uOttawa

L'Université canadienne
Canada's university

FACULTÉ DES ÉTUDES SUPÉRIEURES ET POSTOCTORALES

uOttawa

L'Université canadienne
Canada's university

FACULTY OF GRADUATE AND POSDOCTORAL STUDIES

Doaa Ibrahim-Swailum
AUTEUR DE LA THÈSE / AUTHOR OF THESIS

M.A.Sc. (Electrical Engineering)
GRADE / DEGREE

School of Information Technology and Engineering
FACULTÉ, ÉCOLE, DÉPARTEMENT / FACULTY, SCHOOL, DEPARTMENT

Identifying Risk Factors for Newborn Outcomes using Artificial Neural Networks
TITRE DE LA THÈSE / TITLE OF THESIS

M. Frize
DIRECTEUR (DIRECTRICE) DE LA THÈSE / THESIS SUPERVISOR

CO-DIRECTEUR (CO-DIRECTRICE) DE LA THÈSE / THESIS CO-SUPERVISOR

EXAMINATEURS (EXAMINATRICES) DE LA THÈSE / THESIS EXAMINERS

A. Chan

A. El Saddik

Gary W. Slater
LE DOYEN DE LA FACULTÉ DES ÉTUDES SUPÉRIEURES ET POSTDOCTORALES /
DEAN OF THE FACULTY OF GRADUATE AND POSTDOCORAL STUDIES

Identifying Risk Factors for Newborn Outcomes using Artificial Neural Networks

by

Doaa Ibrahim-Swailum,

M.Sc., Information and System Science, Carleton University, Canada

B. Eng., Biomedical Engineering, Cairo University, Egypt

Thesis submitted to the

Faculty of Graduate and Research

in partial fulfillment of the requirements for the degree of

Master of Applied Science

In Electrical Engineering

Ottawa-Carleton Institute of Electrical and Computer Engineering

School of Information Technology and Engineering

Faculty of Engineering

University of Ottawa

Library and
Archives Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 0-494-14914-0
Our file Notre référence
ISBN: 0-494-14914-0

NOTICE:
The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:
L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

Abstract

The goal of this thesis is to identify the risk factors for caesarean delivery, neonatal mortality and low Apgar score using Artificial Neural Networks (ANNs). The medical domain of interest used is the perinatal database provided by the Perinatal Partnership Program of Eastern and Southeastern Ontario (PPPESO). The ability of the ANNs to generate strong predictive model with the most influential variables was tested. Different ANN techniques for weight extraction and determining the importance of each input variables were applied. The thesis used feedforward ANNs trained by the backpropagation algorithm, as this is a widely used ANN in medical applications. Finally, minimal sets of variables (risk factors) that are important in predicting each outcome without degrading the ANN performance were identified.

In loving memory of my dad, Shawky Ibrahim,
who believed in me and in my ability to succeed,

To my great lovely mom, Samia Khalil, who taught me the secret of
life and encouraged me to challenge myself.

To my husband, Sherif Radwan, for helping me to keep a focus on my
study and work.

Acknowledgments

I would like to acknowledge the support of my thesis supervisor, Dr. Monique Frize, for guiding me through the work of my thesis. Thank you for believing in me from the beginning and helping me getting my first Canadian degree in Information and System Science, Carleton University, 2002. Thank you for your continuous support and guidance throughout this thesis work.

Many thanks to my medical supervisors: Dr. Robin Walker and Dr. Paula Stewart. I appreciated your feedback, helpful discussions and encouragement throughout the duration of my thesis.

I thank Colleen Ennett in our MIRG research group, who, though she was working on her PhD, was always available to provide help and suggestions.

Nomenclature

ANN	Artificial neural network
Apgar1	Apgar score at 1 minute of birth
Apgar5	Apgar score at 5 minutes of birth
ASE	Average Squared Error
CCR	Correct Classification Rate
CP	Constant Predictor
C/S	Caesarean section
err_ratio	Error ratio
lambda (λ)	Weight elimination constant
lambda_inc	Weight elimination constant increment
lambda_dec	Weight elimination constant decrement
lr	Learn rate
lr_inc	Learn rate increment
lr_dec	Learn rate decrement
m	Momentum constant
MIRG	Medical Information-technology Research Group, University of Ottawa and Carleton University.
MLP	Multilayer perceptron

Niday-2001	Perinatal 2001 database provided by PPPESO
PPPESO	Perinatal Partnership Program of Eastern and Southeastern Ontario
ROC	Receiver Operating Characteristic curve
SSE	Sum of Squared Errors
$\mathbf{w_0}$	Weight scale factor

Tables of Contents

1 Introduction

In the current 'information age' there is a belief that analysis of stored information in any field can help decision making and prediction of new facts. Analysing collected data is like using past experiences to learn how to deal with current problems and avoid making the same mistakes.

In many fields, the amount of information being collected and stored increases at an alarming rate. Although there is an increase in the information retrieval and analysis techniques, there is a lot of information being lost in storage. In order to decrease the amount of information lost, the right techniques must be used to search for important information and patterns in the databases. Also, there is a growing demand for studies and researchers with the skills to analyse and extract information from databases. One of the fields that could benefit the most from information analysis is the medical field.

After considering the importance of information extraction and use, the Perinatal Partnership Program of Eastern and Southeastern Ontario (PPPESO) began collecting perinatal data 25 years ago. The perinatal period is traditionally defined as the period of time during pregnancy, birth, and the first month after birth, but has grown to include the preconception period, with the recognition that health status prior to pregnancy also influences maternal and newborn outcomes [PPPESO]

The PPPESO works with hospitals, health departments, community agencies, academic institutions, private practitioners, and consumers to effectively link perinatal care, education, and research. The PPPESO works with its partners to identify issues, develop and implement solutions and produce results that will improve evidence-based regionalised perinatal care for childbearing families in Eastern and Southeastern Ontario. PPPESO's role is to provide support for this process among the partners by collecting and analysing data, dissemination of information, communication, advice, facilitation and education (Figure 1.1).

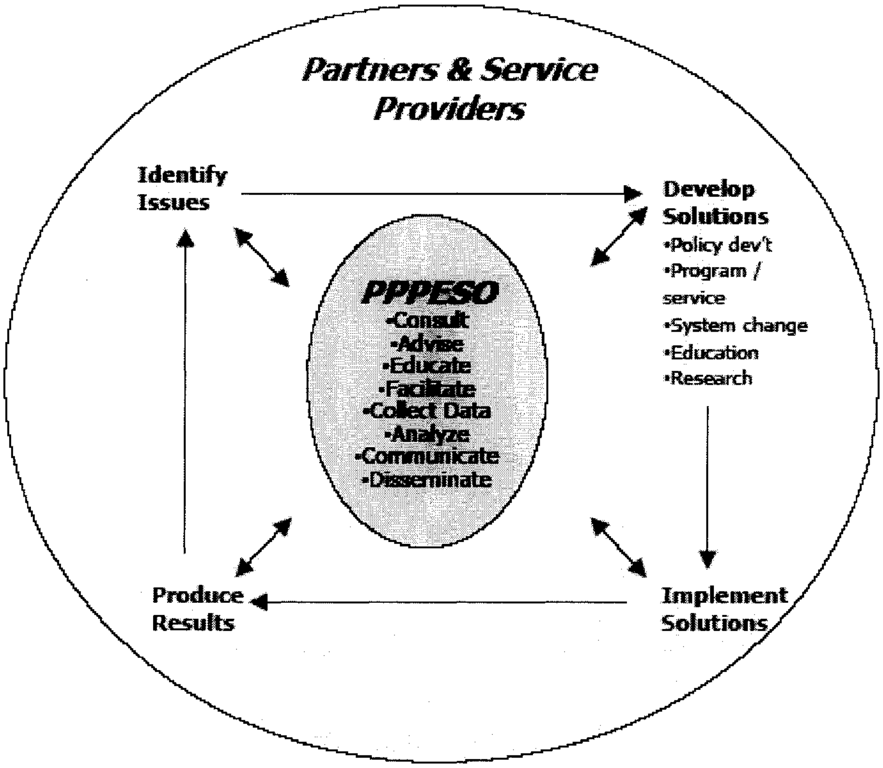
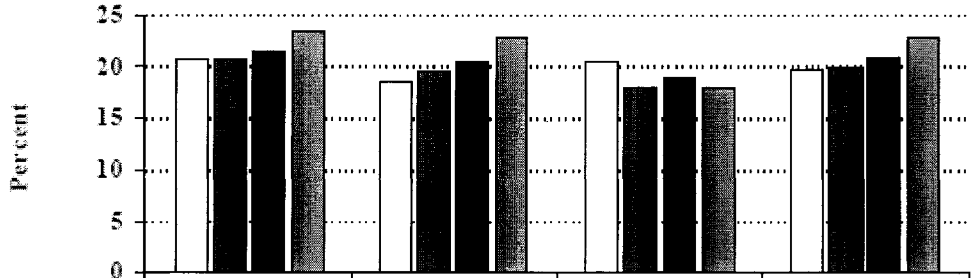


Figure 1.1: Partners and service providers of the PPPESO [PPPESO 2005].

In June of 1999, agencies involved in perinatal care in the region signed a formal partnership agreement to work together in an interdisciplinary approach and to share resources and data. In 2001, 17,406 women gave birth in hospitals in the Eastern and Southeastern Region with an increase of 2.65% compared to the year 2000. The information on these women and babies born was collected in the 2001 Niday Enhanced Perinatal database (Perinatal Niday-2001).

In this thesis, three clinical outcomes for newborns are predicted using the Perinatal Niday-2001 database: a) delivery type (vaginal or caesarean); b) neonatal mortality (probability of the newborn dying within 28 days after birth); c) Apgar5 (Apgar score at 5 minutes after birth). Finally, data sets from the 1999 Niday Enhanced Perinatal database (Perinatal Niday-1999) are used for testing the prediction models built using the Perinatal Niday-2001 database.

Successfully predicting the delivery type can help to improve the medical service that is offered to around 17,000 mothers during delivery. It is important to study the delivery type as the caesarean birth rate continued to increase in 2001 to 22.8% from 20.9% in 2000 and 19.9% in 1999. This increase was apparent in both the teaching and large community hospitals (see Figure 1.2). Debate continues on whether the present caesarean birth rate is too high. Six years ago the caesarean birth rate in Ottawa-Carleton hospitals was 16% [PPPESO 2002].



	Teaching	Large Community	Small Community	Region
□1998	20.7	18.5	20.5	19.7
■1999	20.7	19.5	17.9	19.9
■2000	21.5	20.5	18.9	20.9
▨2001	23.5	22.9	18.0	22.8

Figure 2.1: Proportion of babies who had a caesarean birth by hospital type, Eastern and Southeastern Ontario, 1998-2001 [PPPESO 2002].

Reduction in infant and child mortality is a major goal of the WHO (World Health Organization). During the last quarter of the century emphasis has been placed on reducing child mortality largely through immunization, oral rehydration and control of acute respiratory infections. Consequently, deaths among children over one month of age have sharply declined in the last three decades. These changes, however, did not have a marked impact among neonates (deaths within 28 days of birth), leading to the shifting of infant mortality at early days [James et al. 2000]. Babies who are born early and of low birth weight (less than 2500 grams) or who have congenital anomalies are more likely to die in the first month of life. In Eastern and Southeastern Ontario the death rate has been steady over the last four years. According to the PPPESO annual report 6 out of 1000 newborns died in the region [PPPESO 2002]. Predicting the risk factors of neonatal mortality using the Perinatal Niday-2001 will be a first step in reducing mortality in the region.

The third outcome studied is the Apgar score. It is measured at one minute and five minutes after delivery. The Apgar test is considered to be the first test for the newborn and it is used to quickly evaluate the newborn's condition after delivery. It tests five qualities: Appearance (color); Pulse (heartbeat); Grimace (reflex); Activity (muscle tone); and Respiration (breathing). A score is determined by awarding zero, one or two points in each category. Generally, the higher the score the better the baby's condition. Scores of seven and over indicate the baby is in good condition. This valuable tool was developed in 1952 by the late pediatrician, Dr. Virginia Apgar [Apgar 1952, Apgar 1953, Apgar et al. 1958].

Since the publication of the Rumelhart et al.'s paper in 1986, neural network techniques have been used in different applications. Researchers in many fields, including medicine, have investigated whether ANNs might be able to assist them in analysing large and complex sets of medical data. Because of their nonlinear modeling capabilities, ANNs are widely used for nonlinear modeling. When applied to the problem of medical outcome estimation, ANNs are most often used to identify patterns in large medical databases. The MIRG (Medical Information-technology Research Group), led by Dr. Monique Frize and Dr. Robin Walker has demonstrated in several articles that artificial neural networks (ANNs) are a useful tool to estimate outcomes with large medical databases [Ennett et al. 2001, Frize et al. 2001b, Walker et al. 2003]. This study's approach uses feedforward ANNs trained with backpropagation learning algorithm to estimate the three outcomes selected. Ultimately, the goal of this work is to identify the most influential variables for each of the three outcomes.

Many studies use ANNs to predict medical outcomes for adult patients, but there are only a few studies using ANNs for perinatal and newborns outcomes [McDowell et al. 2001]. The small number of available studies deals mainly with predicting mortality and Length of Stay

(LOS) in pre-term infants [Sargent 2001, Ambalavanan & Carlo 2001, Zernikow et al. 1998, Frize et al. 2004]. This thesis involves ANNs in the estimation of an additional number of the perinatal outcomes. Some of these outcomes such as Apgar score have never been analysed by ANNs.

1.1 Problem Statement

The main problem of this thesis is to find the risk factors for three perinatal outcomes: delivery type, neonatal mortality, and Apgar5. These risk factors can be used by researchers and PPPESO program to improve the perinatal heath care offered to patient (mother and newborn) in Eastern and Southeastern Ontario. Predicting delivery type can help in deciding whether to perform or not perform caeserean sections (C/S). Finding neonatal mortality (mortality) risk factors can help in the diagnosis of critical cases and sick newborns and potentially reduce the number of deaths. Predicting Apgar score can help in assessing the health status of a newborn baby and the health care this baby needs. Our neural networks were designed to achieve these goals. Neural networks are sophisticated techniques capable of modeling complex nonlinear functions with a large number of variables. Because the Perinatal Niday-2001 database, is a huge database with a large number of variables, neural networks were a good choice for analysing this data. The Perinatal Niday-1999 datasets which not presented in the training sets are used as test sets for the ANN models.

1.2 Thesis Objectives and Motivations

The main objective of this thesis is to explore the use of ANN techniques in developing prediction models that identify the most influential input variables for three perinatal outcomes: delivery type, mortality and Apgar5.

Other objectives of this work can be broken down into the following points:

1. Helping the medical professional in the diagnosis of new patient cases (mother or newborn) and in a selection of treatment based on information from previous similar cases.
2. Hoping to increase the survival of newborns and improve the care of the mother.
3. Improving the use of information and the availability of permanent, understandable and reliable perinatal information, especially in the region of eastern and southeastern Ontario.
4. Supporting decision making in perinatal care.
5. Continuing the development of Clinical Decision Support Systems implemented by the MIRG group.

