171-178, May, 1994.

12. Dharmar. C., Srinivasan,S., Mital, Haque. S .D., "*Expert System for the Diagnosis of Neonatal Jaundice for Use by Medical Field Personnel*". Dept. of Health Inf., UMDNJ, NJ, USA; ISBN: 981-04-8364-3, Vol. 2, pp. 1002- 1006, December, 2002.

13. Kendal, S.L.; Creen, M., "*An Introduction to Knowledge Engineering*", London: Springer, ISBN 9781846284755, OCLC 70987401, 2007.

14. Feigenbaum, Edward A., McCorduck, Pamela, "*The Fifth Generation*" (1st ed.), Reading, MA: Addison-Wesley, ISBN 9780201115192, OCLC 9324691, 1983.

15. Buchanan, B. and Shortliffe E. (Eds), "*Rule-Based Expert Systems*", Addison-Wesley, 1984.

16. Ghai. O. P. "*Essential Pediatrics*", 6th Edition, CBS Publisher and Distributors, New Delhi, pp. 136, 2004.

17. K. Orsvarn, O. Olsson, H. Hassan, "*Guidelines for Adapting an Interpretation Model in An Application*", Ute Gappa, Hans Voss (Eds.), Proceedings of Knowledge engineering forum, 1995.

18. Duda, R., and Hart(Eds).P.E, *"Pattern Classification and Scene Analysis"*, Wiley John & Sons 1973.

19. Samanta R. K., "*Expert System and Knowledge Representation Scheme", CSI Communications,* Computer Society of India, Vol. 19, No. 3, 1995.

20. Giarratano, J. and Riley (Eds) G. "*Expert Systems: Principles and Programming",* PWS Publishing Company, Boston, 1994.

21. Gupte Suraj, "*The Short Text Book of Pediatrics*", 10th Edition, Jaypee Publication, New Delhi, pp. 93, 2003.

22. Russel, S. and Norvig (Eds) P., *"Artificial Intelligence: A Modern Approach"*, Prentice-Hall, Inc., 1995.

23. Fred A., Filipe J., Partinen M., Paiva T. "*PSG-Expert: An Expert System for the Diagnosis of Sleep Disorders*", IOS Press, no 78, series Studies in Health Technology and Informatics, pp.127-147, 2000.

24. Bhattacharya Sharbani, *Artificial Intelligence,* University Science Press, pp. 181-183, 2008.

C H A P T E R 5

A Data Mining and Knowledge Unearthing Concept of Designing DSS for Prevalent Neonatal Disease Diagnosis[§]

5.1. Introduction

On chapter 4, the process of designing a decision support system for the neonatal disease diagnosis has been discussed. It is the initial stage of development of such a system which will predict the disease, on the basis of signs and symptoms provided by the patient. Rule based approach using chaining reasoning have been applied. When the diagnosis result was tallied with the domain expert and specialist, we found the optimum accuracy. There the major challenge was to building up the knowledge base. Huge number of clinical data has been processed every day. In this respect, all the data are not relevant and giving actual information. Thus, there is a need of proper management of the data bases and knowledgebase. In this chapter we just initialized the process of data mining on the neonatal database and integrating that with decision support system.

This study is focusing on the neonatal diseases. As we mentioned in the previous chapters that the neonates constitute a large population group, and also vulnerable or special risk group. The risk is related with growth, development, disease pattern and survival. Thus by improving health status of neonates, proper diagnosis of the neonatal diseases is essential. Particularly Health care facilities have at their disposal of vast amount of data. Different analysis of available data on a given problem can lead to more efficient decision-making, which needs extracting relevant knowledge from this data and act upon it in a timely manner. But due to the scarcity of the neonate experts in rural and remote areas increases the mortality rate of neonates [1].

Even sometimes it is very critical to extract and analysis the real data from huge databases. Data mining is one of method to classify and analyze clinical data [2] for this concern. Our study of designing decision support system for prevalent disease diagnosis and management identifies the need of mining the clinical data.

5.1.1. Needs of the Present Study

A few hospitals worldwide are in the advanced stage of developing clinical data repositories for their basic needs of documentation on systems. Very few hospitals are

---

§ This chapter is based on the publication masde by the author entitled “Designing Decision Support System for Prevalent Disease Diagnosis and Management for Neonates : A Data Mining and Knowledge Unearthing Concept”, Proc. Nat. Conf. NCMicroCom-2010, February, 2010.

fully capitalizing on the benefits offered by knowledge base systems; physicians order entry, clinical decision support system, data mining of medical data and rule engines. It has been estimated that the amount of information in the world doubles every twenty months and the size of the number of databases increasing even. Developing new decision support system out of those data is cumbersome if proper data mining concept is not applying properly. Data Mining aims at discovering knowledge out of data and presenting it in a form that is easily comprehensible to humans. Considering our study, majority of the neonates in developing countries are born and cared for in rural homes but the available information is mostly hospital based [3]. Neonates not only constitute a large population group, but also vulnerable or special risk group. The risk is related with growth, development, disease pattern and survival. From the commonly accepted indices, it is evident that mortality rates in this age group are higher than adult population especially in developing countries. Thus by improving health status of neonates, we contribute to the health of the general population. These considerations have led to the formulation of special health services for children all over the world. A study shows that the incidence of the diseases namely, Neonatal Septicemia, HIE, Metabolic Disturbances, Neonatal jaundice etc. are quite high in the North eastern parts of India specially in Tarai region [4]. This has been discussed on chapter 3.

If neonates suffers from these diseases and proper management is not available in proper time the mortality and morbidity will increase considerably. To overcome the problems mentioned above proper treatment planning is required. And for proper treatment planning the basic requirement is Neonatologists. But the pediatric expert distribution in rural region is not at all satisfactory [5]. Not even that, for disease diagnose and treatment planning one has to go through plenty of conditions, sign and symptoms and critical case history. To mitigate the scarcity of the domain expertise, an expert system for prevalent disease diagnosis and management for neonates may help considerably using data mining and data unearthing concepts.

An estimated two-thirds of childhood deaths occur in infancy, and, in turn, two-thirds of infant deaths occur in the first month of life. In addition to these four million neonatal deaths, primarily due to serious bacterial infections, birth asphyxia, and complications of pre-maturity and intrauterine growth restriction, an estimated 3.9 million pregnancies end in stillbirth**[**6**]**. These deaths, at least half, could have been avoided if proper interventions were taken up. Research suggests that integration of maternal and neonatal health care services is key to the health benefits for mothers and their newborns. This is particularly so if a given intervention has potential direct health benefits for both mothers and newborns.

5.1.2. Related Studies

Several studies have demonstrated the value of data mining and knowledge discovery techniques have been used to automate the development of rules that detect clinical conditions by interpreting data generated from huge database which have been used to analyze and structure narrative patient reports in order to provide data for applications, such as automated encoding, decision support, patient management, quality assurance, outcomes analysis, and clinical research [7-10]. A study on Refractive errors in schools going children identifies the need of mining the clinical data shows how knowledge unearthing and data mining process have been used to development of rules and design the knowledge base of clinical data **[**11**]**. Expert System technology has been applied in many medical disciplines also for new born babies whom we call neonates [12-17]. To the best of our knowledge, no such comprehensive work has been reported for neonates for such prevalent disease diagnosis and management.

This chapter presents a rule-based, object-oriented expert system for prevalent disease diagnosis and management of neonates using data mining concept. Firstly, the database was designed on the basis of signs and symptoms provided by the patient. Then we applied data mining clustering approach for classification.

The chapter is summarized as below:

In section 5.2, the common causes of neonatal deaths have been pointed out. Section 5.3 explains the knowledge engineering, knowledge accusation and data mining technique applied, Section 5.4 presents the research methodology for system design and implementation. Section 5.5 describes the performance of the expert system. In the last section, i.e. section 5.6 is about the conclusion of this study.

5.2. Common Causes of Neonatal Deaths

In respect to this study we need to know about the problems with the neonates in the study area. Thus again there is need of discussing common causes of neonatal deaths. In India most common causes of neonatal deaths are systematic bacterial sepsis, birth asphyxia, congenital malformation, pre-maturity, Hyperbilirubenimia and others. Few common diseases of neonates are Neonatal Sepsis, Hypothermia, Birth Asphyxia, HIE, Metabolic Disorders, Inborn Errors of Metabolism, Respiratory Disorders, Neonatal Seizure, Hyperbilirubinimia, Prematurely, Hemorrhagic Disease of New Born and others. Few common disease pattern concerning neonatal heath problems and diagnosis criteria are tabulated below in Table 5.1. This also to note that single most important determinate is LBW i.e. birth weight less than 2500 gms.

Table 5.1. Common Disease Pattern, Heath Problems and Diagnosis Criteria

Sl. No.	***Neonatal Health Problems***	***Diagnostic Criteria***
1.	**Birth asphyxia**	(*i*) Mild: At 1 minute after birth, no cry, or the breath was absent or slow, weak or gasping. (*ii*) Severe: At 5 minutes after birth, the breath was

		absent or slow, weak or gasping. (*iii*) Indirect: In the absence of direct observations by VHWs about newborn's condition at 1 and 5 minutes, presence of following two: (*a*) baby did not cry on its own so the care provider had to make efforts to make the baby cry; (*b*) color of the umbilical cord was green or yellow.
2.	**Neonatal Sepsis**	**(Septicemia, meningitis or pneumonia diagnosed clinically):** Simultaneous presence of any two of the following six criteria any time during 0-28 days: (*i*) Baby which cried well at birth, it's cry became weak or abnormal, or stopped crying; or baby who earlier sucked or licked well, stopped sucking or mother feels that sucking became weak or reduced: or baby who was earlier conscious and alert, became drawsy or unconscious. (*ii*) Skin temperature >99ºF or <95ºF (*iii*) Sepsis in skin or umbilicus (*iv*) Diarrhea or persistent vomiting or distension of abdomen (*v*) Grunt or severe chest indrawing. (*vi*) Respiratory rate (RR) 60 or more per minute even on counting twice.
3.	**Preterm**	Less than 8 months and 14 days (37 weeks) of gestation counted from the onset of the last menstrual period as per the history given by the mother.
4.	**Low birth weight**	Weight less than 2500 g.

5.	**Delayed breastfeeding**	Due to traditional practice, breastfeeding not started in first 24 hours after birth, but baby licked/sucked the sweetened water.
6.	**Problems in breastfeeding**	Presence of any one of the following: (*i*) Baby did not suck breast for more than continuous 8 hours even when offered. (*ii*) –Mother unable to breast feed, or –baby fed on extracted breast milk, or goat, or cow milk, or bottle, or --sweetened water beyond 3 days, or –inadequate breast milk evidenced by continuous crying of baby and failure to gain weight.
7.	**Diarrhea**	Watery, liquid motions 3 or more, or > 9 motions of normal consistency in 24 hours; or mucus or blood in liquid stool.
8.	**Hemorrhage**	bleeding from mouth, anus, eyes, nose or in skin or in urine any time or vaginal bleeding after first week.
9.	**Conjunctivitis**	Mother complained of excessive discharge from the eyes of baby and on examination, eyes were red, and purulent discharge or dried pus
10	**Skin Infection**	(*i*) *Pyoderma:* pus, ulcer, boil, pustule in skin. (*ii*) *Intertrigo:* excoriation with moist, cracked skin at skin folds.
11.	**Abnormal Jaundice**	Skin or eyes yellow on the first day or yellowness persisted at 3 weeks, or when yellowness associated with sepsis.
12.	**Meconium Aspiration**:	History of difficult delivery or presence of birth asphyxia and respiratory distress (RR 60 or more; or severe indrawing of lower chest) started in first 24 hours after birth.
13.	**Hyaline Membrane**	Respiratory distress started within 6 hours after

	Disease (HMD)	birth in preterms baby.
14.	**Pneumonia**	RR 60 or more, persistent even when counted twice (Increased RR when associated with other signs symptoms of sepsis was included in neonatal sepsis).
15.	**Upper Respiratory Infection (URI)**	- Cough or nasal discharge present for three days or more without respiratory distress or increased RR.
16.	**Hypothermia**	Auxiliary temperature <95ºF.
17.	**Umbilical Sepsis**	Pus discharge from umbilicus.
18.	**Tetanus**	Baby which earlier sucked well, stopped taking feeds from 4th day or more; and appearance of seizures, spasm and trismus.
19.	**Convulsive Disorder**	Seizures but baby conscious, alert and feeds well between seizures (excludes tetanus, asphyxia, sepsis)
20.	**Unexplained fever**	Axillary temperature >99ºF without any attributable cause.
21.	**Failure to Gain Weight**	Total weight gain during 0-28 days <300 g.

Source: *Indian Pediatrics, 38: 952-965, 2001.*

5.3. Knowledge Engineering Process

5.3.1. Knowledge Accusation and Data Mining Techniques Applied

In this chapter we applied the same knowledge engineering processes that have been discussed in chapter 4. But for developing an actual expert system and to have the highest performance, knowledge acquisition plays one of the most vital roles during the development of the system. Here the system's knowledge had been acquired through three main sources namely, (a) Medical Experts, (b) Published literature, (c) Real observation at the hospitals and (d) Text Books **[**18,19,20**]**.

Three Experts in Pediatrics, each of them having 12-33 years of experience were the first source of the main knowledge. The main experts are from Dept. of Pediatrics, North Bengal Medical College and Hospital and others from said department and other hospitals were consulted through structured interviews. To record the knowledge extracted from the domain experts, forms were prepared in consultation with them. The experts were asked to give their judgments for different sets of possible real observation and cases. In this phase, it was decided to involve multiple experts in the knowledge acquisition process. Thereafter, knowledge was gathered from various research publications including books and workshop reports published by Indian Pediatrics **[**3**]**, some other journals and text books [18, 19, 20].

For real pragmatic cases, frequent visits were conducted at neonatal unit of Dept of Pediatric Medicine of North Bengal Medical College & Hospitals, Siliguri, West Bengal, India. All real life physical observations were recorded and incorporated in the knowledge base of the system which afterwards creates a repository of knowledge.

5.3.2. Knowledge Repository

The unstructured knowledge as acquired from the above three sources of knowledge was then structured by the knowledge engineer. The knowledge was then represented in an Object Oriented form for later implementation. An Object Oriented Approach to KR (Knowledge Representation) schemes is more structured than other well known schemes and improves consistency, understandability, maintainability and modifiability of the knowledge base [21].

In the knowledge repository, the knowledge in the system is stored in as group of objects. Each group is represented by a class with its attributes. A class defines the general properties of structure of a group of objects. Attributes describe the object's important characteristics. The knowledge library class serves as a database.

5.3.3. Clustering

Afterwards data mining technique-Clustering has applied. This Cluster detection consists of building models that finds data records similar to each other. This is inherently undirected data mining, since the goal is to find previously unknown similarities in the data. Clustering data may be considered a very good way to start any analysis on the data. Self-similar clusters can provide the starting point for knowing what is in the data and for figuring out how to best make use of it.

To identify and discover patterns, Clustering is one of the best tools in data mining. Clustering helps in recognizing exciting distributions inside the data [26]. When there is no class to be predicted, then Clustering techniques may be applied. Here all the instances are divided into groups. Data clustering identifies the sparse and the crowded places. In this way it discovers the overall distribution patterns of the dataset. Clustering is considered one of the most important unsupervised learning techniques. In most of the other problems in similar type, it deals with finding a structure in a collection data which is not labeled.

Clustering is the process of organizing objects into groups whose members are similar in some way. Hence, a cluster is a collection of objects which are "similar" between them and are "dissimilar" to the objects belonging to other clusters. The reasons for making clustering are as follows:

i. **Detection of Patterns**:

The new born babies may have several patterns of disease. Even the there may of different patterns of sign and symptoms. Thus Pattern detection is one of the vital reasons.

ii. **Simple and Robust**:

Clustering techniques are simple to find the similar kind of classes. Hence always gives robust result, if the proper information is given.

iii. **Data Concept Construction**:

In the knowledge base designing process, data concept creation is important. It helps to build similar categorical database. Hence searching process is easy and simple.

iv. **Unsupervised Learning:**

Clustering techniques apply when there is no class to be predicted. The instances are divided into natural groups. A mechanism causes some instances to bear a strong resemblance to each other than they do to the remaining instances.

Figure 5.1 shows the process of clustering:

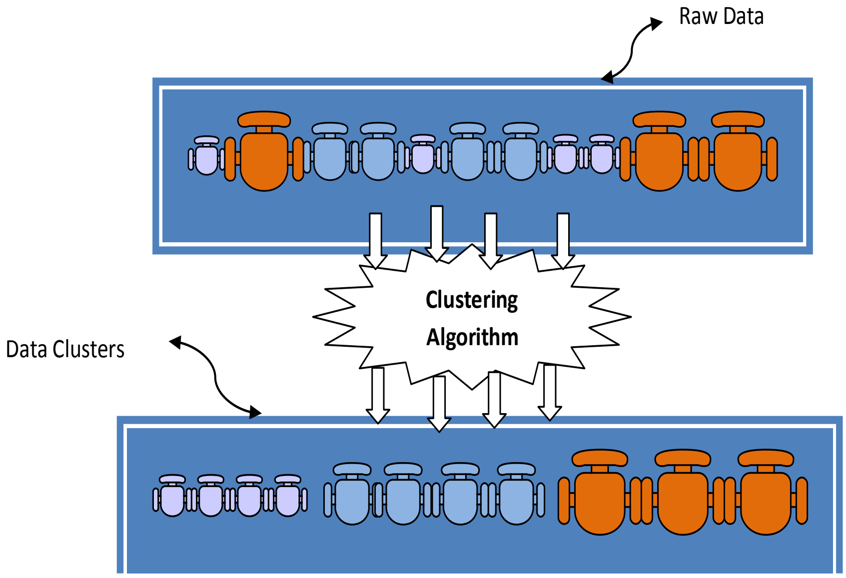

Figure 5.1. Clustering Process.

In respect of our study, various information related to neonatal disease were collected from North Bengal Medical College during field surveys and from experts. Both the experts and medical persons are accustomed with various medical terms like “APGAR”, “HIE” etc. which have been incorporated as attribute values. Then the clustering process started for categorizing the similar group class.

5.3.4 Knowledge Unearthing Concept

Knowledge unearthing is a concept of data mining only. Data mining strategies can be broadly classified as either supervised or unsupervised. Supervised learning builds models by using input attributes to predict output attribute values. Many supervised data mining algorithms only permit a single output attribute. Whereas other supervised learning tools allow us to specify one or several output attributes. Output attributes are known as dependent variables as their outcome depends on the values of one or more input attributes. Input attributes are referred to as independent variables.

When learning is unsupervised, an output attribute does not exist. Therefore all attributes used for model building are independent variables. Supervised learning strategies can be further labeled according to whether output attributes are discrete or categorical, as well as by whether models are designed to determine a current condition or predict future outcome. In this section we examine three supervised learning strategies, take a closer look at unsupervised clustering, and introduce a strategy for discovering associations among retail items sold in catalogs and stores. Thus knowledge unearthing process is really helpful for the followings:

- Text and graphical based data may be used applicable.
- Very useful in data mining as a assisting tools.
- Help simplify data complexity.
- Classification.
- Detect hidden pattern in data. Etc.

5.4. Research Methodology for System Design and Implementation

Data Mining often requires data integration, the merging of data from multiple data sources into one coherent data store. These sources include in our case knowledge repository, flat files, and data entry values. Equivalent real-world entities from multiple data sources must be matched up, for example, *disease_pattern* in one database must be matched up with *cause_of_disease* in another database. Careful integration of the data from multiple sources helped reducing and avoiding redundancies and inconsistencies in the resulting data set. This helped improving the accuracy and speed of the subsequent mining process.

The system is designed to aid the decision-making process for identification of common diseases in neonates. It also selects the controls measure taken/ management for confirmation of disease. A system flow chart of the system is shown on figure 5.2. Basically three phases are there in the expert system. During the first phase, preliminary identification of the sick neonate in the field can be done on the basis of neonatal information observed in the field and clustering them from knowledge repository. This identification is related with complain/ sign-symptoms of the sick baby. This identification is further confirmed with the knowledge related to characteristics such as Birth Weight, Apgar score, and Gestational Age etc. The system uses domain knowledge in this phase. The system asks for the inputs related to the sick neonate history and sign & symptoms for which he/she is suffering. Firstly the system asks for history related with mother and child, like Mothers age, LMP, DOD (for calculating term/preterm/post term baby), Parity(no of children), Mothers Blood Group, History of hereditary diseases, weight of the baby(to find out LBW / VLBW/ ELBW / NORMAL or finding out weather the bay is AGA(appropriate for gestational age/ SGA(small for gestational age)/LGA(large for gestational age) etc.

After that it asks for the primary signs such as lethargy, refusal to suck poor cry, poor weight gain, Hypothermia, Sclerema, excessive jaundice, Bleeding, GI disorder, seizure, sluggish neonatal reflex etc. The production of rules of inference are used a good level of accuracy is achieved in the resulting identification. The forward chaining of reasoning is performed with the rule based knowledge.

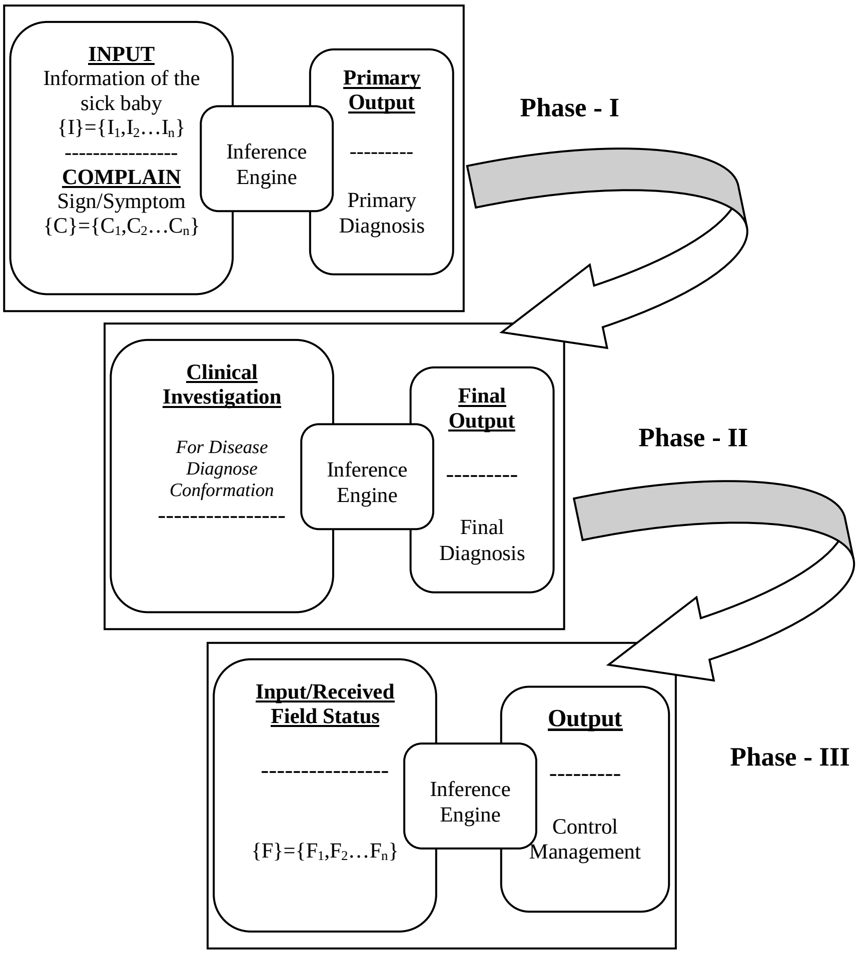

Figure 5.2. System Flow Chart.

For inferences in the system " IF {I} AND {C} THEN {D} CF" rules have been used, where {I }is the information of the sick baby, {C} is complaint or symptoms and {D} is the differential diagnosis. Here CF is the certainty factor which often call confidence factor associated with the rule. The confidence factor CF attached to the most of the rule represents actual confidence of that rule. The value assigned to CF ranges from 0 to 100. This knowledge is taken from expert doctors after few interactions. Multiple experts in the pediatric domain actually have suggested the confidence factor (CF) at the time of knowledge acquisition.

The format chosen for the definition of rules allows flexibility in structuring the knowledge [22,23]. A predecessor of any rule may be a composite of number of clause connected through logical operation OR and AND. In addition to the "if-then" structures, common to all rule-based systems, knowledge representation supports a new type of concept –the *criterion* – that enables a more compact and natural way of expressing rules of the form: "if at least *n* conditions are satisfied then..."[24].

It was implemented in an object- oriented environment by using Level 5 Object for Microsoft Windows (release 3.0 or higher) by the Information Builders Inc, USA. It has an integrated array of powerful tools like GUI development, Forms and Display builders and also has got capability to chain more than one knowledge base together and also have clustering facility by writing code. The facilities provided by the Level 5 Expert System Cell have been discussed on the chapter 4. Considering that facilities we developed this knowledge based decision support system where data mining and knowledge unearthing process has been applied efficiently.

5.5. Performance and Evaluation of the Expert System

In this study, the system had been analyzed for 68 real field cases in different neonatal disease category from the database repository. After applying the data mining techniques comes the job of identifying the obtained results, in form of interesting patterns representing knowledge depending on attention-grabbing measures. Such measures can be used after the data mining step in order to rank the discovered patterns according to their interestingness, filtering out the uninteresting ones. More importantly, such measures can be used to guide and constrain the discovery process, improving the search efficiency by pruning away subsets of the pattern space that do not satisfy pre-specified interestingness constraints.

The introduction of the expert system caused no additional clinical work; the only minor change in working practice required was for the auxiliaries, who normally perform the sampling, to note the mother's name and hospital identification number before analysis. The system was designed to be simple to use, and the users' quick acceptance of the system endorsed this. The multiple sticky labels for patient notes eliminated the time consuming and error prone process of transcription, and this helped significantly with user acceptability. For trail run, some randomized cases taken from the set of real field case used for evaluation.

It was found that 90% accuracy or certainty of the system, of course after being verified by the expert pediatricians of North Bengal Medical College & Hospital, Siliguri, and Darjeeling. Case studies are presented in Appendix B as representative examples. These examples have been analyzed by the domain experts. After applying knowledge unearthing concepts the decision of system were matched with expert opinion. We received an accuracy of 90% for accurate diagnosis of the neonatal disease by the system.

5.6. Conclusion

This chapter presents an integrated environment for the development of diagnosis-oriented knowledge based systems applied to diagnose and manage the common neonatal diseases in line with various medical domains [17]. Data mining approach for knowledge repository has successfully shaped. Out of several sign and symptoms mining approach really helped to find the useful knowledge to be processed for reasoning and development of decision support system.

Based on the work done, the following conclusions were drawn:

1. This system provides graphical support for every phase of the development cycle, from knowledge acquisition and addition to reasoning explanation and knowledge-base validation.

2. Level 5, is very suitable as a mining engine with its interface and manipulating modules that allow data exploration, manipulation and exploration of any interesting knowledge patterns

3. Using the same data sets with different mining techniques such as clustering and comparing results of each technique in order to construct a full view of the resulted patterns and levels of accuracy of each technique may be very useful for this application.

4. Data mining technique is very useful in the process of knowledge discovery in the medical field, especially in the domains where available data have many limitations like inconsistent and missing values.

5. It is an interactive Object Oriented Rule-Based Expert System developed to meet the needs of such places where is scarcity of the medical expert for diagnose of neonatal disease and give proper clinical control management, specially in hospital, and primary health centres at remote villages.

6. Appropriate Graphical User Interface components incorporated in the system a facility to select and deselect multiple options from the easy to use menu. It needs almost no training for its user. Very easy and commonly used linguistic variable inputs and output, which are in natural language and commonly used terms add advantages to a less-trained person seeking expert system guidance for proper neonatal disease diagnosis and treatment plan practices. [Appendix C]

7. The extensive system testing and validation process in conjunction with the overall structured project methodology, has greatly contributed to the expert system reliability and user-acceptance.

References

1. Kumar D. Varma A, Sehgal VK, "*Neonatal Mortality in India*", Rural and Remote Health 7: 833. (Online) 2007. Available: http://www.rrh.org.au [Last accessed 8th October, 2009].

2. Chae Ym, "*Data Mining Approach to Policy Analysis in Health Insurance domain*", International Journal of Medical Informatics, Vol.62, No. (2-3), pp. 103-111, July 2001.

3. Bang Abhay T., Bang Rani A., Baitule Sanjay, Deshmukh Mahesh and Reddy M. Hanimi, "*Burden of Morbidities and the Unmet Need for Health Care in Rural Neonates - A Prospective Observational Study in Gadchiroli, India*", Indian Pediatrics 2001; 38: 952-965.

4. Roy Chowdhury Dilip, Samanta R.K., Chatterjee M., "*A Study of the status of new born in Terai region of West Bengal*", A.M.S.E. France, No. 05 225 (2C), February 2007.

5. "*Government of India National Population Policy–India(2000):National socio-demographic goals for 2010*", (Online) 2000. Available: http://populationcommission.nic.in/npp_obj.htm#box2 [Last accessed on 1st November, 2009].

6. Ministry of Health and Family Welfare, Government of India, "*Government of India. Annual Report 2006-2007*". New Delhi, 2007.

7. S. Bicciato, A. Luchini, C. Di-Bello, "*Marker identification and classification of cancer types using gene expression data and SIMCA*", Methods-of-information-in-medicine, Germany: 2004.

8. Marx K. A., P. O'Neil, Hoffman P., Ujwal M. L., "*Data mining the NCI cancer cell line compound GI(50) values: identifying quinone subtypes effective against melanoma and* leukemia *cell classes*", Journal-of-chemical-information-and-computer-sciences, United-States,2003.

9. Forgionne G. A, Gangopadhyay A., and Adya M., "*Cancer SurveillanceUsing Data Warehousing, Data Mining, and Decision Support Systems*", Health Information Management, Proquest Medical Library, vol. 21(1); August 2000

10. Kuo W., Chang R., Chen D. and Lee C. C., "*Data Mining with Decision Trees for Diagnosis of Breast Tumor in Medical Ultrasonic Images*", Breast Cancer Research and Treatment, Dordrecht, Vol. 66, No. 1, March 2001.

11. Chandra Shekar D.V. and Srinivas V. Sesha, "*Clinical Data Mining–An Approach for Identification of Refractive Errors*" Proc. Int. Multi Conf. of

Engineers and Computer Scientists 2008, Vol. I, IMECS 2008, Hong Kong, 2008.

12. Gil Dollberg, Michael Mimouni, Shaul Dollberg, "*Computerized Decision-Making Assistance for Managing Neonatal Hyperbilirubinemia*", Pediatrics in Review, American Academy of Pediatrics, Vol. 117, No.1, pp. 262-263, January 2006.

13. Srinivasan S., Mital D. P. and Haque S., "*A Point of Care Clinical Decision Support System for the Diagnosis of Neonatal Jaundice by Medical Field Personnel*", Journal of Applied Sciences, Vol. 6, No.5, pp. 1003-1008, 2006.

14. Karayiannis NB, Tao G, Xiong Y, Sami A, Varughese B, Frost J D Jr, Wise MS, Mizrahi EM. Epilepsia "*Computerized Motion Analysis of Videotaped Neonatal Seizures of Epileptic Origin*", NCBI-PubMed, Vol. 46, No. 6, pp. 901-917, June 2005.

15. Singh A. K., Kohli M., Trell E., Wigertz O., Kohli S., "*Bhorugram (India): revisited. A 4 year Follow-up of A Computer-Based Information System For Distributed MCH Services*", Int. J Med. Informatics, Vol. 44, No. 2, pp. 117-125, 1997.

16. Shimomura K., Shono H., Kohara M., Uchiyama A., Ito Y., Sugimori H. "*Neonatal Assessment Using the Apgar Fuzzy Expert System*", Comput. Biol. Med., Vol. 24, No. 3, pp. 171-178, May, 1994.

17. Dharmar. C., Srinivasan,S., Mital, Haque. S .D., "*Expert system for the diagnosis of neonatal jaundice for use by medical field personnel*", Dept. of Health Inf., UMDNJ, NJ, USA; ISBN: 981-04-8364-3, Vol. 2, 1002- 1006, December, 2002.

18. Ghai. O. P. "*Essential Pediatrics*", 6th Edition, CBS Publisher and Distributors, New Delhi, pp. 136, 2004.

19. Gupte Suraj, "*The Short Text Book of Pediatrics*", 10th Edition, Jaypee Publication, New Delhi, pp. 593, 2003.

20. Duda, R., and Hart (Eds).P.E, *"Pattern Classification and Scene Analysis"*, Wiley John & Sons, 1973.

21. Samanta R. K., *"Expert System and Knowledge Representation Scheme", CSI Communications*, Computer Society of India, Vol. 19, No.3, 1995.

22. Buchanan, B. and Shortliffe E. (Eds), *"Rule-Based Expert Systems"*, Addison-Wesley, 1984.

23. Giarratano, J. and Riley (Eds) G. *"Expert Systems: Principles and Programming"*, PWS Publishing Company, Boston, 1994.

24. Russel, S. and Norvig (Eds) P., *“Artificial Intelligence: A Modern Approach”*, Prentice-Hall, Inc., 1995.

25. Fred A., Filipe J., Partinen M., Paiva T. "*PSG-Expert: An Expert System for the Diagnosis of Sleep Disorders*”, IOS Press, Series Studies in Health Technology and Informatics, No. 78, pp. 127-147, 2000.

26. Guha S., Rastogi R., Shim K., “*CURE: An Efficient Clustering Algorithm for Large Databases*", ACM SIGMOD, Vol. 27, No. 2, pp. 73-84, 1998.

CHAPTER 6

An Improved Data Mining for Neonatal Prevalent Disease of North Bengal Districts‡

6.1. Introduction

In the previous chapters, especially chapter 3 and chapter 4, an initial data mining approach has been applied for developing decision support system for the neonates. The concentration was mainly given on the proper knowledge accusation and followed by building a productive knowledge base system, on those chapters. In this chapter applications of improved data mining approaches have been focused. Studies have revealed that in rural and remote areas of India, the mortality and morbidity are high [1]. In rural and remote areas of India, the mortality and morbidity are high [1]. The chapter 2 contains all the statistics relating to neonatal mortality ratio and morbidity ratio, particularly for the rural areas statistics. One of the prime reasons of such high mortality and morbidity is prevalent diseases. There are a number of neonatal diseases and a number of parameters involved. Before taking any suitable measure by any appropriate authority to reduce mortality and morbidity due to different diseases of neonates, it might be useful to know the prevalent disease statistics and pattern of a region.

This work concentrated on North-Bengal districts of India. The aim of this work is to find the prevalent disease pattern of neonates of North Bengal districts from a data base created by the author from field studies during six months or so using a data mining technique for classification. One of the prime reasons found on the concerned area is a prevalent disease which leads to high mortality and morbidity. There are a number of neonatal diseases and a number of parameters involved.

Decision tree approach has been found suitable for the purpose. Since decision tree construction can make use of both symbolic or nominal and real-valued attributes, thus it has been efficiently used in medical domain. The results of this study might be useful for the future course of actions to be taken by an appropriate authority.

Data mining and knowledge discovery techniques are being used after its birth in the year 1993 [2] in connection with a business. It is now applied to interpret huge

‡ This chapter is based on the publication made by the author entitled “Data Mining for Neonatal Prevalent Disease of North Bengal Districts”, Proc. Nat. Seminar on Data Mining and Decision Support, pp. 24-31, March, 2010.

clinical data base(s) all over the world. In order to provide extreme facility, data mining approaches are being applied for many applications development such as automated encoding, decision support, quality assurance, patient management, outcome analysis, and clinical research [3 - 10].

There are different algorithms available for classification problems: decision tree, naïve nays, support vector machine, and feed forward neural networks. Decision tree approach has been found suitable for our purpose since decision tree construction can make use of both symbolic or nominal and real-valued attributes [11] which has been a characteristic of medical domain.

Almost all the institution or business houses are following computerized system of information storing, processing and retrieval. Day by day the volume of database is growing at an unexpectedly higher rate. In the medical domain the situation is much harder. In fact, in medical domain the data base is relatively bigger than the other areas. Every one need sophisticated information out of the data they are getting from the database. All these kinds of problem can be overcome by the use of improved data mining approach. Actually through the data mining process we are to find the important hidden information from the database.

An architecture has been proposed in this chapter which can be used as assistance of data mining to provide decision support for diagnoses of various neonatal disease. Although this system is for the diagnosis of neonatal disease still this can be implemented for the diagnoses of any other age group also, provided the knowledgebase is having strong and useful data. The proposed system is making the process of disease diagnosis faster and with possibly of maximum accuracy. It is not actually substituting the human specialist but is a real useful tool for them. In last two chapters, i.e. chapter 4 and chapter 5 we have discussed regarding the expert system and its merits in respect to our study. The basic difference between a typical 'Expert System' and this proposed system is that the data mining process boost the process of learning and thereby improve decision making process. To the best of our knowledge, no such work of this type has been reported for the region.

This chapter is summarized as below:

In section 6.2, the basic overview of data mining has been pointed out. Section 6.3 gives an idea of different data mining techniques. Section 6.4 presents the concepts of data mining and knowledge-base system. In section 6.5 presents our results and discussions.

6.2. Basic Overview of Data Mining

Database is a system that holds all the data. It is a process of maintaining the structured collection of records or eventual information which is stored in the computer system. In contrast, data mining is the process of extracting hidden patterns from database. In case of knowledgebase also, the same process has been utilized by named knowledge mining. Data mining is a knowledge discovery process. To find out relevant data from a huge database is the basic purpose.

6.2.1. Data Mining Process Factors

This process contains the integration five factors. The factors are domain knowledge, information, people, statistics and computing technology. The figure 6.1 shows the factors of Data Mining.

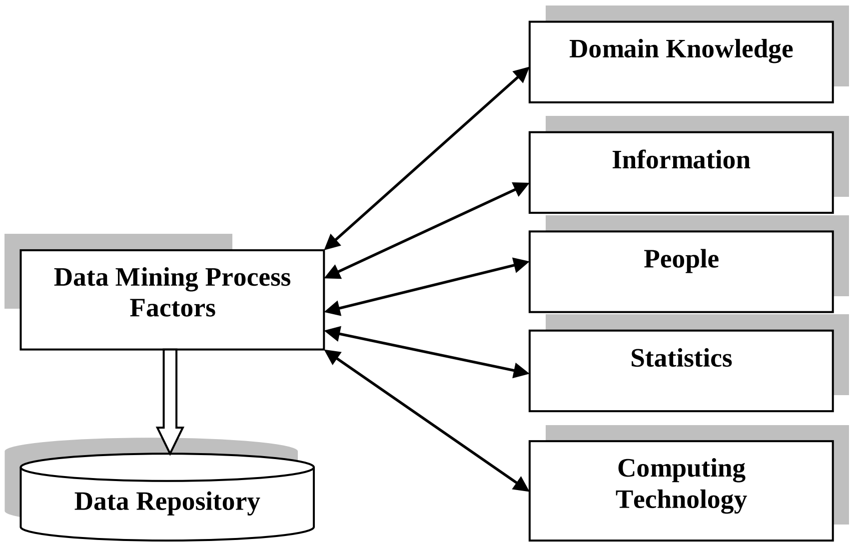

Figure 6.1. Factors of Data Mining.

6.2.2. Domain Knowledge

Data mining process first acquire the domain knowledge from the available source. That knowledge contains the information relating to the objectives of the study. In this regard, our domain knowledge is neonatal disease database. There may be several sign and symptoms which may be taken as inputs. Then that has to process with processing algorithms. Actual knowledge base is created based on the domain knowledge only.

6.2.3. Information

Information related to domain activities is the other criteria of the data mining. The knowledgebase or the data base may contain a huge numbers of data, out of which a good numbers of data is directly related with the domain knowledge. This is due to unnecessary information provided by the users or the patients. That is why there is the ultimate need of getting correct and factual information. Information must be unambiguous and accurate. Any wrong information may cause a massive hamper on the patient.

6.2.4. People

People are the user, domain experts and the knowledge engineer or programmer. Depending on the domain knowledge and factual domain related information data mining process works. Most the entire process of knowledge base development process knowledge engineer and data base administrator is laying the important role.

6.2.5. Statistics

Statistics of mined data from the entire database is always the area of the study and research analysis part. Different types of analytical process have done through this statistics. In our study of this domain also, we have used several utility statistics.

6.2.6. Computing Technology

Computing technology plays the vital role of data mining approach. Different kinds of critical analytical jobs can be easily done through the use of computing technology. There is a huge speed up activities found in the data mining process because of this improved computing technology.

6.3. Different Data Mining Techniques

There are several techniques available for data mining. Selection of these techniques needs immense expertise to get optimal output. Different techniques have their different mode of analytical views. A good way to apply advanced data mining techniques is to have a flexible and interactive data mining tool that is fully integrated with a database or knowledge-base. Using a tool that operates outside of the database is not as efficient. Using such a tool will involve extra steps to extract, import, and analyze the data. When a data mining tool is integrated with the data warehouse, it simplifies the application and implementation of mining results. Furthermore, as the warehouse grows with new decisions and results, the organization can mine best practices continually and apply them to future decisions. The major techniques are discussed below:

i. **Predictive**

It makes the prediction about the values of data using known results found from different data.

ii. **Regression**

Regression assumes that target data fit into some known type of function. And then find the best function of this type that models the given data. In the simplest case, regression uses standard statistical techniques such as linear regression. Unfortunately, many real-world problems are not simply linear projections of previous values.

iii. **Correlation**

Correlation is statistically oriented in nature. It is a fast growing technique of data mining.

iv. **Classification**

Classification is perhaps most commonly applied Data mining technique. Classification maps the data in predefined classes. These classes based on data attribute values. Classification has a special type known as segmentation. In segmentation database divide data into small segment. This process is known as segmentation.

v. **Time Series Analysis**

In this technique, the value of an attribute is examined as it varies over the time. This value is most of the time is taken of evenly spaced time points.

vi. **Descriptive**

It identifies patterns or relationships in data, in our case neonatal data.

vii. **Association**

Association identifies relationships between events that occur at one time, determines which things go together.

viii. **Clustering**

Clustering identifies groups of items that share a particular characteristic segmenting a diverse group into a number of more similar subgroups or clusters. Clustering differs from classification in that it does not rely on predefined classes or characteristics for each group.

ix. Sequence Discovery

Sequence discovery is used to determine sequential patterns in data.

x. Summarization

It maps data into subsets with associated simple description. It represents information about the data base. The tasks mention above may be combined to obtain more sophisticated Data mining application.

6.4. Data Mining and Knowledge-Base

Data Mining is the basis for any successful Knowledge Based System. Conventionally it is very difficult to make better decision making for neonatal disease diagnosis, in contrast data mining uses statistical analysis to develop better decision making on this respect. Even through the use of data mining one can predict the disease may occur in future considering the knowledge base. Many data mining tools on the market today can help you build powerful Knowledge-Based Systems. Even in medical informatics also there are several uses of data mining tools and applications. Data mining theory is developed in correspondence to knowledge based system in mind at first place. If it wouldn't be knowledge base it hadn't been possible to gather the knowledge from the data ware houses. To deal with the large data base or knowledge base, data mining is often used as a successful tool.

6.5. Data Mining Approaches Applied

6.5.1. Decision Tree and the Algorithm

6.5.1.1. General

Decision trees are used to recognize in various way of splitting a set of data into several segments. It is the procedure of creating branches to reach the goal or make the decision. Decision trees are easy, yet powerful form of multiple variable analysis. They provide unique capabilities to addition, complement, and substitute for traditional statistical forms of analysis. Actually segments in the decision tree generated from the root or from the top and then going on splitting into several conditions. The object of analysis is reflected in this root node as a simple, one-dimensional display in the decision tree interface. The name of the field of data that is the object of analysis is usually displayed, along with the spread or distribution of the values that are contained in that field. Knowledge is basically structured in forms of top-to-bottom manner through decision tree for any expert system development. For diagnosing the disease of neonate identification of any objects can be made through this decision tree approach better. First and intermediate nodes correspond to the identities of objects. Each of the object attribute corresponds to a non terminal node in the tree and each branch of the decision tree corresponds to an attribute value or a set of values [12].

Figure. 6.2 shows a simple decision tree structure. This structure tells that the decision tree can reflect both a continuous and categorical object of analysis. The I-Nodes

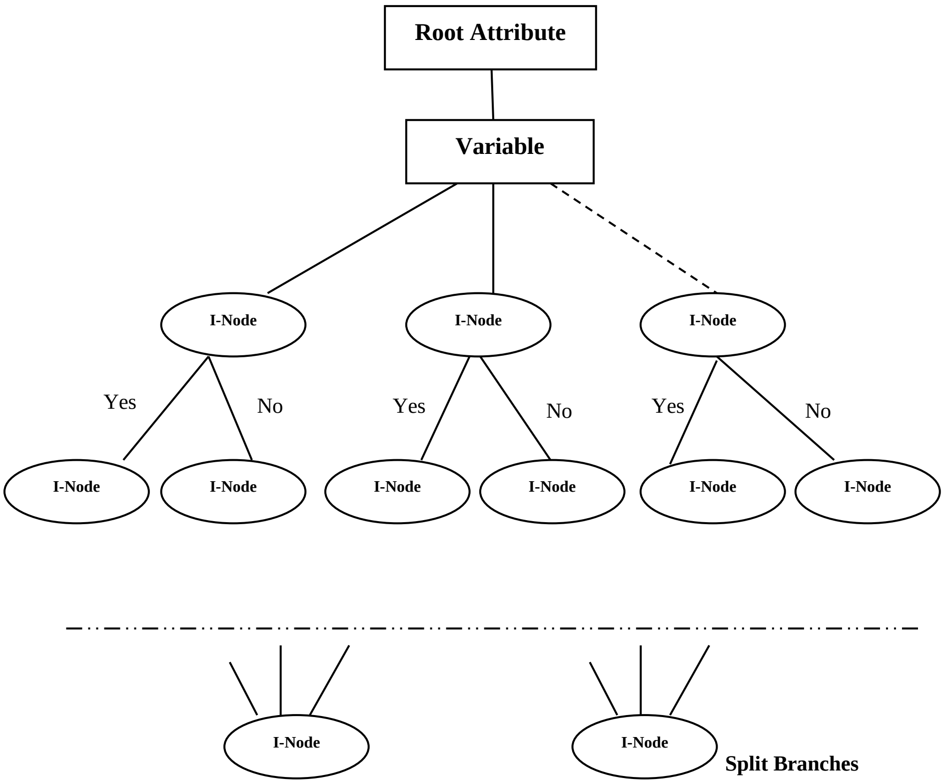

Figure 6.2. Decision Tree Structure.

reflects all the data set records, fields, and field values that are found in the object of analysis. A rule can be generated for the decision making process from the branches or segments below the root node. The values in the input field are used to estimate the likely value in the target field. The target field is also called an outcome, response, or dependent field or variable.