1.3 Thesis Outline

This thesis is organized in five chapters. In chapter two: the Literature Review describes the background of concepts used in this thesis. There is a detailed description of the medical environmental and the database (Perinatal Niday-2001) being investigated, followed by a description of the ANN techniques used for estimating the medical outcomes being studied. It focuses on the measure of performance, generalisation methods, weight-elimination and weight extraction concepts used in this thesis. Lastly, a review of the earlier work by the MIRG research group is illustrated. There is a brief description of the different ANN techniques that other researchers have applied to estimate different medical outcomes. This discussion explains the choice of ANN techniques for the research carried out in the thesis.

Chapter three: Methodology: It discuses the problem statement of the thesis. The first part begins by an overview of the Perinatal Niday-2001 and Perinatal Niday-1999 databases, including their statistical description, and the preprocessing and preparation techniques applied to them. It also describes the choice of the training, validation and test sets. In the second part of this chapter, there is a detailed description of the ANN method applied to this work. It describes different ANN architectures and the ANN techniques used to improve the prediction models. Furthermore, it describes the weight extraction techniques that were employed for identifying the most important predicting variables.

Chapter four: Simulations and Results: First, the MIRG ANN system is evaluated by comparing its results with the Fuzzy logic system of the Coventry Group. Second, the results of each of the three outcomes under study are summarized and an analysis and discussion of the findings of the thesis is presented. The results of the experiments are presented in several steps to show important information found in each step. Finally, this chapter concludes with a discussion of the results and findings.

Chapter five: Conclusions derived from this work and suggestions for future work are presented.

2 Literature Review

2.1 The PPPESO Database

2.1.1 History

Under the direction of the Perinatal Partnership Program of Eastern and Southeastern Ontario (PPPESO), a regional database was created in 1997 to provide perinatal data to PPPESO partners. The database has evolved significantly since its inception, and has become a unique co-operative venture with the 37 partners of PPPESO, consisting of: hospitals, health units, Health Information Partnership – Eastern Ontario (HIP), community health centres, and Ottawa and Queen's Universities, the Ministry of Health and Longterm Care (MoHLTC), and CritiCall (CC). The database was modeled after the existing Ottawa-Carleton database. The Niday Perinatal Database (formerly known as the Eastern and Southeastern Ontario Perinatal Database) has been developed in collaboration with Healthy Babies Healthy Children to ensure that the systems are complementary and to minimize duplication of effort [PPPESO].

The HIP provided funding for the initial development of the database. In 2000, HIP became a partner organization and has provided additional support for the analysis and preparation of reports from the database.

Beginning in January 1^{st}, 2001, the database was enhanced from a stand-alone computer program installed in each hospital, to the web-based "CritiCall Ontario" system. Hospitals now enter data directly into the database and can generate reports independently. The Ministry of Health and Long-term Care, along with the database's hospital and health unit partners, provided funds for the project. These changes have greatly enhanced its usefulness by improving the reporting of data and the retrieval of information [PPPESO].

2.1.2 Description of the Perinatal Databases

Each hospital in the partnership program PPPESO used a computerized data entry program to record information about each baby born in their hospital. Two hospital sites with their own perinatal database contributed a data file with the required variables. At the end of each quarter, the data was forwarded to the PPPESO data analyst. It was checked for consistency and corrections were made with the assistance of the hospital staff [PPPESO 2002].

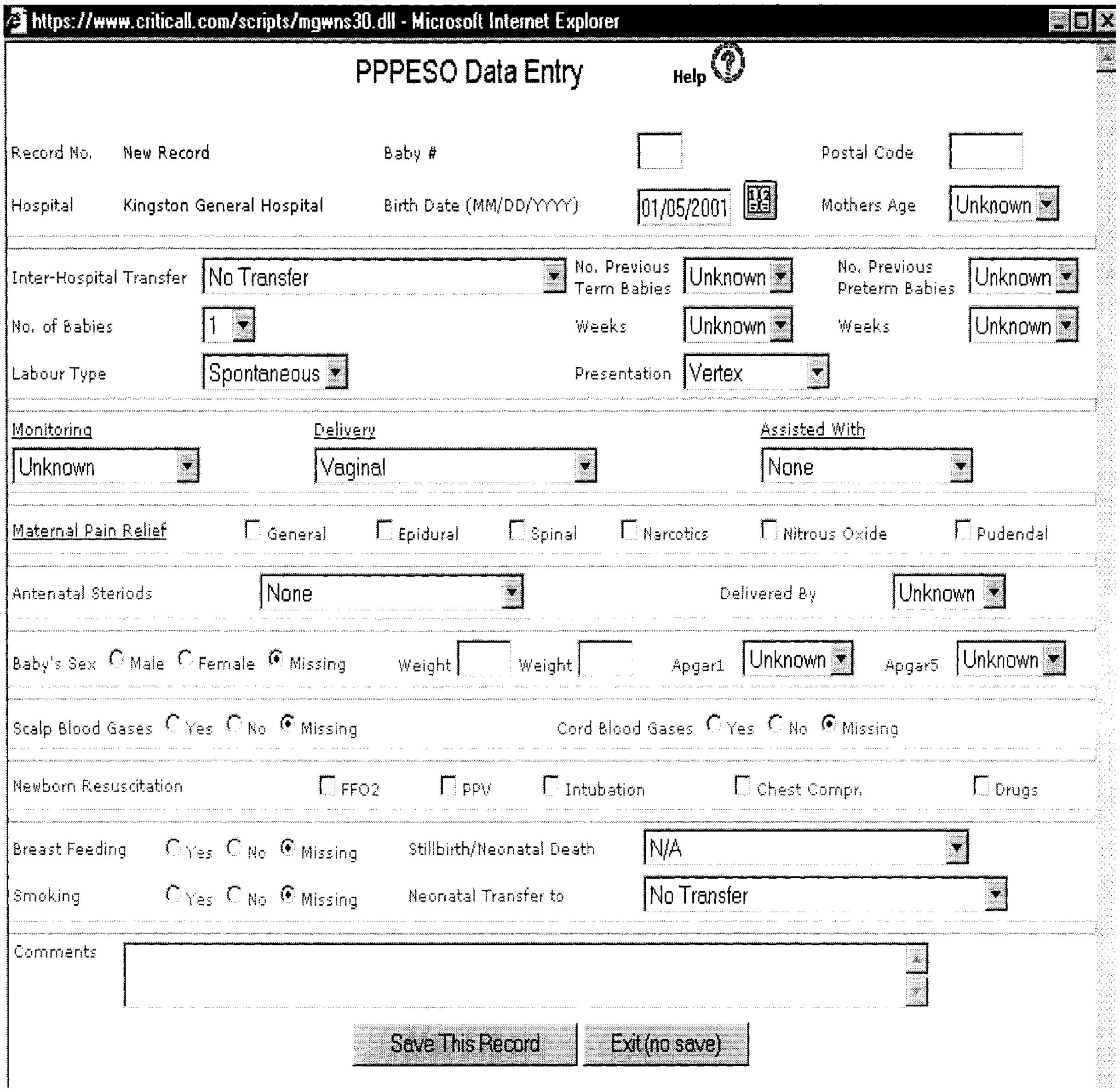
The Perinatal databases contained the following main variables:

- *Birth Date*: indicates the baby's date of birth
- *Postal Code*: the mother's complete postal code
- *Mother's Age*: the mother's age in years (the Range is 12-55 years)
- *Inter-Hospital Transfer*: indicates where the woman was transferred from
- *Previous Term Babies*: number of previous term babies for the mother
- *Previous Preterm Babies*: number of previous preterm babies for the mother
- *Number of Babies*: indicates the number of babies (1, 2, 3, 4, 5, 6, > 6) in this pregnancy.
- *Weeks*: indicates the number of completed weeks (gestational age)
- *Labour Type*: indicates the type of labour (Spontaneous, Induced, No Labour)
- *Presentation*: indicates the type of presentation (Vertex, Breech, Other, Not Available)
- *Monitoring*: indicates which monitoring methods were used (Unknown, No Monitoring, Auscultation only, Electronic only, Auscultation and Electronic)

- *Delivery*: indicates the type of delivery (Vaginal, VBAC, Primary C/S, Repeat with Trial of Labour, Repeat No Trial Labour)
- *Assisted with*: indicates whether the following were utilized to assist either a vaginal or cesarean birth: None, Forceps, Vacuum, Forceps and Vacuum.
- *Maternal Pain Relief*: indicates all methods of anesthesia and/or pain relief utilized for this birth (General, Epidural, Spinal, Narcotics, Nitrous Oxide, Pudendal (for 2001 database))
- *Antenatal Steroids*: indicates whether antenatal steroids were administered
- *Delivered By (for 2001 database)*: indicates who delivered the infant (Unknown, Physician, Midwife)
- *Baby's Sex*: indicates whether the baby is male or female
- *Weight*: indicates the weight of the baby in grams
- *Apgar 1, Apgar 5*: indicates the Apgar scores at 1 and 5 minutes (the range is 0 to 10)
- *Scalp Blood Gases*: indicates whether scalp blood gases were done
- *Cord Blood Gases*: indicates whether umbilical cord blood gases were done
- *Newborn Resuscitation*: indicates all interventions that were utilized, regardless of duration such as FF02, PPV, Intubation, Chest Compression, Drugs
- *Breastfeeding: indicates* whether the woman intends to breastfeed
- *Smoking: indicates* whether any maternal smoking has occurred at any time after 20 weeks gestation
- *Stillbirth/Neonatal Death:* indicates when and where the death occurred
- *Neonatal Transfer*: includes infants transferred during the first 7 days of life only

Refer to Appendix A for the complete list of variables of the enhanced database computer entry program [PPPESO 2001].

The PPPESO data entry form (shown in Figure 2.1) which had been used to enter the data into the Perinatal databases.



PPPESO Data Entry

Record No. New Record
Baby #
Postal Code
Hospital Kingston General Hospital
Birth Date (MM/DD/YYYY) 01/05/2001
Mothers Age Unknown

Inter-Hospital Transfer No Transfer
No. Previous Term Babies Unknown
No. Previous Preterm Babies Unknown
No. of Babies 1
Weeks Unknown
Weeks Unknown
Labour Type Spontaneous
Presentation Vertex

Monitoring Unknown
Delivery Vaginal
Assisted With None

Maternal Pain Relief ☐ General ☐ Epidural ☐ Spinal ☐ Narcotics ☐ Nitrous Oxide ☐ Pudendal

Antenatal Steriods None
Delivered By Unknown

Baby's Sex ○ Male ○ Female ● Missing
Weight
Weight
Apgar1 Unknown
Apgar5 Unknown

Scalp Blood Gases ○ Yes ○ No ● Missing
Cord Blood Gases ○ Yes ○ No ● Missing

Newborn Resuscitation ☐ FFO2 ☐ PPV ☐ Intubation ☐ Chest Compr. ☐ Drugs

Breast Feeding ○ Yes ○ No ● Missing
Stillbirth/Neonatal Death N/A
Smoking ○ Yes ○ No ● Missing
Neonatal Transfer to No Transfer

Comments

Figure 2.1: PPPESO database entry screen [PPPESO 2001]

2.1.3 Data Quality

A user guide is available to ensure consistency of definitions for each variable for all the hospitals. In addition, each hospital has received training to manage the system data entry and reporting capabilities. The 2001 data was audited to assess concordance between the perinatal database and the case room logbook. In addition, the charts of one hundred women from each hospital were reviewed to assess concordance between the case room logbook and the chart. Results have found high concordance (less than 5% error) between the database and the logbook for all but four variables: electronic monitoring (6.6%), induction (7.0%), postal code (8.5%), and antenatal steroid use (>10%) [PPPESO 2001].

2.2 Perinatal Models in the Literature

To better understand the perinatal medical environment, a brief description of the medical environment and several models are discussed. Perinatal, obstetric and neonatal environments are connected to each other. Obstetric variables allow assessing the mothers' condition while neonatal variables collected asses babies' condition. The perinatal environment is situated between theses two environments.

The main question in the search is to identify the factors that most influence the outcomes. For delivery type, after reviewing existing models, using different databases, one discovers that no two models find the same dataset for statistically significant factors or ANN main factors. Several risk factors are either statistically found or clinically selected as important. Such clinically important factors are: cephalopelvic disproportion, maternal medical condition necessitating a cesarean section, arrest of labour, malpresentation of the baby, fetal distress, and failed induction. In other studies some more non clinical factors such as the mothers' views on birthing mode were also found to be important in determining the deliver type [McDowell et al. 2001, Burns 1995]. In 1990, another obstetric study found that risk factors influencing the decision to perform primary C/S delivery after vaginal delivery trial were abnormal labor progression, above 28 years of age, cervical dilatation less than 3 cm on admission, evening shift and birth weight more than 3,100 gm [Yosef 1990]. The challenge faced in this study was to determine which factors most influence the delivery type decision and whether there is a great difference using the Perinatal Niday-2001 database to detect these factors. Or in other words what are the C/S delivery risk factors?

The effective predictors in estimating neonatal mortality risk are birth weight, gestational age, gender, race, Apgar score and mode of delivery [Gray et al. 1992]. Apgar scores are used to quickly assess the risk of complication after delivery [Apgar 1953, Apgar et al. 1958]. In 1994, Stevens et al. reported that clinicians in the NICU tended to overestimate mortality risk for their patients, which impacts patient triage, transfer, initiation and termination of life support and allocation of medical resources. In their study, the physicians predicted mortality with a 90% sensitivity and a 68% specificity [Stevens et al. 1994]. In other words, the physicians accurately predicted death 9 out of 10 times, but over-predicted death in patients who survived about one-third of the time. This study demonstrates the clinicians' tendency to overestimate mortality (i.e. they predict death with a high sensitivity). A model that correctly classifies more patients who will survive (i.e. predicts survival with a high specificity) would be beneficial and complementary to the way that these experts think. Ideally, to ensure confidence in the model, a specificity of close to 100% and a sensitivity of at least 50% would be preferred. These overestimates of risk lead to more aggressive management and increase admission to NICU [Ennett 2003]

Several models are available for predicting mortality. CRIB (Clinical Risk Index for Babies) , a simple scoring system that uses six input variables including birth weight and gestational age to adjust the mortality risk score for the severity of illness (51% sensitivity, 95% specificity) [INN 1993]. Appendix B has details about the variables used and points scored for the Apgar, CRIB and NTISS scores. SNAP (Score for Neonatal Acute Physiology) was developed as a mortality risk model for the NICU using the same modelling technique [Richardson et al. 1993]. SNAP used 37 input variables to assess mortality risk with the points scored for each deviation. The more severe the condition, the higher the points scored. SNAP was also shown to be highly correlated with nursing workload, length of hospital stay and therapeutic intensity [Richardson et al. 1993, Ennett 2003]. This new score called SNAP-II is a modified SNAP, selected 6 of the original 37 variables from SNAP using logistic regression to predict in-hospital mortality, as listed in Table 2.1 (the first 6 variables). Birth weight, small for gestational age and Apgar score at 5 minutes were also included to take these important factors into consideration. This was called SNAPPE-II-SNAP-II with the Perinatal Extension. Excluded patients were moribund babies, those transferred within 24 hours to the regular nursery, those admitted after having been discharged or after more than 48 hours since birth. The time frame for data collection was reduced to the first 12 hours after admission to the NICU. The area under the receiver operating characteristic curves (ROC curves) for the SNAP-II score was 0.91±0.01 with the best classification performance for the group of infants of all birth weights and the worst performance for infants of birth weight less than 1500g [Ennett 2003]

Variable	Range	Points
Lowest blood pressure	MBP[a] 20-29 mmHg	9
	MBP[a] <20 mmHg	19
Lowest temperature	95-96 °F	8
	<95 °F	15
Lowest PO_2/FiO_2 ratio	1.0-2.49	5
	0.3-0.99	16
	<0.3	28
Lowest serum pH	7.10-7.19	7
	< 7.10	16
Multiple seizures	>1	19
Lowest urine output	0.1-0.9 mL/kg/h	5
	<0.1 mL/kg/h	18
Birth weight	750-999 g	10
	<750 g	17
Small for gestational age	< 3rd percentile	12
Apgar score at 5 minutes	< 7	18

[a] MBP= mean arterial blood pressure

Table 2.1: SNAPP-II-SNAP-II variable list and scoring system [Richardson et al. 2001]

In the case of Apger score, there is no predicting models or research for it in the medical environment. Apgar scores used to quickly assess the risk of complication after delivery [Apgar 1953, Apgar et al. 1958]. This score was always found to be one of the most important factors in predating mortality. It is interesting to predict this outcome, for the first time, which indirectly predicts mortality.