In Decision tree, each segment or branch is called a node. A node with all its descendent segments forms an additional segment or a branch of that node. The bottom nodes of the decision tree are called leaves (or terminal nodes). For each leaf, the decision rule provides a unique path for data to enter the class that is defined as the leaf. All nodes, including the bottom leaf nodes, have mutually exclusive assignment rules; as a result, records or observations from the parent data set can be found in one node only. Once the decision rules have been determined, it is possible to use the rules to predict new node values based on new or unseen data. In predictive modeling, the decision rule yields the predicted value. Identification made by traversing a path through the tree until it reaches a unique leaf node which corresponds to the unknown object's identity.

The decision tree algorithm is a common and one of the most popular algorithms used in data mining because it is easy to understand how it makes predictions. The goal is to create a model that predicts the value of a target variable (**dependent**) based on several input variables (**independent**). A tree can be 'learned' by splitting the source set into subsets based on an attribute value test. This process is repeated on each derived subset in a recursive manner called recursive partitioning. The recursion is completed when the subset at a node all has the same value of the target variable, or when splitting no longer adds value to the predictions [13]. Data comes in records of the form:

$$(X, Y) = (x_1, x_2, x_3, \ldots\ldots x_n, Y) \qquad (1)$$

The dependent variable, Y, is the target variable that we are trying to understand, classify or generalize. The vector X is composed of the input variables x_1, x_2, x_3, x4 ……etc., that are used as input for the model. In data mining, there are different categories of trees available in the literature. We use classification tree analysis, as because our predicted outcome is the class to which the data belongs.

6.5.2. C4.5 algorithm [14]

The C4.5 algorithm is based on information gain which is again based on the concept of entropy of information theory.
For a random variable X with N outcomes,

$$\{ x_i : i = 1,2,3,4,\ldots\ldots..N \} \qquad (2)$$

The Shannon entropy, a measure of uncertainty and denoted by H(X), is defined as

$$H(X) = -\sum_{i=1}^{N} P(x_i) \log_2 P(x_i) \qquad (3)$$

where $P(x_i)$ is the probability mass function of outcome x_i. This algorithm C4.5 uses a set of training data S ($S = s_1, s_2, s_3, \ldots\ldots\ldots s_n$) of classified samples for building decision trees. Each sample s_i will be having a set of attributes or features for classification along with a class attribute. Based on the normalized information gain (difference in entropy), at each node of the tree, C4.5 chooses one attribute of the data that most effectively splits its set of samples into subsets enriched in one class or the other. The attribute with the highest normalized information gain is chosen to make the decision. It then uses recursion on the smaller sub lists for further building and completion of decision tree.

Decision trees represent a supervised approach of classification from a data set provided for the said purpose. The problem of missing values can be rectified using this. The reason for selecting C4.5 decision tree algorithm is the algorithm's ability to handle data with missing values and rectification and also avoids overfitting the data and reduce error pruning. Figure 6.3. shows the results of running C4.5/J48 (Weka implemented) Decision tree algorithm and a tree.

6.6. System Design and Implementation

6.6.1 Attributes

In this study the following attributes have taken for the analysis. The first 18 attributes are independent attributes as input, mostly sign and symptoms of a disease and the last attribute i.e., Disease_Confirmation is dependent attribute as output.

Input Attributes (Independent Attributes):

@attribute Birth_Term_Status {Term,Pre_Term,Post_Term}
@attribute Birth_Weight_Status {Normal,LBW,VLBW,ELBW}
@attribute Age_in_Hours>72 {Y,N}
@attribute Lathergy {Y,N}
@attribute Refusual_to_Suck {Y,N}
@attribute Poor_Cry {N,Y}
@attribute Poor_Weight_gain {N,Y}
@attribute Hypothalmia {N,Y}
@attribute Sclerema {N,Y}
@attribute Excessive_Jaundice {N,Y}
@attribute Bleeding {N,Y}
@attribute GI_Disorder {N,Y}
@attribute Seizure {Y,N}
@attribute Sluggish_Neonatal_Reflex {Y,N}
@attribute CTR_HIE {HIE_III,NORMAL,Negetive,HIE_II}
@attribute CTR_MD {NORMAL,MD_HypoCalcimia,Hypo_Thalmia,Negetive, Positive, MD_HypoThermia, MD_HypoGlycimia}
@attribute CTR_SEPTI {NORMAL,Positive,Negetive}
@attribute CTR_OTHERS {NORMAL,HDN,Positive,URI, Hemorrhage,Jaundice, Pneumonia}

Output Attributes (Dependent Attribute):

@attribute Disease_Confirmation {
HIE_III,No_Disease,MD_HypoCalcimia, Septicimia, Hypo_Thalmia, MD_HypoCalcimia, Hemorrhage, Others, Jaundice,MD_Hypothermia, Jaundice_BA, MD_Hypoglycimia, HIE_II, Sizure_Disorder }

6.6.2 Algorithm Settings

Not only the selection of the right algorithm is important but also the proper settings of the parameters from data mining expertise, knowledge of the available algorithms, and often experimentation to determine which algorithm best fits the problem with suitable values of the parameters are equally important. Algorithm settings allow users to exert finer control over the algorithm to attain better results during the build process. Decision tree models can be extremely accurate on the build data if allowed to overfit the build data. This occurs by allowing the algorithm to build deeper trees with rules specific to even individual cases. Hence overfit models give very good accuracy with the build data, but do not generalize well on new data, resulting in decreased predictive accuracy [15]. To avoid overfitting as well as to control tree size, one has to apply pruning techniques and / or stopping criteria for decision tree algorithms. At the same time, goodness of a node split is determined by the information gain. So, in order for our model to generalize well it must not be built around the training data too closely [16]. Different pruning techniques have been proposed along with different splitting criteria, it has been found that there is not much variation in terms of performance [17,18,19]. All these issues have been taken in consideration during the present study in this chapter.

6.6.3. Terms Related with the Algorithm Settings

6.6.3.1. Overfitting

Overfiting generally occurs at the time of classification. A classification algorithm is said to be overfit to the training data, if it generates any representations of data. In this study, decision tree that depends very much on irrelevant features of the training instances. If we verify the outcome then it will produce poor performance using unseen data. Practically, overfitting will always occur because of the training set does not contain all possible instances which are relevant. It only becomes a problem when the classification accuracy on unseen instances is significantly poor. Hence for getting better prediction and classification accuracy we always need to be aware of the possibility of significant over-fitting problem. Even Overgeneralization may also cause poor classification accuracy. The above mentioned Top-Down approach of Decision

Trees algorithm is one of the most commonly used methods of classification. Though this is widely used and well known approach in decision making, still it may have problems of overfitting to the training data. It also produces very large rule set and gives poor predictive accuracy on unseen instances.

In the context of overfitting, we consider a typical rule such as:

> **IF** ***Birth_term_status = Term and Birth_weight_status = LBW and Bleeding = Yes*** **THEN** ***Class_Decision = Septicemia.***

Any addition of an additional term to this rule will specialise it, for example the increased rule

> **IF** ***Birth_term_status = Term and Birth_weight_status = LBW and Bleeding = Yes and Poor_Cry = Yes*** **THEN** ***Class_Decision = Septicemia***

It will normally refer to fewer instances than the original form of the rule maybe the same number, but positively not more than that. In contrast, removing a term from the original rule will generalise it, for example the depleted rule will normally refer to more instances than the original form of the rule possibly the same number, but certainly no fewer.

> **IF** ***Birth_term_status = Term and Birth_weight_status = LBW***
> **THEN** ***Class_Decision = Septicemia***

The main problem with top down approach is that while generating the classification rules problems of overfitting occurs. Tree generation is a repeated process of specialisation.Every time the algorithm splits on an attribute an additional term is added to each resulting rule. If a decision tree is generated from data containing noise or irrelevant attributes it is likely to capture incorrect classification information, which will tend to make it perform imperfectly when classifying unseen instances. If the tree is over-specialised its ability to generalise when classifying unseen instances, overfitting wil be reduced. Even in specific cases, specialising a rule by adding further terms can become productive. The generated rules give a perfect fit for the instances from which they were generated but in some cases are too specific to have a high level of predictive accuracy for other instances of that class. Another effect of too much specificity is that there are often an unnecessarily large number of rules. A smaller number of more general rules may have greater predictive accuracy on unseen data.

The standard approach to reducing overfitting is to sacrifice classification accuracy on the training set for accuracy in classifying unseen test data.

The analyst wants to process the models which are to be more “conservative" in contrast to other model which is to be more “liberal". If both models are specially “conservative" then the implication is that they would only classify new cases that are very closely related to cases they already have seen in the training data. In this situation, the net effect would be many cases to be left as unclassifiable by both systems. Similarly, if both systems are classifying new data in a “liberal" manner, then they may contradict each other, very often when they are presented with new cases. Again, this situation might be undesirable. Thus, a “liberal" behavior by a classification model means that the model has a tendency for overgeneralization. A similar relationship exists between the concept of “conservative" and overfitting [20].

6.6.3.2. Pruning the Decision Tree

To reduce the overfitting problem pruning technique is applied. It is a technique in machine learning which uses for reducing the size of decision trees by removing sections of the tree which is having less capacity to classify the instances. Pruning methods originally suggested in Breiman *et al.*, in 1984 [21]. They developed for solving dilemma of overfitting data on the training set. According to this methodology, a loosely stopping criterion is used, letting the decision tree to overfit the training set. Then the over-fitted tree is cut back into a smaller tree by removing sub–branches that are not contributing to the generalization accuracy. It has been shown in various studies that employing pruning methods can improve the generalization performance of a decision tree, especially in noisy domains.

There are two basic idea of pruning firstly reduced complexity of the final classifier as well as better predictive accuracy by the reduction of overfitting and Secondly, removal of sections of a classifier that may be based on noisy or erroneous data on the training data set provided for classification.

Several pruning methods which are used for classification in data mining are as follows [22]:

- **Cost–Complexity Pruning**

  Cost-complexity pruning (also known as weakest link pruning or error complexity pruning) proceeds in two stages (Breiman *et al.*, 1984). In the first stage, a sequence of trees $T_0, T_1, \ldots, T_k$ is built on the training data where T_0 is the original tree before pruning and T_k is the root tree. In the second stage, one of these trees is chosen as the pruned tree, based on its generalization error estimation. The tree T_{i+1} is obtained by replacing one or more of the sub–trees in the predecessor tree Ti with suitable leaves. The sub–trees that are pruned are those that obtain the lowest increase in apparent error rate per pruned leaf:

$$\alpha = \frac{\varepsilon(\,pruned\,(T, t_i), S) - \varepsilon(T, S)}{|\,leaves(T)\,| - |\,leaves(pruned(T, t))\,|} \qquad (4)$$

where ε(*T*, *S*) indicates the error rate of the tree *T* over the sample *S* and |leaves(T)| denotes the number of leaves in *T*. *pruned(T,t)* denotes the tree obtained by replacing the node *t* in *T* with a suitable leaf. In the second phase the generalization error of each pruned tree T_0, T_1, . . . , T_k is estimated. The best pruned tree is then selected. If the given dataset is large enough, the authors suggest breaking it into a training set and a pruning set. The trees are constructed using the training set and evaluated on the pruning set. On the other hand, if the given dataset is not large enough, they propose to use cross–validation methodology, despite the computational complexity implications.

❖ **Reduced Error Pruning**

A simple procedure for pruning decision trees, known as reduced error pruning, has been suggested by Quinlan (1987). While traversing over the internal nodes from the bottom to the top, the procedure checks for each internal node, whether replacing it with the most frequent class does not reduce the tree's accuracy. In this case, the node is pruned. The procedure continues until any further pruning would decrease the accuracy. In order to estimate the accuracy, Quinlan (1987) proposes to use a pruning set. It can be shown that this procedure ends with the smallest accurate sub–tree with respect to a given pruning set.

❖ **Minimum Error Pruning (MEP)**

The minimum error pruning has been proposed in (Olaru and Wehenkel, 2003). It performs bottom–up traversal of the internal nodes. In each node it compares the l-probability error rate estimation with and without pruning. The l-probability error rate estimation is a correction to the simple probability estimation using frequencies. If *St* denotes the instances that have reached a leaf *t*, then the expected error rate in this leaf is:

$$\varepsilon'(t) = 1 - \max_{c_i \in dom(y)} \frac{|\sigma_{y=c_i} S_t| + l \cdot p_{apr}(y = c_i)}{|S_t| + l} \tag{5}$$

where $papr(y = c_i)$ is the *a–priori* probability of y getting the value ci, and l denotes the weight given to the *a–priori* probability.

The error rate of an internal node is the weighted average of the error rate of its branches. The weight is determined according to the proportion of instances along each branch. The calculation is performed recursively up to the leaves. If an internal node is pruned, then it becomes a leaf and its error rate is calculated directly using the last equation. Consequently, we can compare the error rate before and after pruning a certain internal node. If pruning this node does not increase the error rate, the pruning should be accepted.

❖ **Pessimistic Pruning**

Pessimistic pruning avoids the need of pruning set or cross validation and uses the pessimistic statistical correlation test instead (Quinlan, 1993). The basic idea is that the error ratio estimated using the training set is not reliable enough. Instead, a more realistic measure, known as the continuity correction for binomial distribution, should be used:

$$\varepsilon'(T, S) = \varepsilon(T, S) + \frac{|leaves(T)|}{2 \cdot |S|} \tag{6}$$

However, this correction still produces an optimistic error rate. Consequently, one should consider pruning an internal node t if its error rate is within one standard error from a reference tree, namely (Quinlan, 1993):

$$\varepsilon'(pruned(T, t), S) \leq \varepsilon'(T, S) + \sqrt{\frac{\varepsilon'(T, S) \cdot (1 - \varepsilon'(T, S))}{|S|}} \tag{7}$$

The last condition is based on statistical confidence interval for proportions. Usually the last condition is used such that T refers to a sub–tree whose root is the internal node t and S denotes the portion of the training set that refers to the node t. The pessimistic pruning procedure performs top–down traversing over the internal nodes. If an internal node is pruned, then all its descendants are removed from the pruning process, resulting in a relatively fast pruning.

- **Error–based Pruning (EBP)**

Error–based pruning is an evolution of pessimistic pruning. It is implemented in the well–known C4.5 algorithm. As in pessimistic pruning, the error rate is estimated using the upper bound of the statistical confidence interval for proportions.

$$\varepsilon_{UB}(T,S) = \varepsilon(T,S) + Z_{\alpha} \cdot \sqrt{\frac{\varepsilon(T,S)\cdot(1-\varepsilon(T,S))}{|S|}} \qquad (8)$$

where $\varepsilon(T, S)$ denotes the misclassification rate of the tree T on the training set S. Z is the inverse of the standard normal cumulative distribution and α is the desired significance level.

Let *subtree(T,t)* denote the subtree rooted by the node *t*. Let *maxchild(T,t)* denote the most frequent child node of *t* (namely most of the instances in *S* reach this particular child) and let *St* denote all instances in *S* that reach the node *t*. The procedure performs bottom–up traversal over all nodes and compares the following values:

1. $\varepsilon UB(subtree(T; t); S_t)$
2. $\varepsilon UB(pruned(subtree(T, t), t), S_t)$
3. $\varepsilon UB(subtree(T, maxchild(T, t)), S_{maxchild}(T,t))$

According to the lowest value the procedure either leaves the tree as is, prune the node *t*, or replaces the node *t* with the subtree rooted by *maxchild(T,t)*.

- **Optimal Pruning**

The issue of finding optimal pruning has been studied in (Bratko and Bohanec, 1994) and (Almuallim, 1996). The first research introduced an algorithm which guarantees optimality, known as OPT. This algorithm finds the optimal pruning based on dynamic programming, with the complexity of $\Theta(|leveas(T)|^2)$, where T is the initial decision tree. The second research introduced an improvement of OPT called OPT–2, which also performs optimal pruning using dynamic programming. However, the time and space complexities of OPT–2 are both $\Theta(|leveas(T^*)| \cdot |\mathrm{internal}(T)|)$, where T^* is the target (pruned) decision tree and T is the initial decision tree. Since the pruned tree is habitually much smaller than the initial tree and the number of internal nodes is smaller than the number of leaves, OPT–2 is usually more efficient than OPT in terms of computational complexity.

❖ Minimum Description Length (MDL) Pruning

The minimum description length can be used for evaluating the generalized accuracy of a node (Rissanen, 1989; Quinlan and Rivest, 1989; Mehta *et al*., 1995). This method measures the size of a decision tree by means of the number of bits required to encode the tree. The MDL method prefers decision trees that can be encoded with fewer bits. The cost of a split at a leaf *t* can be estimated as (Mehta *et al*., 1995):

$$\mathrm{Cost}(t) = \sum_{c_i \in dom(y)} |\sigma_{y=c_i} S_t| \cdot \ln \frac{|S_t|}{|\sigma_{y=c_i} S_t|} + \frac{|dom(y)|-1}{2} \ln \frac{|S_t|}{2} + \ln \frac{\pi^{\frac{|dom(y)|}{2}}}{\Gamma(\frac{|dom(y)|}{2})} \qquad (9)$$

where ***St*** denotes the instances that have reached node ***t.*** The splitting cost of an internal node is calculated based on the cost aggregation of its children.

Figure 6.3 shows the C4.5/ J48(Weka implemented) Pruned Tree Structure.

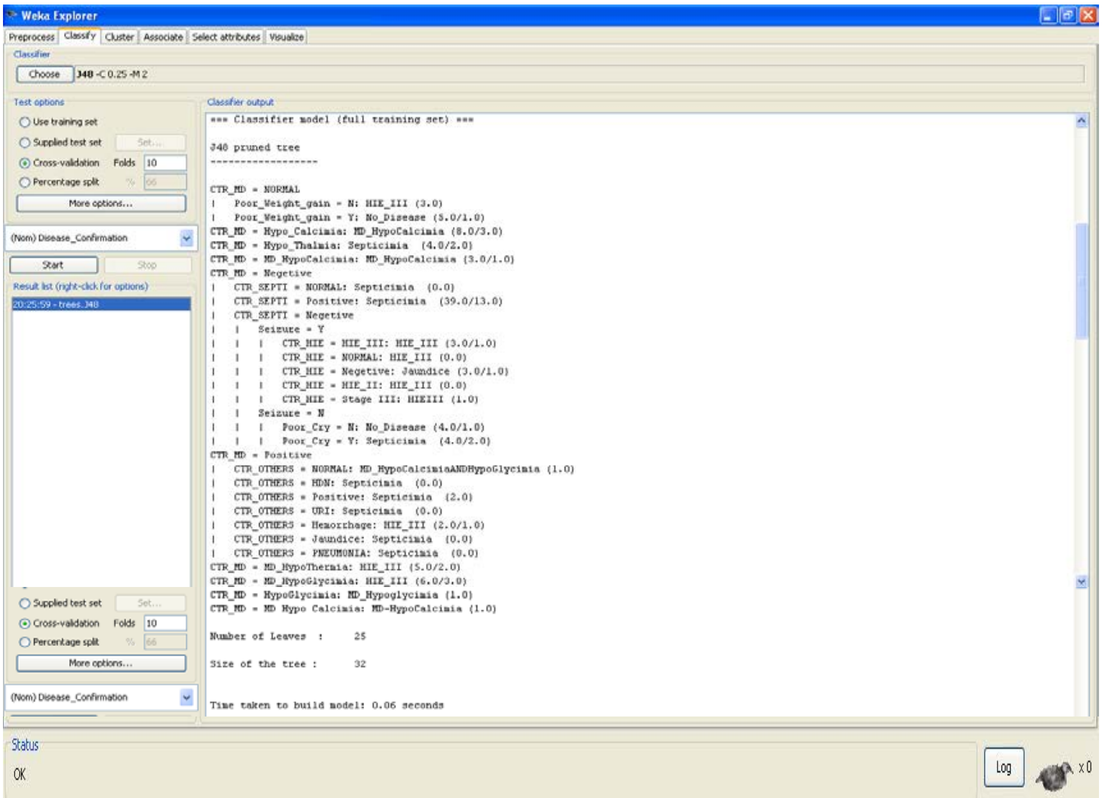

Figure 6.3. C4.5 / J48 Pruned Tree Structure.

6.6.3.3. Receiver Operating Characteristic (ROC)[25]

Receiver operating characteristic (ROC) curves are very useful while findings the accuracy in the prediction. Predicting the disease is now most important in diagnosing the disease. Thus in medical disease diagnosis there is also immense needs of ROC for finding the accuracy the accuracy.

A receiver operating characteristic (ROC) curve allows us to explore the relationship between the sensitivity and specificity of a clinical test for a variety of different cut points, thus allowing the determination of an optimal cut point. The ROC is most helpful when comparing two or more methods. Altman (1998) argues that for a single test the ROC does not really add anything to a table of sensitivity and specificity values for different cut points. However, the graph might be a preferable presentation if there are many possible cut points. Altman (1998) also notes that if the cost of a false positive and a false negative are not equal then the best cut point is not necessarily the point nearest the top left hand corner [25]. Determining the presence and absence of a disease, we generally carry out the tests which gives result in continues measure.

In order to determine the presence or absence of a disease, we often have to carry out a test which provides a result on a continuous measure, for example a blood glucose measurement or a score on the fever of any sick baby. Using this and other information the domain experts decide if the disease is present or absent, so a cut point is selected. Here one of the cut point sides claims that the disease is present and other side reflects the absence of the disease. Using any test we will make diagnostic errors. Commonly used measures of the performance of a test are the sensitivity and specificity. Sensitivity is the probability that we diagnose the disease when it is actually present (the true positive rate). Specificity is the probability that we identify that the disease is absent when it is truly absent (the true negative rate). Ideally we want both sensitivity and specificity to be one. In order to construct a ROC curve we need to calculate the sensitivity and specificity of the test for each possible cut point value. So, for example, suppose the test scale permits scores varying between 10 and 20 and we have measured this score for 500 neonates with known disease status. Then we could explore how these 500 neonates would be diagnosed if the cut point was 10 and we could calculate the sensitivity and specificity. Then we could repeat this exercise using a cut point of 11 and so on.

To make the ROC graph, the X-axis is 1 minus the specificity (the false positive rate) and the Y-axis is the sensitivity (the true positive rate). We draw a diagonal line on the graph from (0,0) in the lower left hand corner to (1,1) in the upper right hand corner. This line reflects the characteristics of a test with no discriminating power.

The closer the graph gets to the upper left hand corner (0,1), the better the test is at discriminating between cases and non cases. An index of the goodness of the test is the area under the curve ± a perfect test has area 1.0, whilst a no discriminating test (one which falls on the diagonal) has area 0.5[25].

An example below is showing the concepts about the prediction forecast.

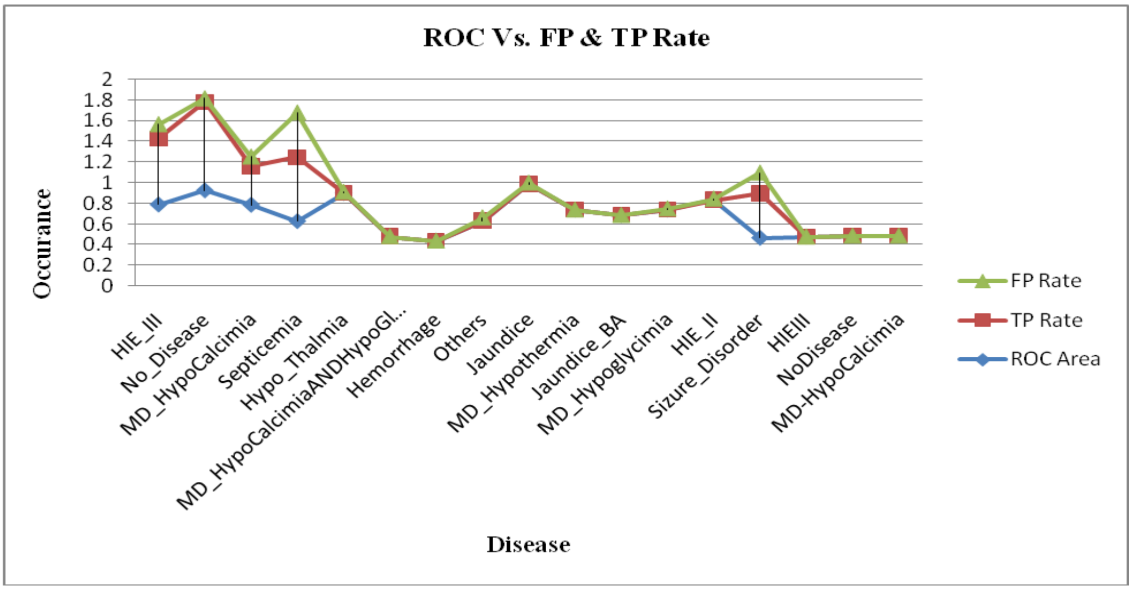

Figure 6.4. ROC Area Vs. FP and TP Rate.

In an Observation, there were 37 neonates when were suffering from Septicemia, and

Table 6.1. Prediction Forecast for Septicemia Diagnosis.

Prediction	**Observation**		
	Diagnosed Septicemia	***Not Diagnosed Septicemia***	***Total***
Diagnosed Septicemia	37	16	53
Not Diagnosed Septicemia	14	28	42
Total	51	44	95

there were 28 neonates who were not suffered from Septicemia. Those two diagonal cells of the 2x2 table, the shaded portion of Table 6.4 represent the two types of correct prediction or forecast. A general terminology is called true positives (TP) and true negatives (TN). The roots of this terminology can be found in medical diagnostic studies when a test is called positive if it shows disease and negative if it does not show disease. Similarly we can consider Diagnosed Septicemia to mean "positive" and Not Diagnosed Septicemia means "negative" in which case we will have 37 true positives and 28 true negatives. There were 16 Not Diagnosed Septicemia found at the time of prediction and none were observed. There were 14 Not Diagnosed Septicemia when it was observed. Now we can easily describe the terminology to call these two cells as false positives (FP) and false negatives (FN). The following Table 6.2 sows a representation of Table 6.1 using the terminology written above.

6.6.3.4. Misclassification Rate (MR)[26]

For prediction and forecasting there are plenty of ways that one can summarize. Below we will be discussing about the various forecast accuracy types. Misclassification rate (MR) is the proportion of all misclassified parts, i.e. Septicemia prediction, the sum of false negative and false positives, out of all diagnosed Septicemia.

$$MR = \frac{FN + FP}{TP + FN + FP + TN} \qquad (10)$$

One minus the misclassification rate is sometimes called "percent correct" or simply "accuracy." MR for the data in Table 6.4 is 40/95=42%.

Table 6.2. Prediction Forecast for Septicemia Diagnosis.

Prediction	**Observation**		
	Positive	Negative	Total
Positive	TP (True Positive)	FP (False Positive)	TP + FP
Negative	FN (False Negative)	TN (True Negative)	FN + TN
Total	TP + FN	TN + FP	TP + FN + FP + TN

In our case of study of the prevalent disease diagnosis using improve data mining concepts also ROC plays very important role of predicting the disease. Almost all sources of neonates including field surveys, domain experts' knowledge and after all patients query provides the neonatal related problems which helps for predicting the disease depending upon sign and symptoms.

There is even dedicated Neonatal Intensive Care Unit (NICU) for getting all the related information which may help predicting the disease. Still forecasting or predicting the neonatal disease goes beyond parents of a little child decision whether the neonates suffering from a particular disease or not. Anything related to neonatal disease if not properly managed then the result may be negative effects.

For this reason collecting data that helps forecasting neonatal conditions and building statistical models to produce predicting from these data have become major in several diagnostic institutions.

Table 6.3. Detailed Accuracy by Class.

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
0.643	0.136	0.45	0.643	0.529	0.786	HIE_III
0.857	0.034	0.667	0.857	0.75	0.924	No_Disease
0.375	0.092	0.273	0.375	0.316	0.784	MD_HypoCalcimia
0.622	0.431	0.479	0.622	0.541	0.625	Septicemia
0	0.011	0	0	0	0.898	Hypo_Thalmia
0	0	0	0	0	0.479	MD_HypoCalcimia AND HypoGlycimia
0	0	0	0	0	0.436	Hemorrhage
0	0.033	0	0	0	0.629	Others
0	0.011	0	0	0	0.987	Jaundice
0	0	0	0	0	0.737	MD_Hypothermia
0	0	0	0	0	0.685	Jaundice_BA
0	0.011	0	0	0	0.737	MD_Hypoglycimia
0	0.011	0	0	0	0.832	HIE_II
0.432	0.201	0.325	0.432	0.371	0.463	Sizure_Disorder
0	0	0	0	0	0.476	HIE_III
0	0	0	0	0	0.484	NoDisease
0	0	0	0	0	0.484	MD_HypoCalcimia
0.432	0.201	0.325	0.432	0.371	0.702	Weighted Avg.

6.7 Implementation

We have used WEKA (Version 3-6-2) [23], an open source, java-enabled, platform independent data mining software. It has different features and capabilities implementing different algorithms, algorithm settings, evaluating model quality, experimentation for the data miners, integrating data bases of different formats etc.

6.7.1. Advantages of Weka as a Data Miner [27]:

Weka is having enormous capacity as a data mining tool. It includes a set of useful algorithms for using data mining jobs easily. Weka consists of data mining tools like filters, association rule learning, clustering, algorithms for attribute selection, regression, classification and many more. Few of them are discussed below:

❖ **Classification**

Several classification algorithms are included in Weka. Bayesian methods include naive Bayes, complement naive Bayes, multinomial naive Bayes, Bayesian networks, and AODE. There are many decision tree learners: decision stumps, ID3, a C4.5 clone called "J48," trees generated by reduced error pruning, alternating decision trees, and random trees and forests thereof. Rule learners include OneR, an implementation of Ripper called "JRip" decision tables, single conjunctive rules, and Prism. There are several separating hyperplane approaches like support vector machines with a variety of kernels, logistic regression, voted perceptrons, Winnow and a multi-layer perceptron.

❖ **Clustering**

Weka is having few standard clustering algorithms which are: KMeans, EM for naive Bayes models, farthest-first clustering, and Cobweb etc.

❖ **Association rule learning**

The standard algorithm for association rule induction is Apriori, which is implemented in the workbench, and there is also Tertius, which can extract first-order rules.

❖ **Regression**

There are implementations of many regression schemes. They include multiple and simple linear regression, pace regression, a multi-layer perceptron, support vector regression, locally-weighted learning, decision stumps, regression and model trees. The standard instance-based learning schemes IB1 and IBk can be applied to regression problems. Moreover, there are additional meta-learning schemes that apply to regression problems, such as additive regression and regression by discretization system.

- **Attribute Selection**

  There are a huge range of filtering criteria are implemented, including correlation-based feature selection, the chi-square statistic, gain ratio, information gain, symmetric uncertainty, and a support vector machine-based criterion Weka. There are also a variety of search methods: forward and backward selection, best-first search, genetic search, and random search. Additionally, principal components analysis can be used to reduce the dimensionality of a problem.

- **Filters**

  Filers are the processes that transform and set instances. Filters may be Supervised and Unsupervised. Unsupervised attribute filters include adding a new attribute, adding a cluster indicator, adding noise, copying an attribute, discretizing a numeric attribute, normalizing or standardizing a numeric attribute, making indicators, merging attribute values, transforming nominal to binary values, obfuscating values, swapping values, removing attributes, replacing missing values, turning string attributes into nominal ones or word vectors, computing random projections, and processing time series data.

Weka is well known data miner as it is having an important aspect of the architecture is its modularity. Different algorithms may be grouped together for better reuse. For an example, we can combine supervised classification, use of J48 algorithm, decision tree learning and arbitrary filters easily with writing up the code. It is having graphical user interfaces (GUI). Thus we may use these tools for data miner.

6.8. Results and Discussions

An excerpt from our model is shown in Table 6.4. The confidence factor is used for pruning; the smaller the value more pruning is done. So, it is required to test and set the model with varying confidence factor value. It is seen from the table that the maximum classification accuracy of the model is ~70% with confidence factor = 0.55 and tested with the training set data. After that there is no increment of accuracy. The size of the tree becomes 38 and the number of leaves becomes 28. The model generated decision tree is shown in Figure 6.5. The Kappa statistics is a measure of accuracy for categorical measurements [24]. In this model the maximum Kappa value is 0.6; a moderate agreement with the test data.

Now, evaluating model quality, there are different popular test metrics for classification models such as confusion matrix, prediction accuracy, receiver operating characteristics (ROC), and lift. We present here a confusion matrix generated by the model as shown in Table 6.5. The right diagonal value shows the accurate classification values for each disease and other values resents the confusions by the model during classifications.

Table 6.4. Excerpt from the Model.

Sl. No	Confidence Factor	No. of Instances	Attributes	No. of Leaves	Size of Tree	Correctly Classified Instances (CCI)	% CCI	Incorrectly Classified Instances (ICCI)	%ICCI	Kappa Statistics
1	0.1	95	19	4	6	50	52.63%	45	47.37%	0.3504
2	0.15	95	19	18	24	61	64.21%	34	35.79%	0.5269
3	0.2	95	19	18	24	61	64.21%	34	35.79%	0.5269
4	0.25	95	19	19	26	62	65.26%	33	34.74%	0.5349
5	0.3	95	19	19	26	62	65.26%	33	34.74%	0.5349
6	0.35	95	19	19	26	62	65.26%	33	34.74%	0.5349
7	0.4	95	19	26	35	65	68.42%	30	31.58%	0.5753
8	0.45	95	19	26	35	65	68.42%	30	31.58%	0.5753
9	0.5	95	19	26	35	65	68.42%	30	31.58%	0.5753
10	0.55	95	19	28	38	66	69.47%	29	30.53%	0.5933
11	0.6	95	19	28	38	66	69.47%	29	30.53%	0.5933
12	0.65	95	19	28	38	66	69.47%	29	30.53%	0.5933
13	0.7	95	19	28	38	66	69.47%	29	30.53%	0.5933
14	0.75	95	19	28	38	66	69.47%	29	30.53%	0.5933

In summary, our model predicts ~70% accuracy while tested with the self data. The maximum Kappa statistics is 0.6. The value of Kappa has to be increased at least to 0.8 for substantial agreement. These findings indicate that we have to do more work either in preprocessing / increasing the data set or in selecting the correct parameters for classifications. Our future work will be towards this goal.

Table 6.5. Confusion Matrix.

=== Confusion Matrix ===

a	b	c	d	e	f	g	h	i	j	k	l	m	n		<-- classified as
13	0	0	1	0	0	0	0	0	0	0	2	0	0		a = HIE_III
0	7	0	1	0	0	0	0	0	0	0	0	0	0		b = No_Disease
1	0	8	0	0	0	0	0	0	0	0	0	0	0		c = MD_HypoCalcimia
1	1	3	32	0	0	0	0	0	0	0	0	0	0		d = Septicimia
0	0	0	2	0	0	0	0	0	0	0	0	0	0		e = Hypo_Thalmia
0	0	0	0	0	1	0	0	0	0	0	0	0	0		f = MD_HypoCalcimiaANDHypoGlycimia
1	0	0	0	0	0	0	0	0	0	0	0	0	0		g = Hemorrhage
1	0	0	4	0	0	0	0	0	0	0	0	0	0		h = Others
0	0	0	0	0	0	0	0	2	0	0	0	0	0		i = Jaundice
3	0	0	0	0	0	0	0	0	0	0	0	0	0		j = MD_Hypothermia
0	0	0	3	0	0	0	0	0	0	0	0	0	0		k = Jaundice_BA
0	0	0	0	0	0	0	0	0	0	0	3	0	0		l = MD_Hypoglycimia
0	0	0	4	0	0	0	0	0	0	0	0	0	0		m = HIE_II
0	0	1	0	0	0	0	0	0	0	0	0	0	0		n = Sizure_Disorder

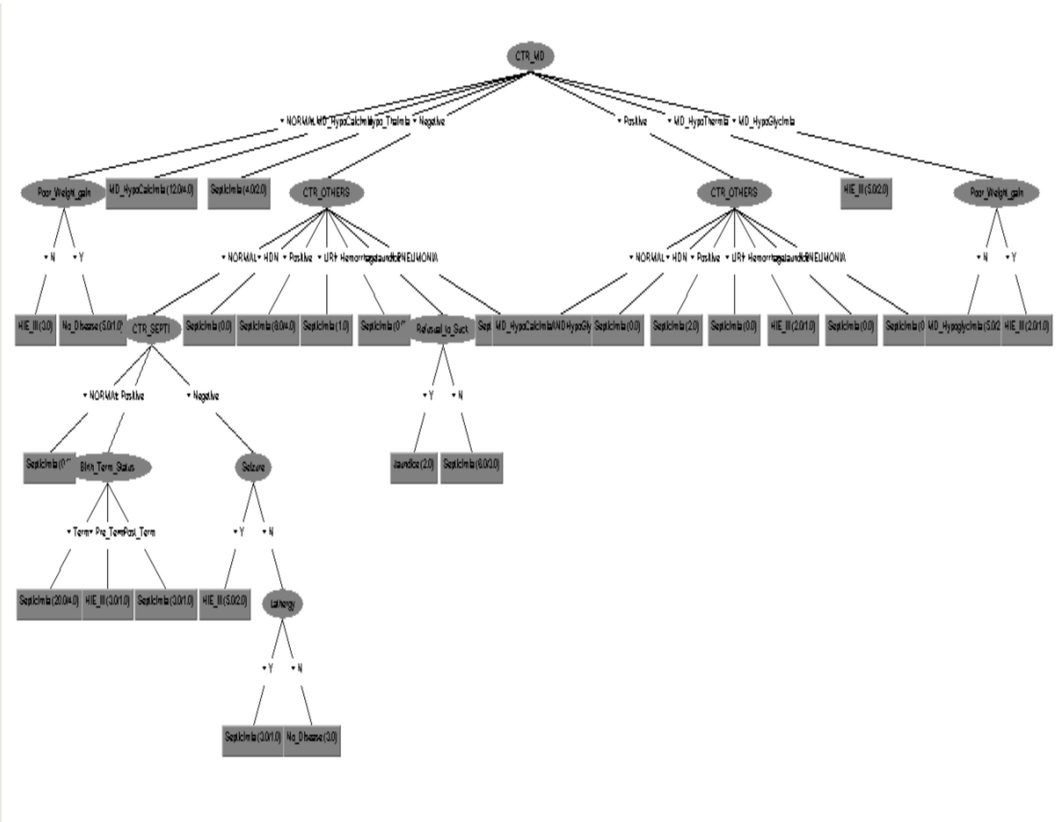

Figure 6.5. Decision Tree

References

1. Kumar D., Verma A., and Sehgal V. K., "*Neonatal Mortality in India*", Rural and Remote Health 7: 833 (online) 2007; Available: http://www.rrh.org.au [Last Accessed 8th October, 2009].

2. Agrawal Rakesh, Imielinski Tomasz, Swami Arun N. "*Mining Association Rules between Sets of Items in Large Databases*". SIGMOD Conference, USA, pp.207-216, 1993.

3. Marx K. A., O'Neil P., Hoffman P., . Ujwal M. L, *"Data mining the NCI cancer cell line compound GI(50) values: identifying quinone subtypes effective against melanoma and* leukemia *cell classes",* Journal-of-chemical-information-and-computer-sciences, United-States, 2003.

4. Forgionne G. A, Gagopadhyay A., and Adya M., "*Cancer Surveillance Using Data Warehousing, Data Mining, and Decision Support Systems",* Health Information Management, Proquest Medical Library, Vol. 21, No. 1, August 2000.

5. Kuo W., Chang R., Chen D. and Lee C. C., "*Data Mining with Decision Trees for Diagnosis of Breast Tumor in Medical Ultrasonic Images*", Breast Cancer Research and Treatment, Dordrecht, Vol. 66, No. 1, March 2001.

6. Chandra Shekar D.V. and Sesha Srinivas V., "*Clinical Data Mining – An Approach for Identification of Refractive Errors*" Proc. Int. Con. of Engineers and Computer Scientists, IMECS 2008, Hong Kong, Vol. I, 19-21 March, 2008.

7. YM Chae, SH Ho, KW Cho , DH Lee, SH Ji, "*Data Mining Approach to Policy Analysis in Health Insurance Domain*", International Journal of Medical Informatics, Vol. 62, No. (2-3), pp. 103-111, July 2001.

8. H. Chen, S.S. Fuller, C. Friedman, W. Hersh, "*Knowledge Management, Data Mining, and Text Mining in Medical Informatics*", Medical Informatics: Knowledge Management and Data Mining in Biomedicine, Springer, pp. 3–33, 2005,

9. Sellappan Palaniappan, Rafiah Awang, "*Intelligent Heart Disease Prediction System Using Data Mining Techniques*", International Journal of Computer Science and Network Security, Vol. 8, No.8, pp. 343-349, August 2008.

10. Nevine M. Labib, and Michael N. Malek, "*Data Mining for Cancer Management in Egypt Case Study: Childhood Acute Lymphoblastic Leukemia*", World Academy of Science, Engineering and Technology, 2005.

11. X. Wu, "*Data Mining: An AI Perspective*", IEEE Computational Intelligence Bulletin, Vol. 4, No. 2, pp. 23-26, December 2004.

12. Pstterson Dan W., "*Introduction to Artificial Intelligence and Expert Systems*", PHI, pp. 340, 2003.

13. http://en.wikipedia.org/wiki/Decision_tree_learning Available: [Last accessed 5th February, 2010].

14. Quinlan J. R., "*C4.5: Programs for Machine Learning*". Morgan Kaufmann Publishers, 1993.

15. Hornick Mark F., Marcade Eric, and Venkayala Sunil, "*Book excerpt: Java Data Mining Concepts*", JavaWorld.com, 22th February 2007.

16. http://www.decisiontrees.net/book/export/html/37 Available: [Last accessed 4th February, 2010].

17. http://www.decisiontrees.net/book/export/html/44 Available: [Last accessed 4th February, 2010].

18. Mingers J., "*An Empirical Comparison of Pruning Methods for Decision Tree Induction*", Machine learning Archive, Vol. 4, No.2, pp. 227-243, November 1989.

19. Espoito F., Malerba D., and Semeraro G., "*A Comparative Analysis of Methods for Pruning Decision Trees*", IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol.19, No.5, pp. 476-491, 1997.

20. Pham Huy Nguyen Anh, Triantaphyllou Evangelos, "*The Impact of Overfitting and Overgeneralization on the Classification Accuracy in Data Mining*", Soft Computing for Knowledge Discovery and Data Mining, Part 4, pp. 391-431, 2008.

21. L Breiman., Friedman J., Olshen R., and Stone C., "*Classification and Regression Trees*", Wadsworth Int. Group, 1984.

22. Rokach Lior, Maimon Oded, "*Decision Trees*", Data and Knowledge Discovery Hand Book, Chapter 9, pp. 165-192.

23. Hall M., Frank E., Holmes G., Pfahringer B., Reutemann P., Witten I. H., "*The WEKA Data Mining Software: An Update*", SIGKDD Explorations, Vol. 11, No. 1, 2009.

24. Landis J. R. and Koch G. G., "*The Measurement of Observer Agreement for Categorical Data*", Biometrics, Vol. 33, pp. 159 – 174, 1977.

25. O'Connell B. and Myers H., "*Information Point: Receiver Operating Characteristic (ROC) Curves*", Blackwell Science Ltd, Journal of Clinical Nursing, Vol. 11, pp. 134-36, 2002.

26. "*Receiver Operating Characteristic (ROC) Curves*", Statistical Data Analysis, Mithat Gönen, Memorial Sloan-Kettering Cancer Center, SUGI, pp. 210-231.

27. Frank Eibe, Hall Mark, Holmes Geoffrey, Kirkby Richard, Pfahringer Bernhard, Witten Ian H., "*WEKA - A Machine Learning Workbench for Data Mining*", Chapter 1, pp. 5-6.

Chapter 7

Soft Computing Approach in Neonatal Disease Diagnosis††

7.1. Introduction

This chapter describes a way of designing a hybrid decision support system in soft computing paradigm for neonatal disease diagnosis. Hybridization includes the application of rough set theory and C4.5 classification algorithm. Using rough set approach we generate rules for classification. C4.5 classification algorithm has been used for the same data set taken in chapter 6. A comparative study has been presented in this chapter for the neonatal domain. This methodology also helps in studying the variation of different classification parameters including confidence factor, folding etc.

7.1.1. Soft Computing Paradigm

Soft computing is an area of Artificial Intelligence, in which, there is process of developing intelligent and wiser systems. This Intelligence is supposed to give the power to derive the answer and not simply arrive to the answer. Even Soft Computing techniques give lots of help towards developing decision support systems. There are few involvement of criteria's like machine intelligence, purity of thinking, liberty to work, dimensions, complexity and fuzziness handling capability etc. for successful development of decision support system. The basic purpose of developing DSS or ES is lies on the working principles as similar to human beings can do. The wisdom of human beings can be replicated in computers in some artificial manner.

Soft Computing is now treated as a new multidisciplinary field. The idea of Soft Computing was initiated in 1981 by Lotfi Zadeh. The main purpose was to construct new generation Artificial Intelligence, known as Computational Intelligence. Zadeh defined Soft Computing in its latest manifestation as the fusion of the fields of Neuro-computing, Fuzzy Logic, Evolutionary and Genetic Computing, and Probabilistic Computing into one multidisciplinary system. The main goal of Soft Computing is to develop intelligent machines and to solve nonlinear and mathematically un-modeled system problems [1]. There are mainly two advantages of soft computing applications. Firstly, for solving non linear problems where no mathematical models are present as such, it provides immense help. Secondly, soft computing provides human knowledge

†† This chapter is based on the publication made by the author entitled "Neonatal Disease Diagnosis with Soft Computing", Proc. International Conf. on Computing and System, ICCS-2010, pp. 27-34, November, 2010.

such as recognition, cognition, understanding, learning, and others into the fields of computing. This basic advantages results a successful intelligent system which is automated in nature and can utilized self-tuning systems.

Soft computing differs from conventional hard computing. Unlike hard computing, it is tolerant of imprecision, uncertainty, partial truth, and approximation. The guiding principle of soft computing is that, it exploits the tolerance for imprecision, uncertainty, partial truth, and approximation to achieve tractability, robustness and low solution cost.

The basic ideas underlying soft computing in its current incarnation have links to many earlier influences of fuzzy sets, analysis of complex system, decision process, probability and possibility theory. The inclusion of neural computing and genetic computing in soft computing came at a later point. At this moment, the principal constituents of soft computing are fuzzy logic, neural computing, evolutionary computation, machine learning and probabilistic reasoning, with the latter subsuming belief networks, chaos theory and parts of learning theory. More importantly, soft computing is not a mixture of techniques, but it is a partnership in which each of the partners contributes a distinct methodology for addressing problems in its domain. In this perspective, the principal constituent methodologies in soft computing are complementary rather than competitive. Furthermore, soft computing may be viewed as a foundation component for the emerging field of conceptual intelligence [2].

7.1.2. Importance of Soft Computing in Medical Domain

There are several efforts that have been taken in artificial intelligence to medical reasoning problems which basically based upon rule-based systems primarily. These kinds of rule based programs are easy to create, because their knowledge is cataloged in the form of if-then rules. In relatively well-constrained domain such programs show skilled behavior. But in real-life situations, there is considerable degradation of performance due to both presence of ambiguity and incomplete information as well as inadequate modeling of the diseases by the rules.

Other conventional methods like Bayes classifier, flow charts and many others are also unable to deal with most complex clinical decision making problems. Thus it is very much necessary to provide health care systems from different domains with software tools combining different machine learning models with the capability of extracting information from large inadequate data sets, without manual intervention. There are currently large repositories of data in several domains of health care. These data may support clinical, bibliographic, administrative, or epidemiological studies and helps a lot in medical decision making also. Medical decision making are dealing with new challenges where various criteria of disease diagnosis plays important role. Medical decision making in different medical tasks concerning both disease treatment comprising of disease diagnosis, disease progression, monitoring the effect of treatments, etc. and research activities like drug discovery, robot-assisted therapy, understanding of human body structure and neural connectivity, etc. are using soft computing techniques for getting better results.

The application of soft computing methodologies in disease diagnosis, particularly for the neonatal disease diagnosis has increased in the last few years due to their capability to deal with the uncertainty and imprecision underlying most of real-world problems and, specifically, medical applications requiring the use of decision support system techniques under soft computing paradigm. Beside, the aims of using soft computing techniques in medical domain are to provide a transverse research formation from different industrial sectors. Among the various sector, scientific researches, technology developments, practical uses in hospitals, serving as useful tools are few of them.

7.1.3. Objective of the Study

The main objective of this study is to design a medical decision support system, for neonatal disease diagnose and management. This decision support system basically uses knowledge based network in combination with rough set theory and genetic algorithms in soft computing paradigm. The projected system is capable of using the self-learning, parallelism, and default tolerance characteristics model development. This is having knowledge encoding capabilities of rough set theory, and the adaptive, parallel and robust searching characteristics of genetic, exhaustive search. The model is built on the data of neonatal signs and symptoms for detecting and predicting the disease in its various stages and shows the control mechanisms also.

Neonatal disease diagnosis is one of the most vital issues in healthcare. Computer aided systems have been developed in order to diagnose diseases by examining the internal human organs through various procedures. Many techniques have been utilized in order to develop these systems. For the sake of our study we need to focus on the areas of neonatal information related to their health status and survival condition again. Among the different phases of the development of the child, Neonatal phase is the most significant phase. This has already been discussed in the chapter 2. Neonatal phases are also belongs to a special risk group; where the risks are related with the prevalent disease, proportionate growth, development, and survival. In India, the mortality and morbidity are high [3] especially in rural and remote areas. The prime reason of such mortality and morbidity is prevalent diseases. An estimated two-thirds of childhood deaths occur in infancy, and, in turn, two-thirds of infant deaths occur in the first month of life, the neonatal period. There are a number of neonatal prevalent diseases have been seen particularly in north eastern tarai region and a number of parameters or factors are involved in this regard. The typical neonatal health problems which are the common causes of neonatal deaths are: Neonatal Septicemia, Birth asphyxia, Hypoxic-Ischemic Encephalopathy (HIE), Preterm, Low birth weight, Failure to Gain Weight, Delayed breastfeeding, Problems in breastfeeding, Diarrhea, Hemorrhage, Conjunctivitis, Skin Infection, Abnormal Jaundice, Me conium Aspiration, Hyaline Membrane Disease (HMD), Pneumonia, Upper Respiratory Infection (URI), Hypothermia, Umbilical Sepsis, Tetanus, Convulsive Disorder, Unexplained fever. This has thoroughly discussed in previous chapters of 2, 3 and 5. Still for the use of the diagnosis under soft computing techniques, this really needs to be taken into considerations.

Before taking any suitable measure by any appropriate authority to reduce mortality and morbidity due to different diseases of neonates, it might be useful to know the prevalent disease statistics and pattern of a region [4]. The study discussed in chapter 2, shows that the incidence of diseases namely, neonatal septicemia, HIE, metabolic disturbances, neonatal jaundice etc. are quite high in the North Eastern parts of India especially in Tarai region. A table was created as a result of that study for further investigations. In a study [5], chapter 6, the authors have applied a data mining technique (J48/ C4.5) on this data set (which included clinical test reports) using WEKA [6] implementation. The model thus developed predicts disease(s) with ~70% accuracy with confidence factor = 0.55 and tested on the training set data with confusion matrix approach along with the Kappa statistics having the value 0.6. As it is observed that the accuracy level is not good enough in spite of setting various external parameters' values needed for optimizing the model. The reasons might be the weakness of the classification algorithm used the nature of the data itself - incompleteness, the other uncertainties involved in the data of the table used. Moreover, from this classification model, it was not possible to find the minimum number of attributes contributing significantly on the decision matrix along with the minimization of rules.

Taking all the above factors in consideration, the present study is planned to obtain a better model for the problem domain using a soft computing technique, namely rough set theory. A hybridization methodology is proposed for a better decision support system development. Also, a comparison study between rough set results and C4.5 algorithm results has presented. The data set is used in this study which excludes the clinical test reports leading to differential diagnosis based on 14 attributes including signs and symptoms (see section 7.4.).

The chapter is summarized as below:

In the section 7.2, the basics of soft computing and rough set theory are discussed. In section 7.3, we discuss decision tree and C4.5 algorithm. Section 7.4 presents implementation and results details of the study. Section 7.5 Rough set and C4.5 comparison and finally section 7.6 presents our conclusion and discussion.

7.2. Soft Computing and Rough Set Theory

Soft computing is a consortium of methodologies which works synergetically and provides, in one form or another, flexible information processing capability for handling real life ambiguous situations [7]. The aim of soft computing is to exploit the tolerance for uncertainty, imprecision, partial truth and approximate reasoning in order to achieve robustness, tractability, low cost solution, and close resemblance with human like decision making. Soft computing paradigm includes Fuzzy Sets (FS), Rough Sets (RS), Artificial Neural Networks (ANN), Genetic Algorithms (GA's), Genetic Programming (GP), Support Vector Machines (SVM), Swarm Optimization (SO), Ant Colony Optimization (ACO), Memetic Algorithms (MA) and others.

7.2.1. Rough Sets

Rough sets theory was first presented by Pawlak in the 1980's [8]. Rough set is a formal approximation of a crisp set in terms of a pair of sets which give lower approximation with positive region and upper approximation with negative region. In between there is a boundary. Let there be an information system I = (U, A) (attribute - value system), where 'U' be the universe of discourse and is a non-empty set of finite objects and 'A' is a non-empty finite set of attributes. With any $P \subseteq A$, there is an associated equivalence relation IND(P). The relation IND(P) is called P–indiscernibility relation. Let $X \subseteq u$ be a target set we wish to represent using attribute subset P. Now, the target set X can be approximated using only the information contained within P by constructing P-lower ($\underline{P}X$) and P-upper ($\overline{P}X$) approximation of X. The tuple ($\underline{P}X$, $\overline{P}X$) is called a rough set. The accuracy of the rough-set representation of the set X can be given [9] by the following:

$$\chi P(x) = \left|\underline{P}X\right| \Big/ \left|\overline{P}X\right| \qquad (1)$$

Rough set theory is an intelligent technique for managing uncertainties that is used for the discovery of data dependencies, reducing redundancies, evaluating the importance of attributes. It also used for discovering the patterns in data, and classify objects. There are several useful features of rough sets such as:

(i) Extraction of rules from data sets in the form of if-then rules;
(ii) It requires no external parameters unlike other intelligent techniques except the data itself;
(iii) It can predict whether the data is complete or not.
(iv) The computation of reduct and core using rough set theory is another important feature.

7.2.2. Rough Set as Decision Set [10]

In rough set theory, representing a decision table is done by T = (U, A, C, D), where U is universe of discourse, A is a set of primitive features, and C, D $\subset$ A are the two subsets of features that are called condition and decision features, respectively. Let a $\in$ A, P $\subseteq$ A. A binary relation IND (P), called the Indiscernibility relation, is defined as follows:

$$\text{IND}(P) = \{ (x, y) \in U \times U: \text{for all } a \in P,\ a(x) = a(y) \} \qquad (2)$$

Let U / IND (P) denote the family of all equivalence classes of the relation IND (P). For simplicity of notation U/P will be written instead of U/ IND (P). Equivalence classes U/INDI and U/IND (D) will be called condition and decision classes, respectively. Let R$\subseteq$ C and X $\subseteq$ U,

$$\underline{R}\,X = \cup\{ Y \in U / R : Y \subseteq X \} \text{ and} \qquad (3)$$

$$\bar{R}\,X = \cup \{ Y \in U/ R:\ Y \cap X \neq \Phi \} \qquad (4)$$

Here $\underline{R}X$ and $\bar{R}X$ are said to be R-lower and R upper approximations of X and ($\underline{R}X$, $\bar{R}X$) is called *R-rough set.* If X is R-definable then $\underline{R}X$ = $\bar{R}X$ otherwise X is R-Rough.

The boundary $BN_R(X)$ is defined as :

$$BN_R(X) = \underline{R}X - \bar{R}X \qquad (5)$$

Hence, if X is R-definable, then $BN_R(X) = \Phi$.

7.2.3. Reduct [11]

A reduct is a subset of attributes that are jointly sufficient and individually necessary for preserving a particular property of a given information table. One of the basic ideas of rough set application is the concept of attribute reduction. The objective of reduct generation is to reduce the number of attributes, and at the same time, preserve a certain property that we want in future applications. For an example, assuming that there is a need of classifying few properties. A reduct should be able to preserve the original classification power provided by the whole attribute set. This power may be interpreted by syntax properties and semantics properties for both positive and boundary rule sets [11].

Let $c \in C$. A feature c is dispensable in T, if $POS_{(C-I)}$ (D) = POS_C (D); otherwise feature c is indispensable in T. c is an independent if all $c \in C$ are indispensable. Taking this consideration we may say that a Reduct is a A set of features $R \subseteq C$ is called a reduct of C, if T' = {U, A, R, D} is independent and POS_r (D). In other words, a reduct is the minimal feature subset preserving the above condition [10].

7.2.4. **CORE** [11]

The core is the intersection of all reducts. No element of the core can be eliminated affecting the classification power of attributes. The core of a set T is the set of all indispensable attributes of T, which can be given as

$$CORE(T) = \cap RED(T) \quad (6)$$

where RED (T) is the set of all reducts of T.

If for a dependency $C \Rightarrow D$, D depends on E where $E \subseteq C$ then E is called as relative D-reduct of C. Relative D-core of *C* is given by

$$CORE_D I = \cap RED_D I \quad (7)$$

where Red_D I is the family of all D-reducts of *C*.

Similarly on the basis of the Reduct explanation, CORE I denotes the set of all features indispensable in C. E have CORE I = ∩ REDI , where REDI is the set of all reducts of C.

7.3. Decision Tree and C4.5 Algorithm

7.3.1. General

The decision tree algorithm is a common and one of the most popular algorithms used in data mining because it is easy to understand how it makes predictions. Even in Soft Computing Applications also it has great importance. The goal is to create a model that predicts the value of a target variable (dependent) based on several input variables (independent). A tree can be 'learned' by splitting the source set into subsets based on an attribute value test. This process is repeated on each derived subset in a recursive manner called recursive partitioning. The recursion is completed when the subset at a node all has the same value of the target variable, or when splitting no longer adds value to the predictions [12]. Data comes in records of the form:

$$(X, Y) = (x_1, x_2, x_3, \ldots\ldots x_n, Y) \quad (8)$$

The dependent variable, Y, is the target variable that we are trying to understand, classify or generalize. The vector X is composed of the input variables x_1, x_2, x_3, x_4 ……etc., that are used as input for the model. In data mining, there are different categories of trees available in the literature. We use classification tree analysis, as because our predicted outcome is the class to which the data belongs.

A Decision Tree is basically built up using a top-down approach. The leafs are the instances which are assigned labels of most regular class in the data set. If it is not generated then it may be placed in to an internal node and this process is continues till the last instance reached. In this approach, it creates partition of several instances of all block. It expands recursively. This decision tree can be pruned. It actually restricts the growth of the tree before it generates.

There may two different kinds of tests done during the decision tree construction. Firstly, Univariate test, when there is only one variable. Secondly, Multivariate test, when variables are more to tests. The decision tree rules are applied for traversing the tree from root to different leaf nodes. Figure 6.2 and 6.4 on chapter 6 depicts the construction of decision tree on the data set of the neonates.

7.3.2. C4.5 algorithm [13]

The C4.5 algorithm is based on information gain which is again based on the concept of entropy of information theory. For a random variable X with N outcomes { x_i : I = 1,2,3,4,……..N}, the Shannon entropy, a measure of uncertainty and denoted by H(X), is defined as,

$$H(X) = - \sum_{i=1}^{N} P(x_i) \log_2 P(x_i) \qquad (9)$$

where $P(x_i)$ is the probability mass function of outcome x_i. This algorithm C4.5 uses a set of training data S (S = s_1, s_2, s_3, ………s_n) of classified samples for building decision trees. Each sample s_i will be having a set of attributes or features for classification along with a class attribute. Based on the normalized information gain (difference in entropy), at each node of the tree, C4.5 chooses one attribute of the data that most effectively splits its set of samples into subsets enriched in one class or the other. The attribute with the highest normalized information gain is chosen to make the decision. It then uses recursion on the smaller sub lists for further building and completion of decision tree.

7.4. Implementation and Results

7.4.1. Attributes

The following first 14 attributes are independent attributes as input and the last attribute i.e., Disease_differential is dependent attribute as output.

@attribute Birth_Term_Status {Term,Pre_Term,Post_Term}
@attribute Birth_Weight_Status {Normal,LBW,VLBW,ELBW}
@attribute Age_in_Hours>72 {Y,N}
@attribute Lathergy {Y,N}
@attribute Refusual_to_Suck {Y,N}
@attribute Poor_Cry {N,Y}
@attribute Poor_Weight_gain {N,Y}
@attribute Hypothalmia {N,Y}
@attribute Sclerema {N,Y}
@attribute Excessive_Jaundice {N,Y}
@attribute Bleeding {N,Y}
@attribute GI_Disorder {N,Y}
@attribute Seizure {Y,N}
@attribute Sluggish_Neonatal_Reflex {Y,N}
@attribute Disease_differential
{HIE_III,No_Disease,MD_HypoCalcimia,Septicimia,Hypo_Thalmia,MD_HypoCalcimia,Hemorrhage,Others,Jaundice,MD_Hypothermia,Jaundice_BA,MD_Hypoglycimia,HIE_II,Sizure_Disorder}

The above attributes are in form of Weka .arff format data.

7.4.2. Implementation on Rough Set

7.4.2.1. Reduct and Core Generation

Reduct is a subset of attributes which can, by itself, fully characterize the knowledge in the database. However, for a problem domain, there might be more than one such reduct. If there are more than one reduct, some attributes might be common to all such reducts; those attributes are called core. The core attribute(s) are indispensable for an information system. We had used initially three algorithms [14][15] namely, Exhaustive algorithm, Genetic algorithm, and Dynamic reducts for finding reduct and core with the help of RSES 2.2 [16], a software tool that provides the means for analysis of tabular data sets with the use of various methods, in particular those based on Rough Set Theory. For a comparative study, we attempted other algorithms [17][18] with the help of ROSETTA [19] software tool. Comparative results are shown in the Table 7.1.

Table 7.1. Comparison among Different Algorithms for Reduct and Core.

Software Tool	Methods/ Algorithms	No. Reduct	Length Reducts	Size of core
RSES	Exhaustive	3	10	8
	Genetic	3	10	8
	Dynamic	3	10	8
	Genetic	3	10	8
ROSETTA	Johnson (with approx. solutions)	1	8	8

7.4.2.2. Attribute dependency

One of the most important aspects of predictive analysis is the discovery of attribute dependencies. This essentially means that one has to discover which attributes are strongly related to which other attributes. These strong relationships need further investigation, and that will ultimately be of use in predictive modeling. Table 7.2 presents the results of such dependencies.

Table 7.2. Core Attributes.

Software Tool	Methods/ Algorithms	Core attributes
RSES	Exhaustive	Birth_Term_Status, Birth_Weight_Status, Lathergy, Refusual_to_Suck, Poor_Cry, Hypothalmia, **Bleeding**, Sluggish_Neonatal_Reflex
	Genetic	Birth_Term_Status, Birth_Weight_Status, Lathergy, Refusual_to_Suck, Poor_Cry, Hypothalmia, **Bleeding**, Sluggish_Neonatal_Reflex
	Dynamic	Birth_Term_Status, Birth_Weight_Status, Lathergy, Refusual_to_Suck, Poor_Cry, Hypothalmia, **Bleeding**, Sluggish_Neonatal_Reflex
ROSETTA	Genetic	Birth_Term_Status, Birth_Weight_Status, Lathergy, Refusual_to_Suck, Poor_Cry, Hypothalmia, **Bleeding**, Sluggish_Neonatal_Reflex

	Johnson (with approx. solutions)	Birth_Term_Status, Birth_Weight_Status, Lathergy, Refusal_to_Suck, Poor_Cry, Hypothalmia, **Seizure**, Sluggish_Neonatal_Reflex

It is observed that the reducts generated by all the algorithms exclude 'Excessive_Jaundice', 'Sclerema', and 'GI_Disorder'. Moreover, it is evident from the above Table 7.2 that Johnson algorithm differs from other algorithms in one attribute. We are to study all the above facts during rule generation and prediction results through confusion matrix.

7.4.3. Rule Generation and Prediction

Table 7.3 shows rules generated using different methods. There are different techniques for testing the prediction accuracy such as confusion matrix, prediction accuracy, receiver operating characteristics (ROC), and lift. This work deploys confusion matrix.

Table 7.3. Rules and Prediction with Confusion Matrix.

Software tool	Methods/Algorithms	No. of rules	Prediction Accuracy (%)
RSES	Exhaustive without reduct	3568	70
	Genetic without reduct	1904	70
	Exhaustive with reduct	192	70
	Genetic with reduct	192	70
	Dynamic with reduct	192	70
ROSETTA	Genetic with reduct	192	71.6
	Johnson (with approx. solutions) with reduct	60	70.5

It is observed that ROSETTA implementation of genetic algorithm with reduct shows the best performance. It is also observed that prediction accuracy does not decrease with reucts where number of rules are as low as 192 compared to 3568 and 1904. There is effectively rather no contribution observed of three attributes 'Excessive_Jaundice', 'Sclerema', and 'GI_Disorder' in decision making. Moreover, Johnson algorithm offers somewhat a comparable accuracy (70.5%) with only 60 rules with only one reduct of length eight. So, Johnson algorithm may be used for further investigations.

It is now to test the comparative contributions of two attributes namely 'Bleeding' , and 'Seizure' using two sets of different core attributes (see table 7.2) with the help of Johnson algorithm. The findings are presented in table 7.4.

Table 7.4. Contributions of 'Bleeding' and 'Seizure'.

Software Tool	**Methods / Algorithms**	**Core attributes**	**No. of Rules**	**Prediction accuracy (%)**
ROSETTA	Johnson (with approx. solutions)	Birth_Term_Status, Birth_Weight_Status, Lathergy, Refusual_to_Suck, Poor_Cry, Hypothalmia, **Bleeding**, Sluggish_Neonatal_Reflex	55	68.4
	Johnson (with approx. solutions)	Birth_Term_Status, Birth_Weight_Status, Lathergy, Refusual_to_Suck, Poor_Cry, Hypothalmia, **Seizure**, Sluggish_Neonatal_Reflex	60	70.5

From the above table 7.4, we see that 'Seizure' plays somewhat a more significant role compared to 'Bleeding'.

7.5. Rough Set and C4.5 : A Comparison

It is now to present a performance comparison between rough set approach and C4.5 algorithm as a predictive model. We have used WEKA [6] software tool implementation for C4.5 approach. The results are shown in Table 7.5.

Table 7.5. Comparison between Rough Set and C4.5

Software tool	Methods/Algorithms	Prediction Accuracy (%)
WEKA	J48 / C4.5 without reduction	65.3
	J48 / C4.5 with 8 core* attributes including ‘Bleeding’	60
	J48 / C4.5 with 8 core* attributes including ‘Seizure’	63.2
	ID3 without reduction	71.6
	ID3 with 8 core* attributes including ‘Bleeding’	68.4
	ID3 with 8 core* attributes including ‘Seizure’	70.5
RSES	Exhaustive without reduct	70
ROSETTA	Genetic with reduct	71.6

*Table 7.4.

It seen that the predictive accuracy of C4.5 is much lower than that of rough set classification. But, however, the rough set classification accuracy is seen to compare favourably with the well-known ID3 [20], the Interactive Dichotomizer 3 classifier algorithm. Looking through the results of the above Table 7.5, it is evident that attribute ‘Seizure’ plays a more significant role than that of ‘Bleeding’.

7.6. Conclusion and Discussion

Soft computing paradigm has been playing vast role in different fields including medical domain, particularly where a number of uncertainties involved. Current scenario reflects that with the developments of society, numbers of population increases also. Due to lake of proper medical infrastructure children are suffering a lot from various kind of disease. Especially if it in case of neonates, the problems are much greater. Hence for proper diagnosing of disease, more sophisticated methods that can model non-linear, not very complicated, real world applications are needed. In this perspective, soft computing methods have been successfully applied to solve non-linear problems in medicine, disease diagnosis and management. Soft computing methods and its methodologies are used to find rough solutions for real-world problems which contain various kinds of inaccuracies and uncertainties can be alternative methods to statistical methods. The soft computing paradigms, with the use of rough set computing and decision tree approach with rule set designing for the development of decision support system very much support.

The objective has been fulfilled in this chapter, by designing a decision support system, for diagnosing neonatal disease, using a knowledge-based network in combination with rough set theory and C.45 Decision tree in soft computing paradigm. The proposed system is able to exploit the parallelism, self-learning, and fault tolerance characteristics of rough set models, knowledge encoding capabilities of rough set theory, and the adaptive, parallel and robust searching characteristics of genetic, exhaustive search algorithms.

This work concentrated on neonatal disease. After explaining the complexity of the domain, it is argued that any automated decision support system certainly should assist the health people in their decision making. We presented here a hybrid approach with rough set and C4.5 classification algorithm. A comparative study is presented here. Appendix D shows the graphical representation of the RSES implementation for Rule set and Reduct generation and various ROC analysis consecutively. We observe that neither all attributes are required nor play significant role. From our study, it is evident that the three attributes, namely, ‘Excessive_Jaundice’, ‘Sclerema’, and ‘GI_Disorder’ do not play any significant role in prediction accuracy. Moreover, it is found that ‘Seizure’ plays a more significant role than ‘Bleeding’. So, a less number of attributes can be used in rational decision making which will reduce the number of rules, and the search space as well.

References

1. Zadeh, Lotfi, "*Fuzzy Logic and Soft Computing*", Plenary Speaker, Proceedings of IEEE International Workshop on Neuro Fuzzy Control. Muroran, Japan, 1993.

2. Zadeh, Lotfi, "*What is Soft Computing*" Soft Computing, Springer-Verlag Germany/USA, 1997.

3. Kumar D., Verma A., and Sehgal V. K., "*Neonatal Mortality in India*.", Rural and Remote Health 7: 833 (online) 2007; http://www.rrh.org.au Available: [Last accessed 8th October, 2009].

4. Roy Chowdhury D., Chatterjee M., and Samanta R. K., "*A Study of the Status of New Born in Terai Region of West Bengal*", Advances in Modeling C, vol. 68, no. 3, pp. 44-52, 2007.

5. Roy Chowdhury D., Chatterjee M., and Samanta R. K., "*Data Mining for Neonatal Prevalent Disease of North Bengal Districts*", Proc. Nat. Sem. on Data Mining and Decision Support, Vidyasagar University, India, pp. 24-31, March 4-5, 2010.

6. Hall M., Frank E., Holmes G., Pfahringer B., Reutemann P., Witten I. H., "*The WEKA Data Mining Software: An Update*", SIGKDD Explorations, Vol. 11, issue 1, 2009.

7. Zadeh L. A., "*Fuzzy logic, neural networks, and soft computing*", Comm. ACM, Vol. 37, pp. 77-84, 1994.

8. Pawlak Z., "*Rough sets*", Int. J. of Parallel Programming, Vol. 11, No. 5, pp. 341-356, 1982.

9. Pawlak Z., "*Rough Sets: Theoretical Aspects of Reasoning about Data*", Dordrecht: Kluwer Academic Publishing, 1991.

10. Yellasiri Ramadevi, Rao C.R., Reddy Vivekchan, "*Decision Tree Induction using Rough Set Theory- Comparative Study*", Journal of Theoretical and Applied Information Technology(JATIT), pp. 110-114, 2007.

11. Pawlak, Z. "Rough Sets", International Journal of Computer and Information Sciences, Vol. 11, pp. 341-356, 1982.

12. http://en.wikipedia.org/wiki/Decision_tree_learning Available: [Last accessed 5th Feb, 2010].

13. Quinlan J. R., "*C4.5: Programs for Machine Learning*". Morgan Kaufmann Publishers, 1993.

14. Bazan J., "*A Comparison of Dynamic and Non-Dynamic Rough Set Methods for Extracting Laws From Decision Table*", In L. Polkowski, A. Skowron (eds.), Rough Sets in Knowledge Discovery, Physica-Verlag, Heidelberg, pp. 321-365, 1998.

15. Bazan J. et.al., "*Rough Set Algorithms in Classification Problems*", In L. Polkowski, S. Tsumoto, and T. Lin (eds.), Rough Set Methods and Applications, Physica-Verlag, Heidelberg, pp. 49-88, 2000.

16. RSES 2.2 User's Guide, Warsaw University, http://logic.mimuw.edu.pl/~rses, Available: [Last accessed 19th January, 2010].

17. Vinterbo S. and Ohrn A., "*Minimal Approximate Hitting Sets and Rule Templates*", Int. J. Approximate Reasoning, vol. 25, no. 2, pp.123-143, 2000.

18. Johnson D. S., "*Approximate Algorithms for Combinatorial Problems*", J. of Computer and System Sciences, Vol. 9, pp. 256-278, 1974.

19. The ROSETTA Homepage, http://www.idi.ntnu.no/~aleks/rosetta/ Available: [Last accessed 19th January, 2010].

20. Pao Y. H., "*Adaptive Pattern Recognition and Neural Networks*", Addison-Wesley, Reading, MA, 1989.

C H A P T E R **8**

Diagnosis of Neonatal Disease using Rough Set Based Model [§§]

8.1. Introduction

Uncertainty is prevalent in medical domain. Rough sets offer an effective approach of managing uncertainties. This can be employed for tasks such as to discover patterns in data, attribute dependency analysis, dimensionality reduction, feature identification and classification. With the help of reduct and core, one can sidetrack the rules which are not that significant for the development of a decision support system. In this chapter we tried to implement a model based on the use of RSES software and ROSETTA software to generate multiple reducts. Different algorithms are implemented using this two such software tools for comparative study. Differential diagnoses of neonatal disease are primarily based on fourteen independent attributes in general. Here we have to explore relative importance of those fourteen attributes by generating reducts and core using different algorithms. At the time of the analysis we find that some of the attributes are excluded, out of fourteen, during the reduct generation. Moreover, we also find few different core attributes thus obtained one attribute, namely, ‘seizure’ plays somewhat a more significant role compared to ‘bleeding’. After that number of rules has been generated for each core sets and finds out prediction accuracy after generating confusion matrix. Results are also compared with ID3 algorithm to compare the overall performance of the proposed model.

Rough Set based model actually prefers for identifying relevant symptoms out of several symptoms that the neonates are suffering from. The complete study in this chapter is a process of rough set based implementation. Though several matter relating rough set we have discussed in the chapter 7. There we have experienced some useful results of diagnosing the prevalent disease amongst the sick neonates in soft computing paradigm. Keeping in a view of that idea we tried for generate this model which actually works better than the previous one. We tried to create a model based on the rough set that would be helpful for finding useful and relevant symptoms and classify them properly. Using this model, a strong rule base would generate to help in decision making process. Not even that, this intelligent technique is having great potential of discovering the data dependencies, findings the importance of the features. Duplicate objects and features can be traced off. Rough Set also used for finding least

[§§] This chapter is based on the publication made by the author entitled “Rough Set Based Model for Neonatal Disease Diagnosis”, Proc. International Conf. on Mathematics and Soft Computing, ICMSCAE-2010, December, 2010.

subsets of features among the various features on the data base. Consequently, neonatal disease diagnosis is another vulnerable area for rough sets to play important roles in resolving problems and providing solutions to properly diagnose the disease.

We have experienced a number of uncertainties in medical domain. With special reference to neonates, they belong to a special risk group having a number of uncertainties. The risk is related with the growth, development, disease, and survival. In the Indian context, the mortality and morbidity are high as discussed in reference [1] especially in rural and remote areas. This has already been discussed on the previous chapters. The prime reason of such mortality and morbidity is prevalent diseases. An estimated two-thirds of childhood deaths occur in infancy, and, in turn, two-thirds of infant deaths occur in the first month of life i.e. at their neonatal phase. There are a number of neonatal diseases and a number of parameters involved. The typical neonatal health problems which are the common causes of neonatal deaths are: Preterm, Low birth weight, Birth asphyxia, Neonatal Sepsis, Hypoxic-Ischemic Encephalopathy, Delayed breastfeeding, Problems in breastfeeding, Diarrhea, Hemorrhage, Conjunctivitis, Skin Infection, Abnormal Jaundice, Me conium Aspiration, Hyaline Membrane Disease, Pneumonia, Upper Respiratory Infection, Hypothermia, Umbilical Sepsis, Tetanus, Convulsive Disorder, Unexplained fever, Failure to Gain Weight. Chapter 2 gives the details study of the fact we mentioned above.

Each of these causes involves uncertainty in decision making and taking appropriate decisions. While developing and using decision support systems, one has to take care of the issue of close resemblance with human like decision making. Peoples are using soft computing for handling real life ambiguous situations. Soft computing is a consortium of methodologies which works synergetically and provides, in one form or another, flexible information processing capability for handling real life ambiguous situations [2]. The aim of soft computing is to exploit the tolerance for uncertainty, imprecision, partial truth and approximate reasoning in order to achieve robustness, tractability, low cost solution, and close resemblance with human like decision making. Soft computing paradigm includes fuzzy sets, rough sets, artificial neural networks, genetic algorithms, genetic programming, support vector machines, swarm optimization, ant colony optimization, memetic algorithms and others. In chapter 7 we have discussed the glimpse of the soft computing techniques for decision making process.

Rough set theory offers an effective approach of managing uncertainties. This can be employed for tasks such as to discover patterns in data, attribute dependency analysis, dimensionality reduction, feature identification and classification. The said tasks are common in medical domain, so one can try using this theory to this domain for some significant results. Rough set theory has been applied in medical informatics like medical data mining [3-5], medical image segmentation [6–9], medical classification

[10-12], and computer assisted medical decision making and decision support system [13, 14].

The objective of this chapter is to develop a rough set based model for neonatal disease diagnosis. This intends to study the relative importance of different attributes during differential disease diagnosis with the help of reducts and core concepts of rough set theory. In the process, we are to find the minimum number of rules which are significant for developing a decision support system. The model is built on the data of a study [15] on prevalent diseases of neonates. The performance of the model is compared with ID3 algorithm.

The chapter is summarized as below:

In the next section, i.e. section 8.2 the basics of rough set theory and ID3 algorithm are discussed. In section 8.3, we have discussed the fundamental statistics of the data set used for analysis. Section 8.4 presents implementation and results details. Lastly section 8.5 presents our conclusion and discussion of the study.

8.2. Rough Set Theory and ID3

8.2.1. Rough Set as Decision Set [16]

In rough set theory, representing a decision table is done by T = (U, A, C, D), where 'U' is universe of discourse and 'A' is a set of primitive features, and C, D $\subset$ A are the two subsets of features that are called condition and decision features, respectively. In chapter 7 a through details of rough set is given.

Let a $\in$ A, P $\subseteq$ A. A binary relation IND (P), called the Indiscernibility relation, is defined as follows:

$$\text{IND}(P) = \{ (x, y) \in U \times U: \text{for all } a \in P,\ a(x) = a(y) \} \quad (1)$$

Let U / IND (P) denote the family of all equivalence classes of the relation IND (P). For simplicity of notation U/P will be written instead of U/IND (P). Equivalence classes U/INDI and U/IND (D) will be called condition and decision classes, respectively. Let R$\subseteq$ C and X $\subseteq$ U,

$$\underline{R}\,X = \cup\{ Y \in U / R : Y \subseteq X \} \text{ and} \quad (2)$$
$$\overline{R}\,X = \cup \{ Y \in U/ R: Y \cap X \neq \Phi \} \quad (3)$$

Here $\underline{R}X$ and $\overline{R}X$ are said to be R-lower and R upper approximations of X and ($\underline{R}X$, $\overline{R}X$) is called *R-rough set.* If X is R-definable then $\underline{R}X = \overline{R}X$ otherwise X is R-Rough. The boundary $BN_R(X)$ is defined as :

$$BN_R(X) = \underline{R}X - \overline{R}X \quad (4)$$

Hence, if X is R-definable, then $BN_R(X) = \Phi$.

8.2.2. Reduct

A reduct is a subset of attributes that are jointly sufficient and individually necessary for preserving a particular property of a given information table. One of the basic ideas of rough set application is the concept of attribute reduction. The objective of reduct generation is to reduce the number of attributes, and at the same time, preserve a certain property that we want in future applications. For an example, assuming that there is a need of classifying few properties. A reduct should be able to preserve the original classification power provided by the whole attribute set. This power may be interpreted by syntax properties and semantics properties for both positive and boundary rule sets [17].

Let $c \in C$. A feature c is dispensable in T, if $POS_{(C-I)}(D) = POS_C(D)$; otherwise feature c is indispensable in T. c is an independent if all $c \in C$ are indispensable. Taking this consideration we may say that a Reduct is a A set of features $R \subseteq C$ is called a reduct of C, if $T' = \{U, A, R, D\}$ is independent and $POS_r(D)$. In other words, a reduct is the minimal feature subset preserving the above condition [16].

8.2.3. CORE

The core is the intersection of all reducts. No element of the core can be eliminated affecting the classification power of attributes. The core of a set T is the set of all indispensable attributes of T, which can be given as

$$CORE(T) = \cap RED(T) \quad (5)$$

where RED (T) is the set of all reducts of T.

If for a dependency $C \Rightarrow D$, D depends on E where $E \subseteq C$ then E is called as relative D-reduct of C. Relative D-core of *C* is given by

$$CORE_D I = \cap RED_D I \quad (6)$$

where Red_D I is the family of all D-reducts of *C*.

Similarly on the basis of the Reduct explanation, CORE I denotes the set of all features indispensable in C. E have CORE I = ∩ REDI , where REDI is the set of all reducts of C.

Rough set theory is an intelligent technique for managing uncertainties that is used for the discovery of data dependencies, to reduce redundancies, to evaluate the importance of attributes, to discover patterns in data, and to classify objects. There are several useful features of rough sets such as (i) extraction of rules from data sets in the form of if-then rules; (ii) it requires no external parameters unlike other intelligent techniques except the data itself; (iii) it can predict whether the data is complete or not. The computation of reduct and core using rough set theory is an important feature.

8.2.4. ID3 (Interactive Dichotomizer 3) [18]

ID3 (Interactive Dichotomizer 3) is an algorithm used to generate a decision tree invented by Ross Quinlan. ID3 can be thought of as an inductive inference procedure for machine learning or rule acquisition. At any point we examine the feature that provides the greatest gain in information or the greatest decrease in entropy. Entropy is defined as $– p \log_2 p$, where p is the probability which is determined on the basis of frequency of occurrence. The ID3 algorithm can briefly be stated as follows:

1. Consider all unused attributes and count their entropy ;
2. Choose that attribute for which entropy is minimum, or information gain is maximum to serve as the root node of the decision tree;
3. Build the next level of the decision tree providing the greatest information gain;
4. Repeat step 1 through step 3. Continue the procedure until all subpopulations are of a single class and the system entropy is zero.

ID3 falls under the category of supervised machine learning, where machine Learning is a branch of Artificial Intelligence. Using these algorithms we can infer and deduce the solution also allows automatic generation of patterns and rules in huge dataset. For predicting the outcome of queries and classifying data ID3 machine learning techniques has the enormous involvement.

8.2.4.1. Classification through ID3

The process which information or documents are grouped together which have the same kinds of properties and used for a meaningful outcome, is called Classification.

If the data are classified in proper manner then it becomes very easy to find or processed the same. Even we can place our logics also based on the classification. It makes the entering process easy for a new data and gives better and faster results. Sometime we treat the information retrieval process and data mining process as the same manner. But this is not same. Information retrieval is the process that we want the data exactly what we are looking for in contrast in data mining process helps to find patterns which will lead us to information that previously not known to us. ID3 algorithm learns from looking at input/output matches of training data to find results for new data for classification. This based on the supervised learning classification process. Where as in unsupervised learning, there are no training data sets available for matching, rather it looks on the inputs patterns of the data and then predict the result for the classification. Another classification is the reinforcement learning, which actually observe the states of the inputs and then predicts depending upon the accuracy prediction accuracy.

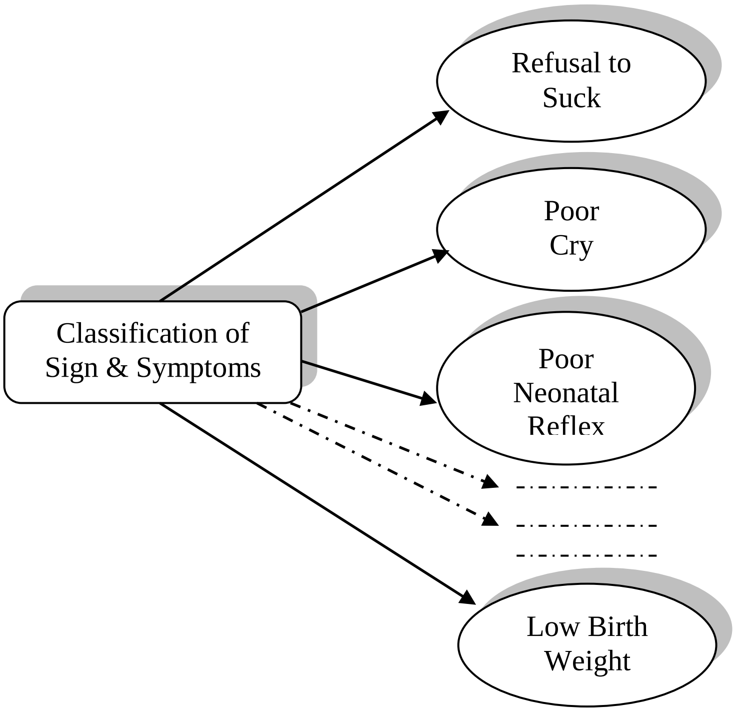

Figure 8.1. Classification of Symptoms.

Decision tree classifies data using its attributes or instances. In this decision tree, there must be a decision node and also the leaf nodes. If there is no node from the leaf node, it has to be treated as homogenous data. Classification at this portion is unnecessary in this homogenous data. In ID3 decision tree algorithm also do the same decision tree unless homogenous leaf nodes are found/reached.

8.2.4.2. ID3 Entropy [19]

For measuring homogeneity in learning set ID3 Entropy used. Entropy measures the amount of information in an attribute and measures the impurity or disorders. Entropy refers to the randomness of the dataset used for decision tree generation. In this case of study the entropy of the output values of a set of training instances. If the output values are split 50%-50%, then the set is impure and disorderly in nature. The entropy is 1 if the output values were all positive where as if the set is quite pure and orderly in nature the entropy is 0 [20].
The formula for entropy is as follows:

$$E(S) = -(p+)*log2(p+) - (p-)*log2(p-) \quad (7)$$

Where,

p+ represents the positive samples,
p- represents the negative samples, and
S represents Set of the sample of attributions

The objective of ID3 algorithm is to classify data using decision trees, in such a way that the resulting leaf nodes are all homogenous with zero entropy. An entropy value of 0 identifies a purely and orderly classified set. Entropy helps in deciding the nodes for splitting in next iteration. Generally for the maximum classification a higher entropy value is required.

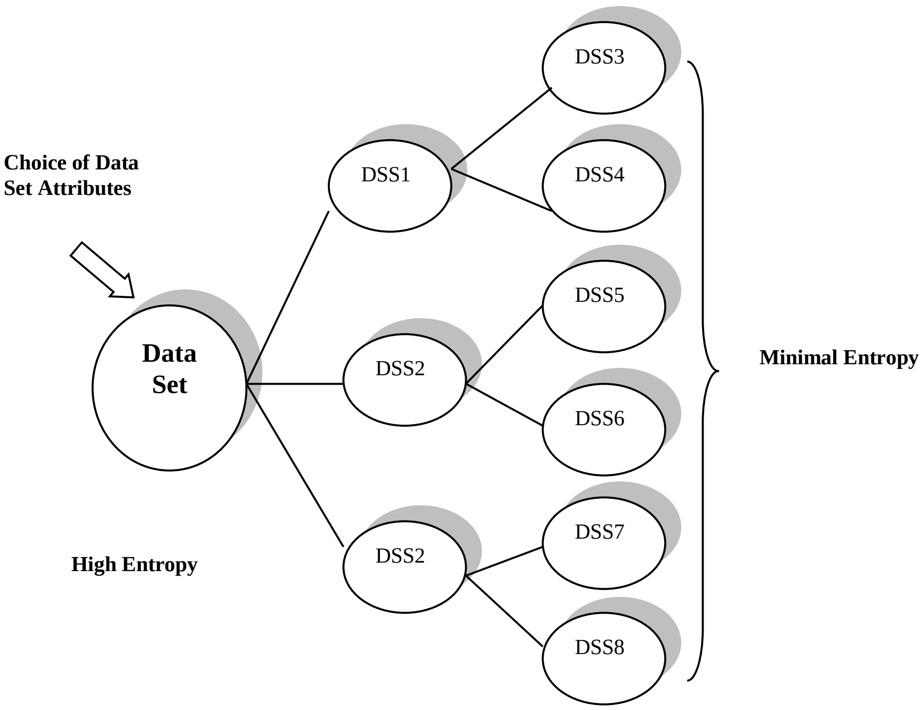

Figure 8.2. ID3 Entropy.

Figure 8.2 shows the ID3 Entropy. Here Data Set is divided in to several Data Subsets (DSS1-DSS8). If the attributes which is having highest gain is chosen then we get the minimized entropy value.

8.2.4.3. Information Gain on ID3

Information Gain is the probable reduction of entropy with a specific attributes of data set. When there is splitting is required on a node of decision tree. The concept of gain which we have used is basically for ranking the attributes for building decision trees where the node is located and the attribute with greatest gain among the attributes [21].

The information gain, Gain(S, A) of an attribute A,

Gain(S,A)= Entropy(S) - Sum for v from 1 to n of (|Sv|/|S|)*Entropy(Sv) (8)

Here, "S" is the Data Set,
"A" is the Attribute,
"Sv" is the Sub set of "S",
"|S|" is the number of elements in the set and
"|Sv|" is the number of elements in subset.

The algorithm chose the highest gain attributes for making decision tree. This will go on creating or splitting the nodes until it gets the entropy value of nodes is 0 or equal. Lastly it creates the homogeneous subsets from the attributes.

8.3. Data Set Description for the Study

A statistical summary of the data used [15] in the present study is given in Table 8.1. Here in this table the total numbers of independent attributes are 14. This independent attributes are taken as the input for the data processing. There is only one dependent attribute which actually confirms the diagnosed disease. All together there are 95 instances. In this table there is no missing value. Every attributes has its occurrence values showed in the brackets.

Table 8.1. Data Set Summary.