2.3 Basic Concepts of Artificial Neural Networks

The artificial neural network is an array of logic components that map an input to an output. These logical units work by combining independent variables (inputs) to a threshold and then producing a dependant variable (output). The functionality of the neural network resembles the operation of neural processes in the human brain. An ANN is like the human brain in that it is composed of a large number of highly interconnected simple processing elements, called nodes or neurons that operate in parallel [Rumelhart et al. 1986].

A multilayer feedforward network, otherwise known as a multilayer perceptron (MLP) consists of a set of neurons that are logically arranged into two or more layers. There is an input layer and an output layer, each containing at least one neuron. Neurons in the input layer are hypothetical in that they do not themselves have inputs, and do not do any processing. Their activation (output) is defined by the network input. The term *feedforward* means that information flows in one direction only. The input to neurons in each layer come exclusively from the outputs of neurons in the previous layers, and the outputs from theses neurons pass exclusively to neurons in following layers [Masters 1993].

Hidden layers are located between the input and output layers. One or more hidden layers may be used. An ANN with no hidden layers is a two-layer ANN. A network with one hidden layer, a three-layer ANN (Figure 2.2), can describe almost any function to any desired degree of accuracy, given that the function and its derivative are continuous, and the ANN uses sufficient hidden units [Fausett 1994, Penny & Frost 1996]. In general, the more neurons there are in hidden layers, the better the network can fit the data; however, when too many neurons are used, overfitting can occur. On the other hand, if training a network for a long time still results in large errors, the problem is most likely a lack of hidden layer neurons [Demuth & Beale 2001]. It has been observed that for the vast majority of practical problems, there is no reason to use more than one hidden layer. Problems that require two hidden layers are rarely encountered in real-life situations [Masters 1993].

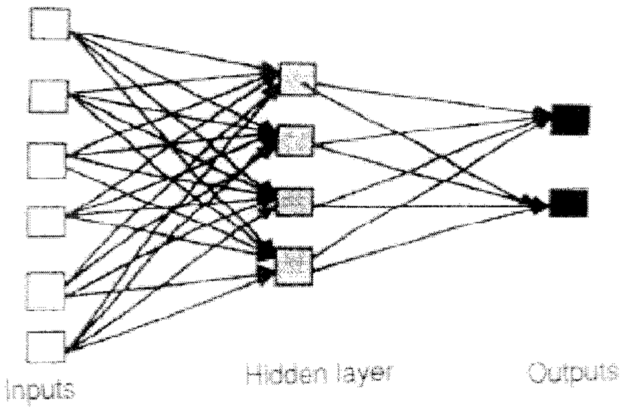


Figure 2.2: A three-layer ANN as an example of a multilayer feedforward network

The output of each neuron in the network is a function of that neuron's inputs. Given an input to such a network, the activation of all output-layer neurons can be computed in one deterministic pass; iterations are not required, and randomness does not play a role. [Masters 1993]

The code for the ANN used in this thesis was implemented in MATLAB (version 6.5). The McCulloch-Pitts model of an artificial neuron (Figure 2.3) and equation 2.1 show how an artificial neuron works. The weights w_i for the inputs p_i and the bias b_j are randomly initialized. W_n is the vector that represents the weights of the inputs to any layer. If n=2, then W_1 is the vector that represents the weights of the inputs to the second layer. While, W_2 represents the weights of the inputs of the third layer (if one hidden layer is used). W_1 was randomly initialized in the range of -1 and +1, while W_2 was randomly initialized between -0.1 and 0.1.

The initial weights for the bias terms were assigned using the same approach such that vector b_1 had random values ranging from -1 to 1 (output layer for two-layer network, hidden layer for three-layer network) and vector b_2 ranged between –0.1 and 0.1 (output layer of three-layer network). The bias vectors are in fact unit vectors. The transfer function of each node was the hyperbolic tangent. The output was calculated according to Equation 2.1 [Demuth & Beale 2001, Ennett 2003].

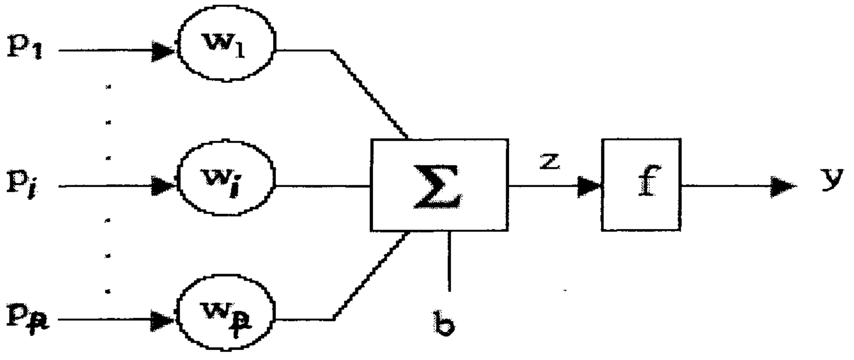


Figure 2.3: McCulloch-Pitts model of an artificial neuron (two-layer ANN)

$$A_n = \tanh\left(\sum_{i=1}^{p} W_n[i]A_{n-1}[i] + b_n[i]\right), n \geq 1 \qquad (2.1)$$

where n corresponds to the *n-th* layer, A_n corresponds to the output of the node of the *n-th* layer and p is the number of inputs to the current node.

The input matrix A_0 to the system represents the input variables of the dataset. A_0 is a matrix whose rows consist of database cases and whose columns contain variables regarding the patient's medical information and the outcomes under investigation. The system output vector A_n represents the network's prediction of a certain outcome. The output is compared with the true output and the error E_0 is the difference between these two values. The cut-off value for the network was typically 0, such that a value in the output vector that was greater than 0 classified the output as +1 type output and an output with a value less than 0 was classified as a –1 type output [Demuth & Beale 2001, Scales 2001, Ennett 2003].

Transfer Function

The sigmoid transfer function is commonly used in backpropagation networks. An example is the log-sigmoid (logistic) transfer function (Equation 2.2). It takes the input, which may have any value between plus and minus infinity, and compresses the outputs into the range 0 to 1. [Demuth & Beale 2001]

$$f(x) = \frac{1}{1+e^{-x}} \tag{2.2}$$

Alternatively, the multilayer networks may use the tan-sigmoid (hyperbolic tangent) transfer function which has a similar curve to the log-sigmoid transfer function but with an output range from -1 to +1 (Equation 2.3).

$$f(x) = \tanh(x) \tag{2.3}$$

The most important characteristic of a sigmoid activation function is its nonlinearity. In a multilayer network, the activation function must be nonlinear, or the computational power will be equivalent to a two-layer network (no hidden layer) and it will not be able to learn nonlinear mapping [Blum 1992]. In this thesis, the hyperbolic tangent transfer function was chosen because of its nonlinearity and its sharp sigmoidal transition between its output values. Like the log-sigmoid transfer function, the hyperbolic tangent and its derivative is continuous, differentiable and monotonically non-decreasing, which simplifies the training algorithm. The hyperbolic tangent (ranging from –1 to 1) offers a sharp sigmoidal transition at zero which results in very few output values near zero, so a nonzero output is more likely [Fausett 1994].

2.3.1 Backpropagation Algorithm

The main procedure in the learning process is to adjust the weight of each unit in the ANN in such a way that the error between the desired output and the actual output is reduced. The error derivative of the weight must be calculated to know how the error changes as each weight is slightly increased or decreased. These error derivatives are computed for all the units in one network layer, and then moves from layer to layer in a direction opposite to the way activities propagate through the network [Anderson 1995]. The backpropagation algorithm, also, known as the generalised delta rule, is based on the minimisation of the average error between the actual output signal and the corresponding desired output signal. This minimisation can be approximately expressed by the average squared error ASE in the training set. The re-development and popularisation of the backpropagation algorithm was done in 1986 by Rumelhart [Rumelhart et al. 1986].

D_2 is defined as proportional to the rate of change of the network output A_2 with respect to the error E and D_1 as is proportional to the rate of change of the output of the first layer A_1 with respect to D_2:

$$D_2 = \frac{dA_2}{dE} \bullet E = \left[\frac{d\tanh(E)}{dE}\right] \bullet E \tag{2.4}$$

$$D_1 = \frac{dA_1}{dD_2} \bullet W_2 \bullet D_2 = \left[\frac{d\tanh(D_2)}{dD_2}\right] \bullet W_2 \bullet D_2 \tag{2.5}$$

The change in weights can be calculated as follows:

$$W = W + dW \tag{2.6}$$

where dW can be found from:

$$dW_n = m * dW_n + (1 - m) * lr * D_n * A_{n-1}^T - 2\lambda \left[\frac{w_0^2 W_{n_{ij}}\left(1 + \frac{W_{n_{ij}}^2}{w_0^2}\right) - W_{n_{ij}}^3}{w_0^4\left(1 + \frac{W_{n_{ij}}^2}{w_0^2}\right)^2}\right] \tag{2.7}$$

Equation (2.7) represents the rate of change of the weights with respect to time. The first term in the equation includes the momentum term m that helps prevent the network from getting caught in local minima. The second term is proportional to the rate of change of the error with respect to the output D_n and includes the learning rate term lr.

Similarly, the bias vectors can be updated by:

$$B = B + dB \tag{2.8}$$

where *dB* can be found from:

$$dB_n = m^* \, dB_n + 1 - 2\lambda \left[\frac{w_0^2 B_{n_{ij}} \left(1 + \frac{B_{n_{ij}}^2}{w_0^2} \right) - B_{n_{ij}}^3}{w_0^4 \left(1 + \frac{B_{n_{ij}}^2}{w_0^2} \right)^2} \right] \tag{2.9}$$

After each training epoch, Equations (2.7) and (2.9) update the weights and bias vectors. To monitor the classification performance of the test set, the measures of performance are calculated for the test set after each training epoch as well [Ennett 2003].

The main drawback of the backpropagation algorithm is the possibility of finding a local minimum rather than the global minimum [Rumelhart et al. 1986b]. Training the network several times with different initial weight settings may be done to overcome this problem [Penny & Frost 1996].

2.3.2 Adaptive Networks

The adaptive networks are the networks which are able to change their weights, i.e. dW/dt not equal 0. This thesis uses adaptive networks and discusses the methods used for their learning.

In order to develop the best ANN model, nine adjustable parameters have to be accurately tuned. These nine parameters are: error ratio err_*ratio*, learning rate *lr*, learning rate increment *lr_inc*, learning rate decrement *lr_dec*, momentum constant *m*, the weight decay constant λ (or lambda), the weight-decay constant increment λ*_inc*, the weight-decay constant decrement

λ_dec and weight scale w_0. The user defined *err_ratio* determines how *lr*, m and λ are updated. These parameters may be incremented or decremented depending on the most recent change in error. The change in the *lr* is controlled by *lr_inc* and *lr_dec*. Similarly, λ_inc and λ_dec are responsible for the appropriate changes in λ. The network performance is very sensitive to these parameter settings.

The following algorithm from the MIRG ANN software shows the adaptive behavior:

```
if TSSE>SSE*err_ratio
   lr=lr*lr_dec;
   λ=λ*λ_dec;
   m=0;
else if TSSE<SSE*err_ratio
   lr=lr*lr_inc;
      λ=λ*λ_inc;
end
```

Where, TSSE refers to the training set errors from the current epoch and, SSE is the error of the previous epoch [Ennett 2003].

As mentioned above, *err_ratio*, *lr*, λ and m are among the user-defined nine variables that have to be tuned to get the best network performance. The *err_ratio* is used to check and observe the error during the training process. It checks if the current epoch error is less than the previous one or not. *err_ratio* judges the improvement (decreasing) in the error, which directly affects the improvement of the network. Next step is decided by adapting *lr*, λ and m for the next epoch. The previous process has to be repeated to ensure the error is decreasing and the network performance is increasing. Typical values of *err_ratio* range from 1.01 to 1.15 to allow for some minor lapses in performance.

The performance of the training algorithm is very sensitive to the proper setting of the learning rate *lr*. If the learning rate is set too high, the algorithm may oscillate and become unstable. If the learning rate is too low, the algorithm will take too long to converge. If the iteration ratio of the current epoch to the pervious epoch is less than the user-defined ratio, the decision is taken to increase the *lr* of the next iteration. It means that the algorithm is going in the right direction. But, if the error ratio is larger than the defined ratio, then the learning procedure must slow down. The learn rate *lr* will decrease by the value of the learning rate decrement *lr_dec*. It is important that the *lr* slows down to prevent the algorithm from oscillating around the global minimum without converging. Care must be taken to ensure that the learning rate is neither too large (causing instability and oscillations) nor too small (resulting in slow convergence) [Demuth & Beale 2001]. It is not practical to determine the optimal setting for the learning rate before training, and, in fact, the learning rate changes during the training process as the algorithm moves across the performance surface.

Momentum allows a network to respond not only to the local gradient, but also to recent trends in the error surface. Acting like a low-pass filter, momentum allows the network to ignore small features in the error surface. Without momentum a network may get stuck in a shallow local minimum. With momentum, a network can slide through such a minimum. The backpropagation algorithm incorporates a momentum term in which a ratio of the previous weight change value is added to the current value [MATLAB]. Adding such momentum to the algorithm through weight space may save it from becoming trapped in local minima.

Momentum can be added to backpropagation learning by making weight changes equal to the sum of a fraction of the last weight change and the new change suggested by the backpropagation rule. The magnitude of the effect that the last weight change is allowed to be mediated by a momentum constant which can be any number between 0 and 1. When the momentum constant is 0, a weight change is based solely on the gradient. When the momentum constant is 1, the new weight change is set to equal the last weight change and the gradient is simply ignored. The gradient is computed by summing the gradients calculated at each training example, and the weights and biases are only updated after all training examples have been presented.