Number of Independent Attributes = 14 (INPUT) Dependent Attribute = Disease_differential (OUTPUT) Number of Instances = 95, Missing Values = Nil	
Attributes	**Values (occurrences)**
Birth_Term_Status	Term (52) ; Pre_Term (22), Post_Term (21)
Birth_Weight_Status	Normal(33); LBW(40);VLBW(2);ELBW(20)
Age_in_Hours>72	Y(48); N(47)
Lethargy	Y(62); N(37)
Refusal_to_Suck	Y(43); N(52)
Poor_Cry	Y(50); N(45)
Poor_Weight_gain	Y(73); N(22)
Hypothalmia	Y(63); N(32)
Sclerema	Y(51); N(44)
Excessive_Jaundice	Y(53); N(42)
Bleeding	Y(57); N(38)
GI_Disorder	Y(77); N(18)
Seizure	Y(47); N(48)
Sluggish_Neonatal_Reflex	Y(27); N(68)
Disease_differential	HIE_III(16), No_Disease(8), Septicemia(37), MD_HypoCalcimia(9),Hypo_Thalmia(2), MD_HypoCalcimiaANDHypoGlycimia(1), Hemorrhage(1), Others(5), Jaundice(2), MD_Hypothermia(3), Jaundice_BA(3), MD_Hypoglycimia(3),HIE_II(4), Sizure_Disorder(1)

8.4. Implementation and Results

8.4.1. Reduct and Core

Reduct is a subset of attributes which can, by itself, fully characterize the knowledge in the database. However, for a problem domain, there might be more than one such reduct. This has already been discussed on the above. If there are more than one reduct, some attributes might be common to all such reducts; those attributes are called core. The core attribute(s) are indispensable for an information system. We had used initially three algorithms [22,23] namely, Exhaustive algorithm, Genetic algorithm, and Dynamic reducts for finding reduct and core with the help of RSES 2.2 [24], a software tool that provides the means for analysis of tabular data sets with the use of various methods, in particular those based on Rough Set Theory. For a comparative study, we attempted other algorithms [25, 26] with the help of ROSETTA [27] software tool. Comparative results are shown in the table 8.2.

Table 8.2. Comparison among Different Algorithms for Reduct and Core.

Tool	Methods / Algorithms	No. of Reducts	Length of Reducts	Size of Core
RSES	Exhaustive	3	10	8
	Genetic	3	10	8
	Dynamic	3	10	8
ROSETTA	Genetic	3	10	8
	Johnson(with approx. solutions)	1	8	8

8.4.2. Attributes Dependency

One of the most important aspects of predictive analysis is the discovery of attribute dependencies. This essentially means that one has to discover which attributes are strongly related to which other attributes. These strong relationships need further investigation, and that will ultimately be of use in predictive modeling. Table 8.3 presents the results of such dependencies.

Table 8.3. Core Attributes.

Tool	Methods/Algorithms	Core Attributes (8)
RSES	Exhaustive	**X, Bleeding**
	Genetic	**X, Bleeding**
	Dynamic	**X, Bleeding**
ROSETTA	Genetic	**X, Bleeding**
	Johnson (with approx. solutions)	**X, Seizure**

Let **X** = { Birth_Term_Status, Birth_Weight_Status, Lathergy, Refusual_to_Suck, Poor_Cry, Hypothalmia, Sluggish_Neo natal_Reflex }

It is observed that the reducts generated by all the algorithms exclude ‘Excessive_Jaundice’, ‘Sclerema’, and ‘GI_Disorder’. Moreover, it is evident from the above Table 8.3 that Johnson algorithm differs from other algorithms in one attribute. We are to study all the above facts during rule generation and prediction results through confusion matrix.

8.4.3. Rule Generation and Prediction

Table 8.4. shows rules generated using different methods. There are different techniques for testing the prediction accuracy such as confusion matrix, prediction accuracy, receiver operating characteristics (ROC), and lift. Testing techniques are already discussed in the previous chapter 7 briefly. This work deploys confusion matrix.

Table 8.4. Rules and Prediction with Confusion Matrix.

Tool	Methods/ Algorithms	No. of Rules	Prediction Accuracy (%)
RSES	Exhaustive without reduct	3568	70
	Genetic without reduct	1904	70
	Exhaustive with reduct	192	70
	Genetic with reduct	192	70
	Dynamic with reduct	192	70
ROSETTA	Genetic with reduct	192	71.6
	Johnson (with approx. solutions) with reduct	60	70.5

It is observed that ROSETTA implementation of genetic algorithm with reduct shows the best performance. It is also observed that prediction accuracy does not decrease with reucts where number of rules are as low as 192 compared to 3568 and 1904. There is effectively rather no contribution observed of three attributes ‘Excessive_Jaundice’, ‘Sclerema’, and ‘GI_Disorder’ in decision making. Moreover, Johnson algorithm offers somewhat a comparable accuracy (70.5%) with only 60 rules with only one reduct of length eight. So, Johnson algorithm may be used for further investigations.

It is now to test the comparative contributions of two attributes namely ‘Bleeding’ , and ‘Seizure’ using two sets of different core attributes (see table 8.4) with the help of Johnson algorithm. The findings are presented in Table 8.5.

Table 8.5. Contribution of 'Bleeding' and 'Seizure'.

Tool	Methods/ Algorithms	Core Attributes	No. of Rules	Prediction Accuracy (%)
ROSETTA	Johnson (with approx. solutions)	**X**, **Bleeding**	55	68.4
	Johnson (with approx. solutions)	**X**, **Seizure**	60	70.5

From the above Table 8.5., we see that 'Seizure' plays somewhat a more significant role compared to 'Bleeding'.

8.4.4. Rough Set and ID3: A Comparison

It is now to present a performance comparison between rough set approach and ID3 algorithm as a predictive model. We have used WEKA [28] software tool implementation for ID3 approach. The results are shown in Table 8.6.

Table 8.6. Comparative Classification Scores of Rough Set Theory and ID3.

Tool	Methods/Algorithms	Classification Scores (%)
WEKA	ID3 without reduction	71.6
	ID3 with 8 core attributes including 'Bleeding'	68.4
	ID3 with 8 core attributes including 'Seizure'	70.5
RSES	Exhaustive without reduct	70
ROSETTA	Genetic with reduct	71.6

The study shows that the predictive accuracy of rough set classification accuracy is seen to compare favourably with the well-known ID3[18], the Interactive Dichotomizer 3 classifier algorithm. Looking through the results of the above Table 8.6, it is evident that attribute ‘Seizure’ plays a more significant role than that of ‘Bleeding’. This comparative study has been shown graphically on the figure 8.3 below.

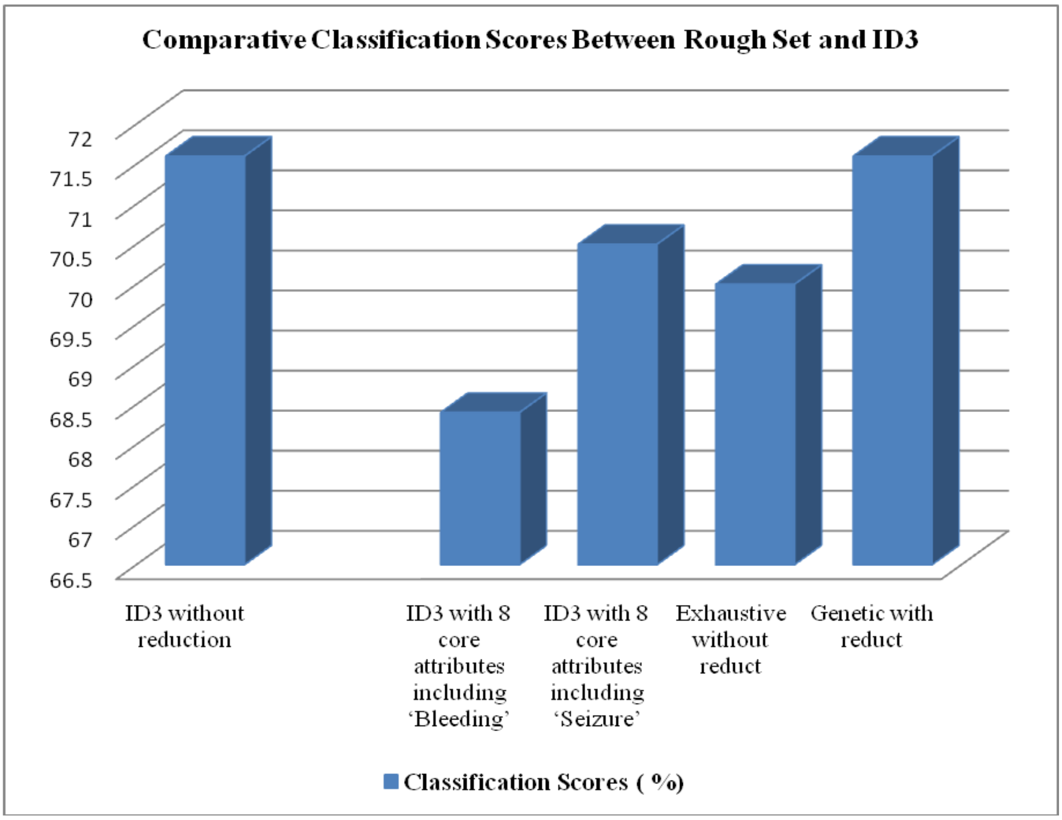

Figure 8.3. A Comparative Study of Rough Set with ID3 Classification.

8.5. Conclusion and Discussion

Rough set theory has been playing vast role in different fields including medical domain, particularly where a number of uncertainties involved. This work concentrated on neonatal disease. After explaining the complexity of the domain, it is argued that any automated decision support system certainly should assist the health people in their decision making. We presented here a rough set based model for neonatal disease diagnosis. We deployed here different algorithms for finding reduct and core with the objective of a comparative study. We observe that neither all attributes are required nor play significant role. From our study, it is evident that the three attributes, namely, ‘Excessive_Jaundice’, ‘Sclerema’, and ‘GI_Disorder’ do not play any significant role in prediction accuracy. Moreover, it is found that ‘Seizure’ plays a more significant role than ‘Bleeding’. So, a less number of attributes can be used in rational decision making which will reduce the number of rules, and the search space as well. Reduction of parameters speeds up the decision making (manual or automated) process.

Rough set based model for differential diagnosis of neonatal disease have shown very useful and helpful tool in dealing with such applications containing data with high degree of uncertainty. In this case study we tried to present and clarify the differential diagnosis for neonatal disease applying rough set after several analyses of data. This study shows that how we overcome the problems of high degree of data uncertainty, likely diagnosis disparity, even increasing treatment expenditures.

References

1. Kumar D., Verma A., and Sehgal V. K., "*Neonatal Mortality in India*.", Rural and Remote Health 7: 833 (online) 2007; http://www.rrh.org.au Available: [Last accessed 8th October, 2009].

2. Zadeh L. A., "*Fuzzy Logic, Neural Networks, and Soft Computing*", Comm. ACM, Vol. 37, pp. 77-84, 1994.

3. Huang X-M. and Zhang Y-H., "*A New Application of Rough Set to ECG Recognition*", Int. Conf. N Machine Learning and Cybernetics, Vol. 3. pp. 1729-1734, 2003.

4. Tsumoto S., "*Mining Diagnostic Rules from Clinical Databases Using Rough Sets and Medical Diagnostic Model*", Information Sciences: an International Journal, Vol. 162, No. 2, pp. 65-80, 2004.

5. Roy Chowdhury D., Chatterjee M., and Samanta R. K., "*Data Mining for Neonatal Prevalent Disease of North Bengal Districts*", Proc. Nat. Sem. on Data Mining and Decision Support, Vidyasagar University, India, pp. 24-31, March 4-5, 2010.

6. Kobashi S., Kondo K., and Hata Y., "*Rough Sees Based Medical Image Segmentation with Connectedness*", 5th Int. Forum on Multimedia and Image Processing, pp. 197-202, 2004.

7. Mohabey A. and Ray A. K., "*Fusion of Rough Set Theoretic Approximations and FCM for Color Image Segmentation*", IEEE Int. Conf. On Systems, Man, and Cybernatics, Vol. 2, pp. 1529-1534, 2000.

8. Peters J. F. and Borkowski M., "*K-means Indiscernibility Relation Over Pixels*", Int. Conf. On Rough Sets and Current Trends in Computing, pp. 580-585, 2004.

9. Widz S., Revett K., Slezak D., "*Application of Rough Set Based Dynamic Parameter Optimization to MRI Segmentation*", 23rd Int. Conf. Of the North American Fuzzy Information Processing Society, pp. 440-445, 2004.

10. Wojcik Z., "*Rough Approximation of Shapes in Pattern Recognition*", Computer vision, Graphics, and Image Processing, Vol. 40, pp. 228-249, 1987.

11. Swiniarski R. and Skowron A., "*Rough Set Methods in Feature Selection and Recognition*", Pattern Recognition Letters, No. 24, pp. 833-849, 2003.

12. Cyran K. A. and Mrzek A., "*Rough sets in hybrid methods for pattern recognition*", Int. J. of Intelligence. Systems, Vol. 16, No. 2, pp. 149-168, 2001.

13. Pondraza R., Dominik A., and Walkiewicz M., "*Decision Support System for Medical Applications*", Applied Simulation and Modelling, 2003.

14. Wakulicz-Deja A. and Paszek P., "*Applying Rough Set Theory to Multi Stage Medical Diagnosing*", Fundamenta Informaticae, Vol. 54, No. 4, pp. 387- 408, 2003.

15. Roy Chowdhury D., Chatterjee M., and Samanta R. K., "*A Study of the Status of New Born in Terai Region of West Bengal*", Advances in Modeling C, Vol. 68, No. 3, pp. 44-52, 2007.

16. Yellasiri Ramadevi, Rao C.R., Reddy Vivekchan, "*Decision Tree Induction using Rough Set Theory- Comparative Study*", Journal of Theoretical and Applied Information Technology(JATIT), pp. 110-114, 2007.

17. Pawlak, Z. "*Rough Sets*", International Journal of Computer and Information Sciences, Vol. 11, pp. 341-356, 1982.

18. Pao Y. H., "*Adaptive Pattern Recognition and Neural Networks*", Addison-Wesley, Reading, MA, 1989.

19. J. R Quinlan, "*Induction of Decision Trees*". Mach. Learn. Vol.1, No. 1, 81-106. March, 1986.

20. MitchelTom M. l, "*Machine Learning*", Singapore, McGraw-Hill, p55, 1997.

21. Peng Wei, Chen Juhua and Zhou Haiping, "*An Implementation of ID3-Decision Tree Learning Algorithm*", School of Computer Science & Engineering, University of New South Wales, Sydney, Australia, http://web.arch.usyd.edu.au/~wpeng/DecisionTree2.pdf, Available: [Last accessed 15th July 2011].

22. Bazan J., "*A Comparison of Dynamic and Non-Dynamic Rough Set Methods for Extracting Laws from Decision Table*", In L. Polkowski, A. Skowron (eds.), Rough Sets in Knowledge Discovery, Physica-Verlag, Heidelberg, pp. 321-365, 1998.

23. J. Bazan et.al., "Rough Set Algorithms in Classification Problems", In L. Polkowski, S. Tsumoto, and T. Lin(eds.), Rough Set Methods and Applications, Physica-Verlag, Heidelberg, pp. 49-88, 2000.

24. "*RSES 2.2 User's Guide*", Warsaw University, http://logic.mimuw.edu.pl/~rses , 19th January, 2005, Available: [Last accessed 14th July 2011]. .

25. Vinterbo S. and Ohrn A., "*Minimal Approximate Hitting Sets and Rule Templates*", Int. J. Approximate Reasoning, Vol. 25, No. 2, pp.123-143, 2000.

26. Johnson D. S., "*Approximate Algorithms for Combinatorial Problems*", J. of Computer and System Sciences, Vol. 9, pp. 256-278, 1974.

27. *The ROSETTA Homepage* [http://www.idi.ntnu.no/~aleks/rosetta/] Available: [Last accessed 15th July 2011].

28. Hall M., Frank E., Holmes G., Pfahringer B., Reutemann P., Witten I. H., "*The WEKA Data Mining Software: An Update*", SIGKDD Explorations, Vol.11, No. 1, 2009.

CHAPTER 9

Differential Diagnosis of Neonatal Disease : A Data Mining Model ††

9.1. Introduction

Data mining basically refers to information elicitation from data warehouse. Since its birth in the year 1993, data mining techniques are being deployed in various disciplines including medical domain. Data mining and knowledge discovery techniques are being used after its birth in the year 1993 [1] in connection with a business. It is now equally being applied to interpret huge clinical data base(s) all over the world in order to provide data for applications such as automated encoding, decision support, quality assurance, patient management, outcome analysis, and clinical research [2][3][4]. There are different algorithms available for classification problems: decision tree, naïve nays, support vector machine, and feed forward neural networks. Decision tree approach has been found suitable for this purpose since decision tree construction can make use of both symbolic or nominal and real-valued attributes [5] – a characteristic of medical domain.

This chapter intends to report some results of differential diagnosis of neonatal disease applying two important data mining algorithms, namely, ID3 and C4.5. We use data base from our earlier study. It is also planned to find relative importance of different disease parameters for differential diagnosis. For the purpose, we use rough set theory having the provision of reduct and core. Results are compared for ID3 and C4.5 without/with reduct and core. The relevant issues related to model development and validation such as algorithm settings, overfiting, confusion matrix etc., have been discussed here.

Out of the different phases of the development of a child, neonatal phase is considered to be a critical one. They are also vulnerable or special risk group; the risk is related with growth, development, disease and survival. In rural and remote areas of the world, the mortality and morbidity are still significant in number. One of the prime reasons of such high mortality and morbidity is prevalent diseases. There are a number of neonatal diseases and a number of parameters involved; a multi-criteria decision making system (MCDMS). For a typical MCDMS, it is useful to take advantage of

†† This chapter is based on the publication made by the author entitled “A Data Mining Model for Differential Diagnosis of Neonatal Disease”, IFRSA’s International Journal of Computing, Vol. 1, Issue 2, pp 143-150, April 2011.

any automated system for rational decision making. Moreover, the relevant data bases created by experienced medical practitioners and researchers should be mined for information elicitation. Moreover, for a domain which mostly vulnerable, like pediatric domain, where uncertainties are involved [6], we are to use a suitable theory for managing such uncertainties. In the recent years, rough set theory of soft computing paradigm has found its applications in various disciplines including medical domain. At the same time, data mining techniques are being deployed to the data bases for useful information elicitation.

In the research period we have experienced few things related to neonatal disease. These are various signs and symptoms of diseases, medical problems, working knowledge with several doctors and medical test reports of several neonates having problems or suffering from diseases. The specialist or the doctor will use all that evidence to arrive at diagnosis for the baby and also informs what is wrong with the neonate. It was noticed that most of the time it works, but other times it may be misdiagnosed or even doctor may fail the diagnosis. In the study area, this misjudging and misdiagnosed frequency rate is very high. Thus there is a need to know how doctors diagnose a child disease and what one can do to confirm before he or she has arrived at the right answer. Doctors basically use information drawn from this description of sign symptoms, medical tests report, knowledge of medicine, experience and additional inputs. Domain specialist or doctors will then make a list of all the possible diagnoses that could explain what is medically wrong with patient. Then, one by one, using that same information, he will begin to narrow down the list by finding clues that don't fit. This process of elimination is called "differential diagnosis." Finally doctor will be left with one diagnosis, and that's the one he gives the patients.

If these procedures use data mining modeling, then there is a high chance of improvements for the disease diagnosing procedure. This has been statistically proved on the study. Medical domain involves a number of inexactnesses. The nature of inexactness can be vagueness, uncertain, partial truth, imprecision. The sources of inexactness can be logical as well as physical type [6]. The typical logical sources are: lack of adequate data, inconsistency of data, inherent human fuzzy concepts, matching of similar rather than identical situations, differing (expert) opinions, imprecision in measurements, lack of available theory to describe a particular situation. Concentrating on the pediatric domain, the typical physical sources are: problem domain itself, child, parents/guardians, doctors, laboratory tests/technicians, symptoms, non-availability of laboratory results. That’s why dealing with the pediatric problems is very sensitive during rational decision making. Moreover, over the years thousands data have been accumulated in different disciplines on medical domain and the process of accumulation is continuing all over the world. For knowledge discovery from such huge data bases, data mining tools are being deployed. On the other hand, soft computing techniques are being deployed which are capable to deal with imprecision and uncertainty especially needed in ill-defined problem areas. Data mining basically refers to information elicitation and soft computing is meant for extracting intelligent behavior. Data mining offers exact output within the error bounds estimated. On the other hand, soft computing offers approximate output but treated as intelligent one. So, it is tempting to use the merits of both the paradigms synergistically in a complementary way leading to knowledge discovery in databases.

This work in the chapter reports the results of a study where some coupling with data mining and rough set theory is proposed for differential neonatal disease diagnosis. The experimental set up was planned to extract the relative importance of neonatal disease attributes; dimensionality reduction; finding core attributes, if any; classification prediction using data mining approach.

The chapter summery is as follows:

Section 9.2 intends to discuss data mining as well as soft computing perspectives with special reference to ID3 and C4.5, and rough set theory. Section 9.3 presents the problem description and experimental set up. Section 9.4 presents the results of the study. In the last section, i.e. section 9.5 presents the discussions and conclusion of the study.

9.2. Data Mining and Soft Computing Paradigms

9.2.1. Data Mining: Decision Tree

The decision tree algorithm is a common and one of the most popular algorithms used in data mining because it is easy to understand how it makes predictions. The goal is to create a model that predicts the value of a target variable (dependent) based on several input variables (independent). A tree can be 'learned' by splitting the source set into subsets based on an attribute value test. This process is repeated on each derived subset in a recursive manner called recursive partitioning. The recursion is completed when the subset at a node all has the same value of the target variable, or when splitting no longer adds value to the predictions [7]. Data comes in records of the form:

$$(X, Y) = (x_1, x_2, x_3, \ldots\ldots x_n, Y) \qquad (1)$$

The dependent variable, Y, is the target variable that we are trying to understand, classify or generalize. The vector X is composed of the input variables x_1, x_2, x_3 ,x_4 …etc., that are used as input for the model. Through discussion regarding this is elaborated in chapter 8.

In data mining, there are different categories of trees available in the literature. We use classification tree analysis, as because our predicted outcome is the class to which the data belongs.

9.2.2. ID3 Algorithm [8]

ID3 (Interactive Dichotomizer 3) is an algorithm used to generate a decision tree invented by Ross Quinlan. ID3 can be thought of as an inductive inference procedure for machine learning or rule acquisition. At any point we examine the feature that provides the greatest gain in information or the greatest decrease in entropy. Entropy is defined as – p $\log_2$ p, where p is the probability which is determined on the basis of frequency of occurrence. The ID3 algorithm can briefly be stated as follows:

- Consider all unused attributes and count their entropy;
- Choose that attribute for which entropy is minimum, or information gain is maximum to serve as the root node of the decision tree;
- Build the next level of the decision tree providing the greatest information gain;
- Repeat step 1 through step 3. Continue the procedure until all subpopulations are of a single class and the system entropy is zero.

In chapter 8 we have broadly discussed about the ID3 algorithms with its all implementation details.

9.2.3. C4.5 Algorithm [9]

The C4.5 algorithm is based on information gain which is again based on the concept of entropy of information theory. For a random variable X with N outcomes $\{ x_i : i = 1,2,3,4,\ldots\ldots.N\}$, the Shannon entropy, a measure of uncertainty and denoted by H(X), is defined as:

$$H(X) = - \sum_{i=1}^{N} P(x_i) \log_2 P(x_i) \qquad (2)$$

where P(xi) is the probability mass function of outcome xi. This algorithm C4.5 uses a set of training data S ($S = s_1, s_2, s_3, \ldots\ldots\ldots s_n$) of classified samples for building decision trees. Each sample si will be having a set of attributes or features for classification along with a class attribute. Based on the normalized information gain (difference in entropy), at each node of the tree, C4.5 chooses one attribute of the data that most effectively splits its set of samples into subsets enriched in one class or the other. The attribute with the highest normalized information gain is chosen to make the decision. It then uses recursion on the smaller sub lists for further building and completion of decision tree. The details of C4.5 tree has been discussed on chapter 7.

9.2.4. Soft Computing and Rough Set Theory

Soft computing is a consortium of methodologies which works synergistically and provides, in one form or another, flexible information processing capability for handling real life ambiguous situations [10]. The aim of soft computing is to exploit the tolerance for uncertainty, imprecision, partial truth and approximate reasoning in order to achieve robustness, tractability, low cost solution, and close resemblance with human like decision making. Soft computing paradigm includes Fuzzy Sets (FS), Rough Sets (RS), Artificial Neural Networks (ANN), Genetic Algorithms (GSs), Genetic Programming (GP), Support Vector Machines (SVM), Swarm Optimization (SO), Ant Colony Optimization (ACO), Memetic Algorithms (MA) and others.

Rough sets theory was first presented by Pawlak in the 1980's [11]. Rough set is a formal approximation of a crisp set in terms of a pair of sets which give lower approximation wit h positive region and upper approximation with negative region. In between there a boundary. Let there be an information system I = (U, A) (attribute - value system), where u be the universe of discourse and is a non-empty set of finite objects; A is a non-empty finite set of attributes. With any P $\subseteq$ A, there is an associated equivalence relation IND(P). The relation IND(P) is called P–indiscernibility relation. Let X $\subseteq$ u be a target set we wish to represent using attribute subset P. Now, the target set X can be approximated using only the information contained within P by constructing P-lower ($\underline{P}X$) and P-upper ($\overline{P}X$) approximation of X. The tuple ($\underline{P}X$, $\overline{P}X$) is called a rough set. The accuracy of the rough-set representation of the set X can be given [11] by the following:

$$\chi_P(x) = |\underline{P}X| \,/\, |\overline{P}X| \qquad (3)$$

Rough set theory is an intelligent technique for managing uncertainties that is used for the discovery of data dependencies, to reduce redundancies, to evaluate the importance of attributes, to discover patterns in data, and to classify objects. There are several useful features of rough sets such as (i) extraction of rules from data sets in the form of if-then rules; (ii) it requires no external parameters unlike other intelligent techniques except the data itself; (iii) it can predict whether the data is complete or not. The computation of reduct and core using rough set theory is an important feature. For the purpose of the current study we have taken these rough set details again, which has already been discussed in chapter 8.

9.3. Problem Description and Experimental Setup

9.3.1 Problem Description

There are different pediatric age groups : Neonates (0 – 4 weeks), Infants (4 weeks – 1 year), Toddler (1 year – 3 years), Pre-school (3 years – 5 years), School-going (5 years – 10 years) and Adolescence (10 years – 18 years). Out of these age groups, neonates belong to a special risk group. The risk is related with the growth, development, disease, and survival. Obviously, mortality and morbidity heavily depends on disease diagnosis and management. We consider thirteen differential disease states (objects). Each object has 14 attributes (sign and symptoms). In addition, each object belongs to a decision class. A summary of the data used[12] in the present study is shown in 9.1.

Table 9.1. Data Set for Study*.

Attributes	Values (occurrences)
Birth_Term_Status	Term (52) ; Pre_Term (22), Post_Term (21)
Birth_Weight_Status	Normal(33); LBW(40);VLBW(2);ELBW(20)
Age_in_Hours>72	Y(48); N(47)
Lethargy	Y(62); N(37)
Refusal_to_Suck	Y(43); N(52)
Poor_Cry	Y(50); N(45)
Poor_Weight_gain	Y(73); N(22)
Hypothalmia	Y(63); N(32)
Sclerema	Y(51); N(44)
Excessive_Jaundice	Y(53); N(42)
Bleeding	Y(57); N(38)
GI_Disorder	Y(77); N(18)
Seizure	Y(47); N(48)
Sluggish_Neonatal_Reflex	Y(27); N(68)
Disease_differential	HIE_III(16), No_Disease(8), MD_HypoCalcimia(9), Septicimia(37), Hypo_Thalmia(2), Hemorrhage(1), Others(5), Jaundice(2), MD_Hypothermia(3), Jaundice_BA(3), MD_Hypoglycimia(3), HIE_II(4), Sizure_Disorder(1)

* The above table also contains the following information:

Number of Independent Attributes = 14 (INPUT)
Dependent Attribute = Disease_differential (OUTPUT)
Number of Instances = 95,
Missing Values = Nil.

From the above table 9.1, it is to be noted that the amount of data what is available to this project is not equally distributed over the different decision classes.

9.3.2. Experimental Setup

For the experimental setup we have used the following steps:

Step I. Find the reducts and core using rough set theory by different algorithms and compare.

Step II. Explore the attribute dependency.

Step III. Classify the instances with ID3 and C 4.5 using the two test options
(i) on training data and
(ii) by folding method.

Three sets of data were tested: (i) with original 14 attributes;
(ii) with attributes from reducts; and
(iii) with core attributes.

9.4. Results

9.4.1. Reduct and Core

Reduct is a subset of attributes which can, by itself, fully characterize the knowledge in the database. However, for a problem domain, there might be more than one such reduct. If there are more than one reduct, some attributes might be common to all such reducts; those attributes are called core. The core attribute(s) are indispensable for an information system. We had used initially three algorithms [13][14] namely, Exhaustive algorithm, Genetic algorithm, and Dynamic reducts for finding reduct and core with the help of RSES 2.2 [15], a software tool that provides the means for analysis of tabular data sets with the use of various methods, in particular those based on Rough Set Theory. For a comparative study, we attempted other algorithms [16][17] with the help of ROSETTA[24] software tool. Comparative results are shown in the table 9.2.

Table 9.2. Comparison among Different Algorithms for Reduct and Core.

Tool	Methods / Algorithms	No. of Reducts	Length of Reducts
RSES	Exhaustive	3	10
	Genetic	3	10
	Dynamic	3	10
ROSETTA	Genetic	3	10
	Johnson (with approx. solutions)	1	10

9.4.2 Attribute Dependency

One of the most important aspects of predictive analysis is the discovery of attribute dependencies. This essentially means that one has to discover which attributes are strongly related to which other attributes. These strong relationships need further investigation, and that will ultimately be of use in predictive modeling. Table 9.3 presents the results of such dependencies. It is observed that the reducts generated by all the algorithms exclude 'Excessive_Jaundice', 'Sclerema', and 'GI_Disorder'. Moreover, it is evident from the table 9.3 that all algorithms suggest { X, Bleeding } as core attributes; so we concentrate on { X, Bleeding } for further study.

Table 9.3. Core Attributes.

Tool	Methods/Algorithms	Core Attributes (8)
RSES	Exhaustive	**X, Bleeding**
	Genetic	**X, Bleeding**
	Dynamic	**X, Bleeding**

Let X = { Birth_Term_Status, Birth_Weight_Status, Lathergy, Refusual_to_Suck, Poor_Cry, Hypothalmia, Sluggish_Neo natal_Reflex }

9.4.3. Applying Data Mining Tools

It is to be stated that originally we had 14 attributes (sign and symptoms). After applying rough set theory, we find that all algorithms do not include 'Excessive_Jaundice', 'Sclerema', and 'GI_Disorder' into their reducts. So we consider rest 11 attributes for further study. Further we consider 8 core attributes for performance evaluation. We now apply ID3 and C4.5 algorithms using WEKA(Version 3-6-2)[18], an open source, java-enabled, platform independent data mining software. It has different features and capabilities implementing different algorithms, algorithm settings, evaluating model quality, experimentation for the data miners, integrating data bases of different formats etc. The characteristics of the model are discussed in the next section.

9.4.3.1. Algorithm Settings

Not only the selection of the right algorithm is important but also the proper settings of the parameters from data mining expertise, knowledge of the available algorithms, and often experimentation to determine which algorithm best fits the problem with suitable values of the parameters are equally important. Algorithm settings allow users to exert finer control over the algorithm to attain better results during the build process. Decision tree models can be extremely accurate on the build data if allowed to overfit the build data. This occurs by allowing the algorithm to build deeper trees with rules specific to even individual cases. Hence overfit models give very good accuracy with the build data, but do not generalize well on new data, resulting in decreased predictive accuracy [19]. To avoid overfitting as well as to control tree size, one has to apply pruning techniques and / or stopping criteria for decision tree algorithms. At the same time, goodness of a node split is determined by the information gain. So, in order for our model to generalize well it must not be built around the training data too closely [20]. Different pruning techniques have been proposed along with different splitting criteria, it has been found that there is not much variation in terms of performance [21[22][23]. All these issues have been taken in consideration during the present study.

Now, we present our results in table 9.4 with analysis. It is observed that the predicted accuracy of ID3 is better than that of C4.5 when tested with self data. Moreover, if we reject 'Excessive_Jaundice', 'Sclerema', and 'GI_Disorder', the prediction accuracy does not change substantially. It indicates that these three parameters are not that significant in rational decision making. Now, if the system is operated with 10-folding for testing, the performance of ID3 degrades substantially. C4.5 prediction accuracy shows much better performance in this situation. The results may be interpreted in the way that ID3 works on unpruned tree structure but C4.5 works with unpruned / pruned structure. As we are aware of data mining concepts till now, recapitulate the whole thing we can say, data mining is all about automating the process of searching for patterns in the data. In this respect we, basically interested on which patters we should intersected for selection for the prediction. Even there was a question, that how it has to be exploited. In our C4.5 algorithm setting we set pruning mechanism so that the model does not lead to overfitting. From the results it is evident that ID3 model tested on self data leads to an overfit model and that leads to substantial degradation on predictive accuracy when tested with 10-folds.

9.4.3.2. Needs of 10 folds Cross Validation

Using 10 folds Cross Validation we have used 90% of full data for training rest and 10% for testing in each fold test. This is a compromise practically motivated by: - 90% is not too far from full 100%, which means that cross-validation produces a fair estimation of test performance when the training model with 100% is tested against another unseen test set. We have not used any other folding method because, say having 5 folds only means the folds are 80% trained, which can be shown to have great effect on the robustness (train » test) of the produced model. Again having more than 10 say, 20 poses two problems: not only is it computationally more demanding, there is an increasing problem with small datasets, i.e. in a dataset of 100 instances, each test fold in 20-fold Cross Validation would have only 5 instances, and it is then increasingly likely that target classes with few instances are tested and not trained i.e. that all the instances wind up in the test fold producing guaranteed zero classification accuracy or that they all end up in train folds in which case that part of the training resource does not 'pay off'.

Although it must be said that Weka enforcing ‘stratified’ which mean class-balanced cross validation does not suffer this problem other than with classes having only one instance in the full training data. Though 10 is not a definite number therefore, but dependent on this singleton problem and the total number of training instances itself. If the data set is larger, the fewer folds are needed to produce a best model.

Table 9.4. A Comparison between ID3 and C 4.5.

Algorithm	Tested on Self			Tested on 10 Folds		
ID3 prediction accuracy (%)	71	71	68	21	22	21
C4.5 prediction accuracy (%)	66	64	63	48	47	43
	14	11	8	14	11	8
	Number of Objects			Number of Objects		

The above comparative study for predicting the accuracy between J48 and C4.5 algorithm has been shown on Appendix E graphically.

9.5. Conclusion and Discussion

With ever increasing medical knowledge as well as ever increasing population, the physicians are sometimes confused and overloaded during decision making. It is then suggested to consult the specialists to resolve confusion. But, however, this facility of consulting specialists might not always be available or not timely available. This is certainly a major problem of medical domain. Moreover, from pediatric domain, especially neonates come to the physicians with incomplete or ambiguous information about patient's history. The situation becomes more acute when they come from lower socio-economic background. These ambiguous and/or incomplete inputs introduce substantial amount of uncertainty in a decision making system.

For managing such uncertainties, a number of methods are deployed such as fuzzy sets, ANNs, rough set theory, which are also dependent on attribute types. We deploy rough set theory for managing uncertainties as well as knowledge extraction. This helps finding reducts and core; thereby reducing the search space. Next, we deploy data mining tools ID3 and C4.5 for classification on the original as well as reduct and core data. This helps understanding the functions and power of this coupled scheme for this problem domain, at least. But, however, this type of coupling scheme may result better performance. Note that this coupling scheme produces somewhat 48% prediction accuracy which seems to be less as expected.

The possible reasons may be are as follows:

(i) The amount of data that is available to this project is not equally distributed over the different classes. Some classes are highly populated and some are really less populated with one or two cases. It will certainly be our efforts in future to have a data base with more or less equal distribution over the different classes.

(ii) One may deploy other algorithms. At the same time algorithms settings with varying parameters need to be tested for improved performance.

This study ensures us that with the use of conventional analysis along with data mining and statistical studies in patient data can improve better disease diagnosing capacity with good accuracy rate. Even using data mining techniques data quality and standard of data, diagnosing plans and treatment procedures and decreases of treatment timings must be improved no doubt. On the other hand, the quality of predicting the disease is usually based upon providing the indices. Mostly, measures have to be taken by physicians in the clinics, and it is their acceptance of the results which is required to ensure that they are put into action. Therefore, understandable results in a clear and adequate presentation are essential. At this juncture this particular study definitely used as a helpful tool for the medical practitioner in real time application. Additionally this work is concerning the statistical consolidation of the results primarily, support for the final conversion as well as the development of an adapted automated data mining process.

The present investigation provides a decision support system for differential diagnosis of neonatal diseases as well as a methodology for coupling rough set theory with data mining tools. Lastly, as this is differential diagnosis, the results might be accepted as first order inference. The next higher order performance is achieved with the results of laboratory tests.

References

1. Agrawal, R., Imielinski T., Swami A. N., "*Mining Association Rules between Sets of Items in Large Databases*", SIGMOD Conference , USA, pp. 207-216, 1993.

2. Chae,Y.M., Ho,S.H,, Cho, K.W., Lee, D. H., Ji, S.H., "*Data Mining Approach to Policy Analysis in Health Insurance Domain*", International Journal of Medical Informatics, Vol. 62(2-3), pp. 103-111, 2001.

3. Chen, H., Fuller, S. S., Friedman, C., Hersh, W., "Knowledge Management, Data Mining, and Text Mining in Medical Informatics", Medical Informatics: Knowledge Management and Data Mining in Biomedicine, Springer, pp. 3–33, 2005.

4. Palaniappan,S., Awang, R., "*Intelligent Heart Disease Prediction System Using Data Mining Techniques*", International Journal of Computer Science and Network Security, Vol. 8, No. 8, pp. 343-349, 2008.

5. Wu, X., "*Data Mining: An AI Perspective*", IEEE Computational Intelligence Bulletin, Vol. 4, No. 2, pp. 23-26, 2004.

6. Saha, A. K., Goswami, M. G., Chatterjee, M., Samanta, R. K., "*Child Growth and Development, Expert Systems and Inexact Information Management*", Proceedings, CSI-95, Hyderabad, India, pp. 258-267, 1995.

7. http://en.wikipedia.org/wiki/Decision_tree_learning Available: [Last accessed 5th Feb, 2010].

8. Quinlan, J. R., "*Induction of Decision Trees*", Machine Learning, Vol. 1, No.1, pp. 81-106, 1986.

9. Quinlan, J. R., "*C4.5: Programs for Machine Learning*", Morgan Kaufmann Publishers, 1993.

10. Zadeh, L. A., "*Fuzzy Logic, Neural Networks, and Soft Computing*", Comm. ACM, Vol. 37, pp. 77-84, 1994.

11. Pawlak, Z., "*Rough Sets*", Int. Journal of Parallel Programming, Vol. 11, No. 5, pp. 341-356, 1982.

12. Roy Chowdhury D, Chatterjee, M., and Samanta, R. K., "*A Study of the Status of New Born in Terai Region of West Bengal*", Advances in Modeling C, Vol. 68(3), pp. 44-52, 2007.

13. Bazan, J. "*A Comparison of Dynamic and Non-dynamic Rough Set methods for Extracting Laws from Decision Table*", In L. Polkowski, A. Skowron (eds.), Rough Sets in Knowledge Discovery, Physica-Verlag, Heidelberg, pp. 321-365, 1998.

14. Bazan, J. et.al., *"Rough Set Algorithms in Classification Problems"*, In L. Polkowski, S. Tsumoto, an d T. Lin (eds.), Rough Set Methods and Applications, Physica-Verlag, Heidelberg, pp. 49-88, 2000.

15. *"RSES 2.2 User's Guide"*, Warsaw University, http://logic.mimuw.edu.pl/~rses. 2005,

16. Vinterbo, S., and Ohrn, A., "*Minimal Approximate Hitting Sets and Rule Templates"*, Int. J. Approximate Reasoning, Vol.25, No.2, pp. 123-143, 2000.

17. Johnson, D. S., "*Approximate Algorithms for Combinatorial Problems"*, J. of Computer and System Sciences, Vol. 9, pp. 256-278, 1974.

18. Hall, M., Frank, E.,Holmes, G., Pfahringer, B., Reutemann, P., Witten, I. H., "*The WEKA Data Mining Software: An Update*", SIGKDD Explorations, Vol. 11(1), 2009.

19. Hornick, M. F., Marcade, E. and Venkayala, S., "*Book Excerpt: Java Data Mining Concepts"*, JavaWorld.com, 22nd Feb., 2007.

20. http://www.decisiontrees.net/book/export/html/37 Available: [Last accessed 4th Nov., 2010].

21. http://www.decisiontrees.net/book/export/html/44 Available: [Last accessed 4th Nov., 2010].

22. Mingers, J., "*An Empirical Comparison of Pruning Methods for Decision Tree Induction*", Machine Learning, Vol. 4, No. 2, pp. 227-243, 1989.

23. Espoito, F., Malerba, D., and Semeraro, G., "*A Comparative Analysis of Methods for Pruning Decision Trees*", IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 19, No. 5, pp. 476-491, 1997.

24. "*The ROSETTA Homepage"*, http://www.idi.ntnu.no/~aleks/rosetta/ Available: [Last accessed 4th Nov., 2010].

CHAPTER 10

Expert System Model Designing in Differential Diagnosis of Neonatal Disease[§§§]

10.1. Introduction

Neonatal phase is one of the vital phases for the development of a child. Many prevalence diseases are the major causes of deaths in the neonates. In India, 30% to 40% babies are Low Birth Weight babies and about 10% to 12% of Indian babies are born less than 37 completed weeks which is preterm stage. It causes physical immaturity of the baby and also reasons for the high neonatal mortality. This mortality problem can prevail over through fast and accurate disease diagnosis and management of the newborn. Development of an Expert System for neonatal diseases diagnosis is a complicated process and requires high level of expertise. Any attempt towards designing and developing of an expert system dealing with differential diagnosis of neonatal disease must be overcome various difficulties.

There has been a growing demand for Neonatal Health Information everywhere. This increased demand for neonatal health information has been accelerated by the introduction of the artificial intelligence and expert system in the field of medical health care. The expert system model can be useful as any information that enables individuals to understand their health and make health-related decisions for themselves or their family and take care of the little child easily. Here we may expect that this model can give a useful and better supporting tool for expert opinions and decision making.

The studies made on prevalence of various disease like Neonatal Septicemia, HIE, Metabolic Disturbances, neonatal jaundice etc. in north eastern parts of India, especially in Tarai region and neonates are suffering with these disease[1]. In the Indian context, mortality and morbidity are high in rural and remote areas [2]. If defects are not diagnosed and treated properly on the early stage of the children, they may have to face many problems in their future. There are a number of neonatal diseases and a number of parameters involved. The typical neonatal health problems which are the common causes of neonatal deaths are: Preterm, Low birth weight, Birth asphyxia, Neonatal Sepsis, Hypoxic-Ischemic Encephalopathy(HIE), Delayed breastfeeding, Problems in breastfeeding, Diarrhea, Hemorrhage, Conjunctivitis, Skin

[§§§] This chapter is based on the publication made by the author entitled “Design and Development of an Expert System Model in Differential Diagnosis of Neonatal Disease”, IFRSA’s International Journal of Computing, Vol. 1, Issue 3, pp 343-350, July 2011.

Infection, Abnormal Jaundice, Meconium Aspiration, Hyaline Membrane Disease (HMD), Pneumonia, Upper Respiratory Infection (URI), Hypothermia, Umbilical Sepsis, Tetanus, Convulsive Disorder, Unexplained fever, Failure to Gain Weight and many more. Each of these involves uncertainty in decision making and taking appropriate decisions. While developing and using decision support systems, one has to take care of the issue of close resemblance with human like decision making.

To overcome the problems mentioned above, proper treatment planning is required. For proper treatment panning the basic requirement is Neonatologists. But pediatric expert distribution in rural region is not at all satisfactory [3]. To mitigate the scarcity of the domain expert, this system development may help considerably. Not even that, in order to improve the diagnosis accuracy and reduce diagnosis time, it has become a demanding issue to designing and developing an efficient and reliable Expert System to support complicated diagnosis decision process. Hence soft computing methods such as Core and Reduct generation, Classification methods with various classifier and Rule generation have shown great potential to be applied in the development of the Expert System for Neonatal Disease diagnosis [4].

Artificial Intelligence is the study of mental facilities through the use of computational models. These models produced several tools which have great significance in design and develop of an expert system. The powerful tools among these are Knowledge Based System (Expert System), Fuzzy Sets (FS), Rough Sets (RS), Artificial Neural Networks (ANN), Genetic Algorithms (GA), Genetic Programming (GP), Support Vector Machines (SVM), Swarm Optimization (SO), Ant Colony Optimization (ACO), Memetic Algorithms (MA) and others [5].

In this chapter different approaches in designing of Expert System and present a new method for developing such system have been discussed. For classification and building the rule base we have tested the real field observation on soft computing environment tools. Comparative study of Johnson Reducer, Exhaustive Reducer and Genetic Reducing Algorithm, Rule Generation and Re-building process are focused for processing the expert system development. The proposed system shows good performance for finding the differential diagnosis of neonatal disease and achieved ~70% accuracy.

This chapter presents a ruled based, object oriented expert system for neonatal disease diagnosis in soft computing paradigm and application of rough set theories.

The present chapter is organized as follows:

Section 10.2 in this chapter describes about the present problem Identification and common cause of death findings. The Model Design and Implementation of Expert system is presented in detail in section 10.3, which includes Data Set Design and Structural design and finally conclusion is given in section 10.4.

10.2. Problem Identification

10.2.1. Common cause of Neonatal Deaths

The common causes of neonatal deaths in India, mostly in Tarai region of west Bengal are systematic bacterial sepsis, birth asphyxia, congenital malformation, pre-maturity, hyperbilirubenimia and others. Few common diseases of neonates are Neonatal Sepsis, Hypothermia, Birth Asphyxia, HIE, Metabolic Disorders, Inborn Errors of Metabolism, Respiratory Disorders, Neonatal Seizure, Hyperbilirubinimia, Prematurely, Hemorrhagic Disease of New Born and others[2]. Few common disease pattern concerning neonatal heath problems and Diagnosis criteria are given below [3]:

i) Neonatal Septicemia

Septicemia, meningitis or pneumonia diagnosed clinically.

Simultaneous presence of any two of the following six criteria any time during 0-28 days:

a. Baby which cried well at birth, its cry became weak or abnormal, or stopped crying; or baby who earlier sucked or licked well, stopped sucking or mother feels that sucking became weak or reduced: or baby who was earlier conscious and alert, became drowsy or unconscious.
b. Skin temperature >99ºF or <95ºF
c. Sepsis in skin or umbilicus
d. Diarrhea or persistent vomiting or distension of abdomen
e. Grunt or severe chest indrawing.
f. Respiratory rate (RR) 60 or more per minute even on counting twice.

ii) Hypoxic-Ischemic Encephalopathy (HIE)

Feature	**Stage I(Mild)**	**Stage II(Moderate)**	**Stage III(Severe)**
Consciousness	Irritable	Lethargy	comatose
Tone	Hypotonia	Marked Hypotonia	Severe Hypotonia
Seizures	No	Yes	Prolonged
Sucking or Respiration	Poor Suck	Unable to Suck	Unable to sustain spontaneous Respiration

iii) Preterm

Less than 8 months and 14 days (37 weeks) of gestation counted from the onset of the last menstrual period as per the history given by the mother.

iv) Delayed breastfeeding

Due to traditional practice, breastfeeding not started in first 24 hours after birth, but baby licked/sucked the sweetened water.

v) Problems in breastfeeding

Presence of any one of the following:

(*i*) Baby did not suck breast for more than continuous 8 hours even when offered.

(*ii*) -Mother unable to breast feed, or
-baby fed on extracted breast milk, or goat, or cow milk, or bottle, or
-sweetened water beyond 3 days, or
-inadequate breast milk evidenced by continuous crying of baby and failure to gain weight.

vi) Hemorrhage

Bleeding from mouth, anus, eyes, nose or in skin or in urine any time or vaginal bleeding after first week.

vii) Skin Infection

(*i*) *Pyoderma:* pus, ulcer, boil, pustule in skin.
(*ii*) *Intertrigo:* excoriation with moist, cracked skin at skin folds.

viii) Abnormal Jaundice

Skin or eyes yellow on the first day or yellowness persisted at 3 weeks, or when yellowness associated with sepsis.

ix) Meconium Aspiration

History of difficult delivery or presence of birth asphyxia and respiratory distress (RR 60 or more; or ever in drawing of lower chest) started in first 24 hours after birth.

x) Upper Respiratory Infection (URI)

Cough or nasal discharge present for three days or more without respiratory distress or increased RR.

xi) Hypothermia

Auxiliary temperature <95ºF.

xii) Umbilical Sepsis

Pus discharge from umbilicus.

xiii) Tetanus

Baby which earlier sucked well, stopped taking feeds from 4th day or more; and appearance of seizures, spasm and trismus.

xiv) Convulsive Disorder

Seizures but baby conscious, alert and feeds well between seizures (excludes tetanus, asphyxia, sepsis)

xv) Birth asphyxia

(*i*) *Mild*: At 1 minute after birth, no cry, or the breath was absent or slow, weak or gasping.
(*ii*) *Severe*: At 5 minutes after birth, the breath was absent or slow, weak or gasping.
(*iii*) *Indirect*: In the absence of direct observations by VHWs about newborn's condition at 1 and 5 minutes, presence of following two:
  (*a*) Baby did not cry on its own so the care provider had to make efforts to make the baby cry;
  (*b*) Colour of the umbilical cord was green or yellow.

10.3. Design and Implementation

Based on the knowledge of the structure and behavior of the devices the system is designed to understand. Developing any model using Model based system is especially useful in diagnosing equipment problems. The systems include a model of the data set from which diseased to be diagnosed. Below there is brief information on the knowledge in terms of data and how these data sets have been designed is discussed.

10.3.1. Data Set Design

A statistical summary of the data used in the present study is given in Table 10.1.

Table 10.1. Data Set Summary.

Number of independent attributes = 14 (INPUT) Dependent attribute -> Disease_differential (OUTPUT) Number of instances = 95 Missing Values = Nil	
Attributes	**Values (occurrences)**
Birth_Term_Status	Term (52) ; Pre_Term (22) ,Post_Term (21)
Birth_Weight_Status	Normal(33); LBW(40); VLBW(2);ELBW(20)
Age_in_Hours>72	Y(48); N(47)
Lethargy	Y(62); N(37)
Refusual_to_Suck	Y(43); N(52)
Poor_Cry	Y(50); N(45)
Poor_Weight_Gain	Y(73); N(22)

Hypothalmia	Y(63); N(32)
Sclerema	Y(51); N(44)
Excessive_Jaundice	Y(53); N(42)
Bleeding	Y(57); N(38)
GI_Disorder	Y(77); N(18)
Seizure	Y(47); N(48)
Sluggish_Neonatal_Reflex	Y(27); N(68)
Disease_differential	HIE_III(16), No_Disease(8), MD_HypoCalcemia(9), Septicemia (37), Hypo_Thalmia(2), MD_HypoCalcemia AND-HypoGlycemia(1), Hemorrhage(1), Others(5), Jaundice(2), MD_Hypothermia (3), Jaundice_BA(3), MD_Hypoglycemia(3), HIE_II(4), Sizure_Disorder(1)

10.3.2. Structural Design

❖ ***Knowledge Base***

Depending upon the data set, first knowledge base is prepared. The knowledge of the system is collected from domain experts, specialized books [6, 7] and from different journals [3, 8]. Initially knowledge base was containing 209, 129 rules on the two consecutive phases**.** All this knowledge are known and incorporated in the system.

As rule based shell, Level 5 stores the knowledge in rules, which are logic based structures as shown below:

```
RULE Sepsis 14
   IF Blood Culture IS Positive THEN
   StrNsepsis := "Neonatal Sepsis Confirmed" AND
   Sepsis CF 80
  IF Blood Culture IS Negative THEN
   SepsisConfirmation := " Neonatal Sepsis not Positive ... "AND
   Sepsis CF 20
  IF Blood Culture IS Negative THEN
   StrNsepsis: = "Neonatal Sepsis Not Diagnosed" AND
   Sepsis CF 20
 END
```

Figure 10.1. Level 5 Representations of Rules.

It was observed that, proper pruning and better classification is required to get the maximum accuracy of the system.

❖ *Inference Engine*

When rules are fired, knowledge base makes inferences by deciding which rules are satisfied by facts stored in the working buffer and executes the rules with highest priority. The rules whose patterns are satisfied by facts in the working buffer are stored in the agenda part of the inference engine. Figure 2 explains the inference process of the system.

```
Rule 1:     IF Birth Term Status IS Pre Term Baby AND Birth Weight Status IS
            Extemely Low Birth Weight ELBW AND Age IS More than 72 hrs
            AND Symptoms IS Lathergy = TRUE AND Symptoms IS Refusal to
            Suck = FALSE AND Symptoms IS Poor Cry = TRUE AND Symptoms
            IS Bleeding = TRUE
        THEN
            DiseaseConfirmation := " HIE III".
Rule 2: WHEN CHANGED
        BEGIN
            IF Consciousness OF HIE IS Irritable = TRUE AND Tone OF HIE IS
            Hypotonia AND Seizures OF HIE IS No AND Sucking or Respiration
            OF HIE IS Poor Suck
      THEN
          StrClassificationHIE OF HIE := "Mild"
            IF Consciousness OF HIE IS Lethargy AND Tone OF HIE IS Marked
            Hypotonia AND Seizures OF HIE IS Yes AND Sucking or Respiration
            OF HIE IS Unable to Suck
      THEN
            StrClassificationHIE OF HIE := "Moderate"
            IF Consciousness OF HIE IS Comatose AND Tone OF HIE IS Severe
            Hypotonia AND Seizures OF HIE IS Prolonged AND Sucking or
            Respiration OF HIE IS Unable to sustain spontaneous respiration
      THEN
          StrClassificationHIE OF HIE := "Severe"
END
```

Figure 10.2. Inference Process of the System.

❖ *Language Selection*

The proposed expert system is implemented in an object- oriented environment by using LEVEL 5 Object for Microsoft Windows (release 3.0 or higher) by the Information Builders Inc, USA. It has an integrated array of powerful tools like GUI development, Forms and Display builders and also has got capability to chain more than one knowledge base together. The data driven nature of domain suggest the use of forward-chaining inference engine. It runs in windows environment, which is the platform of choice among the end users.

❖ *Rough Set Application*

Rough set theory is an intelligent technique for managing uncertainties that is used for the discovery of data dependencies, to reduce redundancies, to evaluate the importance of attributes, to discover patterns in data, and to classify objects. There are several useful features of rough sets such as (i) extraction of rules from data sets in the form of if-then rules; (ii) it requires no external parameters unlike other intelligent techniques except the data itself; (iii) it can predict whether the data is complete or not. The computation of reduct and core using rough set theory is an important feature.
Initially three algorithms [9,10] namely, Exhaustive algorithm, Genetic algorithm, and Dynamic reducts have been used for finding reduct and core with the help of RSES 2.2 [11], a software tool that provides the means for analysis of tabular data sets with the use of various methods, in particular those based on Rough Set Theory.

❖ *Rule Generation*

Table 10.2. shows rules generated using different methods. There are different techniques for testing the prediction accuracy such as confusion matrix, prediction accuracy, receiver operating characteristics (ROC), and lift. This work deploys confusion matrix. These are achieved by using data mining concept in rough set paradigm in our earlier study [12].

Table 10.2. Rules and Prediction with Confusion Matrix.

Tool	Methods / Algorithms	No. of Rules	Prediction Accuracy (%)
RSES	Exhaustive without reduct	3568	70
	Genetic without reduct	1904	70
	Exhaustive with reduct	192	70
	Genetic with reduct	192	70
	Dynamic with reduct	192	70
ROSETTA	**Genetic with reduct**	**192**	**71.6**
	Johnson (with approx. solutions) with reduct	60	70.5

❖ *Re-building of Knowledge Base*

Now we get the Rules generated by the use of Exhaustive algorithm, Genetic algorithm, and Johnson Algorithm and use them as knowledge. Here domain experts and other source of knowledge are not incorporated into inference engine. Rules are re-written and rebuild the knowledge base for differential diagnosis.

❖ **Level 5 representation of Rule:**

RULE 1

IF Birth Term Status IS Pre Term Baby AND Birth Weight Status IS Extemely Low Birth Weight ELBW AND Age IS More than 72 hrs AND Symptoms IS Lathergy = TRUE AND Symptoms IS Refusal to Suck = FALSE AND Symptoms IS Poor Cry = TRUE AND Symptoms IS Bleeding = TRUE

THEN DiseaseConfirmation := " HIE III"

RULE 2

IF Birth Term Status IS Term Baby AND Birth Weight Status IS Extemely Low Birth Weight ELBW AND Age IS More than 72 hrs = FALSE AND Symptoms IS Lathergy = TRUE AND Symptoms IS Refusal to Suck = TRUE

THEN DiseaseConfirmation := " SEPTICIMIA"

10.4. Results and Conclusion

For a comparative study, we attempted other algorithms [13, 14] with the help of ROSETTA [15] software tool. Comparative results are shown in the Table 10.3.

Table 10.3. Comparison: Reduct and Core Generation.

Tool	Methods/ Algorithms	No. Reduct	Length Reducts	Size of Core
RSES	Exhaustive	3	10	8
	Genetic	3	10	8
	Dynamic	3	10	8
ROSETTA	Genetic	3	10	8
	Johnson(with approx. solutions)	1	8	8

It is observed that ROSETTA implementation of genetic algorithm with reduct shows the best performance. It is also observed that prediction accuracy does not decrease with reucts where number of rules are as low as 192 compared to 3568 and 1904. There is effectively rather no contribution observed of three attributes ‘Excessive_Jaundice’, ‘Sclerema’, and ‘GI_Disorder’ in decision making. Moreover, Johnson algorithm offers somewhat a comparable accuracy (70.5%) with only 60 rules with only one reduct of length eight. So, Johnson algorithm may be used for further investigations. In an another study [16] of ours, we have applied a data mining technique (J48/C4.5) on the same data set using WEKA [17] implementation. That model also predicts diseases with ~70% accuracy with a confidence factor 0.55.

After re-building of the knowledge base, 20 different case studies were again taken up for testing the expert system which is written on basis of Johnson algorithm rules. The outputs, i.e. differential diagnosis were verified with domain experts. Out of 20 cases 14 were matched with domain experts. It gives prediction accuracy of 70%, which is exceptionally excellent in terms of number of rules are concerned.

The above experiments in this chapter substantiate that, the proposed expert system is giving better performance by speeding up the process of differential disease diagnosis for the neonates and decrease the diagnosis test and cost as well.

References

1. Roy Chowdhury Dilip, Samanta R.K. Chatterjee M., "*A Study of the Status of New Born in Terai Region of West Bengal*", A.M.S.E. France, No. 05 225 (2C), 2007.

2. Kumar D., Verma A., and Sehgal V. K., "*Neonatal Mortality in India: Rural and Remote Health*", Vol. 7, pp. 833, http://www.rrh.org.au Available: [Last accessed 8th October, 2009].

3. Bang Abhay T., Bang Rani A., Baitule Sanjay, Deshmukh Mahesh and Reddy M. Hanimi, "*Burden of Morbidities and the Unmet Need for Health Care in Rural Neonates-A Prospective Observational Study in Gadchiroli, India*", Indian Pediatrics, Vol. 38, pp. 952-965, 2001.

4. Zadeh L. A., "*Fuzzy Logic, Neural Networks, and Soft Computing*", Comm. ACM, Vol. 37, pp. 77-84, 1994.

5. Agarkar A. M., Ghatol A. A., "*FFANN Based Cost Effective Major Infant Disease Management*", International Journal of Computer Applications, Vol. 7, pp. 29-33, 2010.

6. Ghai O. P., "*Essential Pediatrics*", 6th Edition, CBS Publisher and Distributors, New Delhi, pp. 136, 2004.

7. Gupte Suraj, "*The Short Text Book of Pediatrics*", 10th Edition, Jaypee Publication, New Delhi, pp. 593, 2003.

8. Lahariya Chandrakant and Paul K Vinod, "*Burden, Differentials, and Causes of Child Deaths in India*", Indian Journal of Pediatrics, Vol. 77, No. 11, pp. 1312-1321, 2010.

9. Bazan J., "*A Comparison of Dynamic and Non-dynamic Rough Set Methods for Extracting Laws from Decision Table*", In L. Polkowski, A. Skowron (eds.), Rough Sets in Knowledge Discovery, Physica-Verlag, Heidelberg, pp. 321-365, 1998.

10. Bazan J. et.al., "*Rough Set Algorithms in Classification Problems*", In L. Polkowski, S. Tsumoto, and T. Lin (eds.), Rough Set Methods and Applications, Physica-Verlag, Heidelberg, pp. 49-88, 2000.

11. "*RSES 2.2 User's Guide*", Warsaw University, http://logic.mimuw.edu.pl/~rses, January 19, 2005

12. Roy Chowdhury Dilip, Samanta R.K., Chatterjee M., "*A Data Mining Model for Differential Diagnosis of Neonatal Disease*", IFRSA's International Journal Of Computing, Vol. 1, No. 2, pp 143-150, 2011.

13. S. Vinterbo and A. Ohrn, "*Minimal Approximate Hitting Sets and Rule Templates*", Int. J. Approximate Reasoning, Vol. 25, No. 2, pp.123-143, 2000.

14. Johnson D. S., "*Approximate Algorithms for Combinatorial Problems*", J. of Computer and System Sciences, Vol. 9, pp. 256-278, 1974.

15. "*The ROSETTA Homepage*", [http://www.idi.ntnu.no/~aleks/rosetta/].

16. Roy Chowdhury Dilip, Samanta R.K., Chatterjee M, "*Data Mining for Neonatal Prevalent Disease of North Bengal Districts*", Proc. Nat. Sem. on Data Mining and Decision Support, Vidyasagar University, India, pp. 24-31, 2010.

17. Hall M., E. Frank, Holmes G., Pfahringer B., Reutemann P., Witten I. H., "*The WEKA Data Mining Software: An Update*", SIGKDD Explorations, Vol. 11, No. 1, pp. 10-18. 2009.

CHAPTER 11

Artificial Neural Network Model for Neonatal Disease Diagnosis and Management†††

11.1. Introduction

Artificial Intelligence techniques incorporating computer based decision support system works in such a way, somewhat like a human thinks. Several neural network models are developed which helps doctors in diagnosing the patients more correctly and accurately. Neural network provides a general way of approaching problems. When the output of the network is categorical, it is performing prediction and when the output has discrete values, and then it is doing classification. Neural network based decision support in medicine, particularly for the neonates, has at least the role of enhancing the consistency of neonatal care. The significance of disease diagnosis by artificial intelligence is not obscure now days. The increasing demand of Artificial Neural Network application for predicting the disease shows better performance in the field of medical decision making.

The study in this chapter represents the use of artificial neural networks in predicting neonatal disease diagnosis. The proposed technique involves training a Multi Layer Perceptron (MLP) with a Back Propagation (BP) learning algorithm to recognize a pattern for the diagnosing and prediction of neonatal diseases. A comparative study of using different training algorithm of MLP, Quick Propagation, Conjugate Gradient Descent, shows the higher prediction accuracy. The Backpropogation algorithm was used to train the ANN architecture and the same has been tested for the various categories of neonatal disease. About 94 cases of different sign and symptoms parameter have been tested in this model. This study exhibits ANN based prediction of neonatal disease and improves the diagnosis accuracy of 75% with higher stability.

11.1.1. Neural Network based Architecture [1]

Basically neural networks are networks which contain various processing elements that we call nodes. Depending upon the external input it processes the information and provides the output. In the network the nodes are basic models of neurons. In a neural network, all the information are distributed throughout the inputs to all of the nodes.

††† This chapter is based on the publication made by the author entitled "An Artificial Neural Network Model for Neonatal Disease Diagnosis", International Journal of Artificial Intelligence and Expert Systems (IJAE), Vol. 2, Issue 3, pp. 96-106, August, 2011.

Depending upon the weights inputs nodes are summed up. If the sum of all the inputs to a node exceeds some threshold value *T*, the node executes and produces an output which is passed on to other nodes or is used to produce some output response. In the simplest case no output is produced if the total input is less than *T*. If we think for difficult models, there for the output we need to depend on an activation function.

Like the biological neural system and operations related with networks, artificial neural network system works. This is the era where computing is really advanced, still there are some certain tasks or process which are not possible by the program made for any common microprocessor. Thus by developing software implementation of a neural network can be a solution for this.

Taking immense help from the natural human nervous system, neural networks working principles are decided and processed. They are the simplified models for processing of many intelligent abilities such as learning, generalization and abstraction and more. A typical artificial neural network might have a hundred neurons. In comparison, the human nervous system is believed to have about 3×10^{10} neurons. Thus making similar kinds of thing is not very simple enough.

Figure 11.1(a) and Figure 11.1(b) is showing the schematic view of the biological neuron and artificial neuron respectively. In case of biological neuron, signals from other neurons are conveyed to the cell body by dendrites. Then the signals are sent to the axon for distributing other neurons. On the other hand the operation of the artificial neuron is similar to that, and even quite simple for the operation, as biological neuron does. Here activations from other neurons are summed up at the neuron. The summed up neurons then passed through an activation function. Lastly other neuron receives the value and this process goes on continuing till the operations are reaching at the end.

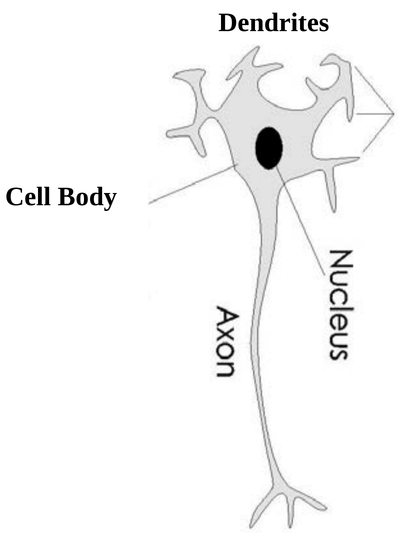

Figure 11.1.(a) Biological Neuron

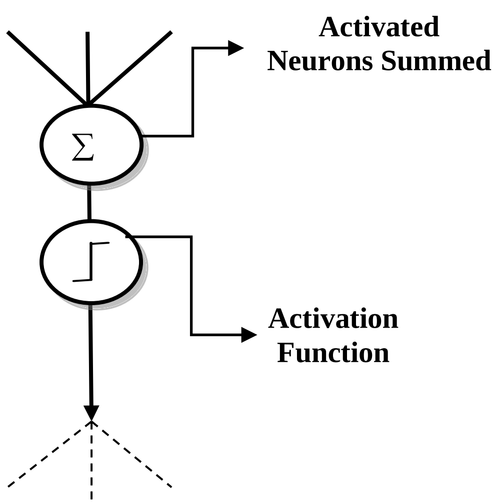

Figure 11.1.(b) Artificial Neuron

11.1.2 Single Neuron Model [2]

An Artificial Neural Network is a nonlinear system which is adaptive in nature that learns to perform an input output function from data. The system parameters are changed during operation, in the training phase. After the training phase the Artificial Neural Network parameters are fixed and the system is deployed to solve the problem at hand, this phase is called testing phase. The Artificial Neural Network is built with a systematic step-by-step procedure to optimize a performance criterion or to follow some implicit internal constraint, which is commonly referred to as the learning rule. The input/output training data are fundamental in neural network technology, because they convey the necessary information to discover the optimal operating point.

Figure 11.2 depicts the single node. Here the Inputs to the nodes are the values of x_1, x_2, x_3, ….., x_n , which takes values of -1, 0, 1 or real values within the range -1 to 1. The weights w_1, w_2, w_3, ….. , w_n , which matches the strengths of neuron. They serve to increase or decrease the effects of corresponding a_i input values. The sum of the products of x_i, w_i, i = 1, 2, 3, ………, n, serve as the total combined input to the node. If this sum is large enough to exceed the threshold amount *T*, the node fires and produces an output ***y***, an activation function-value placed on the node's output links. This output may then be the input and other nodes or final output response from the network [2].

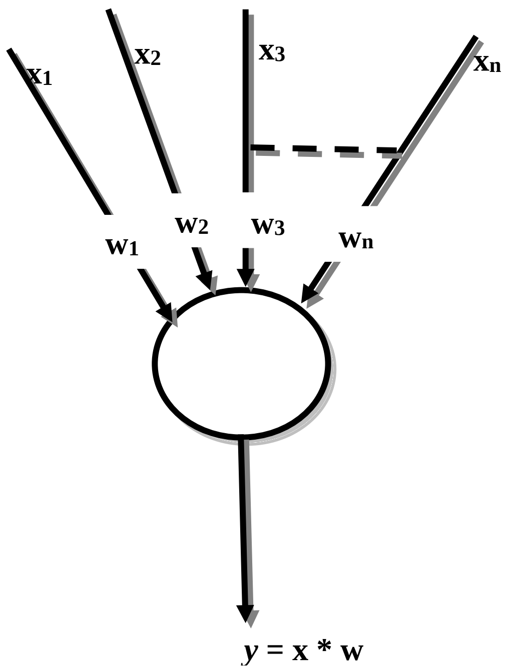

Figure 11.2. Single Neuron

11.1.3. Multi Layer Network Model [2]

Multilayer networks solve the classification problem for non linear sets by employing hidden layers, whose neurons are not directly connected to the output. The additional hidden layers can be interpreted geometrically as additional hyper-planes, which enhance the separation capacity of the network. Figure 11.3 shows typical multilayer network architectures. Here the first layer serves as the input layer, receiving inputs from set of input nodes. The second layer is called hidden layer. It receives inputs from the first layer and produces a pattern of inputs to the third layer, which is the output layer. Input links to layer *j*(*j* = 1, 2, 3) have weights w_{ij} for i = 1, 2, 3, ….. , n. General multilayer network having *n* nodes i.e. number of rows in each of *m* layers i.e. number of columns of nodes will have weights represented as an *n* x *m* matrix *W*. Using this representation, nodes having no interconnecting links will have a weight value of 0. Networks consisting of more than three layers would be more complex than the network illustrated in figure. 11.3.

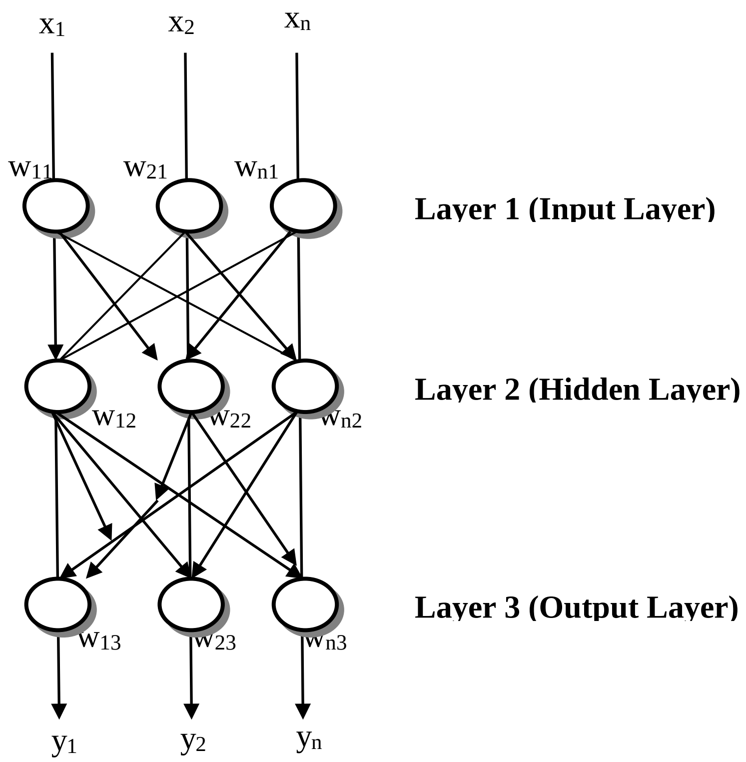

Figure 11.3. Multilayer Neural Network Model.

A neural network can be thought of as a black box that transforms the input vector *x* to the output vector y where the transformation performed is the result of the pattern of connections and weights, according to the values of the weight matrix *w*.

Considering the vector product,

$$x * w = \sum x_i w_i \quad (1)$$

There is a geometric interpretation for this product. It is equivalent to projecting one vector onto the other vector in n-dimensional space. This idea illustrated in figure. 11.4 for the two dimensional case.

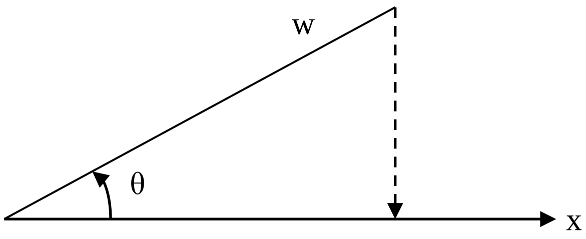

Figure 11.4. Vector Multiplication and Projection.

The magnitude of the resultant vector is given by following expression,

$$x * w = |x| |w| \cos\theta \quad (2)$$

Here, **| x |** denote**s** the norm or length of the vector x. This product is maximum when both vectors point in the same direction. When θ = 180°. This tells how the vector in the weight matrix W influences the inputs to the nodes in a neural network. Figure 11.5 and Figure 11.6 depicts the Two-Layer-Network and the Three-Layer-Network.

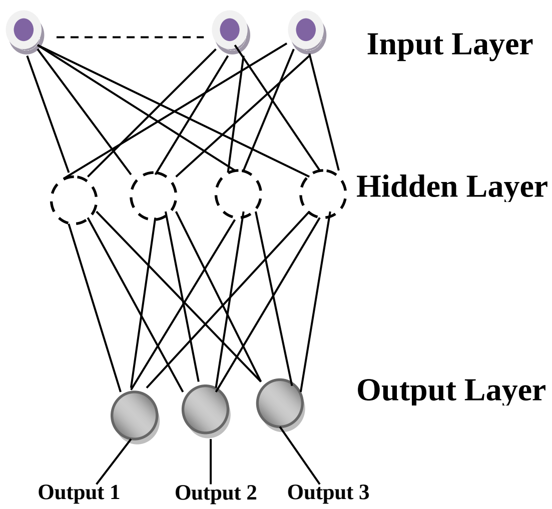

Figure. 11.5. Two Layer Neural Network Model.

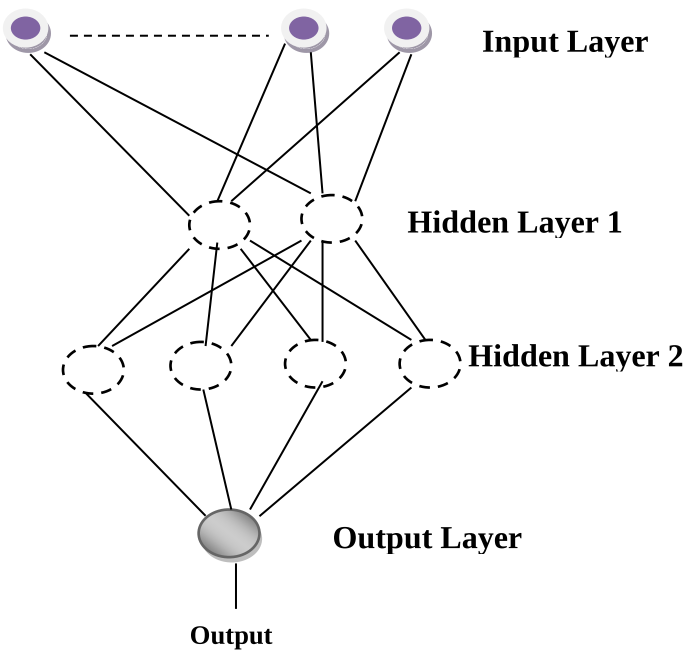

Figure 11.6. Three Layers Neural Network Model.

11.1.4. Conventional Computing Vs. Neural Networks

As we know in conventional computing system, a serial computer has a central processor that can address an array of memory locations where data and instructions are stored. The processor reads the instruction from memory address and then executes the instruction to get the desired result in form of output. These outputs are then saved in memory location specified by the processor. In a serial system the computational steps are deterministic, sequential and logical, and the state of a given variable can be tracked from one operation to another. There is a involvement of complex instruction processing cycle by the processor in conventional computing method

In contrast, artificial neural networks are not sequential or necessarily deterministic. There are no complex central processors, rather there are many simple ones which generally do nothing more than take the weighted sum of their inputs from other processors. Actually programmed instructions are not executed by artificial neural network. In response of input patters as the inputs, the ANN parallel execution done. In fact no separate memory spaces are required for keeping the data, rather information and data is contained in the overall activation state of the network. Hence more knowledge can be represented by the network itself.

ANN provides an analytical alternative to conventional techniques which are often limited by strict assumptions of normality, linearity, variable independence etc. Because an ANN can capture many kinds of relationships it allows the user to quickly and relatively easily model phenomena which otherwise may have been very difficult or impossible to explain otherwise. Depending on the nature of the application and the strength of the internal data patterns you can generally expect a network to train quite well. This applies to problems where the relationships may be quite dynamic or non-linear.

11.1.5. ANN in Neonatal Disease Diagnosis Domain - Justification

Among various phases of child development, Neonatal phase considered to be one of the vital phases. In India, 30% to 40% babies are Low Birth Weight babies and about 10% to 12% of Indian babies are born less than 37 completed weeks (preterm). Thus, these babies are physically immature and cause the high neonatal mortality [3]. In a study, we have described about prevalence diseases are the major causes of deaths in the neonates in Terai region of West Bengal [4]. This mortality problem, especially in rural areas [5], can prevail over through fast and accurate disease diagnosis and management of the newborn. In our earlier studies of data mining model development, several classification techniques have applied to get the maximum accuracy [6]. However, any ANN based model may be useful for classification of disease and even for taking necessary decision. This chapter describes how artificial intelligence, for example artificial neural networks can improve this area of diagnosis.

The proposed model has the potential to cover rare conditions of all the exceptional symptoms of neonatal diseases to diagnose. The increasing range of neonatal patient information makes it feasible to more accurately quantify important experimental indicators, such as the relative likelihood for competing diagnoses or the clinical outcome. It is observed that, in few instances, computer-assisted diagnoses, particularly ANN based model have been claimed to be even more accurate than those decision taken by domain experts [7].

11.2. Related Studies of Artificial Neural Network

There are several studies which have applied neural networks in the diagnosis of different disease. An artificial neural network trained on admission data can accurately predict the mortality risk for most preterm infants. However, the significant number of prediction failures renders it unsuitable or individual treatment decisions. In a study [8], the artificial neural network performed significantly better than a logistic regression model (area under the receiver operator curve 0.95 *vs* 0.92). Survival was associated with high morbidity if the predicted mortality risk was greater than .50. There were no preterm infants with a predicted mortality risk of greater than 0.80. The mortality risks of two non-survivors with birth weights >2000 g and severe congenital disease had largely been underestimated.

In another study [9], an effective arrhythmia classification algorithm used for the heart rate variability (HRV) signals. The proposed method is based on the Generalized Discriminant Analysis (GDA) feature reduction technique and the Multilayer Perceptron (MLP) neural network classifier. At first, nine linear and nonlinear features are extracted from the HRV signals and then these features are reduced to only three by GDA. Finally, the MLP neural network is used to classify the HRV signals. The proposed arrhythmia classification method is applied to input HRV signals, obtained from the MIT-BIH databases. Here, four types of the most life threatening cardiac arrhythmias including left bundle branch block, fist degree heart block, Supraventricular tachyarrhythmia and ventricular trigeminy can be discriminated by MLP and reduced features with the accuracy of 100%.

The study [10] of a functional model of ANN is proposed to aid existing diagnosis methods. This work investigated the use of Artificial Neural Networks (ANN) in predicting the Thrombo-embolic stroke disease. The Backpropogation algorithm was used to train the ANN architecture and the same has been tested for the various categories of stroke disease. This research work demonstrates that the ANN based prediction of stroke disease improves the diagnosis accuracy with higher consistency. This ANN exhibits good performance in the prediction of stroke disease in general and when the ANN was trained and tested after optimizing the input parameters, the overall predictive accuracy obtained was 89%.

As per the artificial neural networks in medicine world map [11], different Universities, Research Centres, Medical Diagnostic Centres are using ANN for medical diagnosis and management. Some studies are carried out using some combined architecture using ANN and different data mining techniques [12].

11.3. MLP Neural Network Model

11.3.1. Structure of MLP

In medical decision making a variety of neural networks used for decision accuracy. MLPs are the simplest and commonly used neural network architectures programs due to their structural litheness, good representational capabilities and availability, with a large number of programmable algorithms [13]. MLPs are feed forward neural networks and universal approximators, programmed with the standard back propagation algorithm. They are supervised networks so they require a desired response to be trained. They are able to transform input data into a desired response, so they are widely used for pattern classification. With one or two hidden layers, they can approximate virtually any input-output map. Generally, an MLP consists of three layers: an input layer, an output layer and an intermediate or hidden layer. In this network, every neuron is connected to all neurons of the next layer, in other words, an

MLP is a fully connected network [14]. Figure 11.7. shows the structure of a MLP network.

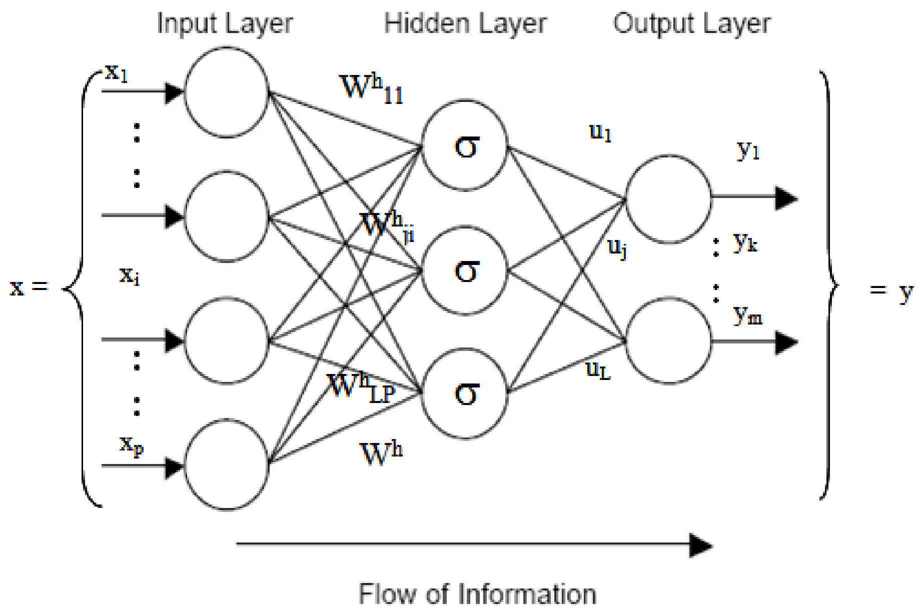

Figure 11.7. A Structure of MLP Network

On the left this network has an input layer with three neurons, in the middle, one hidden layer with three neurons and an output layer on the right with two neurons. There is one neuron in the input layer for each predictor variable ($x_1 \ldots x_p$). In the case of categorical variables, *N*-1 neurons are used to represent the *N* categories of the variable.

11.3.2. MLP Input Layer

A vector of predictor variable values ($x_1 \ldots x_p$) is presented to the input layer. The input layer (or processing before the input layer) standardizes these values so that the range of each variable is -1 to 1. The input layer distributes the values to each of the neurons in the hidden layer. In addition to the predictor variables, there is a constant input of 1.0, called the *bias* that is fed to each of the hidden layers; the bias is multiplied by a weight and added to the sum going into the neuron.

The net calculation of input and output of the j hidden layer neurons are as follows:

$$net^h_j = \sum_{t=1}^{N+1} W_{ji} x_i \qquad (3)$$

$$y_j = f\,(net^h_j) \qquad (4)$$

11.3.3. MLP Hidden Layer

Arriving at a neuron in the hidden layer, the value from each input neuron is multiplied by a weight (w_{ji}), and the resulting weighted values are added together producing a combined value u_j. The weighted sum (u_j) is fed into a transfer function, σ. The outputs from the hidden layer are distributed to the output layer.

11.3.4. MLP Output Layer

The value from each hidden layer neuron is multiplied by a weight (w_{kj}), and the resulting weighted values are added together producing a combined value *u, at time of a*rriving at a neuron in the output layer *j*. The weighted sum (u_j) is fed into a transfer function, σ, which outputs a value y_k. The *y* values are the outputs of the network. If a regression analysis is being performed with a continuous target variable, then there is a single neuron in the output layer, and it generates a single *y* value. For classification problems with categorical target variables, there are *N* neurons in the output layer producing *N* values, one for each of the *N* categories of the target variable.
Calculate the net inputs and outputs of the k output layer neurons are :

$$\mathrm{net}^{o}{}_{k} = \sum_{j=1}^{J+1} V_{kj} y_j \tag{5}$$

$$Z_k = f(\mathrm{net}^{0}{}_{k}) \tag{6}$$

Update the weights in the output layer (for all k, j pairs)

$$v_{kj} \leftarrow v_{kj} + c\lambda(d_k - Z_k) Z_k (1 - Z_k)\, y_j \tag{7}$$

11.4.1. Proposed Model

11.4.1.1. Input Data

The data for this study have been collected from 94 patients who have symptoms of neonatal diseases. The data have been standardized so as to be error free in nature. All the cases are analyzed after careful scrutiny with the help of the pediatric expert. Table 11.1. below shows the various input parameters for the prediction of neonatal disease diagnosis.

Table 11.1. Input Parameters for Prediction Neonatal Disease.

Sl.No.	Parameters	Column Type
1	Birth_Term_Status	Categorical
2	Birth_Weight_Status	Categorical
3	Age_in_Hours>72	Categorical
4	Lethargy	Categorical
5	Refusual_to_Suck	Categorical
6	Poor_Cry	Categorical
7	Poor_Weight_gain	Categorical
8	Hypothalmia	Categorical
9	Sclerema	Categorical
10	Excessive_Jaundice	Categorical
11	Bleeding	Categorical
12	GI_Disorder	Categorical
13	Seizure	Categorical
14	Sluggish_Neonatal_Reflex	Categorical

11.4.2. Feature Selection of Dataset

Data analysis information needed for correct data preprocessing. After data analysis, the values have been identified as missing, wrong type values or outliers and which columns were rejected as unconvertible for use with the neural network [15]. Feature selection methods are used to identify input columns that are not useful and do not contribute significantly to the performance of neural network. In this study, Genetic method is used for input feature selection. Genetic Algorithm has been discussed broadly on Chapter 12. Genetic algorithms method [16] starts with a random population of input configurations. Input configuration determines what inputs are ignored during performance test. At each following step uses a process analogous to natural selection to select superior configurations and use them to generate a new population. Each step successively produces better input configuration. At the last step the best configuration is selected. The method is very time-consuming but good for determining mutually-required inputs and detecting interdependencies. This method use generalized regression neural networks (GRNN) or probabilistic neural networks (PNN) because they train quickly and proved to be sensitive to the irrelevant inputs. The removal of irrelevant inputs will improve the generalization performance of a neural network. Table 11.2 shows the finalized input parameters after applying feature selection method.

Table 11.2. Percentage of Importance of Input Data after Feature Selection

Code	**Name of the Input Column**	**Input state**	**Importance %**
C3	Age_in_Hours>72	Two-state	0.551381
C4	Lethargy	Two-state	12.344225
C6	Poor_Cry	Two-state	0.832139
C7	Poor_Weight_gain	Two-state	18.140229
C8	Hypothalmia	Two-state	15.23048
C9	Sclerema	Two-state	0.088902
C10	Excessive_Jaundice	Two-state	14.179179
C11	Bleeding	Two-state	4.159191
C12	GI_Disorder	Two-state	8.745518
C13	Seizure	Two-state	22.076618
C14	Sluggish_Neonatal_Reflex	Two-state	3.652138

11.4.3. Development of Neural Network Architecture

In this study, the multilayered feed-forward network architecture with 11 input nodes after feature selection of the input data, 5 hidden nodes, and 13 output nodes have been used for the neural network architecture. The numbers of input nodes are determined by

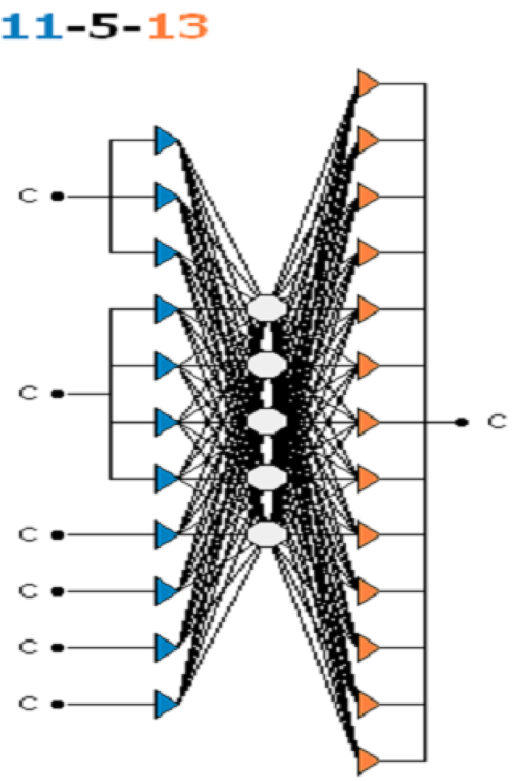

Figure 11.8. ANN Architecture for Neonatal Disease Diagnosis.

the finalized data; the numbers of hidden nodes are determined through trial and error; and the numbers of output nodes are represented as a range showing the disease classification. The most widely used neural-network learning method is the Back Propagation algorithm [17]. Learning in a neural network involves modifying the weights and biases of the network in order to minimize a cost function. The cost function always includes an error term a measure of how close the network's predictions are to the class labels for the examples in the training set. Additionally, it may include a complexity term that reacts a prior distribution over the values that the parameters can take. The activation function considered for each node in the network is the binary sigmoidal function defined (with $\sigma = 1$) as output = $1/(1+e^{-x})$, where x is the sum of the weighted inputs to that particular node. This is a common function used in many back propagation network. This function limits the output of all nodes in the network to be between 0 and 1. All neural networks are basically trained until the error for each training iteration stopped decreasing. Figure 11.8. shows the architecture of the specialized network for the prediction of stroke disease. The complete sets of final data (11 inputs) are presented to the generic network, in which the final diagnosis corresponds to output units.

The following are the results generated from the input given to the neural network after going through the process of careful training, validation and testing using Neuro Intelligence tool [18]. Table 11.3. shows the various categories of neonatal diseases and their classification and probability statistics.

Table 11.3. Category Weights (Prior Probabilities)

Category	Probability
HIE_III	0.1702128
Hemorrhage	0.0106383
HIE_II	0.0425532
Hypo_Thalmia	0.0212766
Jaundice	0.0212766
Jaundice_BA	0.0319149
MD_Hypocalcimia	0.0957447
MD_Hypoglycimia	0.0319149
MD_Hypothermia	0.0319149
No_Disease	0.0851064
Others	0.0531915
Septicemia	0.3936170
Sizure_Disorder	0.0106383

11.4.4. Training Process of MLP Networks

In this context, our objectives of the training process was to find the set of weight values which will cause the output from the neural network to match the actual target values as closely as possible. We have faced several issues concerned in designing and training a multilayer perceptron network model. Some of the issues are:

i. To select the number of hidden layers to use in the network.
ii. To decide the number of neurons to be used in each hidden layer.
iii. Converging to an optimal solution in a reasonable period of time.
iv. Finding a globally optimal solution that avoids local minima.
v. Validating the neural network to test for overfitting.

11.4.5. Hidden Layers Selection

In our study one hidden layer is sufficient for the network. Two hidden layers are required for modeling data with discontinuities such as a saw tooth wave pattern. As we found that using two hidden layers rarely improves the model, and it may introduce a greater risk of converging to a local minima. So, three layer models with one hidden layer are recommended for our study.

11.4.6. Deciding How Many Neurons to be Used in the Hidden Layers

The most significant characteristics of a multilayer perceptron network is to decide the number of neurons in the hidden layer. The network may be unable to model complex data, and the resulting fit will be poor, If an inadequate number of neurons are used in the network. Similarly, If too many neurons are used, the training time may become excessively long, and, worse, the network may over fit the data. When overfitting occurs, the network will begin to model random noise in the data. The result is that the model fits the training data extremely well, but it generalizes poorly to new, unseen data. Validation must be used to test for this. In view of the above our model consist of 5 neurons with one hidden layer.