As with momentum, if the new error ratio exceeds the old error by more than a predefined error ratio, the new weights and biases are discarded ($m=0$). In addition, the learning rate is decreased. Otherwise, the new weights are kept. If the new error is less than the old error, the learning rate is increased. This procedure increases the learning rate, but only to the extent that the network can learn without large error increases. When a larger learning rate could result in stable learning, the learning rate is increased. When the learning rate is too high to guarantee a decrease in error, it gets decreased until stable learning resumes [MATLAB].

2.3.3 Network Structure and Hidden Layers

To define the network structure, we need to decide the number of hidden layers and how many nodes will be in each layer. The number of neurons in a hidden layer affects the complexity and the performance of the network. If there are too few nodes, there will not be sufficient resources to solve the problem and this will result in underfitting results. On the other hand, too many nodes may cause long training times, and possibly memorization rather than generalization (overfitting) [MATLAB].

Nonlinearity

Not every medical data set contains nonlinearities, but many do. Neural networks will offer better results than linear methods only if there are nonlinearities in the data. If a network with no hidden units performs as well as a network with a large number of hidden units, then the problem appears to be linear [Penny & Frost 1996].

The majority of clinical studies use networks with one hidden layer, as these are sufficient to define the nonlinearity in the data.

There are many theories that roughly compute the number of the hidden units in the hidden layer. The number of neurons in each layer follows a geometric progression. For example, if we have a three-layer network with *i* input neurons and *o* output neurons, the number of the hidden nodes *d* in the hidden layer would be calculated using Masters' method (Equation 2.10) [Masters 1993].

$$d = \sqrt{i * o} \tag{2.10}$$

Thus, if there are 25 input neurons and 1 output neuron, the network should have at least 5 hidden nodes in the hidden layer.

Other methods exist for selecting the number of the hidden nodes of an ANN such as: using 75% of the number of the input nodes, using 50% of the number of the input nodes, or using 2n+1 hidden nodes, where n is the number of input nodes in the input layer.

Another method obtains the number of the nodes indirectly by calculating the number of the weights in the network, and thus they can get the number of the hidden units in the hidden layer to the number of the training cases. Livingstone and Manallack's empirical work [Livingstone & Manallack 1993], for example, suggests that

$$d = \frac{m*o}{w} \tag{2.11}$$

where m is the number of training cases, o is the number of network outputs, and w is the total number of network weights, be greater than 3 ($d>3$) to ensure good generalization and avoid memorization of the training set.

Thus, if there are 360 training cases and a single network output, the network should not have more than 120 weights. In a network with 9 inputs this corresponds to having a single hidden layer with 12 units.

It is obvious that the last two methods will produce more hidden nodes in the hidden layer than other mentioned methods. One of these methods depends on the number of the input nodes (2n +1) which will produce a very large number if it applies to the Perinatal Niday-2001 database. Also, Manallack's method, which depends on the size of the training sets, will produce a very large number if it applies to any of the datasets in this thesis (the average of the training sets is 10,000 cases). In general, the analysis of large-sized databases, which covers most of the training cases, will need less hidden nodes (less complex ANN) than the analysis of small-sized database. Our goal is to find the simplest ANN architecture which has the best performance. Therefore, the last two methods are not suitable for the selection of the best

ANN architectures in this thesis. Other methods are chosen in this thesis: Masters, or 50% or 75% of the input nodes in the input layers.

2.3.4 Generalisation

The next step after training a network is generalisation. Generalisation ability of a network is a measure of its performance on data not presented in the training set [Penny & Frost 1996]. In order to discover the ability of a network to analyse a new data, the original data set has to be partitioned from the beginning into two independent sets; a training set and a generalisation set. A generalisation set is an independent set that tests the performance of the network, which does not always reflect the performance achieved on the training set for the same network.

Unfortunately, neural networks may learn features in the training set that are not present in the wider population of cases. This is known as fitting to noise or overfitting. Overfitting will occur if the complexity of the network is greater than the complexity of the function being estimated [MATLAB]. There is a relation between the network structure and generalisation performance.

2.3.5 Weight-elimination and Weight-decay

Weight-elimination and a special case of this method known as "weight-decay", have been found experimentally to be effective in improving the generalisation of various feedforward networks [Weigend et al. 1990, Korgh & Hertz 1992]. Korgh and Hertz in 1991 demonstrated that the weight-decay methodology creates a new hybrid cost function, *E(W)*, by adding a penalty term to one's best error measure, *Eo* [Trigg 1997] .

$$E(W) = \underset{0}{E}(W) + \frac{1}{2}\lambda \sum w_{ij}^{2} \qquad (2.12)$$

where, *W* represents the weights vector, λ is a constant and w_{ij} represents the individual weights of the network.

In Equation 2.12, the penalty term in weight-decay penalises large weights and the constant, λ, is an adjustable parameter of the network which is used to control how strongly large weights are penalized. The main benefit of adding a term to the cost function in Equation 2.12 is that the weights will converge to smaller absolute values than they would otherwise [Korgh & Hertz 1992].

On the other hand, weight-elimination can be calculated with the following hybrid cost function, *E(W)*:

$$E(W) = \sum_{i=1}^{p} (E_0(W))^2 + \lambda \sum_{ij} \frac{\frac{w_{ij}^{2}}{w_0^{2}}}{1 + \frac{w_{ij}^{2}}{w_0^{2}}} \qquad (2.13)$$

The penalty term added to *Eo* in this case measures the size of the ANN. The weight decay constant, λ, is used to control how strongly the weights are penalized. The difference between the two equations is w_o, a scale parameter used to define large and small weight values [Weigend et al. 1991].

The main difference between weight-decay and weight-elimination is that weight-decay tends to shrink the large coefficients more than the small ones using the sum of squared weights while weight-elimination tends to shrink the small coefficients more than the large ones [Sarle 2005].

Actually, weight-decay is contained within the weight-elimination cost function as the special case of a large w_o. The scale parameter w_o enables an ANN developer to express a preference for fewer large weights (w_o small) or many small weights (w_o large). For $|w_{ij}|>>w_o$, the cost of a weight approaches unity multiplied by λ while for $|w_{ij}|<<w_o$, the cost function of a weight is very close to zero. [Weigend et al. 1990, Trigg 1997] .The scale parameter, w_o, is a user defined parameter which defines the sizes of large and small weights. Choosing w_o as a small value will force the small weights to zero resulting in fewer large weights (weight-elimination). On the other hand, choosing a large value for w_o will cause many small weights to remain, and limits the size of large weights. [Weigend et al. 1991, Ennett 2003]

Large values of λ mean that a weight must be closer to zero to be consider a part of the noise distribution. So, larger values of λ enforce the smaller weight to further reduce their size. When dealing with λ one must be very carefully since the network is sensitive to its value. It is recommended to start the network training with λ at zero, so that the ANN could initially take advantage of all input variables, and then slowly increase the value of λ until the network's performance begins to decline. [Weigend et al. 1990, Ennett 2003]

2.4 Early Work by MIRG Group

Former MIRG graduate student Tim Buskard [Buskard et al. 1994] described ANNs, which were used to estimate "Mortality", "Length of Stay" and "Duration of Artificial Ventilation" for 1322 patients of Doctor Everett Charmers Hospital, Intensive Care Unit (DECH ICU) database with best test set correct classification rates (CCRs) of 89.4%, 66.1% and 69%, respectively. Compared to the performance that could have been obtained using a constant predictor (CP) for each of these outcomes (CCRs of 88.8%, 61% and 66.5%, respectively), Buskard's networks showed a marginal level of improvement [Trigg 1997].

In 1997, Trigg greatly improved the database previously used by Buskard and obtained excellent results predicting the same outcomes. She used the DECH ICU database, with additional cases (1491 instead of 1322 cases). Weight-elimination and high/low node representation approaches were used to classify the data. The goal of the high/low node data presentation technique was to see if classifying the parameter as higher-than-normal, normal, or lower-than-normal would improve the classification accuracy [Trigg 1997]. This idea was based on the fact that several physiological parameters have different health effects depending on these characteristics [Frize et al. 1996, Frize et al. 1997]. Trigg used the weight-elimination to prevent overfitting by removing irrelevant weights. She compared the performance of the ANNs using standard backpropagation and backpropagation with weight-elimination cost function developed by Weigend [Weigend et al. 1990, Weigend et al. 1991] using Tong's [Tong 1983] sunspot data. The results confirmed that both algorithms were working properly [Trigg 1997, Frize et al. 1997]. Trigg divided the ICU database into two databases, medical and surgical ICU patients. Since the surgical POSTOP (postoperative) patients have more common physiological characteristics than medical NON_POSTOP (non-postoperative) patients, she chose to analyze the POSTOP database (883 patients). Working with 51 input variables and the "Duration of Ventilation" outcome, specificity, less than 8 hours of artificial ventilation, and greater than or equal to 8 hours of artificial ventilation (VENT8 outcome), she discovered that the three-layer ANNs improved the performance. The optimal network performance occurred using a three-layer ANN with weight-elimination (CCR = 91.8%). Compared to the minimum distance classifier (CCR= 86.1%) and the CP (CCR = 71%). Although, Trigg discovered that the high/low node representation did not improve the CCR of the ANN, she found that the three-layer ANNs slightly improved the performance of these experiments. Also, the combination of the high/low node representation and weight-elimination techniques with a double-layer network allowed the most important variable for predicting artificial ventilation of more than eight hours to be extracted from the network [Trigg 1997].

In 1999, Ennett studied how changing the distribution of the training and the test set by increasing the representation of the under represented class, would affect the ANN classification performance. Working with the coronary artery bypass graft (CABG) surgery at the San Francisco Heart Institute (SFHI), 32 input variables and "Mortality" outcome, she discovered that training the ANN with a higher prevalence than that found in the original dataset produced higher sensitivity values (correct classification of the non-survivors) [Ennett 1999]. This finding agreed with Baxt who achieved better correct classification when the network was trained on a higher-than–normal prevalence [Baxt & Skora 1996]. These results highlight the significance of incorporating the a priori probabilities of the training set with the model performance [Ennett 1999]. Ennett discovered that three-layer ANNs outperformed double-layer ANNs when estimating in-hospital mortality prediction. Three-layer ANNs had higher sensitivities, thereby correctly classifying more non-survivors. In his project Scales concentrated on investigating the capability of a neural network to classify an outcome that is extremely underrepresented in the data. Using the SFHI and POSTOP ICU databases with low mortality rate of 3.7 and 3.2, respectively he tried to predict the "Mortality" outcome. He used log-sensitivity index to create a better balance between the sensitivity and specificity obtained. The specific network architecture used was a feedforward three-layered neural network, with 2 and 7 hidden nodes, and implementing the weight-elimination cost function in its backpropagation algorithm [Scales 2001, Ennett et al. 2002]. His experiments on the SFHI database show an improvement to the performance of mortality classification.

The lengthy process of manually optimising a backpropagation feedforwrd ANN provided the incentive to develop an automated system that could fine-tune the network parameters without user supervision [Ennett et al. 2004]. In 2000 Charette implemented an automated version of the MIRG ANN software [Charette 2000]. The experimental results showed that the automated networks performed equally well or better than the manually optimized ANNs [Ennett et al. 2004]. MIRG research group has successfully estimated a variety of medical outcomes in complex clinical environments using feedforward backpropagation ANN. Using MATLAB Neural Network Toolbox as a base, the MIRG constructed three-layer ANNs to estimate: in-hospital mortality, length of stay and duration of artificial ventilation for adult intensive care unit (ICU) patients [Buskard et al. 1994, Frize et al. 1995, Trigg et al. 1997, Frize et al 2001,] and neonatal intensive care unit (NICU) patients [Walker et al. 1999, Ennett et al. 2001, Tong et al. 2002] and mortality for coronary artery bypass grafting (CABG) surgery patients [Ennett et al. 2004]. This previous work by MIRG group, led to the choice of the packpropagation feedforward ANN to estimate the perinatal outcomes of this thesis using the automated MIRG ANN software.

3 Problem Formulation and Methodology

This chapter focuses on the problem statement and describes the methodology carried out in this thesis. It also outlines the ANN approach for estimating the three perinatal outcomes being studied and describes in details the development of new predicting models that use the most influential variables.

3.1 Problem Formulation and Challenges

The main question in the search for the three perinatal models for delivery type, mortality and Apgar5 was: what are the most important factors for each outcome? Do the most influential factors found match other factors used by physician or found by other researchers? This was the main problem and challenge of this thesis. Working with the Perinatal Niday-2001 and testing with Perinatal Niday-1999 databases were another challenge as that was the first time these databases were preprocessed to be used for any estimation or prediction method.

Another challenge of this thesis was to optimize the performance of the ANNs to be able to predict each of the three outcomes. In order to get the best ANN prediction models for each of the three outcomes the following steps were carried out:

- Applied the feedforward ANNs trained with backpropagation algorithm to the Niday-2001 Enhanced Perinatal database obtained from the PPPESO in order to estimate the three outcomes being studied: delivery type, mortality and Apgar5
- Improved the results by using weight reduction techniques (weight-elimination or weight-decay cost functions)
- Used variable reduction methods to estimate the factors or variables of the Perinatal Niday-2001 database that are the most important in predicting each of the three outcomes and build new models based on the result
- Tested the ANN models using the Perinatal Niday-1999 database

3.2 Data Preparation

3.2.1 Missing Values for Niday-2001

The original Niday-2001 Enhanced Perinatal database provided by PPPESO had 17,688 cases and 38 variables. There were 1,505 cases in the database with missing values. Appendix C shows all the variables in the Perinatal Niday-2001 database. The cases with missing values make up 8.5% of the overall database. This means that 91.5% (16,183 cases) were complete cases. Different methods are available to deal with missing values such as deleting the entire variable with the most missing values or deleting the whole case with missing values. Because the Perinatal Niday-2001 is a large database with only 8.5% missing values, the latter method was chosen.

3.2.2 Deleting Cases for Niday-2001

Since this thesis is a discussion of the prediction of three medical outcomes, some cases and variables had to be deleted from estimating each of the three outcomes after consultation with the medical specialists. In estimating the delivery type, the total number of cases was reduced to 14,766 (1,417 deleted cases) since all the cases of "No Labour" type (one of the categories of the variable *labtype*) had to be deleted. This category "No Labour" gives the possibility of C/S delivery, which contradicts the meaning of the variable *labtype*. The variable *labtype* indicates the type of vaginal delivery as either spontaneous or induced.

Also, for mortality, the focus was to predict death within 28 days of birth; therefore cases of "Death before Birth" (one of the category of the variable *neodeath*) were deleted. It slightly reduced the total number of cases to 16,110 (73 deleted cases) for that outcome.

3.2.3 Deleting Variables for Niday-2001

The total number of variables was reduced to 36 (the total number was 38 variables) after deleting the *idnum* (identification number for each baby birth) and *dob* (baby's date of birth) variables. These two variables are important to keep track of the data but not for analysis. Other variables were deleted for medical reasons. In delivery type analysis, the following variables were deleted form the analysis: *apgar1*, *apgar5*, *cord*, *resffo2*, *resppv*, *resintub*, *reschest*, *resdrugs*, *assisted*, *neodeath* and *neotrans*. Theses 11 variables take place after delivery so they can not play role in the delivery decision. In the case of Apgar5, the following variables were deleted: *apgar1*, *cord*, *resffo2*, *resppv*, *resintub*, *reschest*, *resdrugs*, *neodeath* and *neotrans*. The first variable in the deleted list, *apgar1,* means the measure of Apgar score after 1 minute of birth while *apgar5* (the outcome) measures same thing after 5 minutes. The difference of a few minutes between measuring the two variables made them identical most of the time. The variable *apgar1* is likely to be a cofounder variable for *apgar5* and it should be deleted from analysis. Other variables were deleted from the Apgar5 estimation list because they happen after the Apgar5 score is measured, so, they can not play any role in deciding the risk factors of having a low Apgar score.