11.5. Result and Discussion

During data analysis, the column type is recognized. The last column is considered as the target or output one and other columns will be considered as input columns. The dataset is divided in to training, validation and test sets. The Data have been analyzed using NeuroIntelligence Tool [17].

Table 11.4. Data Partition Set.

Partition Set Using	Records	Percentage (%)
Total	94	100
Training Set	64	68
Validation Set	15	16
Test Set	15	16
Ignore Set	0	0

To train a neural network is the process of setting the best weights on the inputs of each of the units. It has been proved that Genetic Algorithm and Back-Propagation neural network hybrids in selecting the input features for the neural network reveals the performance of ANN can be improved by selecting good combination of input variables [15]· Training set is considered to be the part of the input dataset used for neural network training and network weights adjustment**.** The validation set is parts of the data are used to tune network topology or network parameters other than weights. The validation set is used to choose the best network we have changed the number of units in the hidden layer. The test set is a part of the input data set used to test how well the neural network will perform on new data. The test set is used after the network is trained, to test what errors will occur during future network application.

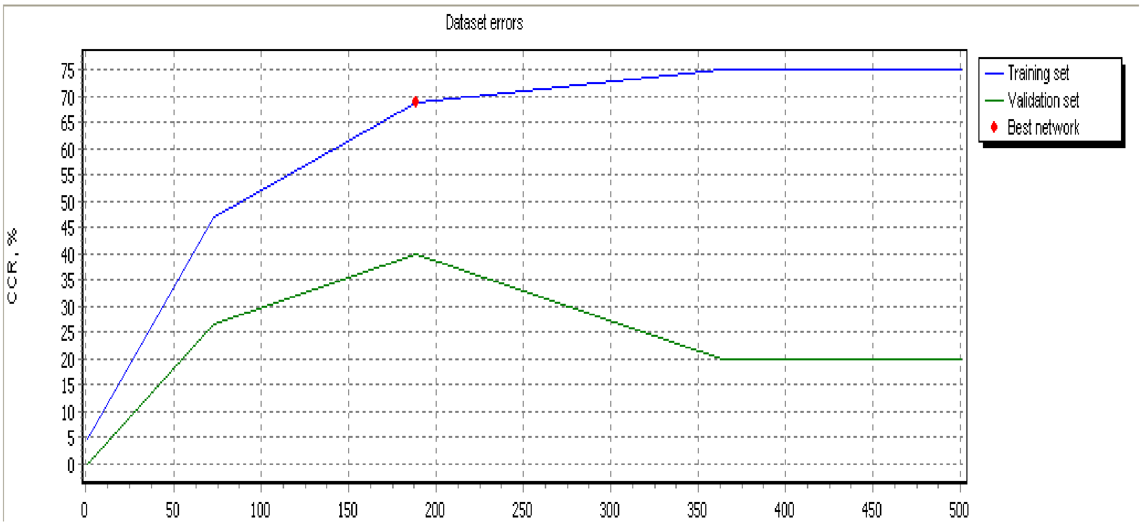

Figure 11.9. Errors in Data Set.

Figure 11.9. shows the various data set errors with respect to training set, validation set and the best network. It accomplishes the level of best network after training through repeated iterations. Correct Classification Rate for training and validation has done to

find the best network after a number of iteration. Table 11.5 shows the number of Iterations and CCR for training and validation as well.

Table 11.5. Best Network on Iterations.

Iteration	CCR (training)	CCR (validation)
73	46.875	26.666666
189	68.75	40
364	75	20

The Network errors have been shown graphically in Figure 11.10. We have tested the trained network with a test set, in which the outcomes are known but not provided to the network. We used diagnostic criteria and disease pattern status to train a neural network to classify individuals as diagnosed with disease name by several categories of neonatal disease.

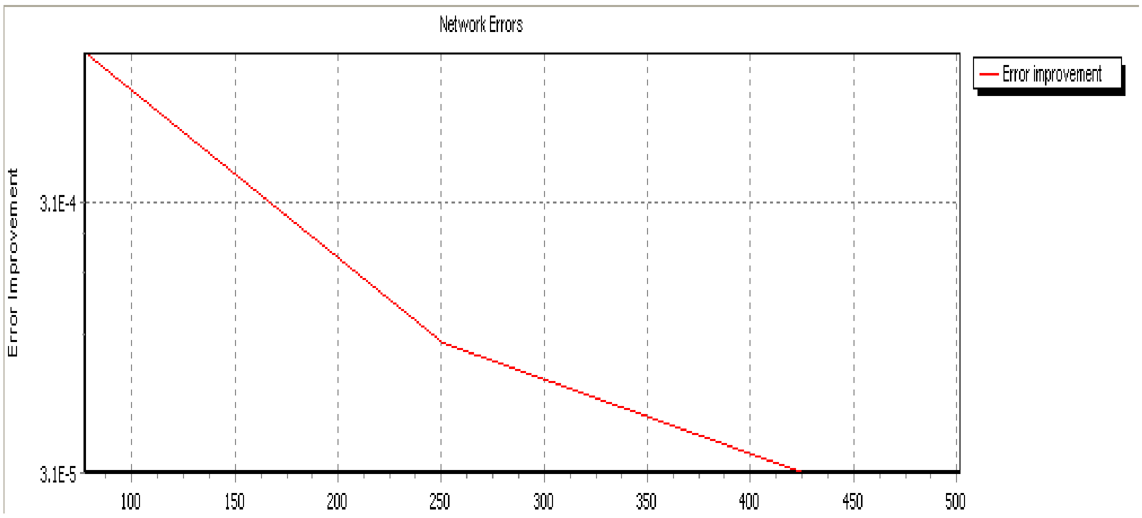

Figure 11.10. Network Error.

The study shows that 39.36% of the respondents have the symptoms of Septicemia; 17.02% have the symptoms of HIE III; and 9.57% of the patients have the symptoms of Metabolic Disorder - Hypocalcaemia. These are the most prevalent disease in the Terai region of North Bengal [4]. Table 11.6 shows the disease conformation percentage with category.

Table 11.6. Disease Conformation Set.

No of cases	Name of Disease Conformation	Percentage (%)
1	Hemorrhage	1.06%
4	HIE_II	4.26%
16	HIE_III	17.02%
2	Hypo_Thalmia	2.13%
2	Jaundice	2.13%
3	Jaundice_BA	3.19%
9	MD_Hypocalcimia	9.57%
3	MD_Hypoglycimia	3.19%
3	MD_Hypothermia	3.19%
8	No_Disease	8.51%
5	Others	5.32%
37	Septicemia	39.36%
1	Sizure_Disorder	1.06%

11.6. Conclusion

Neural network has been established of their potentials in many domains related with medical disease diagnosis and other application. Although, Neural networks never replace the human experts instead they can helpful for decision making, classifying, screening and also can be used by domain experts to cross-check their diagnosis. In our earlier studies on rough set based computing model [19] and soft computing model [20], we have established the accuracy of 71% for decision making of prevalence neonatal disease. This ANN MLP model proves the better results and helps the domain experts and even person related with the field to plan for a better diagnose and provide the patient with early diagnosis results as it performs realistically well even without retraining. As clinical decision making requires reasoning under uncertainty, expert systems and fuzzy logic will be suitable techniques for dealing with partial evidence and with uncertainty regarding the effects of proposed interventions. Neural Networks have been proven to produce better results compared to other techniques for the prediction tasks. Our study concludes with higher prediction result and when the Network has trained and tested after optimizing the input parameters, the overall predictive accuracy acquired was 75%.

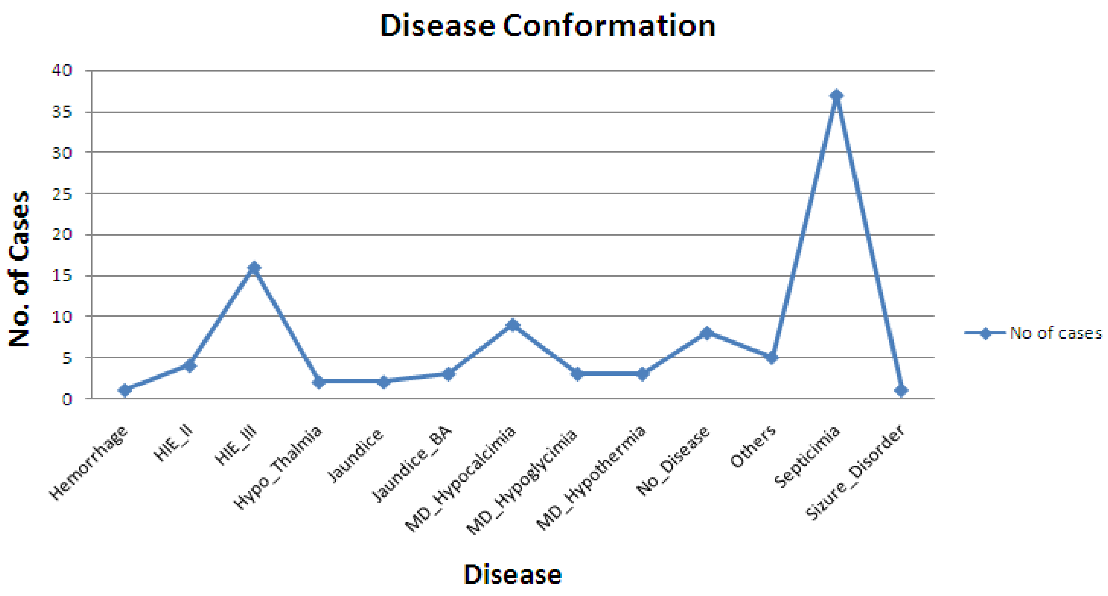

Figure 11.11. Various Neonatal Disease with No. of Cases.

A comparative study [21] is being presented in table. 11.7 to establish the relative suitability of ANN technique with other techniques such as RSES[22] and ROSSETTA[23]. The result of the table clearly demonstrates the superiority of ANN technique over other techniques explained earlier.

Table 11.7. A Comparative Study of Different Techniques.

Tools	**Methods/ Algorithms**	**Prediction Accuracy (%)**
RSES[22]	Exhaustive without Reduct	70
	Genetic without Reduct	70
	Exhaustive with Reduct	70
	Genetic with Reduct	70
	Dynamic with Reduct	70
ROSETTA[23]	Genetic without Reduct	71.6
	Johnson(with approx. solutions) with Reduct	70.5
NEURO INTELLIGENCE [18]	ANN with MLP	75

References

1. Bhattacharjee S., "*Artificial Intelligence*", University Science Press, First Edition, pp. 192-193, 2008.

2. Patterson Dan W., "*Introduction to Artificial Intelligence and Expert Systems*", Prentice Hall, pp. 343-345, 1990.

3. Kumar D., Verma A., and Sehgal V. K. (2007). "*Neonatal mortality in India.*" Rural and Remote Health. [On-line]. 833(7). www.rrh.org.au Available: [Last accessed 8th October, 2009].

4. Roy Chowdhury Dilip, Samanta R.K., Chatterjee M., "*A Study of the Status of New Born in Terai Region of West Bengal*", A.M.S.E. France, No. 05 225 (2C), 2007

5. Bang Abhay T., Bang Rani A., Baitule Sanjay, Deshmukh Mahesh and Reddy M. Hanimi, "*Burden of Morbidities and the Unmet Need for Health Care in Rural Neonates-A Prospective Observational Study in Gadchiroli, India*", Indian Pediatrics, Vol. 38, pp. 952-965, 2001.

6. Roy Chowdhury Dilip, Samanta R.K., Chatterjee M., "*A Data Mining Model for Differential Diagnosis of Neonatal Disease*", IFRSA's International Journal Of Computing, Vol. 1, No. 2, pp 143-150, 2011.

7. Qiua X., Taob N. and Tana Y., et al. "*Constructing of the Risk Classification Model of Cervical Cancer by Artificial Neural Network*", Expert Systems with Applications, An International Journal Archive. Vol. 32, No. 4, pp. 1094-1099, 2007.

8. Zernikow B., Holtmannspoetter K., Michel E., Pielemeier W., Hornschuh F., Westermann A., and Hennecke K, "*Artificial Neural Network for Risk Assessment in Preterm Neonates.*", Arch Dis Child Fetal Neonatal Ed. vol. 79(2), pp. F81–F82, 1998.

9. Yaghouby F., Ayatollahi A. and Soleimani R., "*Classification of Cardiac Abnormalities Using Reduced Features of Heart Rate Variability Signal.*", World Applied Sciences Journal, Vol. 6, No. 11, pp. 1547-1554, 2009.

10. Shanthi D., Sahoo G. and Saravanan N., "*Designing an Artificial Neural Network Model for the Prediction of Thrombo-embolic Stroke.*", International Journals of Biometric and Bioinformatics (IJBB), Vol. 3, No.1, pp. 10-18, 2009.

11. "*Artificial Neural Networks in Medicine World Map*", USENET: http://www.phil.gu.se/ann/annworld.html, Available: [Last accessed July 21st, 2011].

12. Zhou Z. H. and Jiang Y., "*Medical Diagnosis with C4.5 Rule Preceded by Artificial Neural Network Ensemble*", IEEE Transaction on Information Technology in Biomedicine, Vol. 7, No. 1, pp. 37-42, 2003.

13. Baxt WG, "*Application of Artificial Neural Networks to Clinical Medicine*" Lancet. Vol. 346 (8983), pp. 1135-1138, 1995.

14. Narasingarao MR. Manda R., Sridhar GR., Madhu K. and Rao AA., "*A Clinical Decision Support System Using Multilayer Perceptron Neural Network to Assess Well Being in Diabetes*", Journal of the Association of Physicians of India, Vol. 57, pp. 127-133, 2009.

15. Shanthi D., Sahoo G. and Saravanan N., "*Input Feature Selection using Hybrid Neuro-Genetic Approach in the diagnosis of Stroke*", International Journal of Computer Science and Network Security. ISSN 1738-7906. Vol. 8, No. 12, pp. 99-107, 2008.

16. Ahmad F., Nor A. M., Hussain Z., Boudville R. and Osman K. M., "*Genetic Algorithm-Artificial Neural Network (GA-ANN) Hybrid Intelligence for Cancer Diagnosis*", In Proc. Second International Conference on Computational Intelligence, Communication Systems and Networks, IEEE Computer Society, pp. 78-83, 2010.

17. Blais A. and Mertz D., "*An Introduction to Neural Networks – Pattern Learning with Back Propagation Algorithm*", Gnosis Software Inc. 2001.

18. "*Neuro Intelligence using Alyuda*", http://www.alyuda.com, 2008, Available: [Last accessed May 11, 2011].

19. Chowdhury D. R., Chatterjee M. and Samanta.R.K, "*Rough Set Based Model for Neontal Disease diagnosis*", International Conf. on Mathematics and Soft Computing, ICMSCAE, 2010.

20. Chowdhury D. R., Chatterjee M. and Samanta.R.K, "*Neonatal Disease Diagnosis with Soft Computing*", in Proc. International Conf. on Computing and System, ICCS, University of Burdwan, India, pp. 27-34, 2010.

21. Chowdhury D. R., Samanta R.K and Chatterjee M., "*Design and Development of an Expert System Model in Differential Diagnosis for Neonatal Disease*", International Journal of Computing, Vol, 1, No. 3, pp. 343-350, 2011.

22. "*RSES 2.2 User's Guide*", Warsaw University, USENET: http://logic.mimuw.edu.pl/~rses, Available: [Last accessed January 19th, 2010].

23. "*The ROSETTA homepage*". Internet: http://www.idi.ntnu.no/~aleks/rosetta/, Available: [Last accessed January 25th, 2010].

24. Patterson Dan W., "*Introduction to Intelligence and Expert System*", PHI, pp. 344-345, India, 2011.

CHAPTER 12

Neuro-Genetic Fusion Approach for Neonatal Disease Diagnosis: A Decision Support System†

12.1. Introduction

In the field of Artificial Intelligence (AI) neural network based applications are one of the most useful and vastly used data mining tools for making decision and prediction of disease in medicine. Feature selection of attributes and preprocessing are the vital parts of model selection in Artificial Neural Network (ANN). It may be prolonged also due to the selection of input features in Multi Layer Perceptron (MLP) Model. Input attributes also determines the number of layers to be used and number of neurons requires for each layer. Thus selecting the best features among the various inputs is of immense significance in this field, without avoiding the accuracy. This chapter represents the fusion approach of integrating Batch Back Propagation Neural Network (B-BPNN) and Genetic Algorithm (GA) towards developing the decision support system for neonatal disease diagnose. Selection for attribute subsets has done by the Attribute Subset Evaluator (ASE) using genetic algorithm. Selected subset then preprocessed and analyzed for choosing best network architecture. Batch back propagation training algorithm used in this study to train the network using various training algorithm parameters. Experimental results reflect better classification accuracy after comparing different studies.

In medical decision making, a vast amount of data is collected and used for processing. In case of huge data classification, human or domain expert decision making some time gives the poor results. Even the information provided by the patient or the person related with the patients may contain wrong, not-related or redundant information relating to sign and symptoms. It is also seen that, suffering from a particular disease may contains several sign and symptoms and also with same sign and symptoms, patient may different disease. The domain expert may not diagnose the disease correctly, particularly, this happens with neonates in Tarai region of west Bengal, India [1]. Thus it is absolutely necessary to identify and select the correct features of a disease. The mortality problem in rural areas [2] can prevail over through fast and accurate disease diagnosis and management of the newborn. A comprehensive decision

† This chapter is based on the publication made by the author entitled “Neuro-Genetic Fusion Approach towards Developing a Decision Support System for Neonatal Disease Diagnosis”, on NaCCS–2012, National Conference on Computing and Systems, pp. 242-247, March, 2012.

support system for diagnosing the neonatal prevalent disease might helpful for the place, where there is a scarcity of domain experts. Development of an Experts System for neonatal disease diagnosis is complicated process and requires high level of expertise. Any attempt towards designing and developing of an expert system dealing with different diagnosis of neonatal disease has to overcome various difficulties [3].

Neural networks has been found one of the most efficient data mining analytical tools which can be used for making prediction with an optimal accuracy. Back propagation network uses gradient based approach. It trains the network slowly. Due to this problem of slow leaning techniques, we may use optimization methods like GA, PSO, Ant Colony Optimization (ACO) techniques to find the network weights. GA basically is used for searching features from huge search space. It is also used for back propagation neural network determining the parameters like hidden nodes and layers with back propagation neural network. GA also selects relevant features subsets and connection weights.

The main aim of this work is to establish the correct input features, sign and symptoms which were defined by categories of the network. If the information is huge, then it may degrade the classification performance and thereby lose the prediction accuracy. The main advantages of selection the proper attributes are cost reduction for data accusation, increase efficiency in classifier based systems [4]. The study discussed in chapter 9, reports that the results where some coupling with data mining and rough set theory is proposed for differential neonatal disease diagnosis. The experimental set up was planned to extract the relative importance of neonatal disease attributes; dimensionality reduction; finding core attributes, classification prediction using data mining approaches [5].

In this chapter, a neuro-genetic fusion approach has been proposed to find and select the best input features for diagnosis of neonatal disease. The study has been arranged in the following way. Section 12.2 elaborates the related field study. Back Propagation including Batch Back Propagation Neural Network and Genetic algorithm has been discussed in Section 12.3 and 12.4 respectively. The Neuro-Genetic Fusion approach has been explained in section 12.5. Experimental Result analysis is highlighting the in the section 12.6 and followed by conclusion in section 12.7.

12.2. Related Studies

During the analysis of several studies many factors related to this chapter has been found useful and comparable with our study. A study reveals that genetic algorithms are optimization events which were added to intrusion detection system to make them more efficient. Self Organizing Migrating Genetic Algorithm (SOMGA) was integrated into intrusion detection system to obtain a more efficient intrusion detection system [6]. Neuro-Genetic approach has also been proposed for feature selection in diagnosis of stroke disease. Here MLP is used whose inputs are automatically selected using GA and the experimental results shows better classification accuracy with fewer inputs as features [7].

In a study, GA and back propagation neural network were combined, as GAs have better chance of getting to the global optimum and ANN are faster which improves the accuracy of prediction [8]. In another study, a hybrid algorithm of the BPN-ANN is used to optimize the initial weights, and make fast convergence of the BP-ANN. The algorithm has been applied on data fusion of Landsat ETM+ and ERS-2 SAR for classification of urban terrain surface [9]. A study of artificial neural networks (ANN) coupled with genetic algorithms has been applied to evolve combinations of clinical variables optimized for predicting urinary tract infection [10].

12.3. Back Propagation

The Back Propagation (BP) algorithm is widely used for training the network. For training a network several steps needs to be taken efficiently and stepwise weights has to be calculated simultaneously [11]. There are few people who consider the Back

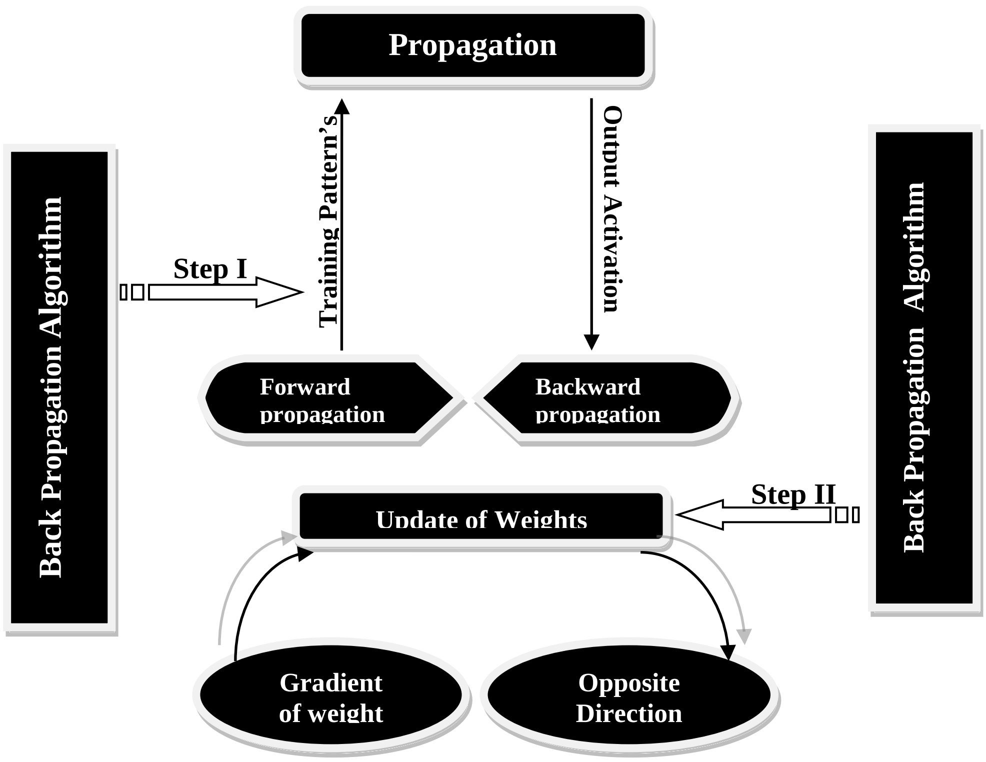

Figure 12.1. Back-Propagation.

Propagation network as the prototypical neural net. Basically, Back Propagation is the training or learning algorithm rather than the network itself. Until now the network used is generally simple types, shown in figure 11.7. in chapter 11 and in the examples. These networks are called Multi-Layer Perceptrons *(MLPs)* or occasionally Feed-

Forward Networks. Back Propagation comes under a supervised learning method which is a simplification of delta rule. This network does not contain any feedback loop, which means there should not be any loop in the connection. Thus it is most popular and effective feed forward network. Actually it is the backward propagation of the errors. For the purpose of training set BP requires the complete dataset as input data set for getting output. It also requires the activation function which basically used by the nodes or the artificial neurons. A Back Propagation in the artificial neural network learns by example. It has to provide algorithm examples of what we want the network to do and it changes the network's weights so that, when training is finished, it will give the required output for a particular input.

The Back Propagation learning algorithm is basically having two steps for getting the desired output. The steps are as follows [12]:

- **Step I. Propagation**

  1. For generating propagation's output activations, Forward Propagation of a training pattern's input through the neural network is used.
  2. For generating deltas of all output and hidden neurons, Backward Propagation of the propagation's output activations through the neural network is used.

- **Step II. Update of Weights**

  - For updating each and every weight-synapse follow the following rules:

    1. For getting the gradient of the weight it first multiply its output delta and input activation.
    2. Then bring the weight in the opposite direction of the gradient by subtracting a ratio of it from the weight.

- The algorithm repeats Step I and Step II till the performance of the network is satisfactory and get the desired output.

Back Propagation networks are ideal for in this study for simple pattern recognition of the neonatal disease diagnosis. To train the network we need to give it examples of what we want. This output is called ***Target*** output.

Figure 12.2 shows the Input patterns and of BP Training Set and the Target with probable output.

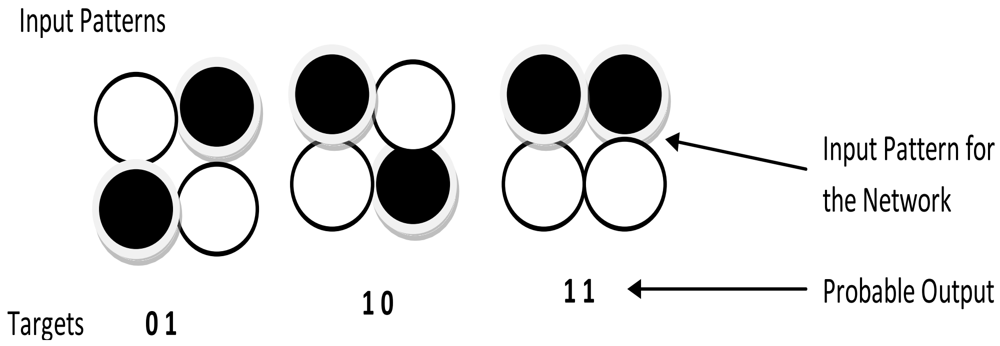

Figure 12.2 Back Propagation Training Set

Analysing the above training set in figure 12.2, if we put in the first pattern to the network, we would like the output to be 0 1 as shown in figure 12.3. The input and its corresponding target are called a Training Pair**.** Here, white circle is represented by 0 and black circle is represented by 1, here in this study they are the signs and symptoms of the disease.

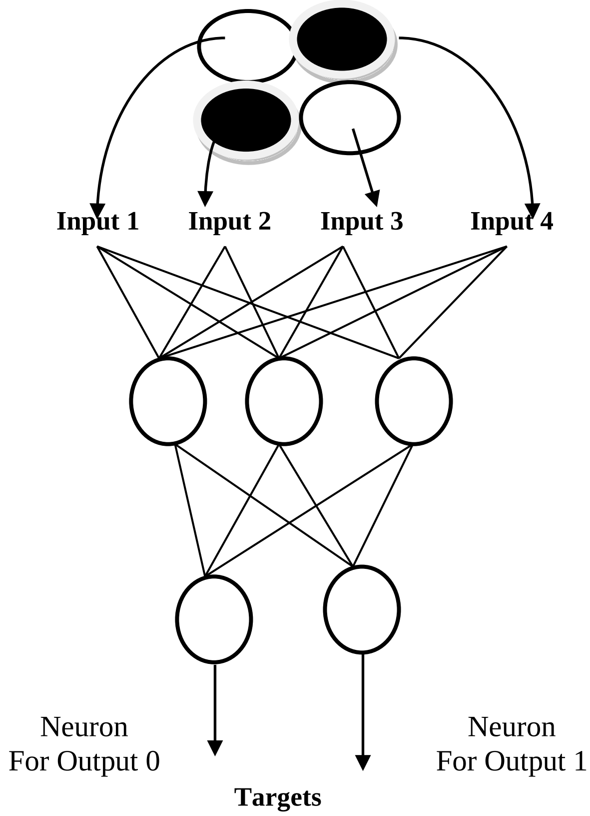

Figure 12.3. Applications of Training Pair to a Network.

The network will provide the desired output for any of the input patterns when the network is trained successfully. Firstly, the network is initialised by setting up all its weights to be small random numbers between –1 and +1. Then applying forward pass using input pattern and the output calculated. Since all the weights are random, the calculation gives an output which is completely different to what we want. After that, calculation of the Error of each neuron has done by applying formula:

Target Output – Actual Output

This error is then used mathematically to change the weights in such a way that the error will get smaller. In other words, the Output of each neuron will get closer to its Target that is reverse pass. The process is repeated again and again until the error is minimal in the running network.

12.3.1. End of Training Process

Training process may end or conclude only when the network can recognise all the letters successfully, but in practice it is usual to let the error fall to a lower value first. We even may stop it once the network can recognise all the patterns of neonatal diseases successfully, but in practice it is usual to let the error fall to a lower value first. This ensures that the patterns are all being well recognised. One can evaluate the total error of the network by adding up all the errors for each individual neuron and then for each pattern in turn to give us a total error as shown in figure 12.4 [13].

The network keeps training all the patterns repeatedly until the total error falls to some pre-determined low target value and then it stops. Note that when calculating the final error used to stop the network which is the sum of all the individual neuron errors for each pattern we need to make all errors positive so that they add up and do not subtract an error of -0.5 is just as bad as an error of +0.5. Once the network has been trained, it should be able to recognise not just the perfect patterns, but also corrupted or noisy versions. In fact if we deliberately add some noisy versions of the patterns into the training set as we train the network, we can improve the network's performance in this respect. The training may also benefit from applying the patterns in a random order to the network.

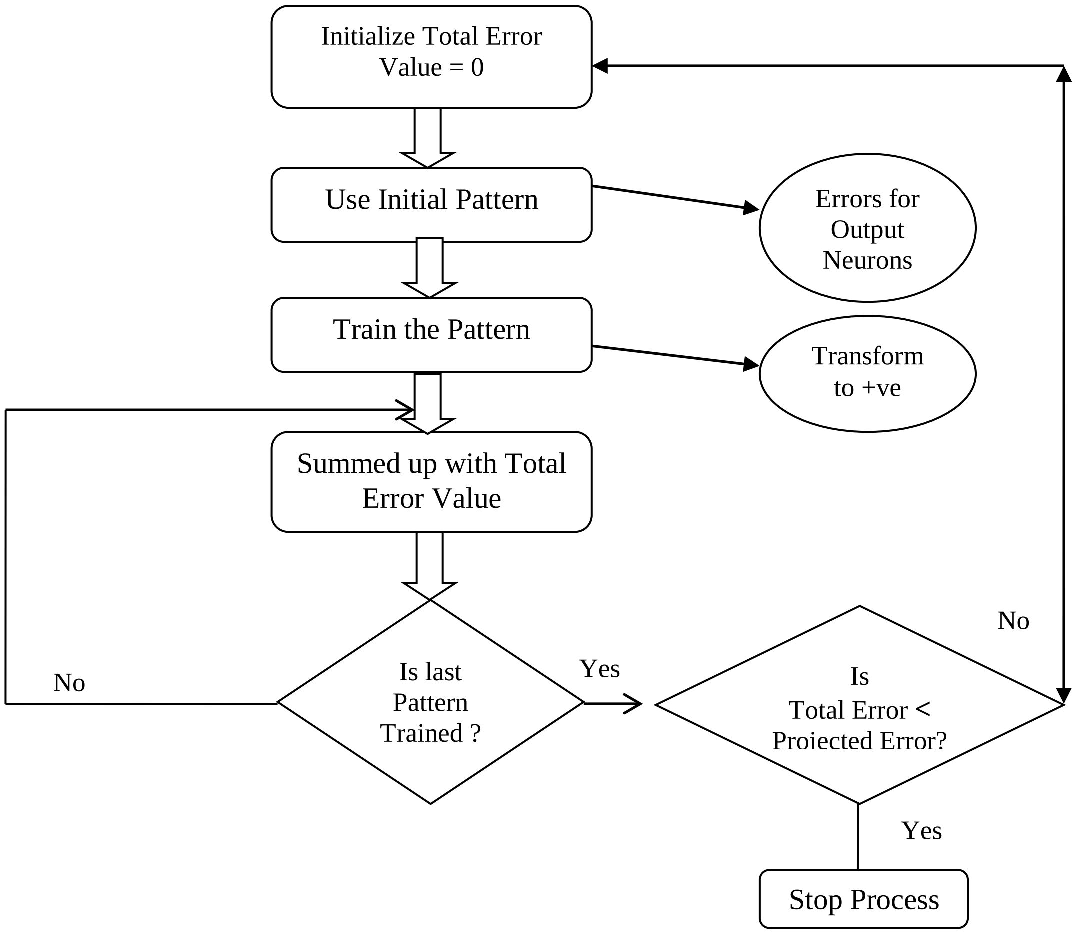

Figure 12.4. Total Error in Network.

12.3.2. Size of the Network

There is certainly need of finding the actual size of network for a specific problem. Generally common neural networks are consisting input layer, hidden layer and output layer. Depending upon the types of pattern or the inputs, input layer size likely to be sets up. In our case, the sign and symptoms of the disease are the inputs. The input layer size is set by the type of pattern or input we want the network to process for the training. In the above example, figure 12.3, the network must have 4 inputs because there are four circles in the pattern, each of them is a sign/symptom of disease. Likewise, the size of the output layer is set by the number of patterns we want to recognise and how one want to code these outputs, here for this problem two output neurons are needed. There is only one hidden layer neuron is required. No hard and fast rules are there for this and the network typically works well over a range of this variable. The number of hidden layer neurons needs to be experimented with for the best results at the time of training.

12.3.3. Better Way of Stopping Network Training

Use of Validation Set is probably the better way of working out when to stop network training. This stops the network overtraining which becomes optimal accurate and can lessen its performance. It does this by having a second set of patterns which are noisy versions of the training set. Each time after the network has trained; this Validation Set is used to calculate an error. When the error becomes low the network stops. When the network has fully trained, the Validation Set error reaches a minimum. When the network is overtraining the validation set error starts rising. If the network over trains, it won't be able to handle noisy data so well. Figure 12.5. shows the use of a validation set on the network training procedure.

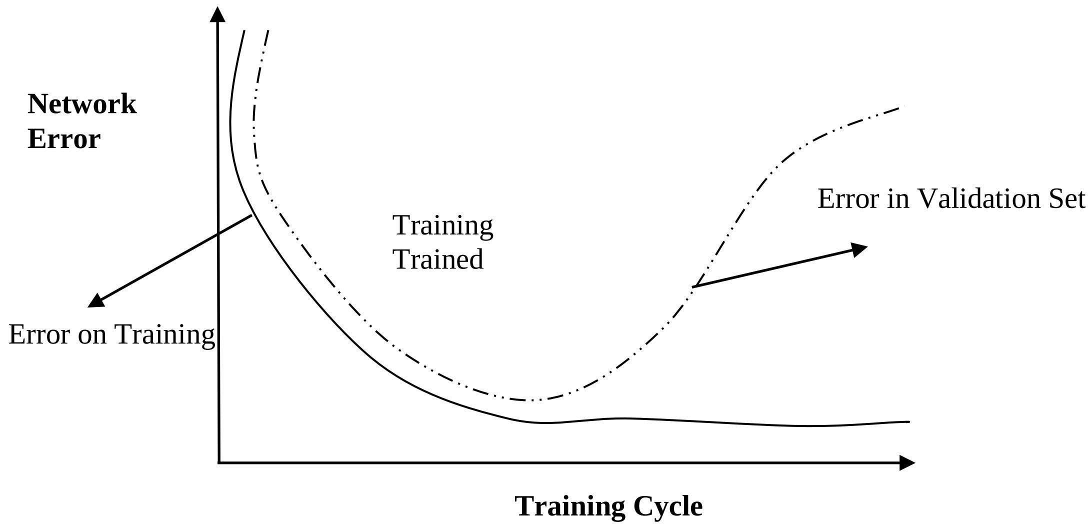

Figure 12.5. Using Validation Sets.

12.3.4. Backpropagation Problem Area

12.3.4.1. Local Minimum

There are some problems associated with Backpropagtion. One of the most frequent problems is the "Local Minimum". Due to frequent changes in the algorithm, it changes the weights in such a way as to cause the error to fall, thus the local minimum problem occurs. But the error might briefly have to rise as part of a more general fall, as shown in figure 12.6, the algorithms may not be able to go upward and get jammed and therefore error will not decrease further in the training process.

12.3.4.2 Solution of the Problem

Many solutions are there for this local minimum problem. They are as follows:

i) ***Resetting the Weight***:

One of the simple solutions is resetting the weights to different random numbers and tries training repeatedly.

ii) ***Momentum***

The other solution is to adding up “momentum” to the weight change at the time of training. This means that the weight change this iteration depends not just on the current error, but also on previous changes.

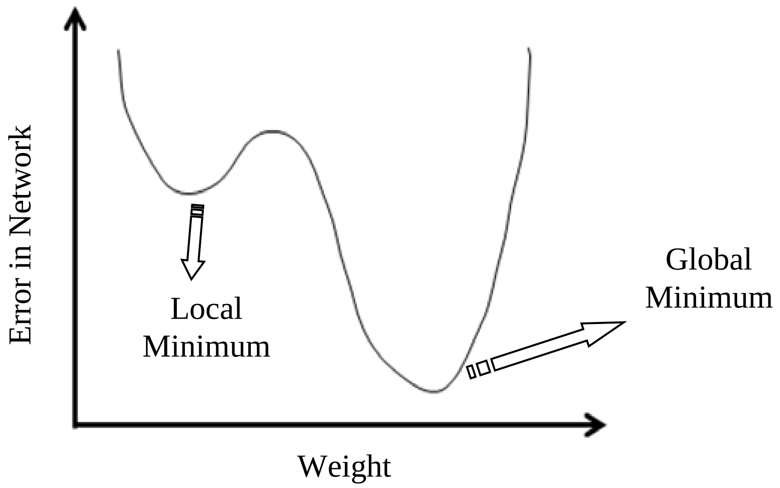

Figure 12.6. Local and Global Minimum.

iii) Re-initialization of Weights

One can overcome the problem by re-initializing the weights to different starting values. These tend to note themselves as the network gets larger.

12.3.4.3. Batch Back Propagation Neural Network

Feed-forward networks are often trained using a back propagation-learning scheme. Back propagation learning works by making modifications in weight values starting at the output layer then moving backward through the hidden layers of the network. In Neural network noisy data can be of high tolerance and can be applied for classifying patterns without being trained on the network [14, 15].

Training algorithms of a multi-layer perceptron can be divided into 2 types based on the condition.

- Incremental Algorithm and
- Batch Algorithm.

- The Incremental algorithms predict the correction of network weights after processing of each record, where as in Batch algorithms predict the correction of network weights after presenting of the entire training set.

- Batch Back propagation algorithm is used for networks of any size. Back propagation algorithm is the most popular algorithm for training of multi-layer perceptrons and is often used by researchers and practitioners. Back propagation can be efficient for large datasets if you properly select the learning rate and momentum. In Batch backpropagation all weight changes are summed over a full presentation of all training patterns which is termed as one epoch. Only then, an update with the accumulated weight changes is performed. This update behavior is appropriate for training pattern parallel implementations where communication costs are important [16].

  In batch mode the weights and biases of the network are updated only after the entire training set has been applied to the network. The gradients calculated at each training example are added together to determine the change in the weights and biases [17].

Considering the scenario of this study, we have used a Batch algorithm for the training of the proposed network.

12.4. Genetic Algorithm [18]

If there is a talk about evolutionary computing then Genetic Algorithms (GAs) are the important part of it. Now this becomes a rapidly growing area of artificial intelligence. This algorithm is basically inspired by Darwin's theory about evolution. Genetic Algorithms (GAs) are a way of solving problems by mimicking the same processes Mother Nature uses. They use the same combination of selection, recombination and mutation to evolve a solution to a problem. In the computational arena also this algorithm works efficiently now a days.

In brief, we may say a genetic algorithm is a programming technique that mimics biological evolution as a problem-solving strategy. Given a specific problem to solve, the input to the GA is a set of potential solutions to that problem, encoded in some way, and a metric called a “fitness function” that allows each candidate to be quantitatively evaluated. These candidates may be solutions already known to work, with the aim of the GA being to improve them, but more often they are generated at random. According to the fitness function GA evaluates each of the candidates. In a group of randomly generated candidates, these will be deleted. A few of them may show activity, even if only weak and imperfect activity, toward solving the problem. Then the candidates which show potential are kept and allowed to reproduce. Multiple copies are made of them, and when the copies are not perfect then frequent changes are introduced during the copying process in the algorithm. Again these winning individuals are selected and copied over into the next generation with random changes, and the process repeats till all the individuals reached. The expectation is that the average fitness of the population will increase each round, and so by repeating this process for hundreds or thousands of rounds, thus we may expect better solutions to the problem [18].

12.4.1. Term Associated with Genetic Algorithm

❖ ***Chromosome / genome***

A Chromosome is a representation of an individual solution for a specific problem. Often chromosome is called as genome also. One may have to redefine the chromosome representation for each particular problem, along with its fitness, mutate, reproduce, and seed methods. All living organisms consist of cells. In each cell there is the same set of chromosomes. Chromosomes are strings of DNA and serves as a model for the whole organism. Each cell in our body contains the same set of chromosomes, strings of DNA that function as a blueprint for making one of us. Figure. 12.7 show the chromosome structure.

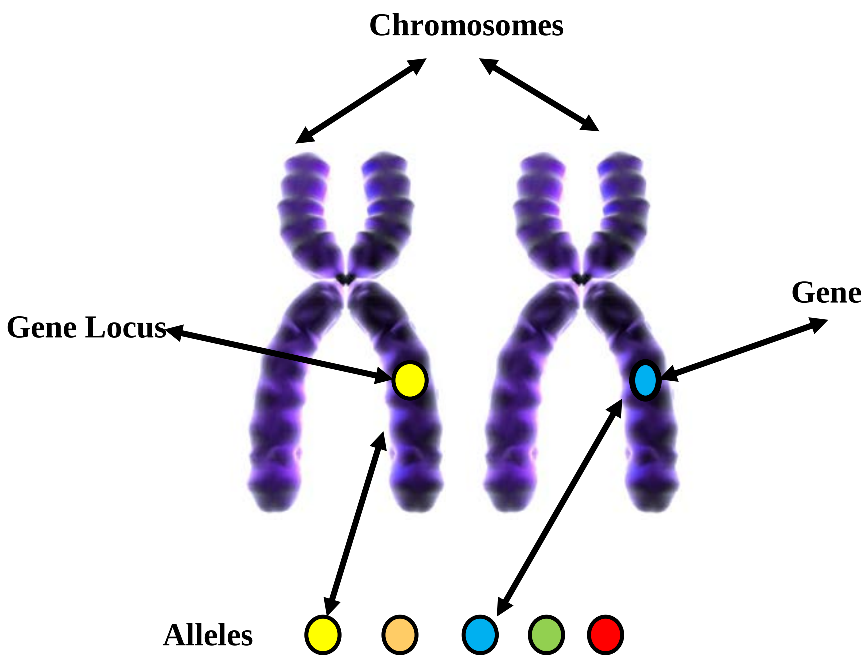

Figure 12.7. Chromosome Structure.

❖ ***Genes***

In the blocks of DNA, chromosome consists of genes. Each gene encodes a particular protein. Basically we can say that each gene encodes characteristics.

❖ ***Locus***

Each gene has its own position in the chromosome. This position is called locus. The location of a gene on a chromosome or on a linkage map is the locus position. A variant of the DNA sequence at a given locus is called an allele. Loci is the plural form of the allele. Genetic map is the ordered list of loci known for a particular genome.

For a particular biological trait, mapping of these Gene is the procession of determining the locus. Below Figure. 12.7. shows the Gene Locus and Alleles.

- ***Genotype***

Complete set of genetic material (all chromosomes) is called genome. Particular set of genes in genome is called genotype. When applied to a single gene locus, the term genotype is used to distinguish one allele, or combination of alleles, from another. The genotype is with later development after birth base for the organism's phenotype, its physical and mental characteristics, such as eye color, intelligence etc. The genotype of an organism determines its phenotype.

12.4.2. Genetic Algorithm Operator

For maintaining the genetic diversity and combining best solutions into others we may use a genetic operator. These operators are:

- Selection
- Crossover
- Mutation

12.4.2.1. Selection

Selection is a genetic operator that chooses a chromosome from the current generation's population for inclusion in the next generation's population. Before making it into the next generation's population, selected chromosomes may undergo crossover or mutation depending upon the probability of crossover and mutation in which case the offspring chromosomes are actually the ones that make it into the next generation's population.

This is the one of the stage of genetic algorithm where individual genomes are chosen from a population. Then that is used for further reproduction and cross over process.

Implementation steps of the selection procedure [19]:

Step I. The fitness function is evaluated for each individual, providing fitness values, which are then normalized. Normalization means dividing the fitness value of each individual by the sum of all fitness values, so that the sum of all resulting fitness values equals 1.

Step II. The population is sorted by descending fitness values.

Step III. Accumulated normalized fitness values are computed the accumulated fitness value of an individual is the sum of its own fitness value plus the fitness values of all the previous individuals). The accumulated fitness of the last individual should be 1.

Step IV. A random number *R* between 0 and 1 is chosen.

Step V. The selected individual is the first one whose accumulated normalized value is greater than *R*.

- **Selection Procedures**

  - ***Roulette-Wheel Selection:***

    One of the selection procedures is Roulette-Wheel Selection. Here procedure is repeatedly process till there are enough selected individuals for the fitness proportionate.

  - ***Stochastic Universal Sampling:***

    In another approach namely stochastic universal sampling, a single pointer spun multiple times, there are multiple, equally spaced pointers on a wheel that is whirls once.

  - ***Tournament Selection:***

    Tournament selection is the process by which it repeatedly selects the best individual of a randomly chosen subset is tournament selection.

  - ***Truncation selection:***

    Truncation selection is the process of taking the best half, third or another proportion of the individuals are truncation selection.

  - ***Elitism* or *Elitist Selection:***

    Retaining the best individuals in a generation unchanged in the next generation, is called elitism or elitist selection. It is a successful variant of the general process of constructing a new population.

12.4.2.2. Crossover

After we have decided what encoding we will use, we can make a step to crossover. Crossover selects genes from parent chromosomes and creates a new offspring. The simplest way how to do this is to choose randomly some crossover point and everything before this point copy from a first parent and then everything after a crossover point copy from the second parent.

In genetic algorithms, crossover is a genetic operator used to vary the programming of a chromosome or chromosomes from one generation to the next. It is analogous to reproduction and biological crossover, upon which genetic algorithms are based. Cross over is a process of taking more than one parent solutions and producing a child solution from them. There are methods for selection of the chromosomes.

12.4.2.3. Mutation

After a crossover is performed, mutation takes place. This is to prevent falling all solutions in population into a local optimum of solved problem. Mutation changes randomly the new offspring. The mutation depends on the encoding as well as the crossover. For example when we are encoding permutations, mutation could be exchanging two genes. It is a genetic operator used to maintain genetic diversity from one generation of a population of algorithm chromosomes to the next. It is similar to biological mutation. Mutation alters one or more gene values in a chromosome from its initial state. In mutation, the solution may change entirely from the previous solution. Hence genetic algorithm can come to better solution by using mutation. Mutation occurs during evolution according to a user-definable mutation probability. This probability should be set low. If it is set too high, the search will turn into a primitive random search.

To introduce the diversity the mutation process works. This process works to avoid local minima by preventing the population of chromosomes with same characteristics. Due to this problem the process of evolution slows down or even stopped. The example of a mutation operator involves a probability that an arbitrary bit in a genetic sequence will be changed from its original state. One of the very useful process is implementing the mutation operator involves generating a random variable for each bit in a sequence. There are several types of mutation process. They are single point, inversion and floating point mutation.

This operator alters one or more gene values in a chromosome from its initial state. This can result in entirely new gene values being added to the groups of gene. With these new gene values, the genetic algorithm may be able to arrive at better solution than was previously possible. Mutation is an important part of the genetic search as help helps to prevent the population from stagnating at any local optima. Mutation occurs during evolution according to a user-definable mutation probability. Using of the probability has to low initially, i.e. 0.01, because in the high values primitive random search turns into effect.

12.4.3. Summary of GAs

For finding the approximate solution to optimization problems Genetic Algorithm is used. It is a global search technique and a particular class of evolutionary algorithms. From biological sciences, evolutionary processes have been borrowed and translated to efficient search and design strategies. Genetic Algorithms use these strategies to find an optimum solution for any multi-dimensional problem [20]. As we mentioned already that, GAs is basically inspired by Darwin's theory of evolution- Survival of the Fittest. GA algorithms are the ways of solving problems by Selection, Crossover, Mutation and Accepting to progress of a problem solution.

Each element in a genetic space is termed as chromosome. It is collection of genes, the basic building block of the chromosome. Locus is the position of particular gene in the chromosome. In every generation the total population is evaluated using fitness function and there by selecting procedure. In this procedure high fitness chromosomes are used to eliminate low fitness chromosomes using Boltzmann selection, Roulette-wheel selection etc. methods. After selection crossover and mutation process starts their job by which two-selected chromosome with high fitness values exchange part of the genes to generate new pair of chromosomes. Crossover may of different techniques, namely, uniform crossover, average crossover, multipoint crossover, one point crossover and two-point crossover. The mutation processes like, heuristic mutation, random gene mutation and creep mutation randomly changes the value of a gene for checking and preventing early convergence to local optima. This whole process will continue till the termination criteria are not satisfied for new population generation [21].

12.5. The Neuro-Genetic Fusion Approach

The fusion process of Artificial Neural Network with Genetic Algorithm has been established by using GA to optimize the parameters for an ANN with specific topology architecture. Back propagation neural network learning done by frequent changing of the weights at the output layer. When the Gradient Descent Algorithm is used to modify the weights of the artificial neural network, it searches for the number of hidden nodes and connection weights in a specific network topology.

There are many search algorithms and several soft computing techniques [22] to perform the domain attribute selection like greedy local search, GA performs a comprehensive search. In this concern GA has been applied to utilized and optimize the connection weight of ANN. This for the improvement of the overall performance of Neural Network for predicting neonatal disease diagnoses.

In this fusion approach GA has been applied in such a way that it can reduce the size of the training set and select the most relevant features from the data set and using that design a proper network structure. By using trial and error method, we select the hidden layer of the network. Over here the constraints is that there should have at least one layer between input and output layer of the network. We have used a dataset of having 95 instances and 15 categorical attribute as different sign and symptoms.

GA search classifier used to select the mostly required attributes. The size of the neural network is particularly depends on the number of input variables. To get best performance and avoid over-fitting problem in the network, the input variables should be less. Thus we use pre-processing on the entire input data. This pre-processing actually eliminates the irrelevant and redundant data. This is requires for optimization of the network structure.

The procedure for attribute selection follows the below mention steps:

Individual Representation consists of a String of N bits, and N is the number of the original attributes. The i th bit can take not more than two values, i.e. 1 or 0. It represents whether the attributing is selected (1) or not (0).

Hence, I = 1 to …. N. For an example taken from our study, in the 15-attribute data set, the individual " 1 0 0 0 1 1 0 0 1 0 0 0 1 1 0" represents a solution where the data set selects the attributes no 1st, 5th, 6th , 9th ,13th and 14th are selected. Here we applied crossover and mutation operations.

The dataset are partitioned into 3 parts, namely Training, Validation and Test. The above Figure 12.8 shows the proposed model of Neuro-Genetic fusion.

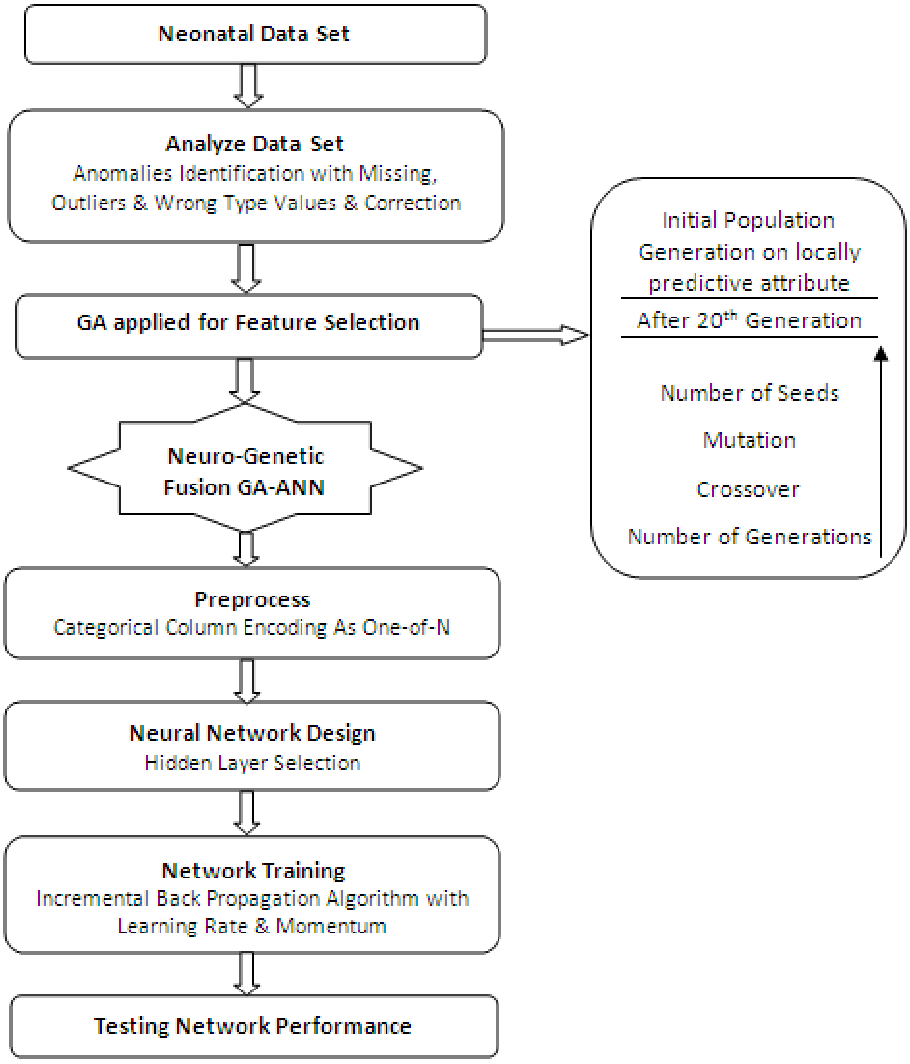

Figure 12.8. Neuro-Genetic fusion Model.

12.5.1. Input Feature Selection

Input feature selection methods are used to identify input columns that are not useful and do not contribute significantly to the performance of neural network. We can remove insignificant inputs and improve the generalization performance of a neural network, in spite of losing some input information.

Taking the training data set, we go for feature selection by using GA algorithm method. Depending upon generation, cross over and mutation rate, features has been selected. The method starts with a random population of input configurations. Input configuration determines what inputs are ignored during performance test. At each following step (called generation) it uses a process analogous to natural selection to select superior configurations and use them to generate a new population. Each step successively produces better input configuration. At the last step the best configuration is selected.

12.5.2. Analyzed Selected Features

Analyzing the dataset is required to define column parameters and detect data anomalies. Data analysis information needed for correct data preprocessing. After data analysis we can see which values have been identified as missing, wrong type values or outliers and which columns were rejected as unconvertible for use with the neural network. Outliers are column values that are far away from the majority of the column data. Outliers can be just extreme cases, measurement errors or other anomalies. Outliers prevent correct neural network training and significantly degrade neural network performance. The study recommends removing outliers from data before using neural networks and model designing.

12.5.3. Data Preprocessing and Post-Processing

Neural networks performance is better on numeric data. But we are having a categorical data to be processed by Neural Network. Therefore, before that data to the network additional transformation of such data is required as an input and after their reception at the output.

Data preprocessing means modification of the data before it is fed to a neural network. Preprocessing transforms the data to make it suitable for neural network.

Post-processing means modifying the neural network output to make it understandable by user or suitable for real world environment. Textual values can be used only if they represent a final set of values. Columns with such values are called categorical columns.

Numeric values should be scaled before feeding them to the network input, because artificial neurons have a limited range of operating values. Therefore, input information should be reduced to this range and output information should be received from network output by a reverse transformation.
In this context, Categorical columns automatically encoded during data preprocessing using One-of-N, encodings method. The One-of-N encoding means that a column with N distinct categories (values) is encoded into a set of N numeric columns, with one column for each category. For example, for the Birth Term Status column with values "Term", "Pre Term" and "Post Term", has been represented, Term as {1,0,0}, Pre Term as {0,1,0}, and Post Term as {0,0,1}.

12.5.4. Designing Network

After preprocessing, the data has been feed to neural network for training. But for designing a network, we need to specify the network architecture. This architecture includes searching number of hidden layer and units of each layers and network properties which includes error and activation functions. Though Network properties are defined automatically, for our experiment, we have changed them manually, for improving network performance. Here feed-forward fully-connected neural network MLP algorithm has been applied for searching best network architecture and also train the network. The most significant characteristics of a multilayer perceptron network is to decide the number of neurons in the hidden layer. Deciding the same we have proposed and discussed in our earlier study [23].

In this context, our objectives of the training process was to find the set of weight values which will cause the output from the neural network to match the actual target values as closely as possible. We have faced several issues concerned in designing and training a multilayer perceptron network model. Some of the issues are:

i. To select the number of hidden layers to use in the network.
ii. To decide the number of neurons to be used in each hidden layer.
iii. Converging to an optimal solution in a reasonable period of time.
iv. Finding a globally optimal solution that avoids local minima.
v. Validating the neural network to test for overfitting.

12.5.5. Hidden Layers Selection

In this fusion study one hidden layer is sufficient for the network. Two hidden layers are required for modeling data with discontinuities such as a saw tooth wave pattern. As we found that using two hidden layers rarely improves the model, and it may introduce a greater risk of converging to a local minima. So, three layer models with one hidden layer are recommended for our study.

12.5.6. Deciding How Many Neurons to be Used in the Hidden Layers

The most significant characteristics of a multilayer perceptron network is to decide the number of neurons in the hidden layer. The network may be unable to model complex data, and the resulting fit will be poor, If an inadequate number of neurons are used in the network. Similarly, If too many neurons are used, the training time may become excessively long, and, worse, the network may over fit the data. When overfitting occurs, the network will begin to model random noise in the data. The result is that the model fits the training data extremely well, but it generalizes poorly to new, unseen data. Validation must be used to test for this.

12.5.7. Training the Network

Train the selected architecture using Batch Back Propagation Training algorithm with the selected architecture. This has been studied with the Test data for finding better performance. We have used Batch Back propagation for networks training as the size might be of different. Back propagation algorithm is the most popular algorithm for training of multi-layer perceptrons and is often used by researchers and practitioners.

12.5.8. Performance Testing

After the network training process has completed, performance testing of the selected network has been done on training set. Here Confusion Matrix displays a square matrix whose rows and columns represent the target column categories or sub-ranges for the real world target and network outputs, respectively. The matrix rows contain the target column categories/sub-ranges and the matrix columns contain network output categories/sub-ranges. The value in the (i,j) position of the matrix is the number of records for which the target column value is in the i-th category/sub-range and whose network output is within the j-th category/sub-range. We got the mean CCR % values in this context.

12.5.9. Query

Lastly, we have use query procedure for the trained network using records from your input dataset using Query Dataset. One can also query the trained network using records from input dataset using query dataset. In the query dataset we use filter for a particular set. The available sets for the filter are: training, test, validation, ignored set and missing targets set. There are two more cases when the querying from dataset feature is very helpful. First of all, input data file contains records used for network training together with records intended for querying the trained network.

The records for querying will have missing values in the target column. In the end of training our goal will be to query network with these records to fill these missing targets with network outputs. Secondly, in some cases we may want to leave some records to query the network. Then we can assign them to Ignored set during analysis and in the query from dataset window select Ignored set to query network. We have got a good response by the output using query builder.

12.6. Experiment and Results Analysis

12.6.1 Data Input for the Study

A total number of 94 patient's data have been collected for these studies that have basic symptoms of neonatal disease. For making as much as possible error free data, they have been standardized. All the cases are analyzed after careful scrutiny with the help of the pediatric expert.

Table 12.1. below shows the various input parameters of the neuro-genetic fusion for prediction of neonatal disease diagnosis.

Table 12.1. Input Parameters for Neuro-Genetic Fusion.

Sl.No.	Column Type	Parameters
1	Categorical	Birth_Term_Status
2	Categorical	Birth_Weight_Status
3	Categorical	Age_in_Hours>72
4	Categorical	Lethargy
5	Categorical	Refusual_to_Suck
6	Categorical	Poor_Cry
7	Categorical	Poor_Weight_gain
8	Categorical	Hypothalmia
9	Categorical	Sclerema
10	Categorical	Excessive_Jaundice
11	Categorical	Bleeding
12	Categorical	GI_Disorder
13	Categorical	Seizure
14	Categorical	Sluggish_Neonatal_Reflex

12.6.2. Attribute Selection Based on GA

Initially we consider all the 14 symptoms for neuro-genetic fusion approach. Genetic Algorithm optimizes 14 inputs and the genotype is represented by the sequence of symptoms. GA algorithm is an adaptive optimization algorithm for simulating the inheritance and evolution of biology in the nature [24]. Genes are represented by real number encoding method. The chromosomes are generated randomly and the original population is a set of N chromosome. Minimum optimization method has been applied for computing the fitness of each chromosome. Fitness is given by Fitness (Ci) = 1/ E for each chromosome of the population, where E is the error computed as root mean square error at the output layer, where summation is performed overall output nodes pj and tj is the desired or target value of output o_j for a given input vector. The equation for Error is:

$$E = ½ \sum_{p} \sum_{j} (t_{pj} - O_{pj})^2 \qquad (1)$$

After computing the fitness for all chromosomes, the worst fit chromosomes are replaced by best-fit chromosomes. Crossover steps are experimented using single point, two-point and multi point crossover term. In this study, the number of individual and initial population is 14. Maximum numbers of generation are fixed at 20. The probability of crossover is 0.6 and mutation is 0.035. This has been chosen in trial and error basis till the best performance achieved. Table 12.2 shows the GA parameters for the feature or attribute selection.

Table 12.2. GA Parameters Used for Feature or Attribute Selection.

Attribute Selection on all Input Data	
Search Method	Genetic search
Start set	No attributes
Population size	14
Number of generations	20
Probability of crossover	0.7
Probability of mutation	0.035
Report frequency	20
Random number seed	1

Finally mutation is applied as the last step to generate the new population. The new population is given as input to the classifier to compute the fitness of each chromosome, followed by process of selection, reproduction, crossover and mutations to generate the next population. This process is repeated till more or less all the chromosomes converge to the same fitness value. The categories of the initial population are showing on the Table 12.3.

Table 12.3. Initial Population Categories.

Merit	**Scaled**	**Subset**
0.09297	0.00646	12
0.12137	0.04982	7
0.34722	0.39471	3 5 6 7 8 9 12 14
0.31215	0.34115	7 9 13 14
0.38434	0.45138	1 2 3 5 6 8 11 12 13 14
0.08874	0	3
0.31063	0.33882	2 3 7 8 10 11 12
0.19273	0.15879	4
0.36413	0.42053	1 2 3 4 5 8 12 14
0.3017	0.32518	2 4 8 11 12
0.33616	0.37781	2 5 7 9 10 14
0.18158	0.14176	14
0.34124	0.38557	1 2 7 8 9 11 14
0.22485	0.20783	1 4

After computing the 20th generation, the subset categories using GA are reflecting on Table 12.4. The irrelevant features are identified here which reduce the misuse classification. The basic idea was to find the best combination out of all inputs features and provide little computational process and maximum accuracy. The selected attributes using Genetic Algorithm are 1,2,3,4,5,6,7,9,10,11,12,13,14. Surprisingly, it is noticed that attribute 8, that is Hypothalamia is missing on the selected attributes set. This reflects the less importance in the feature selection for decision making through neural network.

Table 12.4. After 20th Generation Subset Category.

Merit	Scaled	Subset
0.44418	0.49641	1 2 3 4 5 6 7 9 10 11 12 13 14
0.44418	0.49641	1 2 3 4 5 6 7 9 10 11 12 13 14
0.44418	0.49641	1 2 3 4 5 6 7 9 10 11 12 13 14
0.41041	0	1 2 3 5 7 9 10 11 13 14
0.44418	0.49641	1 2 3 4 5 6 7 9 10 11 12 13 14
0.44418	0.49641	1 2 3 4 5 6 7 9 10 11 12 13 14
0.43019	0.29085	1 2 3 4 5 6 7 8 9 11 13 14
0.44385	0.49158	1 2 4 5 6 7 9 10 11 12 13 14
0.44418	0.49641	1 2 3 4 5 6 7 9 10 11 12 13 14
0.44418	0.49641	1 2 3 4 5 6 7 9 10 11 12 13 14
0.44418	0.49641	1 2 3 4 5 6 7 9 10 11 12 13 14
0.44027	0.43897	1 2 3 4 5 6 7 8 9 10 11 13 14
0.44279	0.47602	1 2 3 4 5 6 7 8 9 10 11 12 13 14
0.44418	0.49641	1 2 3 4 5 6 7 9 10 11 12 13 14

12.6.3. Data Partition Set and Preprocessing

Data sets are data used for neural network training. The data set is subdivided into several data sets: training, validation test sets and ignore set.

- Training set is a part of our input data set used for neural network training, i.e. to adjust network weights for maximizing predictive ability and minimizing forecasting error.
- Validation set is a part of our data used to tune the network topology or network parameters other than weights. For example, it is used to define the number of hidden units of to detect the moment when predictive ability of neural network started to deteriorate.

- Test set is a part of our input data set used to test how well the neural network will forecast on new data. Test set is used after network is ready or trained, to test what errors will occur during future forecasting. This set is not used during training and thus can be considered as though it consists of the new data entered by user for forecasting.

- Ignored set is a part of our input dataset that wasn't used during training or testing. You can assign records to the Ignored set manually or automatically during data partition. Ignored set can be used to query the trained network or to remove specific records from the scope of the network.

Selected categorical input data are encoded by [-1..1] scaling range for the input columns and [0..1] scaling range for the output column at the time of analyzing. The last column is considered as the target or output one and other columns are considered as input columns. The dataset is divided in to training, validation and test sets. The Data have been analyzed using Neuro-intelligence tool [25].

Table 12.5 shows the number of records which are partitioned for training, validation and test set.

Table 12.5. Data Partition Set.

Partition set using	**Records**	**Percentage (%)**
Total	94	100
Training Set	65	68.42
Validation Set	15	15.79
Test Set	15	15.79
Ignore Set	0	0

Data encoding parameters have been depicted in Table 12.6.

Table 12.6. Data Encoding Parameters.

Data Preprocessing Information	**Range**
Rows for preprocessing	94
Columns before preprocessing	14
Columns after preprocessing	35
Input columns scaling range	[-1..1]
Output column(s) scaling range	[0..1]
Categorical Column Encoding Parameters	**Range**
Birth Term Status	One-of-3
Birth Weight Status	One-of-4
Age in Hours>72	Two-state
Lethargy	Two-state
Refusal to Suck	Two-state
Poor Cry	Two-state
Poor Weight gain	Two-state
Sclerema	Two-state
Excessive Jaundice	Two-state
Bleeding	Two-state
GI Disorder	Two-state
Seizure	Two-state
Sluggish Neonatal Reflex	Two-state
Disease Confirmation	One-of-17

For searching the best network architecture numbers of weights, fitness training error, validation error, test error and AIC (Akaike Information Criterion) have been considered. Basically, AIC is used to compare different networks with different weights (hidden units). With AIC used as fitness criteria during architecture search, simple models are preferred to complex networks if the increased cost of the additional weights (hidden units) in the complex networks do not decrease the network error. It also determines the optimal number of weights required for the neural network.

Heuristic searching method has been applied in this study. Out of 9 verified network architectures, [18-3-14] architecture had the best fitness with less number of weights. Table 12.7. shows the architecture search results.

Table 12.7. ANN Architecture Search Results.

ID	Architecture	# of Weights	Fitness	Train Error	Validation Error	Test Error	AIC
1	[18-3-14]	125	1.5	0.369231	0.6	0.333333	-51.288161
2	[18-45-14]	1637	1.363636	0.446154	0.6	0.266667	2964.258386
3	[18-28-14]	1025	1.5	0.476923	0.6	0.333333	1736.543089
4	[18-18-14]	665	1.5	0.415385	0.6	0.333333	1023.772755
5	[18-12-14]	449	1.5	0.415385	0.6	0.333333	591.772755
6	[18-8-14]	305	1.363636	0.415385	0.6	0.266667	303.772755
7	[18-15-14]	557	1.5	0.415385	0.6	0.333333	807.772755
8	[18-16-14]	593	1.5	0.430769	0.6	0.333333	878.039319
9	[18-17-14]	629	1.5	0.6	0.533333	0.333333	927.10593

In this study, Neural Network contains one input, one hidden and one output layer. Each ANN has 18 input neurons corresponding to the inputs, 3 hidden neurons and the output layer has 17 neurons. Here the input neurons represent the symptoms for the neonatal disease diagnosis. The output layer neurons represent the predicted types of disease. This network has been trained using Batch Back Propagation algorithm. Table 12.8. depicts those total available records which are relevant during the Recall or search. Septicemia has been found to be most significant with a recall value 0.946. Here harmonic mean of Precision and Recall is specified as F-Measure for each class.

- **F-measure** is a measure of a test's accuracy. It considers both the precision ***p*** and the recall ***r*** of the test to compute the score: ***p*** is the number of correct results divided by the number of all returned results and ***r*** is the number of correct results divided by the number of results that should have been returned. The F Measure can be interpreted as a weighted average of the precision and recall, where it reaches its best value at 1 and worst score at 0.

- **True Positive (TP)** rate is the proportion of examples which were classified as class *x*, among all examples which truly have class *x*, i.e. how much part of the class was captured. It is equivalent to *Recall*. In the confusion matrix, this is the diagonal element divided by the sum over the relevant row.

- **False Positive (FP)** rate is the proportion of examples which were classified as class *x*, but belong to a different class, among all examples which are not of class *x*. In the matrix, this is the column sum of class *x* minus the diagonal element, divided by the rows sums of all other classes.

- **Precision** is the proportion of the examples which truly have class *x* among all those which were classified as class *x*. In the matrix, this is the diagonal element divided by the sum over the relevant column.

Table 12.8. Detailed Accuracy by Class.

TP Rate	**FP Rate**	**Precision**	**Recall**	**F-Measure**	**ROC Area**	**Class**
0.8	0.05	0.75	0.8	0.774	0.94	HIE III
0.875	0	1	0.875	0.933	0.999	No Disease
0.625	0.023	0.714	0.625	0.667	0.986	Hypo Calcemia
0.946	**0.31**	**0.66**	**0.946**	**0.778**	**0.911**	**Septicemia**
0.5	0	1	0.5	0.667	0.997	Hypo Thalmia
0	0	0	0	0	1	Hypo Calcemia
0.5	0.022	0.5	0.5	0.5	0.977	HIE II
0.5	0	1	0.5	0.667	0.997	Jaundice
0.75	0.022	0.6	0.75	0.667	0.988	Hypothermia
0	0	0	0	0	0.995	Hemorrhage
0	0	0	0	0	0.971	Others
0	0	0	0	0	0.984	Jaundice-BA
0.333	0	1	0.333	0.5	0.993	Hypo_Glycemia
0	0	0	0	0	0.989	Sizure Disorder
0	0	0	0	0	0.995	HIE II
0.705	0.133	0.64	0.705	0.653	0.95	Weighted Average

In Table 12.9, the confusion matrix for the different categories of neonatal disease and Prediction classification using Neuro-Genetic fusion approach has been reflected.

Table 12.9. Confusion Matrix for Neonatal Disease Diagnose.

Target output:	Network output:														
	HIE III	No Disease	Hypo Calcimia	Septicimia	Hypo Thalmia	Hypo Calcimia	HIE II	Jaundice	Hypothermia	Hemorhage	Others	Jaundice-BA	Hypo Glycimia	Sizure Disorder	HIE II
HIE III	8	0	0	0	0	0	0	0	1	1	0	0	0	0	0
No Disease	0	7	0	0	0	0	0	0	0	0	0	0	0	0	0
Hypo Calcimia	0	0	2	1	0	0	0	0	0	0	0	0	0	1	0
Septicimia	1	1	0	23	0	0	0	0	0	0	0	1	0	0	0
Hypo Thalmia	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0
Hypo Calcimia	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
HIE II	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0
Jaundice	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0
Hypothermia	2	0	0	0	0	0	0	0	1	0	0	0	0	0	0
Hemorhage	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
Others	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0
Jaundice-BA	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0
Hypo Glycimia	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Sizure Disorder	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
HIE II	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Correct Classification Rate for training and validation has done to find the best network after a number of iteration. A comparative study has been done with the previous study [23] on chapter 11 base on traditional NN approach and with this Neuro-Genetic Fusion approach. Table 12.10. shows the comparative study.

Table 12.10. Prediction Accuracy.

Approach	Iteration	CCR (training)	CCR (validation)
Traditional NN	73	46.875	26.666666
	189	68.75	40
	364	**75**	**20**

Approach	Iteration	CCR (training)	CCR (validation)
Neuro-Genetic Fusion	80	75.384613	40
	252	78.46154	33.333332
	414	**78.46154**	**33.333332**

The comparative study evidently tells that the current Neuro-Genetic approach endow with better accuracy and swift convergence due to complexity of the network. In this fusion approach the maximum prediction accuracy has been found **78.46154** with reduced feature of Hypothalmia.

12.7. Conclusion

In this chapter, application of neuro-genetic approach has been experimented for classification and selection of the input features for the neural network. This study has presented two combinations of methods of neural networks and genetic algorithm for producing an improved performance on real-world classification problem, particularly this neonatal disease diagnosis. The experimental results showed that, the performance of the network can be improved by the proper selection of the input features. The reduction of input features sometime gives poor performance and accuracy sometimes. Because of, it reduces the input neurons and size of the network. The experimental results for classifying a large set of sign and symptoms show that it improves the generalization capability efficiently. The significant parameters such as number of hidden layer nodes, gain term, speed factor, crossover rate and the number of generations have a great impact on the recognition of disease diagnosis performance of the proposed system. The best possible values of the above parameters have been selected effectively to find out the best performance.

The major contribution of this study lies in showing the possibility of the soft computing techniques using fusion approach are ease of use in contrast to the general probabilistic or mathematical approaches. The neuro-genetic fusion approach sows substantial improvement in predicting the neonatal disease. However, we need to work further to extend some limitations of the current approach. Furthermore, future efforts will concentrate on refining the feature extraction to capture more information, and testing the efficacy of the soft computing techniques.

Considering the fusion techniques, in the specific case of recognizing the standards, it is seen that, if the complexity is increases in training phase, it may produced a lack of quality of results generated by ANN trained by backpropagation. In contrast the proposed system represents a stable performance. It always checks the uniformity of a system based in collaboration between Artificial Neural Network and Genetic Algorithm. Even this study advised that if the training process and cases are difficult then we may shift to these kinds of fusion model approach in the development of decision support system.

In comparison with our previous study in chapter 6, where the aim was to find the prevalent disease pattern of neonates of North Bengal districts from a data base created by me from field studies and decision tree approach has been found suitable [26]. The new neuro-genetic fusion approach shows substantial improvement in predicting the neonatal disease for the development of decision support system.

References

1. Roy Chowdhury Dilip, Samanta R.K., Chatterjee M., "*A Study of the Status of New Born in Terai Region of West Bengal*", A.M.S.E. France, No. 05 225 (2C), 2007

2. Bang Abhay T., Bang Rani A., Baitule Sanjay, Deshmukh Mahesh and Reddy M. Hanimi, "*Burden of Morbidities and the Unmet Need for Health Care in Rural Neonates - A Prospective Observational Study in Gadchiroli, India*", Indian Pediatrics, Vol. 38, pp. 952-965, 2001.

3. Chowdhury D. R., Samanta R.K and Chatterjee M., "*Design and Development of an Expert System Model in Differential Diagnosis for Neonatal Disease*", IFRSA's International Journal of Computing, Vol. 1, No. 3, pp. 343-350, 2011.

4. Sugumaran (ed), Vijayan, "*Several Approaches to Variable Selection by Means of Genetic Algorithms*", Intelligent Information Technologies: Concepts, Methodologies, Tools, and Applications. IGI Publishing, 2008.

5. Roy Chowdhury Dilip, Samanta R.K., Chatterjee M., "*A Data Mining Model for Differential Diagnosis of Neonatal Disease*", IFRSA's International Journal Of Computing, Vol. 1, No. 2, pp. 143-150, 2011.

6. Olusegun Folorunso, Oluwatobi O, dewale O. Ogunde. "*ID-SOMGA: A Self Organising Migrating Genetic Algorithm-Based Solution for Intrusion Detection*", Computer and Information Science, Vol. 3, No. 4, pp. 80-92, 2010.

7. Shanthi D., Sahoo G. and Saravanan N., "*Input Feature Selection Using Hybrid Neuro-Genetic Approach in the Diagnosis of Stroke Disease*", IJCSNS International Journal of Computer Science and Network Security, Vol. 8, No. 12, pp. 99-107, 2008.

8. Janson, D. J. and Frenzel, J. F., "*Training Product Unit Neural Networks with Genetic Algorithms*", IEEE Expert, Vol. 8, No. 5, pp. 26–33, 1993.

9. Jin Ya-Qiu and Cao Guangzhen, "*The BP-ANN Algorithm for Data Fusion of Landsat ETM and ERS SAR in Urban Area*", ISPRS, Vol. XXXVI-8/W27, 2005.

10. Heckerling P. S., Canaris G., Flach S. D., Tape T. G., Wigton R. S. and Gerber B. S., "*Predictors of Urinary Tract Infection Based on Artificial Neural Networks and Genetic Algorithms*", International Journal of Medical Informatics, Vol. 76, No.4, pp. 289-296, 2007.

11. Rumelhart D. E., Hinton G. E. and Williams R. J., "*Learning Internal Representations by Error Propagation*", Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 1, Cambridge: MIT Press, pp. 318–362, 1986.

12. http://en.wikipedia.org/wiki/Backpropagation Available: [Last accessed 18th Jan, 2012].

13. http://www4.rgu.ac.uk/files/chapter3 Available: [Last accessed 18th Jan, 2012].

14. H. Lu, R. Setiono and H. Liu. "*Effective Data Mining using Neural Networks*", IEEE Trans, International Journal on Soft Computing (IJSC), Vol. 2, No.2, 2011.

15. Roy A., "*Artificial Neural Networks–a Science in Trouble*", SIGKDD Explorations.

16. Tveter Donald R., "*Donald Tveter's Backpropagator's Review*", http://www.dontveter.com/bpr/bpr.html, Available : [Last Accessed on 1st September, 2011].

17. Paulin F., Santhakumaran A., "*Classification of Breast Cancer by Comparing Back Propagation Training Algorithms*", International Journal on Computer Science and Engineering (IJCSE), Vol. 3, No. 1, pp. 327-332, 2011.

18. Marczyk Adam, "*Genetic Algorithms and Evolutionary Computation*", The Talk Origins Archives, 2004.

19. http://en.wikipedia.org/wiki/Selection_(genetic_algorithm) Available: [Last accessed 19th Jan, 2012].

20. Goldberg D. E., "*Genetic Algorithms in Search, Optimization, and Machine Learning*". London, Addison-Wesley, 1989.

21. Karegowda Asha Gowda , Manjunath A.S., Jayaram M.A., "*Application of Genetic Algorithm optimized Neural Network connection weights for Medical Diagnosis of PIMA Indians Diabetes*", International Journal on Soft Computing (IJSC), Vol. 2, No. 2, pp. 15-23, 2011.

22. Chowdhury D. R., Samanta R. K., Chatterjee M., "*Neonatal Disease Diagnosis with Soft Computing*", Proc. International Conf. on Computing and System, ICCS-2010, The University of Burdwan, India, Vol. 1, pp. 27-34, 2010.

23. Chowdhury D. R., Chatterjee M., Samanta R. K., "*An Artificial Neural Network Model for Neonatal Disease Diagnosis*", International Journal of Artificial Intelligence and Expert Systems (IJAE), Vol. 2, No. 3, pp. 96-106, August, 2011.

24. Hansen J.V., McDonald J.B., "*Some Experimental Evidence on the Performance of GA-Designed Neural Networks*", Journal of Experimental & Theoretical Artificial Intelligence, Vol. 13, No. 3, pp. 307-321, 2001.

25. “*Neuro Intelligence Using Alyuda*”, http://www.alyuda.com, 2008, Available: [Last accessed 19th Jan, 2012].

26. Chowdhury D. R., Samanta R. K., Chatterjee M., “*Data Mining for Neonatal Prevalent Disease of North Bengal Districts*”, Proc. Nat. Seminar on Data Mining and Decision Support, Vidyasagar University, India, pp. 24-31, 2010.

CHAPTER 13

An Intelligent Diagnostic System for the Diagnosis of Neonatal Disease using ANN in Soft Computing Environment‡

13.1. Introduction

Intelligence is the skill to accomplish the objectives of any particular subject. The system is called intelligent when it accomplishes the objective more easily, accurately and more speedily. To achieve this it needs the ability to learn the objectives and best possible way to reach the goal. Considering any systems, it has many inter-related components. They may be highly cohesive in nature. Likewise, intelligence of a system is basically a property of its mind where brain instructs and controls the whole system.

In medical decision making, especially disease diagnosis, soft computing techniques provide a good platform. Artificial neural network (ANN) is a computational approach towards intelligent behavior, associating ambiguity, complexity, indecisiveness, contradictory nature, and uncertainty. In this paper, an approach of integrating conjugate gradient descent neural network and genetic algorithm has been deployed towards developing an intelligent diagnostic system for neonatal disease diagnoses. Genetic algorithm has been used for selection of attribute subsets and conjugate gradient descent training algorithm for classification.

There is vital importance of diagnosing the disease accurately before it starts affecting the body. Particularly if we are concerning about the neonatal patient, then the risk is much more than the other age group patients. Today there are many intelligent systems which are used for decision making and expert system development in the various fields including medical informatics. Although this intelligent system may differs from their prediction and diagnosis result. This even be there and differ from different medical specialist or domain experts opinion or diagnosis. In reality, in real world specialists' opinions complement one another and when integrated they usually form a better solution [1].

In this study, we have gathered different opinions from real world's experts; merge them together by proposed Intelligent System in the soft computing paradigm. Experimental results reflect better classification accuracy after comparing different studies.

---

‡ This based on the publication by the author entitled "An Intelligent Diagnostic System for the Diagnosis of Neonatal Disease using ANN in Soft Computing Environment", National Conference on Research and Higher Education in Computer Science and Information Technology-*RHECSIT-2012*, pp. 117-122, February, 2012.

The chapter is organized as follows:

Section 13.2 represents the concepts of the system. Section 13.3 describes the problem area of neonatal disease. Section 13.4 discusses the use of soft computing in intelligent systems. The proposed methodology of backpropagation neural network model is given in section 13.5. The design and implementation of intelligent system for neonatal prevalent disease diagnosis using artificial neural network using conjugate gradient descant algorithm is presented in section 13.6 and finally conclusion is given in section 13.7.

13.2. Concepts of a System

The word “System” basically derived from the Greek word “systema”. This means, an organized relationship among the functioning units or components. In our daily life we come into contact with several systems like, reservation system, transportation system, computer system, blood circulation system, disease diagnosis system etc. There are several definitions of systems. But the most common idea suggest about system is, an orderly grouping of interdependent components linked together according to a plan to achieve a specific goal. Here components are the physical parts of system.

13.2.1. Characteristics of the System

Depending upon the above concepts of the system, we may observe that there are several important characteristics presents in all system. They are as follows:

- Central Objective
- Organisation
- Interaction
- Interdependence
- Integration

❖ **Central Objectives**

This is the most important characteristics of the system. Any objectives may be real or stated. There must have a central goal to reach. Using the intelligence one has to achieve the central objective or the goals. The organization of any system may set one objective to operate and achieve. The user of the system must have previous knowledge of central objective before its starts workings.

❖ **Organization**

Organisation refers the structures and order. It is basically arrangement of components that helps to achieve the central objectives. Foe an example, disease diagnosis system is designed around an Inputs procedure, an inference engine to process as processing unit, an Output procedure which predicts or diagnose the disease and few soft computing artificial intelligence tools for classification, selection etc. jobs. When these units linked together, they work as a whole system for generating disease diagnosis prediction and outcome.

❖ **Interaction**

Interaction is the procedure in which each component functions with other components of the system. In the above example of disease diagnosis system, there must have exhaustive interaction between each module or components. Otherwise we may expect different results. The inter relationship between these components enable the system to perform the job accurately.

❖ **Interdependence**

Interdependence means that there should have dependency of components with each others. They are linked and coordinated together in a planned way to achieve the goal or the central objectives.

❖ **Integration**

Integration is a method of assimilating and tiding up the whole components of a system in such a way that, they may give their best performances individually as well as in a group. Any successful integration will typically produce a better result as a whole rather than works independently.

13.2.2. Intelligent System

Intelligence is the ability to grasp; to comprehend and better result from experience. It includes other abilities like, ability to acquire and retain knowledge; mental ability; the ability to respond faster and effectively to a new situation. Intelligent systems provide a standardized procedural approach to solve significant and fairly complex problems and acquire consistent and reliable results over time [2].

An intelligent system must have its own central objective, proper senses to select and actuators. To reach its objective it chooses an action based on its experiences. It can learn by generalizing the experiences it has stored in its memories [3]. In the computational perspective the intelligence of a system can be characterized by its flexibility, adaptability, memory, learning, temporal dynamics, reasoning, and the ability to manage uncertain and imprecise information [4].

An intelligent system is a system that follows some aspects of intelligence exhibited by nature. These include learning, adaptability, and robustness across problem domains, improving efficiency over time and space both, information compression from data to knowledge and check reasoning [5].

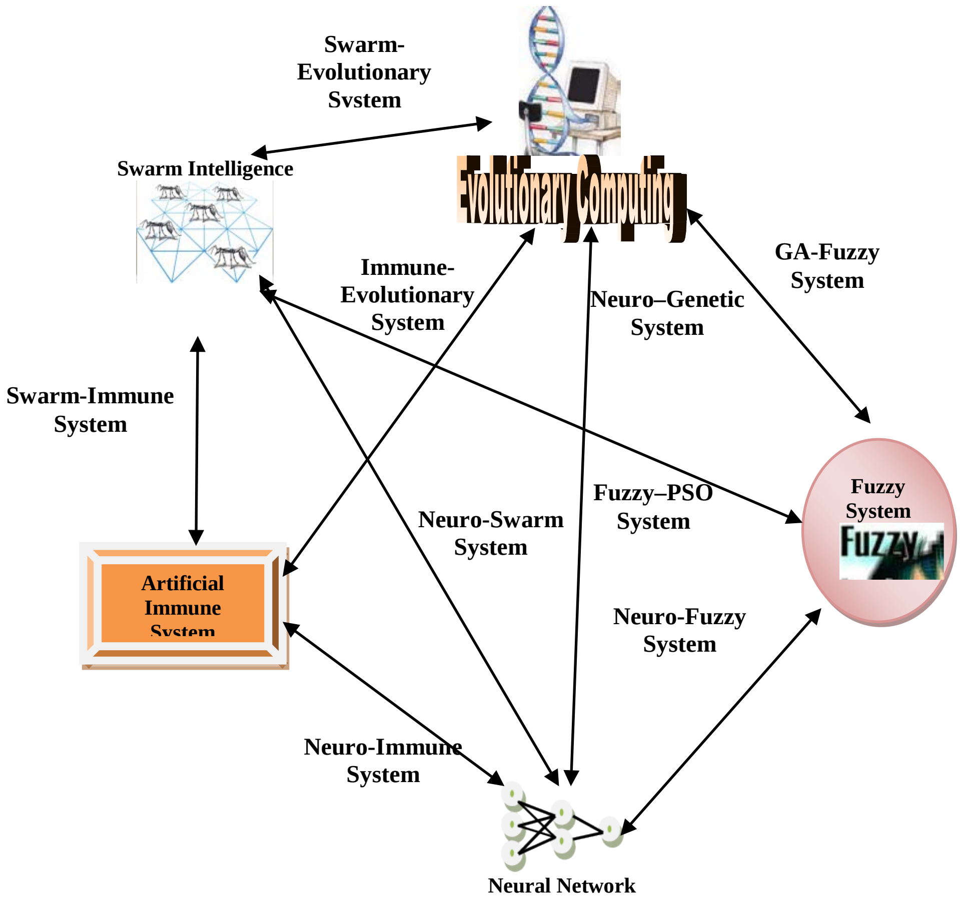

Figure 13.1. Computational Intelligence Paradigms.

13.2.3. Computational Intelligence

The name computational intelligence is a `message', according to scientific folklore it is chosen to indicate the link to and the difference with artificial intelligence. Computational Intelligence is a fairly new name covering a fairly new field [6]. The areas covered by the term computational intelligence are also known under the name soft computing. Computational intelligence techniques are generally bottom-up, where order and structure emerges from an unstructured beginning. While some techniques within computational intelligence are often counted as artificial intelligence techniques, likes genetic algorithms, neural networks etc. There is a clear difference between these techniques and traditional, logic based artificial intelligence techniques. In general, typical artificial intelligence techniques are top-to-bottom, where the structure of models, solutions, etc. is imposed from above.

Computational intelligence is the study and working procedure by which one can design of intelligent agents. In this regard, an agent is something that acts in an environment and it does something. An intelligent agent is a system that acts intelligently. That means what it does is appropriate for its circumstances and its goal, it is flexible to changing environments and changing goals, it learns from experience, and it makes appropriate choices given perceptual limitations and finite computation. The main purpose of this computational intelligence is to understand the principles that make intelligent behavior possible, in natural or artificial systems.

In Computational these interest generally focus on problems that only humans and animals can solve, problems requiring intelligence. Specific interests also focus on methods and tools that are applicable to this type of problems. Starting with seminal papers, special sessions, growing into separate conferences and specialized journals, different branches of Computational Intelligence evolve in many directions, frequently quite far from original roots and inspirations. New communities are formed and need to establish their identity by defining borders distinguishing them from other scientific communities.

Computational Intelligence (CI) is a sub-branch of Artificial Intelligence. It is the study of adaptive mechanisms to enable or facilitate intelligent behavior in complex and changing environments. These mechanisms include those AI paradigms that exhibit an ability to learn or adapt to new situations, to generalize, abstract, discover and associate. There are different opinions on the context of computational intelligence among different researchers. Depending on the working principles the computational intelligence paradigm is covered following procedures. They are: Evolutionary Computation (EC), Artificial Neural Networks (ANN), Artificial Immune Systems (AIS), Fuzzy System (FS) and Swarm Intelligence (SI). Though the independent techniques of this paradigm can work efficiently, still there is a current trend on the development on fusion based paradigm, discussed in chapter 12, which already shows the better results in this computational intelligence field. Here we may get the power of unitization achieving goals by using the cumulative forces of all components of the system and reduce the limitation of individual components [7]. Figure 13.1. shows various components of computational intelligence. It is also depicting individual component as well as hybrid component models in the computational intelligence paradigm.

13.2.4. Why Intelligent Diagnostic System?

Computer aided intelligent diagnostic systems has become a useful and mostly used tools which helps the doctor or domain experts as an assistant for performing better and more accurate disease diagnosis. This becomes vastly useful with the progression of artificial intelligent based system. There are several situations when doctor diagnosed a patient, if found problem, then as ‘diseased’ otherwise if there is no problem, then as ‘normal’ with full of information which is rather hard to the general patient. Even this information is not all the time appropriate. In terms of depth of the disease, they give gradation. It is, therefore, more meaningful to provide the description of the patient’s state in terms of degree of severity of being affected by disease, in our study, the neonates.

To do the proper diagnosis, a knowledgebase has to be developed based on the domain specific information about the disease. After getting all the information, it is being prepared for utilization for the system. At this juncture, system need better intelligence power to solve problem and diagnosis properly. The systems have to be smart enough to learn and even upgrade the information which is required at the time of diagnosis [8].