The final numbers of cases and variables used for estimating each of the three outcomes were:

- Delivery Type: 14,766 cases with 24 input variables.
- Mortality: 16,110 cases with 35 input variables.
- Apgar5: 16,183 cases with 26 input variables.

Variable Number	**Niday-2001 Variables**	**Delivery Type**	**Mortality**	**Apgar5**
1	IDNUM			
2	DOB			
3	GENDER	*	*	*
4	BF	*	*	*
5	SMOKING	*	*	*
6	CORD		*	
7	SCALP	*	*	*
8	GENERAL	*	*	*
9	EPIDURAL	*	*	*
10	SPINAL	*	*	*
11	NARCOT	*	*	*
12	NITOXIDE	*	*	*
13	PUDENDAL	*	*	*
14	RESFFO2		*	
15	RESPPV		*	
16	RESINTUB		*	
17	RESCHEST		*	
18	RESDRUGS		*	
19	DELBY	*	*	*
20	MOMAGE	*	*	*
21	PARITY	*	*	*
22	APGAR1		*	
23	MONITOR	*	*	*
24	LABTYPE	*	*	*

25	PRESENT	*	*	*
26	ASSISTED		*	*
27	STEROIDS	*	*	*
28	NEOTRANS		*	
29	MOMTRANS	*	*	*
30	HOSPTYPE	*	*	*
31	TERM	*	*	*
32	PRETERM	*	*	*
33	NBABIES	*	*	*
34	WEIGHT1	*	*	*
35	GEST	*	*	*
36	DELTYPE		*	*
37	APGAR5		*	
38	NEODEATH			

* Variable used as estimator for this outcome

Table 3.1: List of the input variables of Perinatal Niday-2001 database for each of the three outcomes

3.2.4 Missing Values for the Niday-1999

The original Niday-1999 Enhanced Perinatal database provided by PPPESO had 17,455 cases and 43 variables. There were 3,692 cases in the database with missing values. Appendix D shows all the variables in the Perinatal Niday-1999 database. This means that 78.9% (13,763 cases) were complete cases. Similar to the Perinatal Niday-2001 all the missing cases has been deleted in the Niday-1999 database.

3.2.5 Deleting More Cases for Niday-1999

Cases of "Death before Birth" were deleted which reduced the total number of cases to 13,687 (76 deleted cases) .Also, for delivery type prediction, the total number of cases was reduced to 12,546 (1,141 deleted cases) after deleting all the of “No Labour” type.

3.2.6 Testing Set Construction

More work had to be done on the Niday-1999 database variables in order to make it similar the 2001 variables. *idnum* (identification number for each baby birth) and *dyear, month* and *day* (three variables to indicate the baby's date of birth) were deleted. *Labtype* (spontaneous, induced or no labour) and *spontdel* (spontaneous delivery) are two variables that indicate the same situation. So, *spontdel* was deleted. The variables *forceps* and *vacuum* are merged together to form a new variable called *assisted* (using forceps or vacuum to assist delivery). The merging of these two variables in one variable produced *assisted* variable which has the same definition as the *assisted* variable in the Niday-2001 database. Also, *deltype* variable was built using *primcs*, *tol* ,*vbac*, *repeatcs* and *birthmod* to produce a variable similar to *deltype* of Niday-2001 database. Three variables on Niday-2001 have no similarity to the Niday-1999 database: *parity*, *delby* and *pudendal.* These three variables have been deleted from the analysis and the testing set using Niday-1999 database.

Table 3.2 shows the updated Perinatal Niday-1999 database (Perinatal Updated-1999 database) used for testing. The same input variable combinations as Table 3.1 were used for the testing of the three outcomes using the Perinatal Updated-1999 database.

Variable Number	Niday-1999 Variables
1	GENDER
2	BF
3	SMOKING
4	CORD
5	SCALP
6	GENERAL
7	EPIDURAL
8	SPINAL
9	NARCOT
10	NITOXIDE
11	RESFFO2
12	RESPPV
13	RESINTUB
14	RESCHEST
15	RESDRUGS
16	MOMAGE
17	APGAR1
18	MONITOR
19	LABTYPE
20	PRESENT
21	ASSISTED
22	STEROIDS
23	NEOTRANS
24	MOMTRANS
25	HOSPTYPE
26	TERM
27	PRETERM
28	NBABIES
29	WEIGHT1
30	GEST
31	DELTYPE
32	APGAR5
33	NEODEATH

Table 3.2: Variables of Perinatal Updated-1999 database

The size of the test sets of the Perinatal Updated-1999 used for each outcome was 5000 cases. All the low occurrence cases available in the Perinatal Updated-1999 database for each outcome were included in the test sets.

The following is a description of the test sets used for each outcome:

- Delivery Type: 1,520 cases of caesarean delivery (CP=69.6%)
- Apgar5: 179 cases of low Apgar score (CP=96.4%)
- Mortality: 32 cases of neonatal death (CP=99.4%)

3.2.7 Data Preprocessing

In order to have a uniform magnitude for the values of the variables, these were standardised. The values of each continuous and integer variable were assumed to follow a normal distribution. The mean was subtracted from each input value and the resulted differences were divided by 3 standard deviations over the entire database. This standardisation uses the z-score formula:

$$z_n = \frac{(x_n - X)}{3\sigma} \tag{3.1}$$

where z_n is the standardised value of variable x for observation number n; x_n is the original value of the variable for observation n; and X and σ are the mean and standard deviation of the variable x. It is essential to standardize the input variables so that same percentage change in the weighted sum of the inputs causes a similar percentage change in the unit output [Olden & Jackson 2002]

3.3 ANN Optimisation

As discussed in chapter 2, based on past experience of MIRG, the neural network chosen for this thesis work was the backpropagation feedforward neural network using the hyperbolic tangent transfer function. In order to use the best ANN architecture, a variety of architectures of two-layer and three-layer ANNs were tested. Many automated and manual initial experiments with different initial ANN parameters were performed. In the case of three-layer ANNs, different numbers of hidden nodes in the hidden layer were used.

Also, based on the previous discussion in section 2.2.3, the following steps were followed:

1. Using a network with no hidden layers.
2. Computing the number of the hidden nodes using mainly Masters' method
3. Using three-layer ANNs with different numbers of hidden nodes in the hidden layers (starting from 1 to the computed number in step 2)
4. Choosing the ANN architectures which achieved high performances

After deciding which ANN architectures could be used in the experiments, weight-elimination and weight-decay (weight reduction) were applied at that point to improve the resulting models. For the standard backpropagation trained ANN, the cost function was the sum of squared errors (SSE) between the target values and the network's outputs. Using the automated program of Charette [Charette 2000, Frize et al. 2000, 2001b] weight reduction was applied. This weight reduction technique reduced the "weight noise" in the models.

The experimental results provided information on CCR, sensitivity, specificity, log-sensitivity index, ASE and ROC curve values. Many graphs were produced during the experiments to show the log-sensitivity index, sensitivity, specificity, ASE and ROC curves.

3.4 Generalisation and Stopping Criteria

Among many generalization methods that are used to identify the optimal performance of a network, two are commonly used: *regularisation* and *early stopping.*

3.4.1.1 Regularisation

Using regularization, it is important to train the network until it reaches convergence. The regularization technique modifies the performance function in order to use it as stopping criterion. The commonly chosen stopping criterion is the minimum sum of squares of the network errors on the training set. Other performance functions that can be used as stopping criteria are the maximum sensitivity, the maximum correct classification rate and the best logarithmic-sensitivity index can be used. Our MIRG group has been using the logarithmic-sensitivity with automated network analysis. This criterion limits the requirement for user-supervision and works well even when the outcome has a low rate of occurrence [Ennett et al. 2003]. It had been shown in pervious that logarithmic-sensitivity index (log-sensitivity index) as an optimal performance function achieves the best generalisation [Scales 2001, Ennett et al. 2002]. The automated network automatically monitored the classification performance to determine when was the best time to stop training after no improvement in the performance measure (e.g. highest CCR, lowest ASE or highest log-sensitivity index value) occurred in the subsequent 500 epochs [Ennett et al. 2004]. The log-sensitivity index stopping criterion helps the automated networks to achieve a better balance between the sensitivity and specificity of the classification results. It attempts to achieve optimal sensitivity and specificity classification while slightly favouring higher sensitivity [Scales 2001].

$$\text{log- sensitivity index} = -\text{sensitivity}^{n} * \log 10(1 - \text{sensitivity} * \text{specificity}) \qquad (3.2)$$

This function expands towards infinity as sensitivity and specificity approach one, but is slightly weighted in favor of sensitivity. The degree of weighting can be controlled by the power *n*. The results showed that an automated neural network which optimises the log-sensitivity index can classify outcomes better than manually-optimised networks, and can classify rare outcomes equally well or better than networks that use the maximum correct classification rate or the minimum average squared error as the stopping criteria [Ennett et al. 2002, 2003, 2004].

3.4.1.2 Early Stopping

In the early stopping technique, the available data is divided into three subsets: training, validation and prediction sets. The first subset is the training set, which is used for computing the gradient and updating the network weights and biases. The second subset is the validation set. The validation set should be representative of all points in the training set. The error on the validation set is monitored during the training process. The validation error will normally decrease during the initial phase of training, as does the training set error. However, when the network begins to overfit the data, the error on the validation set will typically begin to rise. When the validation error increases for a specified number of iterations, the training is stopped, and the weights and biases at the minimum of the validation error are returned [MATLAB]. Finally, the prediction set is used to test the resulted model.

Both generalization techniques can ensure good results if trained using different initial conditions. It is possible for either method to fail in certain circumstances. Testing several different initial conditions can verify robust network performance [MATLAB]

Because this thesis used a skewed database, log-sensitivity index was selected as the stopping criteria. This performance measurement tool is very effective with skewed datasets where the outcome under interest has a very low occurrence in the dataset [Ibrahim 2002, Ennett et al. 2003].

3.5 Optimisation ANN Parameters

As Mentioned before in section 2.3.2, nine parameters have to be accurately tuned in order to improve the performance of the network. The preliminary simulations required many trials to explore the effects of tuning these parameters. From these first simulations, reasonable ranges were identifies to achieve "acceptable" performance. Then these parameters were fine tuned for each network to find the optimal settings [Ennett 1997].

There are two ways to run the MIRG ANN program, the simple version and the automated version. The first version of the program takes the initial parameters value from the user and runs these parameters values through the network. The *lr* and λ will change only with the user defined values (e.g. *lr_inc*, *lr_dec*) trying to converge to the solution. It is a programmed version of the conventional manual method. On the other hand, the automated version of the program, will begin similarly to the simple version using the user-defined values, but depend on the user selection it can change one of the ANN parameter at a time trying to get the best solution by optimising this selected parameter. The automated version uses the "divide and conquer" technique for one variable at a time until the network converges and the maximum network performance is obtained. The "divide and conquer" technique begins by a two user defined values (the parameter range) to calculate the midpoint. It tries this new value of this variable through the ANN then starting at the value it repeats the dividing step with each side of the parameter range to get two new midpoints. It repeats the previous steps recursively until the network converges upon the network parameter value that gives the maximum network performance [Frize et al. 2000, Charette 2000]

No doubt, the manual ANN is a time consuming method of finding optimal network parameters. It relies on the user to change one parameter at a time, then the user has to repeat this step many times until finding the best combination of ANN parameters. Using the automated version of our ANN program can help the user to optimise and try all the network parameters within the defined ranges. If more tuning is needed after that the simple ANN run can be used [Frize et al. 2000].

Table 3.2 shows the ANN adjustable parameters and its available ranges used in this thesis. In the first experiments, reasonable ranges were identified then these parameters were fine-tuned for each network to find the optimal settings.

Parameter	Range
Learning rate (lr)	0.00005-0.005
Learning rate incerement (lr_inc)	1-1.3
Learning rate decrement (lr_dec)	1-0.7
Weight-decay constant (λ)	0.00001-0.01
Weight-decay constant increment (λ_inc)	1-1.3
Weight-decay constant decrement (λ_dec)	1-0.7
Weight scale factor (w_0)	0.001-1
Momentum (m)	0-0.99
error-ratio (err_ratio)	1.001-1.05

Table 3.2: Optimisation parameters and their ranges

3.6 Weight Reduction

Weight-elimination and weight-decay help to avoid overfitting during network training. Overfitting occurs when the network begins to memorize from prior outcomes for prediction, instead of learning. Weight-elmination and weight-decay were the first two methods used in this thesis for weight extraction. Weight-elimination reduces the very small connection weights to zero. These connections play a nonsiginficat rule in the outcome prediction. On the other hand, weight-decay restricts the weighting on connections in the network and limits the variance of the output in order to produce a more stable network.

The weight-decay constant (λ) and the weight scale factor (w_o) are key factors when applying weight-elimination or weight-decay. Larger values of λ needed to be applied to forces the weighted connections that are closer to zero to be eliminated. Also, smaller values of w_0 result in a higher cut-off weight and greater weight-elimination. On the other hand, trying larger values of w_o helps in applying the weight-decay cost function. Weight-decay results in a greater amount of smaller weights and restricts the size of the large weights.

3.7 Variable Reduction using Garson Algorithm

In 1991 Garson proposed a method, later modified by Goh (1995), for partitioning the neural network connection weights in order to determine the relative importance of each input variable in the network [Olden & Jackson 2002]. The Garson algorithm uses the absolute values of the connection weights when calculating variable contributions, and therefore does not provide the direction of the relationship between the input and output variables [Garson

1991]. The algorithm consists of partitioning the connections weights of the network to determine the relative importance of the various inputs. The algorithm essentially involves partitioning the hidden output connection weights of each hidden neuron into components associated with each input neuron. Garson, in his original paper [Garson 1991], formulated one equation that summarized his idea (Equation 3.3).

$$S_i = \frac{\sum_{h=1}^{m} \frac{|w_{hi}|}{\sum_{x=1}^{n} |w_{hx}|} * |w_{oh}|}{\sum_{i=1}^{n} \sum_{h=1}^{m} \frac{|w_{hi}|}{\sum_{x=1}^{n} |w_{hx}|} * |w_{oh}|} \tag{3.3}$$

where S_i is the importance of the input variable i, based on the weights at the input node for that variable;. w_{oh} is the connection weight from output o to hidden node h and w_{hi} is the connection weight from hidden node h to input i and m is the number of hidden nodes. Also, w_{hx} is the connection weight from hidden node h to each input x and n is the number of the input nodes.

Garson's equation was too complex to be applied in one step. Goh later divided this equation into several simple equations. The following equations (Equations 3.4-3.7) were formulated by Goh in order to calculate the importance S_i of the input variable i [Goh 1995].

$$S_i = \sum_{h=1}^{m} Q_{hi} \tag{3.4}$$

where m is the total number of hidden nodes in the hidden layer and Q_{hi} is given by Equation 3.5.

$$Q_{hi} = \frac{P_{hi}}{\sum_{x=1}^{n} P_{hx}} \tag{3.5}$$

where n is the total number of input variables in the input layer and P_{hi}, is the absolute weights product for the connection from hidden node h to input i. P is given by the following Equation (3.6):

$$P_{hi} = |w_{oh}| * |w_{hi}| \tag{3.6}$$

where w_{oh} is the connection weight from output o to hidden node h and w_{hi} is the connection weight from hidden node h to input i.

The last step after calculating S_i for each input variable is to obtain its relative importance RS_i (Equation 3.7). Expressed as a percentage, this gives the relative importance or distribution of all output weights attributable to a given input variable [Goh 1995].

$$RS_i = \frac{S_i * 100}{S_{Total}} \tag{3.7}$$

In the case of ANNs with more than one hidden layer, the above equations can be generalized and the calculation can be done for each hidden layer. In general, calculating the importance of each node in a certain hidden layer will start backwards from the first hidden layer (before the output) until reaching the input layer. Generally, n is the total number of nodes in the layer being calculated, m is the number of the nodes in the layer following it, and S_i is the importance of the node i in the current layer [Goh 1995].

In the case of a two-layer ANNs (with no hidden layer), the importance of each input variable i is equal to its absolute input-output weight:

$$S_i = P_{hi} = |w_{oi}| \tag{3.8}$$

In the application of Garson's algorithn by Goh (Equations 3.4-3.7) and all reaserchers following Goh since 1995[Goh 1995, Olden & Jackson 2002], applying Equation 3.5 after 3.6 cancels the term w_{oh} , which means that it does not count the effect of the hidden output connections in the final calculation of the relative importance. This contradicts the main objective of Gason algorithm to partition the hidden output connection weights of each hidden neuron into components associated with each input neuron.

3.8 Partitioning Weight Algorithm

In this section, we updated Goh's equations, which we now call partitioning weight algorithm. Goh's Equations 3.5 and 3.6 were replaced by Equation 3.9 and 3.10 respectively. Other calculation steps were the same.

$$Q_{hi} = P_{hi} * | w_{oh} |$$

(3.9)

where w_{oh} is the connection weight from output o to hidden node h and P_{hi} is calculated by Equation 3.10.

$$P_{hi} = \frac{| w_{hi} |}{\sum_{x=1}^{n} | w_{hx} |}$$

(3.10)

where n is the total number of input variables in the input layer and w_{hi}, is the connection weight from hidden node h to the input i.

Equation 3.4 can be used as before to calculate the importance S_i of the input variable i and equation 3.7 to calculate the relative importance RS_i which, expresses in percentage the importance of each input variables.

Summarising the above work, applying updated equations (Equations 3.9 and 3.10) and Goh's equation (Equation 3.4), finally provides a correct equivalent of Garson's equation (Equation 3.3) to calculate the importance of an input variable. Finally, equation 3.7 can be applied to get the relative importance in percentage form.

3.9 Summary of the Methodology

The methodology of this thesis can be summarized in the following steps:

1. Use automated program to run the backpropagation ANNs preliminary experiments using optimization parameters ranges
2. Apply different ANN architectures (i.e. no hidden layer versus one hidden layer)
3. Apply weight reduction techniques (weight-elimination and weight-decay) to get the best prediction model
4. Calculate the relative importance for each input variables using partitioning weight algorithm
5. Apply variable reduction depending on the results of step 4 to obtain a minimum variable set while maintaining acceptable performance measures
6. Test the model of step 3 and the simplest model of step 5 on Perinatal Updated-1999 database

4 Simulations and Results

This chapter presents the ANN experimental results of estimating delivery type, mortality and Apgar5 using the Perinatal Niday-2001 database. First, a comparison of the results of two artificial techniques (MIRG ANN and BIOCORE fuzzy logic) is presented. Then, the previous methodology steps of building the best predictive model for each outcome are discussed. The calibration of the ANN parameters and selection of the ANN architecture are considered to achieve the optimal network for each of the three outcomes. Weight and variable reductions are applied to the networks in order to improve the results, simplify the predictive models and obtain the minimum variables sets. Finally, the predictive models are tested using the Perinatal Niday-1999 database.

4.1 Comparison of the MIRG ANN with BIOCORE Fuzzy Logic Technique

Fuzzy K-Nearest Neighbour algorithm (FK-NN) has been shown to be a powerful fuzzy pattern classifier [Kuncheva 2000]. Seker et al., from the BIOCORE (Biomedical Computing Research) Group in the UK, further enhanced the algorithm by developing a fuzzy measurement method [Seker 2003]. In order to evaluate the MIRG ANN software the Perinatal Niday-2001 database was sent to the BIOCORE Group to be tested with their fuzzy logic (FL) method. A comparison was done between the two artificial techniques [Frize et al. 2004]. The performance for both tools was measured in the same manner by recording, for each set of experiments, the sensitivity, specificity and CCR.

4.1.1 Data Structures

The Perinatal Niday-2001 database used for this comparison. The following three tables show the data structure used for estimating each outcome:

Delivery Type	Type (+1) Caesarean	Type (-1) Vaginal	Total
Training Samples	2394	8389	10783
Test Samples	1200	4200	5400
Total Samples	3594	12589	16183

Table 4.1: Data structure for estimating delivery type

Apgar5	Type (+1) {1-6}	Type (-1) {7-10}	Total
Training Samples	235	10450	10685
Test Samples	118	5380	5498
Total Samples	353	15830	16183

Table 4.2: Data structure for estimating Apgar5

Mortality	Type (+1) Nonsurvivor	Type (-1) Survivor	Total
Training Samples	33	10700	10733
Test Samples	20	5357	5377
Total Samples	53	16057	16110

Table 4.3: Data structure for estimating mortality

4.1.2 Results of the MIRG ANN Software

This approach used two and three-layer feedforward ANNs trained using the backpropagation algorithm, with 2/3 of the data for the training set and 1/3 for the test set. Three types of training sets were used: Original training set (same ratio as the original data) trOrg, and two artificial training sets. Artificial training set of 20% (tr20%) and 50% (tr50%) of the low occurrence outcome in the original data. The hyperbolic tangent transfer function was used and output variables were classified into an output of -1 or 1. A negative output (-1) predicts a vaginal delivery, a high Apgar score, and survival of the newborn, whereas a positive output (+1) predicts a Caesarian section delivery, a low Apgar score, or death of the infant [Frize et al. 2004]. Two of the estimations used artificial training sets. In the case of delivery type estimation, a training set of 50% was used (tr50%) and in the case of mortality estimation, a training set of 20% was used (tr20%). Artificial training sets were used because the datasets used in estimating these outcomes were skewed. Tables 4.4, 4.5 and 4.6 show the best experimental results of estimating of the three outcomes using the MIRG ANN software.

ANN Type	CCR%	Sensitivity%	Specificity%
2-layer, tr50%	Train:85.34 Test:86.8	Train:81.64 Test:80.57	Train:88.95 Test:88.45

Table 4.4: delivery type predictive performance

ANN Type	CCR%	Sensitivity%	Specificity%
2-layer, trOrg	Train:98.53 Test:98.22	Train:40 Test:31.36	Train:99.83 Test:99.72

Table 4.5: Apgar5 predictive performance

ANN Type	CCR%	Sensitivity%	Specificity%
2-layer, tr20%	Train:98 Test:98.88	Train:94.18 Test:90	Train:98.96 Test:98.92

Table 4.6: Mortality predictive performance

4.1.3 Results of the BIOCORE FL Method

The FK-NN is defined as a function of the number of neighbourhoods (K), class membership degrees (between 0 and 1), and distances between a pattern to be classified and patterns for which the class membership degrees were previously determined. A class membership degree between 0 and 1 is computed using the first K minimum distances and the known class membership degrees of the patterns. The FK-NN not only gives a class to which the pattern is assigned, but also the class membership degree that provides information about the certainty of the classification decision.

Tables 4.7, 4.8 and 4.9 below summarise the experimental results of prediction of the three outcomes using the BIOCORE fuzzy method.

CCR (%)	Sensitivity (%)	Specificity (%)
89.17	71.25	94.29

Table 4.7: Delivery type predictive performance

CCR (%)	Sensitivity (%)	Specificity (%)
97.62	19.49	99.33

Table 4.8: Apgar5 predictive performance

CCR (%)	Sensitivity (%)	Specificity (%)
99.81	55	99.98

Table 4.9: Mortality predictive performance

This study was helpful to compare the performances of two different tools to estimate clinical outcomes of newborns using a large data set. The results show that both artificial techniques provide an excellent CCR and a high specificity, that is, in predicting a vaginal delivery, a high Apgar score, and survival of the newborns. Where the two approaches differed slightly was in predicting the positive outcomes (a Caesarean section delivery, death of the newborns and a low Apgar score). In the prediction of mortality and Apgar5 MIRG ANN outperformed the FL method. In the case of Apgar5, the ANN method achieved sensitivity and CCR of 31.36% and 98.22% compared to 19.49% and 97.62% in the case of the FL method. Also, in mortality prediction, the ANN method achieved a sensitivity of 90% compared to 55% using the fuzzy method, but with slightly less CCR comparable to FL (less 0.93%). In the case of delivery type the two methods had similar results; they obtained high sensitivities at a slightly cost to specificity and the CCR values.

The above evaluation of the MIRG ANN software has been done using Perinatal Niday-2001 database and was considered as a starting point for this thesis work. But, more ANN techniques have been applied in advanced work of this thesis to estimate the same three outcomes.

In the following sections of this chapter the experimental results of each step mentioned in chapter 3 (section 3.9) will be discussed.

4.2 Step 1 & 2: Optimisation and Choosing ANN Architectures

As mentioned in section 2.3.3 the following procedure was applied in order to find the best ANNs architecture:

1. Tried network with no hidden layers (to check linearity).
2. Computed the number of the hidden nodes possible (to use it as a guide for step 3)
3. Tried three-layer network with different hidden nodes (started form 1 and going up)
4. Chose the ANN architectures which achieved high performances

The optimal parameter settings were used with the above procedure to find the best ANN architecture for each of the three outcomes.

First, the optimal settings were applied to two-layer ANN in order to check the linearity of the data. Then, three-layer ANNs with different number of hidden nodes were applied. Because certain problems are too complex than to be solved by linear networks, the number of hidden nodes in the hidden layers can mean make a difference between success and failure [Masters 1993].

4.2.1 Delivery Type

For Delivery type estimation the dataset of 14,766 cases was divided into 2/3 (9,840 cases) training set and 1/3 (4,926 cases) test set. The delivery datasets contain 14.75% of caesarean section cases (+1 type) and 85.25% vaginal delivery cases (-1 type). The CP using this skewed dataset was 85.25%.

Table 4.10 shows the results obtained with different ANN architectures for the prediction of delivery type. The ANN architectures missing from Table 4.10 were unstable (did not converge). One of the unstable cases occurred when using a two-layer ANN, which means that this analysis is nonlinear.

The same table shows that, three-layer ANN with 13 hidden nodes in the hidden layer achieved the best test log-sensitivity index of 0.167 and best sensitivity of 52.79%. Other simpler three-layer ANNs with 4, 3 and 5 hidden nodes achieved similar sensitivities of 52.51%, 52.37% and 52.23% respectively. The three-layer ANNs with 4 and 5 hidden nodes resulted in higher specificities and CCRs than the 3 and 13 hidden nodes. These two ANNs achieved the first and second highest CCRs among all architectures used. On the other hand, the small improvement in the sensitivity of the 13 hidden nodes ANN (sensitivity increased by 0.28%; over the 4 hidden nodes ANN) affected its specificity and CCR. Although, high sensitivity (which is also measured by a high log-sensitivity index) is an effective factor in the selection of the best ANN, this is does not have to be with a cost to specificity and a CCR. Therefore, the three-layer ANNs with 4, 5 and 13 hidden nodes in the hidden layer were selected for more testing. In step 3, weight reduction was applied to these selected ANNs in order to improve the prediction results.

As mentioned in section 2.2.3, there are several methods to calculate the number of the hidden nodes in the hidden layer. Using the first method of Masters, the number of the hidden nodes was expected to be between 4 and 5 (the square root of 24). The experimental results for delivery type were good using the Masters' method. With other methods, the best number of hidden nodes could be 12 or 18 (50% or 75% of the number of the input nodes: 24. Greater instability was observed in the experiments using ANNs with more than 9 hidden nodes in the hidden layer. This instability continued for ANNs with more than 13 hidden nodes. So, Masters' method seemed to be the best guide for designing the ANN architecture.

Performance	Three-layer ANNs with different number of hidden nodes								
	1	2	3	4	5	6	8	9	13
Test Sens	51.67%	51.67%	52.37%	52.51%	52.23%	50.84%	50.98%	50.98%	52.79%
Train Sens	52.02%	53.11%	55.78%	52.50%	53.79%	52.50%	54.89%	54.89%	56.32%
Test Spec	98.41%	98.27%	97.96%	98.43%	98.34%	98.57%	98.17%	98.17%	98.03%
Train Spec	98.31%	98.38%	98.23%	98.35%	98.27%	98.42%	97.97%	97.97%	98.13%
Test CR	91.60%	91.47%	91.31%	91.74%	91.62%	91.62%	91.29%	91.29%	91.43%
Train CR	91.42%	91.65%	91.92%	91.53%	91.66%	91.60%	91.57%	91.57%	91.91%
Test ASE	27.53%	27.90%	28.49%	28.14%	28.11%	27.44%	28.47%	28.47%	28.41%
Train ASE	28.86%	28.05%	27.29%	28.40%	28.20%	27.06%	28.10%	28.10%	27.57%
Test ROC	89.40%	89.30%	90.06%	89.61%	89.80%	90.43%	90.11%	90.11%	89.41%
Train ROC	86.99%	88.86%	90.39%	88.23%	89.54%	90.51%	89.78%	89.78%	90.04%
Test Log Index	0.159	0.159	0.164	0.166	0.163	0.154	0.154	0.154	0.167
Train Log Index	0.162	0.170	0.192	0.166	0.176	0.166	0.184	0.184	0.197

Table 4.10: Delivery type prediction using different ANN architectures

4.2.2 Mortality

For mortality estimation, the dataset of 16,110 cases was divided into 2/3 (10,733 cases) training set and 1/3 (5,377 cases) test set. The original mortality dataset was skewed. It contained 3.27% death cases (+1 type) and 96.73% survivals (-1 type). Using the original training set with such a low death ratio produced a very low sensitivity. Although the resulted model had a good CCR, it could not predict death. The MIRG group had tried an effective method using artificial training sets. Using artificial training sets with more of the low occurrence cases (+1 type) always improves the sensitivity. Increasing the number of the trained cases of a certain type will definitely improve the prediction of that type. Previous MIRG studies used artificial training sets of 20%, 30% or even 50% of the low occurrence type. In mortality prediction, an artificial training set tr20% was created to contain 20% of death cases (+1 type). Training the ANN with tr20% training set clearly improved the ANN performance in terms of sensitivity.