Figure 13.2. below is depicting the differences between Computational Intelligence and Artificial Intelligence. Although both of them are parts of the Machine Intelligence.

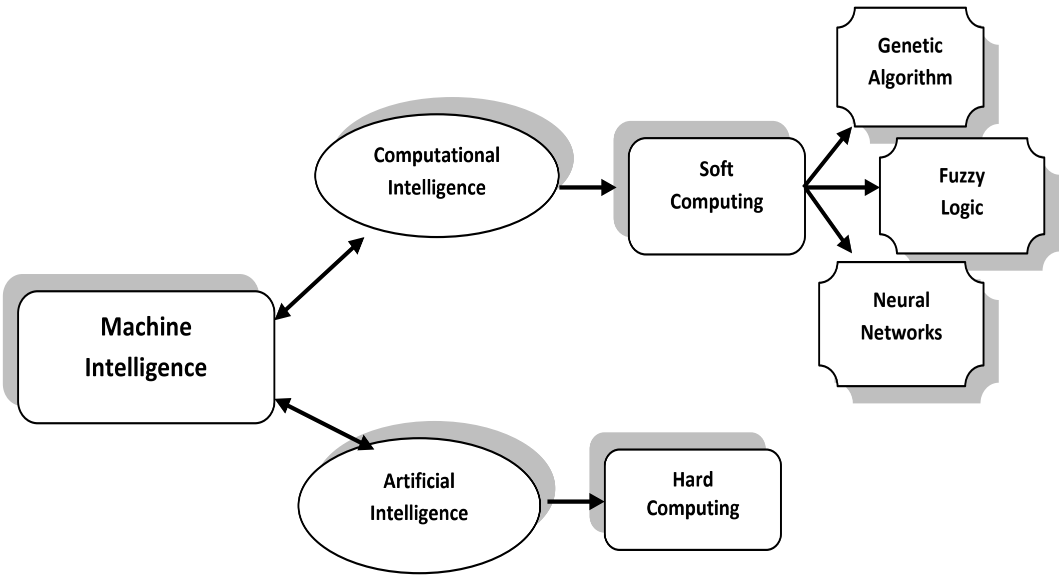

Figure 13.2. Computational and Artificial Intelligence

13.2.5. Neonatal Domain Specific Information

Neonates occupy a large portion of our population also they are also vulnerable or special risk group. This risk is related with improper growth, development, disease pattern and survival. Mortality rates in this age group are higher than adult population especially in developing countries [9]. This has already been discussed in previous chapters. A study shows the disease status of neonates in the North Bengal and Tarai Region of West Bengal [10]. We may re-visualize the Chapter 3 for this particular issue. For improving the diagnosis accuracy and serving as a scientific tool, any attempt towards developing intelligent diagnostic system would be effective in field of medical research particularly for neonatal disease diagnosis. Various soft computing methods have been applied in the medical research. Out of them neural networks, genetic algorithms have shown great potential towards development of intelligent decision support system for medical diagnosis. In this work, we deploy an approach of integrating conjugate gradient descent neural network and genetic algorithm. Genetic algorithm has been used for selection of attribute subsets and conjugate gradient descent network for classification.

13.3. Problem Area of Neonatal Disease

The total pediatric age group is subdivided as: Fetal(Conception – Birth), Neonates(0 - 4 weeks), Infants (4 weeks – 1year), Toddler(1 year – 3 years), Pre-school(3 years – 5 years), School going (5 years – 10 years) and Adolescence(10 years – 19 years). It is to be pointed out that the disease pattern, drug selection, diet, common rearing technologies are different for the different pediatric age groups. Neonatal phase is measured to be a significant one. Neonates are also in special risk group; the risk is related with growth, development, disease and survival. In India, especially in rural and inaccessible or remote areas, the mortality and morbidity are high [11]. The main reason may be of such high mortality and morbidity is prevalent diseases. An estimated two-thirds of childhood deaths occur in infancy, and, in turn, two-thirds of infant deaths occur in the first month of life. In addition to these four million neonatal deaths, primarily due to serious bacterial infections, birth asphyxia, and complications of pre-maturity and intrauterine growth restriction, an estimated 3.9 million pregnancies end in stillbirth [12].

The common causes of neonatal deaths are Birth asphyxia, Neonatal Sepsis, Preterm, Low birth weight, Delayed breastfeeding, Problems in breastfeeding, Diarrhea, Hemorrhage, Conjunctivitis, Skin Infection, Abnormal Jaundice, Me conium Aspiration, Pneumonia, Upper Respiratory Infection (URI), Hypothermia, Umbilical Sepsis, Tetanus, Convulsive Disorder, Unexplained fever, Failure to Gain Weight. Beside this Table 13.1. shows the typical neonatal health problems. This has already been discussed on previous chapters.

Table 13.1. Neonatal Health Problems.

Problems / Disease
LBW (Low Birth Weight)
RDS (respiratory Distress Syndrome
Hemolytic Disease
HIE(Hypoxic Ischemic Encephalopathy)
Congenital Malfunction
Aspirations syndromes
Metabolic Disturbances of Neonates
Neonatal Jaundice
Neonatal Septicemia
Birth Trauma
Neonatal Seizure
Prematurity
HMD(Hyaline Membrane Disease)

13.4. Soft Computing in Intelligent System

In traditional hard computing, the prime facts are precision, certainty and rigidity. In contrast, soft computing is the process that, correctness and certainty carry a cost and that computation, reasoning, and decision making should use wherever it is possible irrespective of having the tolerance for uncertainty and vagueness [13][14]. Soft computing uses the human mind as a role model and aiming effective performance of decision making. The detailed information regarding Soft Computing and its techniques has already been discussed in Chapter 7.

13.4.1 Quick Propagation and Conjugate Gradient Descent Neural Network

If any network has a large number of weights, quick propagation and conjugate gradient descent may be applied for training the data set. They have nearly the convergence speed of second-order methods. The memory requirements are proportional to the number of weights. Conjugate gradient descent and quick propagation are general-purpose training algorithms of choice [15].

13.4.2. Correlation-based Feature Reduction

Feature extraction and reduction is one of the important steps for pattern recognition since even the best classifier may perform poorly if the features are not chosen well

[16][17]. The reduced feature vector includes most of the useful information of the original feature vector. This reduced dimensionality helps reducing the database size as well as speeds up the inference procedure especially for a large data base. Moreover, it is essential for a class of complex pattern recognition algorithms. There are different algorithms for the purpose. Correlation-based feature subset selection (CFS) [18], Principle Component Analysis (PCA) [16], Association Rules (AR) [19] are some of the algorithms to mention. This work uses CFS. The central hypothesis is "A good feature subset is one that contains features highly correlated with (predictive of) the class, yet uncorrelated with (not predictive of) each other" [18].

A feature evaluation formula, based on ideas from test theory [20], provides an operational definition of the above hypothesis as follows:

$$r_{fc} = \frac{k\overline{r_{fc}}}{\sqrt{\{ k + k(k-1)\overline{r_{ff}} \}}} \qquad (1)$$

where r_{fc} is the correlation between the summed features and the class variable, k is the number of components, $\overline{r_{fc}}$ is the average of the correlation between the features and class variable, and $\overline{r_{ff}}$ is the average inter-correlation between features. CFS is an algorithm that couples this evaluation formula with an appropriate correlation measure and a heuristic search strategy.

13.4.3. Conjugate Gradient Descent Neural Network

If any network has a large number of weights, conjugate gradient descent may be applied for training the data set. It has nearly the convergence speed of second-order methods, while avoiding the need to compute and store the Hessian matrix. Its memory requirements are proportional to the number of weights. Conjugate gradient descent and quick propagation are general-purpose training algorithms of choice [15].

In our previous chapters, Chapter 11 and 12, we have discussed about the conventional two layer backpropagation algorithm and found better classification accuracy of neonatal disease data. Still it showed little inefficiency in training the network. To overcome this problem another useful approach based on conjugate-gradient optimization may be used for classification of multisource neonatal disease and sign symptom data, which is basically high dimensional. The conjugate-gradient neural networks give excellent performance in classification of multisource data. Still it may not be comparable with statistical methods in classification of high-dimensional data in soft computing paradigm. Backpropagation using gradient descent often converges very slowly or not at all. On large-scale problems its success depends on user-specified learning rate and momentum parameters. Parameters may not be selected automatically there and if incorrect values are specified the convergence may be exceedingly slow, or it may not converge at all. While backpropagation with gradient descent is still used in many neural network programs, it is no longer considered to be the best or fastest algorithm.

Conjugate Gradient algorithm is basically adjust weight values using the gradient during the backward propagation of errors through the network. Compared to gradient descent, the conjugate gradient algorithm takes a more direct path to the optimal set of weight values. Usually, conjugate gradient is significantly faster and more robust than gradient descent. Conjugate gradient also does not require the user to specify learning rate and momentum parameters. The conventional conjugate gradient algorithm uses the gradient to compute a search direction. It then uses a line search algorithm such as Brent's Method to find the optimal step size along a line in the search direction. The line search avoids the need to compute the Hessian matrix of second derivatives, but it requires computing the error at multiple points along the line. The conjugate gradient algorithm with line search has been used successfully in many neural network programs, and is considered one of the best methods invented still now [21].

The conjugate gradient algorithm selects the successive direction vectors as a conjugate version of the successive gradients obtained as the method progresses. Thus, the directions are not specified beforehand, but rather are determined sequentially at each step of the iteration. At step k one evaluates the current negative gradient vector and adds to it a linear combination of the previous direction vectors to obtain a new conjugate direction vector along which to move. There are three primary advantages to this method of direction selection.

Firstly, unless the solution is attained in less than n steps, the gradient is always nonzero and linearly independent of all previous direction vectors. Indeed, as the corollary states, the gradient gk is orthogonal to the subspace B_k generated by $d_0, d_1, \ldots, d_{k-1}$. If the solution is reached before n steps are taken, the gradient vanishes and the process terminates.

Second, a more important advantage of the conjugate gradient method is the especially simple formula that is used to determine the new direction vector. This simplicity makes the method only slightly more complicated than steepest descent.

Third, because the directions are based on the gradients, the process makes good uniform progress toward the solution at every step. This is in contrast to the situation for arbitrary sequences of conjugate directions in which progress may be slight until the final few steps. Although for the pure quadratic problem uniform progress is of no great importance, it is important for generalizations to non-quadratic problems [22].

Algorithm for Conjugate Gradient is as follows:

Step 1. $\boldsymbol{g}_0 \leftarrow \boldsymbol{Q}x_0 + \boldsymbol{b}$

Step 2. $\mathbf{d}_0 \leftarrow -\boldsymbol{g}_0$

Step 3. for $\boldsymbol{k} = 0, \ldots\ldots, n-1$ **do**

a) $\boldsymbol{\alpha}_k \leftarrow -\boldsymbol{g}^T{}_k\boldsymbol{d}_k \;/\; \boldsymbol{d}^T{}_k\boldsymbol{Q}\boldsymbol{d}_k$

b) $\boldsymbol{x}_{k+1} \leftarrow -\boldsymbol{x}_k + \alpha_k\boldsymbol{d}_k$

c) $\boldsymbol{g}_{k+1} \leftarrow \boldsymbol{Q}x_{k+1} + \boldsymbol{b}$

d) $\boldsymbol{\beta}_k \leftarrow \boldsymbol{g}^T{}_{k+1}\boldsymbol{Q}\boldsymbol{d}_k \;/\; \boldsymbol{d}^T{}_k\boldsymbol{Q}\boldsymbol{d}_k$

e) $\boldsymbol{d}_{k+1} \leftarrow -\boldsymbol{g}k+1 + \boldsymbol{\beta}_k\boldsymbol{d}_k$

Step 4. return $\boldsymbol{x}_n$

In Step 3 b) when k = 0 is a steepest descent. Each subsequent step moves in a direction that modifies the opposite of the current gradient by a factor of the previous direction. Step 3a) to e) gives us the Q-orthogonality of the descent vectors $d_0, d_1. . . , d_{n-1}$.

13.4.4 Converging to the Optimal Solution

If we provide a set starting weight values which has been selected randomly, we may apply the conjugate gradient algorithm to optimize the weight values. In most of the cases, training algorithms follow a cycle to improve the weight values. Figure 13.3. below shows the how weight values are improved on training in a cycle. Here every cycle is called an epoch. Because the error information is propagated backward through the network, this type of training method is called backward propagation [23].

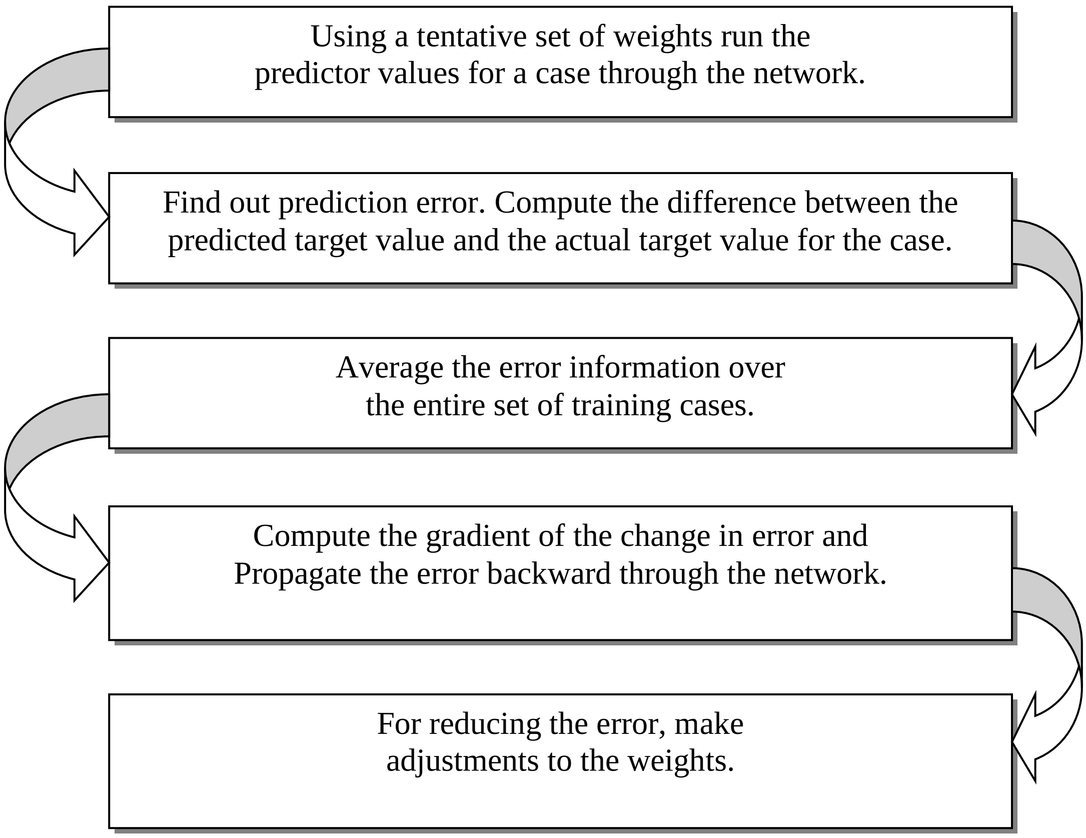

Figure. 13.3. Training Algorithm Improving the Weight Values.

13.4.5 Genetic Algorithm

Genetic Algorithm is basically inspired by Darwin's theory of evolution- Survival of the Fittest. GA algorithms are the ways of solving problems by selection, crossover, mutation and accepting to progress of a problem solution. It is also called global optimization algorithm which are based on the basis of natural selection and natural genetics. It applies randomize and probabilistic searching strategy for attributes selects for the high fitness [23]. A complete discussion about genetic algorithm already is in chapter XII.

13.5. Methodology

13.5.1. Data Set

The data of 95 patients have been collected for our studies having basic signs and symptoms of neonatal disease. They have been standardized for making, as much as, possible error free data. All the cases are analyzed after careful scrutiny with the help of the pediatric expert. Table 13.2. shows the various input parameters of the study.

Table 13.2. Input Data Parameters.

Sl.No.	Parameters	Column Type
1	Birth_Term_Status	Categorical
2	Birth_Weight_Status	Categorical
3	Age_in_Hours>72	Categorical
4	Lethargy	Categorical
5	Refusal_to_Suck	Categorical
6	Poor_Cry	Categorical
7	Poor_Weight_gain	Categorical
8	Hypothalmia	Categorical
9	Sclerema	Categorical
10	Excessive_Jaundice	Categorical
11	Bleeding	Categorical
12	GI_Disorder	Categorical
13	Seizure	Categorical
14	Sluggish_Neonatal_Reflex	Categorical

13.5.2. Attribute Selection based on GA

We have applied CFS as attribute evaluator and GA for searching the data space. All the 14 symptoms have been optimized by genetic algorithm and the genotype is represented by the sequence of symptoms. GA algorithm is an adaptive optimization algorithm for simulating the inheritance and evolution of biology in the nature [24]. GA parameters of the study for the feature or attribute selection have been shown in Table 13.3. The populations after 20th generation are showing on the Table 13.4.

Table 13.3. GA Parameters.

Attribute Selection on all Input Data	
Search Method	**Genetic search**
Start set	No Attributes
Population size	14
Number of generations	20
Probability of crossover	0.6
Probability of mutation	0.033
Report frequency	20
Random number seed	1

Table 13.4. Population after 20th Generation Subset Category.

Merit	**Scaled**	**Subset**
0.44418	0.49641	1 2 3 4 5 6 7 9 10 11 12 13 14
0.44418	0.49641	1 2 3 4 5 6 7 9 10 11 12 13 14
0.44418	0.49641	1 2 3 4 5 6 7 9 10 11 12 13 14
0.41041	0	1 2 3 5 7 9 10 11 13 14
0.44418	0.49641	1 2 3 4 5 6 7 9 10 11 12 13 14
0.44418	0.49641	1 2 3 4 5 6 7 9 10 11 12 13 14
0.43019	0.29085	1 2 3 4 5 6 7 8 9 11 13 14
0.44385	0.49158	1 2 4 5 6 7 9 10 11 12 13 14
0.44418	0.49641	1 2 3 4 5 6 7 9 10 11 12 13 14
0.44418	0.49641	1 2 3 4 5 6 7 9 10 11 12 13 14
0.44418	0.49641	1 2 3 4 5 6 7 9 10 11 12 13 14
0.44027	0.43897	1 2 3 4 5 6 7 8 9 10 11 13 14
0.44279	0.47602	1 2 3 4 5 6 7 8 9 10 11 12 13 14
0.44418	0.49641	1 2 3 4 5 6 7 9 10 11 12 13 14

After the application of CFS and GA, the system rejects 'Hypothalmia' from the input features. We apply the rest 13 features for classification test through ANN.

13.5.3. Data Partition Set and Preprocessing

The Data have been analyzed using Neuro-intelligence tool [15]. Table 13.5. shows the number of records which are partitioned for training, validation and test set. Categorical input data are encoded by [-1..1] scaling range for the input columns and [0..1] scaling range for the output column at the time of analyzing after selection. All the columns are considered as input, except the last column which is considered as the target or output column for the process.

Table 13.5. Data Partition Set.

Partition Set Using	Records	Percentage (%)
Total	**95**	**100**
Training Set	**65**	**68.42**
Validation Set	**15**	**15.79**
Test Set	**15**	**15.79**
Ignore Set	**0**	**0**

13.5.4. Multilayer Feed Forward Network

Fig.13.4. depicted the architecture of the network which is referred as a multilayer feed forward network having input, hidden and output layers.

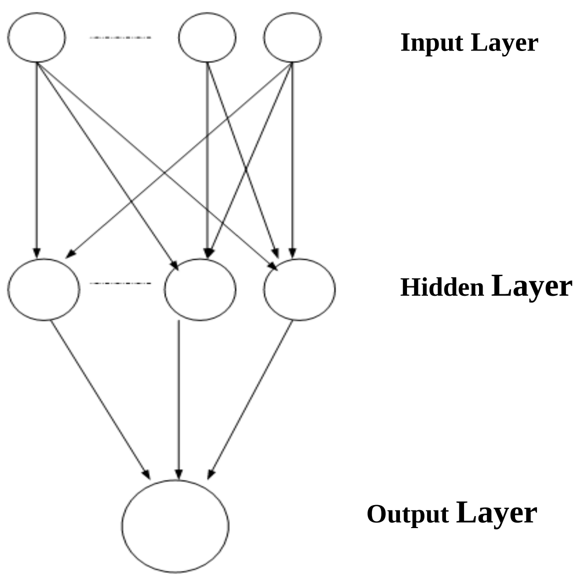

Figure 13.4. Multilayer Feed forward Network.

We consider weights, fitness, training error, validation error, test error and AIC (Akaike Information Criterion) for searching the best network architecture. Simple models are preferred to complex networks and determine the optimal number of weights required for the neural network. Heuristic searching method has been applied in this study. In chapter XII details about multilayer feed forward network have been discussed.

13.5.5. Akaike Information Criterion (AIC)

In this context Akaike Information Ccriterion (AIC) is tool which measures the best fit relatively in any statistical methods. AIC was first developed by Hirotsugu Akaike. He gave the name as "an information criterion" in the year 1974. AIC basically based on the concept of information entropy. When any model is used to describe the reality of usefulness, then AIC measures the entropy and information lost at that period. Data analysis often requires selection over several possible models that could fit the data, in our case also the data set requires proper analysis for the selection the best network.

In every normal cases AIC can be referred as follows:

$$AIC = 2k - 2\ln(L) \quad (2)$$

Here,

k = Number of parameters in the statistical model
L = Maximized value of the likelihood function.

If the data set having minimum AIC value then it is best fit of AIC. It also includes few awful consequences like increasing of function with estimated parameters. This causes the problems of over fitting.

In an another formulation of Akaike Information Ccriterion (AIC) is as under [25]:

$$AIC = -2\ln(L(\theta)) + 2k \quad (3)$$

Here,

L(θ) = Maximized likelihood value and

k = Number of estimable parameters in the model.

At any time when **k** is large, relative to number of data points n that is ($n / k < 40$), then bias-adjustment is required. After this biasing adjustment the corrected Akaike information criterion is,

$$AIC = -2\ln(L(\theta)) + 2k + (2k(k-1)/(n-k-1)) \quad (4)$$

There is absolutely no problem using this even when the condition is not true, specially when the bias-adjustment term becomes too small and size of **n** increases relative to **k**.

An AIC value provides a way of best possible model selection. AIC remain undisclosed any warnings even the model actually fits the data well. If all the candidate models fit poorly, AIC will not give any information regarding the best fit operations. AIC is asymptotically optimal in selecting the model with the least mean squared error, under the assumption that the exact best fit model is not in the candidate set. AIC converges to the optimum value with the best possible fit in the specific model.

13.6. The Design and Implementation of Intelligent System

13.6.1. ANN Results

In this study, neural network contains one input, one hidden and one output layer. 18 numbers of neuron have been taken as input, 6 hidden neurons and the output layer has 14 neurons. Here the input neurons represent the symptoms for the neonatal disease diagnosis. The output layer neurons represent the predicted types of disease. This network has been trained using conjugate gradient descent algorithm. Here in the Table 13.6. we find the best fit value using AIC is -75.7416, which is relatively smaller the other model having the fitness value 1.875. Hence it is chosen the best fit network model for training among the other models. Exhaustive searching method has been applied in this study.

Table 13.6. Best Fit Network Architecture Search.

Architecture	# of Weights	AIC	Fitness	Train Error
[18-12-14]	423	491.2038	1.666667	0.723077
[18-18-14]	627	906.0523	1.666667	0.692308
[18-28-15]	967	1571.548	1.666667	0.753846
[18-6-14]	**117**	**-75.7416**	**1.875**	**0.446154**
[18-4-14]	151	-31.4434	1.875	0.615385
[18-45-14]	1545	2759.106	1.666667	0.6
[18-5-14]	185	46.20388	1.666667	0.553846
[18-6-14]	219	93.22361	1.666667	0.676923
[18-8-14]	287	229.2236	1.666667	0.676923

We have considered weights, fitness, training error, validation error, test error and AIC (Akaike Information Criterion) for searching the best network architecture in this study. Simple models are preferred to complex networks and determine the optimal number of weights required for the neural network.

Table 13.7. Training Performances.

Approach	Network	CCR % (training)	CCR %(validation)	CCR% (Test)	Network Error
Quick Propagation	**18-6-14**	**75.38**	**40**	**40**	**0.021**
Conjugate Gradient Descent	**18-6-14**	**64.61**	**40**	**40**	**0.034**

It is seen from the Table 13.7. that quick propagation training has the lower error than conjugate gradient descent training. Correct classification rate (CCR %) for training is better in quick propagation than that of Conjugate gradient descent method.

The above relative study shows a better prediction accuracy achievement on training data using Intelligent Diagnostic System processed by Conjugate Gradient Descent learning methods in neural network over the traditional Neural Network processed by Back Propagation method.

13.7. Conclusion

In view of the consideration that the increasing population and abrupt changes environment associated with changeable socio-economic condition, the neonatal problems increases day by day. Also the unavailability of neonatologist or domain experts and lack of accuracy diagnosing the neonatal disease would affect very much. Some of the soft computing and data mining techniques of our earlier studies, which has already been discussed in previous chapters, have also suggested about diagnosing neonatal diseases in a better way. This paper would helpful for the domain experts as the soft computing techniques applied here caused better prediction accuracy on the training data.

All the researchers and scientists are mostly familiar with the statistical approach to data analysis. If any particular hypothesis is given for the analysis, few statistical tests are applied to the dataset and they try to find out any relationships between different parameters. But in the Intelligent System environment, machine learning systems can go much further to analysis this in better way. In soft computing paradigm, Intelligent System look at unrefined data and then attempt to hypothesis relationships within the data, and newer learning systems are able to produce quite complex characterizations of those relationships i.e. they attempt to discover humanly understandable concepts.

The innovation in our present study lies in the combination of Intelligent System with Artificial Intelligence and Soft Computing approach architectures. This system and improved technology may used for managing and taking away un-structured biomedical research results into structured disease information for end users and development of decision support system for disease diagnosing for the neonates. This system may be used as the true guide for the domain experts, doctors, specialists and even medical students and researchers for the diagnostic system. Use of soft computing GA, ANN tool to integrate dispersed knowledge sources into a powerful mechanism for dynamic building of new knowledge, on the spot, as new epidemics or neonatal disease come into view. So far this development is showing better results, still lots of work remains, such as implementation of on spot verification of disease, use of global agent, proper security mechanism, online updating procedure for testing and validation etc.

In view of the consideration that the increasing population and abrupt changes of environment associated with changeable socio-economic condition, the neonatal problem increases day by day. Considering the complexity of the neonatal domain and the scarcity of neonatologist or domain experts, an automated intelligent system may be welcome. This work suggests that quick propagation is better than conjugate gradient descent method for the data set we consider. In concise words Intelligent System for the neonatal disease diagnosis is an immense help for the society as well as for the medical practitioners.

References

1. Kokol Peter, Povalej Petra, Stiglic Gregor and Dinevski Dejan, "*Improving Medical Decision Making by Self Organizing Intelligent Systems*", ISCID '08, Proceedings of the 2008 International Symposium on Computational Intelligence and Design, Vol. 1, pp. 267-270, 2008.

2. Byrd T.A. and Hauser R.D., "*Expert Systems in Production and Operations Management: Research Directions in Assessing Overall Impact,*" Int. J. Prod. Res., Vol. 29, pp. 2471-2482, 1991.

3. Fritz Walter, "*Intelligent Systems and their Societies*", Machine Learning and Cognitive Science, NHP Publisher, 2012.

4. Duch W., "*What is Computational Intelligence and Where Is It Going?*", W. Duch and J. Mandziuk, Eds., Challenges for Computational Intelligence, Springer Studies in Computational Intelligence, Vol. 63, pp. 1-13, 2007.

5. Krishnakumar K., "*Intelligent Systems for Aerospace Engineering–An Overview*", NASA Technical Report, Document ID: 20030105746, 2003.

6. Craenen B.G.W., Eiben A.E., "*Computational Intelligence*", Encyclopedia of Life Support Sciences, EOLSS; EOLSS Co. Ltd.

7. Engelbrecht Andries P., "*Computational Intelligence: An Introduction*", Second Edition, John Wiley & Sons Ltd., pp. 4-6, 2007.

8. Chakraboarty Chandan, Mitra Tamoghna, Mukharjee Amaradri, Roy Ajay K., "*CSIDSA: Computer-Aided Intelligent Diagnostic System for Bronchial Asthma*", Expert System with Application, Elsevier, Vol. 36, No 3(1), pp. 4958-4966, 2009.

9. USAID. "*Maternal and Child Health–2004.USAID Battles Neonatal Deaths in India*",(online)2005:48.Available: http://www.usaid.gov/our_work/global_health/home/News/ghachievements.html, Available: [Last accessed 1st November, 2007].

10. Roy Chowdhury Dilip, Samanta R.K., Chatterjee M., "*A Study of the Status of New Born in Terai Region of West Bengal*", A.M.S.E. France, No. 05 225 (2C), 2007.

11. Kumar D., Verma A., and Sehgal V. K., "*Neonatal Mortality in India*", Rural and Remote Health, 7:833 (online) 2007; http://www.rrh.org.au , Available: [Last accessed 8th October, 2009].

12. Ministry of Health and Family Welfare, Government of India, "*Government of India. Annual Report 2006-2007*", New Delhi, 2007.

13. Zadeh, "*Fuzzy Logic, Neural Networks and Soft Computing*", Commun. ACM, Vol. 37, No.3, pp. 77–84, 1994.

14. Zadeh, Kacprzyk, "*Fuzzy Logic for the Management of Uncertainty*", Wiley, New York, 1992.

15. *"Neuro Intelligence using Alyuda*", http://www.alyuda.com, 2008, Available: [Last accessed 11th May, 2011].

16. Calisir D., and Dogantekin E., "*A New Intelligent Hepatitis Diagnosis System", PCA-LSSVM*, Expert systems with Applications, Vol. 38, pp. 1075-10708, 2011.

17. Avci E., "*A New Optimum Feature Extraction and Classification Method for Speaker Recognition"*, GWPNN, Expert Systems with Applications, Vol. 32, No. 2, pp. 485-498, 2007.

18. Hall M. A., "*Correlation-based Feature Subset Selection for Machine Learning"*, Hamilton, New Zealand, 1998.

19. Karabatak M., and Ince M. Cevdet, "*An Expert System for Detection of Breast Cancer Based on Association Rules and Neural Network"*, Expert Systems with Applications, Vol. 36, pp. 3465 – 3469, 2009.

20. Ghiselli E. E., "*Theory of Psychological Measurement*", McGraw-Hill, New York, 1964.

21. http://www.dtreg.com/mlfn.htm Available: [Last accessed 1st Dec. 2011].

22. http://www.cs.iastate.edu/~cs577/handouts/conjugate-gradient.pdf Available: [Last accessed 1st Dec. 2011].

23. Sherrod Phillip H., "*Predictive Modeling Software*", DTREG, pp. 255-256, 2011.

24. Yuan, Y., and Suarga, S. "*On the Integration of Neural Networks and Fuzzy Logic Systems"*, .IEEE International Conference on Systems, Man and Cybernetics, pp. 452-7, 1995.

25. Hansen J.V., McDonald J.B., "*Some Experimental Evidence on the Performance of GA-Designed Neural Networks*", Journal of Experimental & Theoretical Artificial Intelligence, Vol. 13, No.3, pp. 307-321, 2001.

26. Turkheimer, Hinz, Cunningham, "*On the Un-decidability Among Kinetic Models from Model Selection to Model Averaging*", Journal of Cerebral Blood Flow & Metabolism, Vol. 23, pp. 490-498, 2003.

CHAPTER 14

Accomplishment of the Objectives and Future Scope

14.1. Introduction

To accomplish the research objectives and reaching the goal we have placed a numbers of research analyses. Research objectives set the purpose and focus of our research with the fundamental questions that have already been addressed in different chapters in the thesis. Defining our research objectives actually means defining what do we need to explore, investigate and how we are going to accomplish that. This is one of the most important aspects of research design and implementation. It includes individual, substantial steps that will be taken in research design. Our individual steps gyrate around a wider question or problem that we have defined in different chapters in whole study.

14.2. Accomplishment of the Objectives of the Thesis

During past few years of research study and analysis we have come to our conclusions and have been reached the goal that we sets in our objectives in the chapter 1 section 1.3. In this thesis, we planned to do these in steps for which we have tried to explore that status of the domain for understanding the needs of the sphere of persuade. To meet the needs of the domain and achieve our goals, we have developed and tested a knowledge based system intended for neonatal disease diagnosis and management. The decision support system named after, ***NeoExpert,*** has been designed for demonstrating the usability and acceptability of medical knowledge available in a different places for medical help seekers including health professional and domain experts. Even this system might be useful for teaching and training purpose for the students as well as experts. Afterwards this system focuses on the significant features of input neonatal-medical knowledge in a complicated real life decision making problem for helping in estimating the number of retained cases for a problem domain.

Accomplishment of the objectives of the thesis has been drawn out of the experimental review, are as follows:

Using survey methods of the neonatal domain, we found the statistics and present scenario of neonatal problems in global and local respects. We have been studied possible every corner of information related with this study. We have investigated the reason of high neonatal mortality rate, real facts of neonatal disease and also about distribution of domain experts in the region. Using survey methods each component or question must link back to our defined objectives and thus research design. Sometimes we realized that the collected data was not relevant for our hypothesis. At this juncture we have used a particular elegant point at the objective setting step is inadequately assessing the required scope. And finally our findings were ended up with meaningful unambiguous information. Thus we may say that we reached the objectives successfully for designing the decision support system out of the present knowledge [chapter 2 for information].

On finding the objectives included in chapter 3, about studying and surveying the clinical profile and outcome of neonates in Terai area of West Bengal and also to highlight the importance and the problems in neonatal care of the said region, we have successfully reached to get all this useful and most important data. The study in this chapter concludes with searching of prevalence neonatal disease of the said region and probable cause of it.

Very often, objectives may be based on the findings of other research done already. We may get the information that someone else has already investigated or hypothesized and focusing on a specific aspect of their findings to either strengthen or challenge them. But in my case, there is no substantial evidence regarding neonatal disease diagnosis has been recorded yet.

On the very next objective for developing a knowledgebase decision support system depending on expert system technology we found that among the various systems in the field of diagnostic expert system our system is also working efficiently diagnosing neonatal disease and also giving valuable suggestion as the management. The system generates the list of diseases by looking at each sign and symptom, and then match with lab findings to determine how strongly each of the physical findings is associated with the disease under consideration for the neonatal disease. We have tested the results produced by the system with the domain experts and found the satisfactory results [discussed in chapter 4]. It shows good performance as apparent from its performance evaluation. Hence successful accomplishment is focused undoubtedly.

There was a challenge against developing knowledgebase decision support system for the neonatal domain because of building up the knowledgebase base. Every day a huge number of clinical data has to process for diagnosis and that too having data which is irrelevant. Thus, there is a need of properly managed the data base and knowledgebase as well. So we set our goals and objective towards a knowledge unearthing and data mining concepts. Using the same data sets with different mining techniques such as clustering and comparing results of each technique in order to construct a full view of the resulted patterns and levels of accuracy of each technique may be very useful for this application. In the study of chapter, we just initialized the process of data mining on the neonatal database and integrating that with decision support system. In the subsequent chapters afterwards we reached the objectives of extracting the useful knowledge and build up the knowledgebase for accurate decision making as much as possible.

Basically follow-up research involves more than repeating research that's already been done. There are several follow up research on diagnosing disease of neonates concentrated on any particular disease, but in our case our aims was to improve the understanding of a specific topic through asking what else needs to be evidenced before the research is meaningful, or what knowledge could be acquired from a more focused exploration and analysis, or scrutiny of the existing findings.

We have used improved data mining approaches such as Decision Tree approach, which has been found suitable for the purpose since decision tree construction can make use of both symbolic or nominal and real-valued attributes – a characteristic of medical domain. Although this system is for the diagnosis of neonatal disease still this can be implemented for the diagnoses of any other age group also, provided the knowledgebase is having strong and useful data. The proposed system is making the process of disease diagnosis faster and with possibly of maximum accuracy. It is not actually substituting the human specialist but is a real useful tool for them. Here we clarify the basic difference between a typical ‘Expert System’ and the data mining model. This proposed system is that the data mining process boost the process of learning and thereby improve decision making process. This improved data mining analysis has helped a lot to reach the objectives [discussed in chapter 6].

There is few more intercession which has been given throughout the study. Before setting concrete and well-considered research objectives it is essential to ensure that the research is going to be meaningful, and that efforts we have taken already, were not wasted on objectives that don't make any sense or are impractical for decision making. Furthermore, our findings might not also be too wide to finish off anything meaningful or to mold into a persuasive argument.

Now a day’s hybridization of several techniques gives a good result predicting the disease. We also described a way of designing a hybrid decision support system in soft computing paradigm for neonatal disease diagnosis [discussed in chapter 7. Hybridization includes the application of rough set theory and C4.5 classification algorithm. Using rough set approach we generate rules for classification. C4.5 classification algorithm has been used for the same data set. After a comparative analysis we found this methodology also helps in studying the variation of different classification parameters including confidence factor, folding etc. In this perspective, soft computing methods have been successfully applied to solve non-linear problems in medicine, disease diagnosis and management particularly neonatal disease diagnosis. Soft Computing methods which indicate a number of methodologies used to find rough solutions for real-world problems which contain various kinds of inaccuracies and uncertainties can be alternative methods to statistical methods. Soft Computing paradigms with the use of rough set computing and decision tree approach helped a lot for the rule set designing and development of decision support system. Thus we were reaching our objectives very efficiently manner undoubtly.

If the research-scope is too slender and we may find that we haven't explored the subject adequately, then we must think of meaningful information lacks. In our study also there were numbers of uncertainties involved in disease prediction. Thus for overcoming this specific problem we have further set the objective to reduce the uncertainties. We have implemented Rough Set based model because rough sets offer an effective approach of managing uncertainties. This can be employed for tasks such as to discover patterns in data, attribute dependency analysis, dimensionality reduction, feature identification and classification. With the help of reduct and core, one can sidetrack the rules which are not that significant for the development of a decision support system. We have deployed different algorithms for finding reduct and core with the objective of a comparative study. Another important observation focused that neither all attributes are required nor play significant role. We found that the three attributes, namely, ‘Excessive_Jaundice’, ‘Sclerema’, and ‘GI_Disorder’ do not play any significant role in prediction accuracy. Moreover, it is found that ‘Seizure’ plays a more significant role than ‘Bleeding’. So, a less number of attributes can be used in rational decision making which will reduce the number of rules, and the search space as well. Reduction of parameters speeds up the decision making process [discussed in chapter 8]. This really pushed the flow of work in a meaningful one which has utilized to achieve the goal and helps to accomplish the task easily and effectively.

We have trimmed down the needs to answer those specific questions or gather that specific information which ensures that each relates directly to our initial problem or question for each objectives.

Development of an Expert System for neonatal diseases diagnosis is a complicated process and requires high level of expertise. Any attempt towards designing and developing of an expert system dealing with differential diagnosis of neonatal disease has to overcome various difficulties. This is our objective of the study. We have already discussed in chapter **10** that the expert system model can be useful as any information that enables individuals to understand their health and make health-related decisions for themselves or their family and take care of the little Childs easily. Thus this model may give a useful and better tool for expert opinions and decision making. There is the comparative study of Johnson Reducer, Exhaustive Reducer and Genetic Reducing Algorithm, Rule Generation and Re-building process which are focused for processing the expert system development. The proposed system shows good performance for finding the differential diagnosis of neonatal disease and achieved 70% accuracy [discussed in chapter 10]. The above experiments in this chapter substantiate that, the proposed expert system is giving better performance by speeding up the process of differential disease diagnosis for the neonates and decrease the diagnosis test and cost as well.

The present era heading towards the fifth generation computation where computing is really very advanced, still there are some certain tasks or process which are not possible by the program made for any common microprocessor. Thus by developing software implementation of a neural network can be a solution for this. Taking immense help from the natural human nervous system, neural networks working principles are decided and processed. They are the simplified models for processing of many intelligent abilities such as learning, generalization and abstraction and more. Although, Neural networks never replace the human experts instead they can helpful for decision making, classifying, screening and also can be used by domain experts to cross-check their diagnosis. Artificial Intelligence techniques consist of developing a computer based decision support system does somewhat that it were done by a human being. Since we were dealing with categorical values, we need to convert them and preprocessed to get the binary values. This was also a very big challenge making DSS using Artificial Intelligence and Neural Networking techniques. Our study overcomes this problem [discussed in chapter 11] and concludes with higher prediction result and when the Network has trained and tested after optimizing the input parameters, the overall predictive accuracy acquired was 75%. This prediction accuracy is more than the previous one. We have been successful reaching objectives even using of neural network techniques.

We tried Neuro-Genetic fusion approach for establishing the correct input features selection, sign and symptoms which were defined by categories of the network. If the information is huge, then it may degrade the classification performance and thereby lose the prediction accuracy. The main advantages of selection the proper attributes are cost reduction for data accusation, increase efficiency in classifier based systems. Our study presented two combinations of methods of neural networks and genetic algorithm for producing an improved performance on real-world classification problem, particularly this neonatal disease diagnosis. GA also selects relevant features subsets and connection weights. The experimental results showed that, the performance of the network can be improved by the proper selection of the input features. Use of Genetic Algorithm is basically for searching features from huge search space Back Propagation Neural Network is used for determining the parameters like hidden nodes and layers with Back Propagation Neural Network [discussed in chapter 12].

We have used another fusion approach of developing an intelligent diagnostic system which works successfully by giving optimum accuracy. We tried to develop such kinds of system which is having ability to learn and utilize that afterwards. There are few inter-related components, which are highly cohesive in nature. We tried to develop intelligence system which is basically a property of its knowledge and system actually instructs and controls the whole units [discussed in chapter 13]. The study evidently shows the accomplishment of the task reaching the objectives.

Finally, using our decision support system, the user can interact with a computer to solve a certain problem related with neonatal problems. Because of storing heuristic knowledge this system works in better way. Our system can be used in consultation as a helping hand or assistant, since it shows the result of diagnosis faster and in addition, it offers explanations of the obtained results, being very helpful to the professional and related field workers and also for the domain experts. The system concludes the diagnosis based on answers by the user to specific question related with neonatal illness that the system asks the user. The questions provide the system for explanation for the symptoms of the patient that helps the expert system for diagnosis the neonatal disease by inference engine. It processes the information in order to extract rules, which completes the knowledge base. The diagnosed results are matched with expert's opinions, which show maximum accuracy in this context.

For the fulfillment of our study we considered the best means our objectives. For successfully completion of the objects and reaching goals we had given enormous efforts. We have done face to face interviews; regular consultation with pediatrician and experts, surveys of different technologies for better prediction accuracy, analysis various case studies, analysis of other research, observation, several experiments, develop several models and comparative study and development of knowledgebase decision support system for treatment planning for neonatal prevalent disease diagnosis.

14.3. Future Scope and Further Works

As we mentioned earlier chapters that our goal is to develop a knowledgebase decision support system for neonatal disease diagnosis and treatment planning. Though we successfully reached the goals, still we must convey this message that our system is not suppose to replace the domain expert's knowledge, rather it is being helpful for those personnel related with neonatal health care management by giving reliable assistance.

It is now palpable that the fast growing fields in the intelligent system development for any organization, Decision Support System becomes one of the most vital, important, and strategic and most demanded tool. Still it is being implemented as voluntary tool not as compulsory tools. As a researcher of this filed, we should focus on needs of shifting the research focus from user-related factors to the contextual variables. Thus the future scope of this tool has to be extended to make it compulsory instead of voluntary, particularly in the field of disease diagnosis system. We would also focus on the particular points of shifting the implementation research from user-related factors to task-related factors. Even domain specific organizational, and external environmental factors will also be included proper diagnosis and treatment planning.

We would be concentrating on Web Enable Services of this neonatal disease diagnosis decision support system which should contain comprehensive knowledge base as well as more up-to-date inference mechanism. Internet based-GUI DSS architecture may be modeled with a capability of showing analytical reports instantly. Even there will be scope of interaction with national and regional level languages while using Web Enabled DSS for neonatal disease diagnosis.

Another broader scope is to implement of this DSS on mobile platform. There is absolute crisis of experienced domain experts and proper and basic establishments like computer peripheral and accessories in rural areas of India. Thus a handy, portable system which can be available in mobile and with a capability of using short messaging services to get the proper decision would be a great attempt in this field. Because a good numbers of people are using mobile phone in rural areas now a days. It will reduce the above mentioned establishment cost also.

Uncertainty is one of the import term which leads to incomplete and inaccurate decision making process. The future DSS may be developed of such kind of algorithm where there should have the tendency to ignore “If, Then..”, i.e. rule based approach. The problem may not be clearly, logically or mathematically solved al the times. If a neonate has a group of symptoms, then she or he has a particular disease. But many of these symptoms are shared by other diseases. Occasionally, some of these symptoms may also be absent for the disease. The real world presents the problem of uncertainty. Most of the DSS is using forward and backward chaining to search condition for reaching the goal. DSS faced problems with huge databases. In theory, the search process took twice the time for each newly added symptom. Suppose, to reach a prediction the search needs 1 second for a single sign/symptom, 2 seconds for 2 single signs/symptoms and so on. When there is 12 symptom it will take more than four hours to reach the prediction. But domain experts or specialists can take decisions with in fractions of seconds. They defiantly use a technique which eliminates unnecessary items in evaluating the disease. If a symptom is present, all diseases, which positively do not present the symptom, are eliminated. If a symptom is absent, all diseases, which positively present the symptom are eliminated. This process can swiftly evaluate any reasonable database and effectively handle uncertainty. So, future scope of this system may be of handling the uncertainty by using fuzzy logic and rule elimination process.

There is vast scope of using hybridization of this DSS with Robotics. The days are coming where robots can even be replaced with domain experts. Future process that we will be exploring is that how we can incorporates machine intelligence and DSS system into robots. There will be a possibility of shifting simple DSS to man and machine relationships with the help of robotics.

INDEX