Table 4.11 shows the results of testing different ANN architectures for the prediction of mortality using an artificial training set (tr20%). As in the case of delivery type, the two-layer ANN was unstable. This showed the nonlinearity of the mortality estimation problem. Three-layer ANNs with different numbers of hidden nodes in the hidden layer were tested. The best log-sensitivity index resulted from the training of three-layer ANNs with 3 and 5 hidden nodes. These ANNs achieved 97.98% and 98.04% specificities, and 97.97% and 98.03% CCRs. The ANNs with 2, 4, 6 and 7 hidden nodes achieved very high CCRs of 99.7%, 99.42%, 99.72% and 99.39% respectively. The 7 hidden nodes ANN attained a high sensitivity value of 90%. Many experiments were performed to further test the 3, 5 and 7 hidden nodes ANNs. Our goal was to use these three-layer architectures in building the best prediction model for mortality.

Performance	**Three-layer ANNs with different number of hidden nodes**								
	1	2	3	4	5	6	7	8	9
Test Sens	90.00%	85.00%	95.00%	85.00%	95.00%	85.00%	90.00%	90.00%	90.00%
Train Sens	91.52%	94.18%	97.11%	97.11%	85.56%	97.11%	97.11%	88.54%	84.21%
Test Spec	98.04%	99.76%	97.98%	99.48%	98.04%	99.78%	99.42%	98.38%	98.02%
Train Spec	98.01%	99.81%	98.08%	99.62%	98.27%	99.88%	99.52%	98.31%	98.09%
Test CR	98.01%	99.70%	97.97%	99.42%	98.03%	99.72%	99.39%	98.34%	97.99%
Train CR	96.71%	98.69%	97.89%	99.11%	95.72%	99.33%	99.04%	96.36%	95.31%
Test ASE	6.19%	1.21%	7.31%	1.90%	6.80%	1.11%	2.10%	5.57%	12.56%
Train ASE	12.24%	5.20%	8.05%	4.10%	12.88%	2.66%	3.76%	13.05%	22.46%
Test ROC	96.66%	92.39%	96.68%	99.03%	98.98%	94.87%	94.81%	93.51%	97.87%
Train ROC	96.29%	96.99%	97.85%	98.29%	97.81%	98.50%	98.43%	97.19%	97.23%

Table 4.11: Mortality prediction using different ANN architectures

It is suitable to mention that the architecture of 3, 5 and 7 hidden nodes worked well with the Masters' Method. Using Masters' method on 35 input nodes and one outcome suggested the best ANN architecture of hidden nodes was around 5 or 6 (the square root of 35). Other methods with a larger number of hidden nodes performed less well.

4.2.3 Apgar5

For Apgar estimation, the dataset of 16,183 cases was divided into 2/3 (10,685 cases) training set and 1/3 (5,498 cases) test set. The original mortality dataset was skewed. It contained 2.18% low Apgar score which means less than 7 Apgar score value (+1 type) and 97.82% high Apgar values which includes Apgar values of 7 to 10 (-1 type). Similar to mortality prediction, using the original training set with such a low +1 type ratio produced a very low sensitivity. Artificial training set tr20% was created to contain 20% of low Apgar5 cases (+1 type). Training the ANN with tr20% training set clearly improved the ANN performance in terms of sensitivity. Table 4.12 shows the results of testing different ANN architectures for the prediction of mortality using an artificial training set (tr20%). The two-layer ANN achieved good sensitivity of 40.52. This showed that predicting Apgar5 has more linearity characteristics than prediction of delivery type and mortality. Three-layer ANNs with different number of hidden nodes in the hidden layer were tested. Using the same technique as in delivery type and mortality, ANNs with 6, 7 and 13 hidden nodes were selected for the next step.

Performance	Two-layer ANN	Three-layer ANNs with different number of hidden nodes							
		4	5	6	7	8	11	13	15
Test Sens	40.52%	39.66%	41.38%	42.24%	43.10%	42.24%	40.52%	43.10%	39.66%
Train Sens	44.81%	52.04%	45.00%	48.52%	61.68%	44.72%	49.21%	45.69%	43.28%
Test Spec	96.89%	96.42%	96.04%	97.42%	95.87%	97.90%	96.10%	92.04%	94.36%
Train Spec	96.70%	97.27%	95.60%	97.09%	96.98%	97.98%	96.62%	92.03%	94.73%
Test CR	95.68%	95.20%	94.87%	96.24%	94.74%	96.70%	94.90%	90.99%	93.18%
Train CR	86.32%	88.22%	85.47%	87.37%	89.91%	87.33%	87.13%	82.76%	84.44%
Test ASE	20.00%	21.66%	22.16%	20.57%	19.98%	14.99%	21.65%	28.24%	23.54%
Train ASE	43.59%	39.47%	45.54%	42.83%	35.69%	41.57%	43.20%	59.88%	55.70%
Test ROC	80.90%	78.47%	77.26%	75.03%	74.32%	79.80%	81.58%	70.51%	70.20%
Train ROC	79.31%	82.45%	76.58%	78.20%	83.63%	80.41%	79.80%	70.86%	73.19%
Test Log Index	0.088	0.083	0.091	0.097	0.100	0.098	0.087	0.095	0.081
Train Log Index	0.111	0.159	0.110	0.134	0.244	0.112	0.138	0.108	0.099

Table 4.12: Apgar5 prediction using different ANN architectures

4.3 Step3: Weight Reduction

As explained in section 3.6, the weight-decay constant (λ) and the weight scale factor (w_0) play a main role in applying weight-elimination or weight-decay.

In order to apply weight-elimination to the ANNs resulting from step 2, larger values of λ needed to be applied to force the weighted connections that are closer to zero to be eliminated. Applying different small values of w_0 with higher values of λ helps the weight-elimination cost function to play its role in reducing non-significant weights.

Also, trying larger values of w_0 helps in applying the weight-decay cost function on the ANNs resulting from step 2. Weight-decay results in a greater amount of smaller weights and restricts the size of the large weights.

The following tables show the result of applying different combinations of λ and w_0 for each of the three outcomes.

4.3.1 Delivery Type

Weight-elimination		**Three-layer (4 hid) ANNs**			**Three-layer (5 hid) ANNs**			**Three-layer (13 hid) ANNs**		
λ	**ω0**	**Test Sens**	**Test Spec**	**Test CCR**	**Test Sens**	**Test Spec**	**Test CCR**	**Test Sens**	**Test Spec**	**Test CCR**
0.01	*0.01*	49.16%	98.65%	91.43%	48.89%	98.43%	91.21%	48.47%	98.55%	91.25%
0.001	*0.01*	51.53%	98.00%	91.23%	52.23%	98.31%	91.60%	53.34%	97.53%	91.09%
0.0001	*0.01*	50.56%	98.36%	91.39%	52.37%	98.29%	91.60%	51.81%	98.17%	91.41%
0.05	*0.01*	49.16%	98.65%	91.43%	48.89%	98.43%	91.21%	48.47%	98.55%	91.25%
0.005	*0.01*	49.16%	98.65%	91.43%	48.89%	98.43%	91.21%	48.47%	98.55%	91.25%
0.0005	*0.01*	52.37%	98.27%	91.58%	52.65%	98.12%	91.49%	53.34%	98.31%	91.76%
0.01	*0.001*	49.44%	98.22%	91.11%	49.16%	98.46%	91.27%	51.81%	98.50%	91.70%
0.001	*0.001*	49.86%	98.41%	91.33%	51.81%	98.36%	91.58%	50.56%	98.31%	91.35%
0.0001	*0.001*	51.81%	98.62%	91.80%	52.09%	98.24%	91.51%	**54.18%**	**97.65%**	**91.31%**
0.05	*0.001*	49.30%	98.48%	91.31%	51.53%	98.24%	91.43%	52.09%	97.67%	91.03%
0.005	*0.001*	51.81%	98.24%	91.47%	52.79%	98.15%	91.53%	49.72%	98.55%	91.43%
0.0005	*0.001*	51.67%	98.41%	91.60%	51.95%	98.36%	91.60%	52.37%	98.00%	91.35%
0.01	*0.0001*	50.14%	98.60%	91.53%	52.09%	98.34%	91.60%	48.19%	98.60%	91.25%
0.001	*0.0001*	50.14%	98.60%	91.53%	49.72%	98.46%	91.35%	49.16%	98.31%	91.15%
0.0001	*0.0001*	**52.51%**	**98.55%**	**91.84%**	52.37%	98.34%	91.64%	48.33%	98.53%	91.21%
Weight-decay										
0.00001	*0.01*	50.56%	98.43%	91.45%	51.95%	98.31%	91.56%	53.48%	97.96%	91.47%
0.00005	*0.01*	51.25%	98.36%	91.49%	**53.76%**	**98.10%**	**91.64%**	51.95%	98.27%	91.51%
0.00001	*0.1*	49.86%	98.43%	91.35%	52.93%	98.36%	91.74%	51.81%	98.10%	91.35%
0.00005	*0.1*	50.14%	98.46%	91.41%	52.65%	98.36%	91.70%	51.25%	98.17%	91.33%

Table 4.13: Applying weight reduction for ANNs used in delivery type prediction

Table 4.13 shows that tuning the values λ and w_0 increased the sensitivity of the simulations to higher values. It reached 54.18% with a three-layer 13 hidden nodes ANN. This ANN also achieved the highest log-selectivity index of 0.177.

Using more simple ANNs such as the three-layer ANNs with 4 or 5 hidden nodes resulted in slightly less sensitivitiy (52.51% or 53.76% respectively) and less log-sensitivity index (0.16617 and 0.17499 respictively) . These two ANNs achieved a higher specificity and CCR than the 13 hidden nodes ANN. Because the ANN with 5 hidden nodes had the second best sensitivity and log-sensitivity index after the 13 hidden nodes ANN, this ANN was chosen for applying variable reduction. The differences in sensitivity and log-sensitivity index from the best 13 hidden nodes ANN were 0.42 and 0.27 respectively. These differences were considered very small compared to the advantage of reducing the complexity of the ANN by using 5 nodes instead of 13. For the above reasons, the three-layer ANN with 5 hidden nodes was selected for the next step of variable reduction.

The performance of the ANN was measured using 10 test sets. These test sets were of the same size and were created randomly from the original test set. They were created from 1/3 of the database not used in the training set. Table 4.14 shows the ANN performance results of the 10 test sets and the means and standard deviations for the ten results.

Test set	**Sens.**	**Spec.**	**CCR**	**ASE**	**ROC**
1	53.76%	98.10%	91.64%	28.70%	89.70%
2	53.53%	97.93%	91.80%	27.67%	90.19%
3	52.15%	98.17%	91.43%	29.81%	88.88%
4	51.96%	97.95%	91.05%	29.94%	89.09%
5	54.66%	97.93%	91.62%	28.87%	89.07%
6	53.10%	97.82%	91.37%	29.33%	89.08%
7	53.47%	97.98%	90.95%	30.84%	88.14%
8	52.96%	97.71%	91.11%	30.13%	89.13%
9	49.93%	97.61%	90.38%	31.06%	88.35%
10	52.54%	97.93%	91.21%	29.81%	88.78%
Mean	*52.81%*	*97.91%*	*91.25%*	*29.62%*	*89.04%*
STD	*1.29*	*0.17*	*0.41*	*1.02*	*0.59*

Table 4.14: ANN Performance results of 10 test sets for delivery type outcome

4.3.2 Mortality

Table 4.15 shows that tuning the values λ and w_0 increased the sensitivity (by increasing log-sensitivity index) of the experiments to higher values. It reached 97% two times with three-layer 5 hidden nodes ANNs. These two ANNs achieved high specificities and CCRs. But the best Specificities and CCRs resulted from the three-layer ANNs with 7 hidden nodes simulations.

Although three-layer ANNs with 3 hidden nodes in the hidden layer achieved good results for sensitivity, they produced a lower specificity and CCR than the 5 and 7 hidden nodes ANNs. The three-layer ANNs with 3 hidden nodes ANNs could reach the same prediction power of more complicated ANNs with precise tuning of the values λ and w_0. This is a very important result because simplicity is an important factor in designing any prediction model. Since log-sensitivity index and sensitivity play a main role in the selection of the best model, the three-layer ANN with 5 hidden nodes with 97% sensitivity was chosen to continue with the next step of variable reduction.

Weight-elimination		**Three-layer (3 hid) ANNs**			**Three-layer (5 hid) ANNs**			**Three-layer (7 hid) ANNs**		
λ	**ω0**	**Test Sens**	**Test Spec**	**Test CCR**	**Test Sens**	**Test Spec**	**Test CCR**	**Test Sens**	**Test Spec**	**Test CCR**
0.01	*0.01*	95	97.98	97.97	95	98.04	98.03	90	99.42	99.39
0.001	*0.01*	**95**	**99.18**	**99.16**	**97**	**98.82**	**98.83**	**90**	**99.53**	**99.50**
0.0001	*0.01*	95	97.98	97.97	95	98.04	98.03	90	99.42	99.39
0.05	*0.01*	95	97.98	97.97	95	98.04	98.03	90	99.42	99.39
0.005	*0.01*	95	97.98	97.97	95	98.04	98.03	90	99.42	99.39
0.0005	*0.01*	95	97.98	97.97	**97**	**98.53**	**98.53**	**90**	**99.61**	**99.57**
0.01	*0.001*	95	97.98	97.97	95	98.04	98.03	90	99.42	99.39
0.001	*0.001*	95	97.98	97.97	95	98.04	98.03	90	99.42	99.39
0.0001	*0.001*	95	97.98	97.97	95	98.04	98.03	**90**	**99.70**	**99.67**
0.05	*0.001*	95	97.98	97.97	95	98.04	98.03	90	99.42	99.39
0.005	*0.001*	95	97.98	97.97	95	98.04	98.03	90	99.42	99.39
0.0005	*0.001*	95	97.98	97.97	95	98.04	98.03	90	99.48	99.44
0.01	*0.0001*	95	97.98	97.97	95	98.04	98.03	**90**	**99.61**	**99.57**
0.001	*0.0001*	95	97.98	97.97	95	98.04	98.03	90	99.42	99.39
0.0001	*0.0001*	95	97.98	97.97	95	98.04	98.03	90	99.42	99.39
Weight-decay										
0.00001	*0.01*	95	97.98	97.97	95	98.04	98.03	90	99.42	99.39
0.00005	*0.01*	95	97.98	97.97	95	98.04	98.03	90	99.42	99.39
0.00001	*0.1*	95	97.98	97.97	95	98.04	98.03	90	99.42	99.39
0.00005	*0.1*	95	97.98	97.97	95	98.04	98.03	90	99.42	99.39

Table 4.15: Applying weight reduction for ANNs used in mortality prediction

Table 4.16: shows the ANN performance results of the 10 mortality test sets and their mean and standard deviation.

Test set	Sens.	Spec.	CCR	ASE	ROC
1	97%	98.77%	98.73%	3.96%	99.74%
2	90%	98.71%	98.72%	4.18%	99.79%
3	79%	98.79%	98.79%	4.20%	99.66%
4	92%	98.79%	98.79%	3.36%	99.76%
5	84%	98.77%	98.77%	4.28%	99.69%
6	75%	98.64%	98.64%	4.38%	99.64%
7	75%	98.73%	98.74%	4.43%	99.59%
8	88%	98.77%	98.73%	4.30%	99.64%
9	87%	99.01%	99.01%	3.17%	99.75%
10	75%	98.99%	99.00%	3.42%	99.78%
Mean	*84%*	*98.80%*	*98.79%*	*3.97%*	*99.70%*
STD	*6.53*	*0.12*	*0.12*	*0.47*	*0.07*

Table 4.16: ANN Performance results of 10 test sets for mortality outcome

4.3.3 Apgar5

Table 4.17 shows that the best prediction results achieved by using three-layer ANN with 13 hidden nodes at 0.0001 λ and 0.001 w_0. It achieved the highest sensitivity of 45.69% with high specificity and CCR of 96.36% and 95.27%. Following the same method described in delivery type and mortality, the three-layer ANN with 13 hidden nodes was chosen to continue with the next step of variable reduction. The performance of the ANN was measured using 10 test sets as before. Table 4.18 shows the ANN performance results of the 10 Apgar5 test sets and their mean and standard deviation.

Weight-elimination		**Three-layer (6 hid) ANNs**			**Three-layer (7 hid) ANNs**			**Three-layer (13 hid) ANNs**		
λ	**ω0**	**Test Sens**	**Test Spec**	**Test CCR**	**Test Sens**	**Test Spec**	**Test CCR**	**Test Sens**	**Test Spec**	**Test CCR**
0.01	*0.01*	40.52%	97.54%	96.31%	38.79%	96.65%	95.40	42.24%	97.69%	97.69
0.001	*0.01*	41.38%	97.16%	95.96%	**43.10%**	**96.00%**	**94.87**	**43.97%**	**95.83%**	**94.72**
0.0001	*0.01*	40.52%	97.54%	96.31%	38.79%	97.82%	96.55	42.24%	97.69%	96.50
0.05	*0.01*	40.52%	97.54%	96.31%	38.79%	96.65%	95.40	42.24%	97.69%	96.50
0.005	*0.01*	40.52%	97.54%	96.31%	38.79%	96.65%	95.40	42.24%	97.69%	96.50
0.0005	*0.01*	40.52%	97.54%	96.31%	41.38%	95.72%	94.55	42.24%	97.69%	96.50
0.01	*0.001*	40.52%	97.80%	96.57%	38.79%	96.93%	95.68	41.38%	93.26%	92.14
0.001	*0.001*	40.52%	97.80%	96.57%	42.24%	96.53%	95.37	41.38%	93.26%	92.14
0.0001	*0.001*	40.52%	97.80%	96.57%	41.38%	95.76%	94.59	**45.69%**	**96.36%**	**95.27**
0.05	*0.001*	40.52%	97.80%	96.57%	38.79%	96.93%	95.68	41.38%	93.26%	92.14
0.005	*0.001*	40.52%	97.80%	96.57%	38.79%	97.56%	96.29	41.38%	93.26%	92.14
0.0005	*0.001*	40.52%	97.80%	96.57%	39.66%	96.69%	95.46	41.38%	93.26%	92.14
0.01	*0.0001*	**42.24%**	**97.31%**	**96.13%**	38.79%	96.95%	95.70	43.10%	92.21%	91.16
0.001	*0.0001*	**42.24%**	**97.31%**	**96.13%**	38.79%	97.01%	95.76	43.10%	92.21%	91.16
0.0001	*0.0001*	**42.24%**	**97.31%**	**96.13%**	38.79%	97.05%	95.79	43.10%	92.21%	91.16
Weight-decay										
0.00001	*0.01*	40.52%	97.54%	96.31%	38.79%	97.73%	96.46	42.24%	97.69%	96.50
0.00005	*0.01*	40.52%	97.54%	96.31%	38.79%	97.84%	96.57	42.24%	97.69%	96.50
0.00001	*0.1*	41.38%	97.42%	96.22%	41.38%	95.23%	94.07	31.03%	88.29%	87.06
0.00005	*0.1*	41.38%	97.42%	96.22%	38.79%	97.48%	96.22	31.03%	88.29%	87.06

Table 4.17: Applying weight reduction for ANNs used in Apgar5 prediction

Test set	**Sens.**	**Spec.**	**CCR**	**ASE**	**ROC**
1	45.69%	96.36%	95.27%	18.41%	79.20%
2	40.98%	94.84%	93.62%	23.77%	75.38%
3	40.31%	95.20%	93.88%	23.33%	74.98%
4	43.10%	94.96%	93.85%	23.08%	78.07%
5	42.75%	95.75%	94.46%	20.86%	72.23%
6	34.17%	95.30%	93.94%	22.96%	78.49%
7	40.00%	93.19%	91.86%	55.87%	68.30%
8	39.23%	95.10%	93.75%	22.70%	74.46%
9	35.90%	94.56%	93.29%	24.54%	78.25%
10	40.31%	95.20%	93.88%	23.33%	74.98%
Mean	*40.24%*	*95.05%*	*93.78%*	*25.88%*	*75.43%*
STD	*3.19%*	*0.78%*	*0.82%*	*10.13%*	*3.18%*

Table 4.18: ANN Performance results of 10 test sets for Apgar5 outcome

4.4 Step 4: Computing Relative Importance

The relative importance of the input variables was calculated for each of the chosen ANN architecture from the previous step using partitioning weight algorithm.

4.4.1 Delivery Type

Table 4.19 shows the relative importance for each input variable in the delivery type model, ranked in order from highest to lowest. As mentioned in section 4.3.1, the three-layer ANN with 5 hidden nodes was the ANN used for building this model. The ANN achieved sensitivity, specificity and CCR of 53.76%, 98.1% and 91.64% respectively (Table 4.13).

Variable	Rank	R-Importance
PUDENDAL	1	19.64%
GENERAL	2	13.35%
SPINAL	3	12.96%
NITOXIDE	4	5.82%
TERM	5	5.02%
PRESENT	6	4.53%
WEIGHT1	7	4.24%
GEST	8	3.99%
STEROIDS	9	3.26%
NARCOT	10	3.13%
EPIDURAL	11	2.96%
MONITOR	12	2.95%
NBABIES	13	2.91%
DELBY	14	2.45%
SMOKING	15	2.12%
HOSPTYPE	16	2.08%
LABTYPE	17	1.91%
GENDER	18	1.76%
MOMAGE	19	1.73%
PARITY	20	1.44%
SCALP	21	0.96%
MOMTRANS	22	0.78%
PRETERM	23	~0.0%
BF	24	~0.0%

Table 4.19: Relative importance for the input variables in delivery type model

4.4.2 Mortality

Table 4.20 shows the relative importance for each input variables in mortality model, ranked in order from highest to lowest. As mentioned before in section 4.3.2, the three-layer ANN with 5 hidden nodes was the ANN used for building this mortality model. It achieved sensitivity, specificity and CCR of 97%, 98.82%, and 98.83% respectively (Table 4.15).

Variable	Rank	R-Importance
NITOXIDE	1	3.90%
RESDRUGS	1	3.90%
MONITOR	1	3.90%
MOMTRANS	1	3.90%
GEST	1	3.90%
EPIDURAL	6	3.67%
RESPPV	7	3.55%
PARITY	7	3.55%
ASSISTED	7	3.55%
PRETERM	7	3.55%
NARCOT	11	3.34%
SMOKING	12	3.24%
RESCHEST	13	3.16%
APGAR5	13	3.16%
NEOTRANS	13	3.16%
WEIGHT1	13	3.16%
RESFFO2	17	2.75%
MOMAGE	17	2.75%
PRESENT	17	2.75%
TERM	17	2.75%
RESINTUB	21	2.59%
APGAR1	21	2.59%
STEROIDS	21	2.59%
NBABIES	21	2.59%
SCALP	25	2.27%
CORD	26	2.20%
BF	27	2.14%
GENERAL	28	2.08%
SPINAL	29	2.05%
PUDENDAL	30	1.98%
DELBY	30	1.98%
LABTYPE	30	1.98%
HOSPTYPE	30	1.98%
DELTYPE	30	1.98%
GENDER	35	1.39%

Table 4.20 Relative importance for the input variables in mortality model

4.4.3 Apgar5

Table 4.21 shows the relative importance for each input variable in Apgar5 prediction model, ranked in order from highest to lowest. As mentioned before in section 4.3.3, the three-layer ANN with 13 hidden nodes was the ANN used for building this Apgar5 model. It achieved sensitivity, specificity and CCR of 45.69%, 96.36%, and 95.27% respectively (Table 4.17).

Variable	Rank	Relative Importance
GEST	1	6.40%
SPINAL	2	5.89%
SCALP	3	5.54%
BF	4	5.44%
GENERAL	5	5.41%
WEIGHT1	6	5.13%
SMOKING	7	5.12%
PUDENDAL	8	4.70%
EPIDURAL	9	4.38%
NITOXIDE	10	4.03%
STEROIDS	11	3.72%
PRETERM	12	3.55%
MOMTRANS	13	3.51%
HOSPTYPE	14	3.40%
PARITY	15	3.30%
GENDER	16	3.26%
MONITOR	17	3.12%
LABTYPE	18	3.10%
ASSISTED	19	3.05%
PRESENT	20	2.83%
DELBY	21	2.74%
MOMAGE	22	2.69%
TERM	23	2.68%
DELTYPE	24	2.50%
NARCOT	25	2.48%
NBABIES	26	2.03%

Table 4.21 Relative importance for the input variables in Apgar5 model

4.5 Step 5: Variable Reduction

In this step the relative importance was used to find the minimum input variable set required to predict each of the three outcomes. The input variables with the lowest relative importance were removed one at a time until ANN performance degraded.

4.5.1 Delivery Type

Table 4.22 below shows the progress of input node reductions. The initial ANN consisted of 24 input nodes in the input layer. Each following reduction removed an input node with the lowest relative importance ratio in the previous reduction. The first two reductions were done by removing *bf and preterm* variables, which had the least relative importance (Table 4.19). The third to eighth reductions removed *momtrans, scalp, parity, age, gender and labtype* at relative importance less than 2%. Reductions were continued until there was a degradation of ANN performance. The removal of variable *hosptype* (ninth reduction) resulted in a degradation of ANN performance because the log-sensitivity index went below 0.15 for the fist time. Also, the sensitivity went below 50% for the first time: therefore, the decision was made to keep the *hosptype* variable in the model. It must be mentioned here that if reduction was continued the size of the input layer could be reduced to a very small number of input nodes number such as 7 nodes. This reduction resulted in a decrease of 4.18% sensitivity over the original ANN.

The remaining seven variables were: *pudendal*, *general*, *spinal*, *nitoxide*, *term*, *present* and *weight1*. Birth weight *weight1* can found within the C/S delivery risk factors determined by previous research studies [Yosef 1990]. Because of the difference in the perinatal database used in this thesis and the obstetrical databases used in the previously mentioned studies, the resultant risk factors were different. Some of the obstetric variables were not collected in this database. This work is the first in which the C/S delivery (outcome delivery type) was predicted using perinatal data. The new findings from the perinatal database refer to pain relievers taken by the mothers. Four out of six pain relievers were found to be risk factors for C/S delivery. Almost all mothers get pain therapy in labour (not just those with a problem). Knowing that pain therapy is an important factor of predicting delivery type may allow changes in pain management in labour to prevent the increasing ratio of C/S delivery in the region hospitals [Walker 2005]. The baby's position *present* during labour was expected to be a risk factor. Having a baby with a breech position during labour could increase the risk of C/S delivery.

Reduction	**Number of input nodes**	**Sensitivity**	**Specificity**	**CCR**
0	24	53.76%	98.10%	91.64%
1	23	51.81%	98.43%	91.64%
2	22	52.09%	98.31%	91.58%
3	21	50.70%	98.53%	91.56%
4	20	51.11%	98.60%	91.68%
5	19	51.67%	98.48%	91.58%
6	18	51.39%	98.43%	91.58%
7	17	51.11%	98.43%	91.53%
8	16	51.67%	98.48%	91.66%
9	15	49.30%	98.53%	91.35%
10	14	50.98%	98.53%	91.60%
11	13	49.72%	98.62%	91.49%
12	12	50.00%	98.53%	91.45%
13	11	50.56%	98.10%	91.17%
14	10	0.50836	98.65%	91.68%
15	9	51.11%	98.69%	91.76%
16	8	50.98%	98.46%	91.53%
17	**7**	**50.98%**	**98.53%**	**91.60%**
18	6	49.58%	98.67%	91.51%
19	5	38.58%	99.10%	90.28%
20	4	48.47%	98.74%	91.41%
21	3	48.33%	98.74%	91.39%
22	2	10.86%	99.64%	86.70%

Table 4.22: Variable reduction for delivery type model

4.5.2 Mortality

Table 4.23 shows the progress of input variable reductions. The initial ANN consisted of 35 input nodes in the input layer. The first reduction removed gender variable, which is the only variable with relative importance 1.39% (Table 4.20). The second to sixth reductions removed the *deltype, hosptype, labtype, delby* and *pudendal* variables at 1.98% relative importance ratio. The reduction process continued until reduction 28 (removal of *resppv* variable) which, resulted in a degradation of ANN performance because the log-sensitivity index and sensitivity went below 0.87 and 95% and never improved again. Keeping *reppv* in the list of the important factors leaded to keep *parity, assisted and preterm*. Theses three variables have exactly the same relative importance as *resppv* (ranked 7). Also, under the medical advice of the MIRG medical supervisor newborn weight *weight1* and Apgar score *apgar5* were kept in the final model. Keeping these two variables were resulted in keeping all the variable with the same or higher rank of relative importance, i.e. *narcot*, *smoking*, *reschest* and *neotrans* (Table 4.20). Reduction number 19 with 16 variables left in the model was selected to be the last variable reduction (Table 4.23). The resulted risk factors were: *nitoxide, resdrugs, monitor, momtrans, gest, epidural, resppv, parity, assisted, preterm, narcot, smoking, reschest, apgar5, neotrans* and *weight1*.

Comparing this prediction model with other mortality models is a long and difficult process. There are a large number of neonatal mortality models that have been found by researchers in perinatal and neonatal environments. Our risk model included: three maternal pain relievers, using nitrous oxide *nitoxide*, epidural *epidural* and narcotics *narcot*. This gave an alert for the doctors against using these pain-relieving methods, especially nitrous oxide, which came as the first risk factor in the list. Also, another alert concerns to the newborn resuscitation methods used. Three out of the five methods involved in the analysis, using resuscitation drugs *resdrugs* and using bag and mask ventilation *resppv* and using chest compression *reschest.* More attention must be paid to drugs used for resuscitation for a newborn as this came as the second risk factor after using nitrous oxide pain relief for the mother. Some interesting factors were, *monitor* and *momtrans*, these require further research to tell us how using different monitoring method can affect the baby's chances for survival. Also, the common reasons for inter-hospital transfer for the mother, *momtrans*, was found to be because either she is a critical case requiring more care and facilities this can only be provided by certain hospitals or is a home birth planned mother who decides to transfer to hospital because of a complication during delivery. Mother transfer can impact a critical delivery and the possibility of a critical newborn health condition. Variable *gest, preterm, apgar5* and *weight1* were expected to be in the model as it tests the maturity of the baby and the chance to survive and it was always there in other comparable models. As mentioned in section 2.2, weight, gestational age and Apgar score are risk factors of predicting mortality found by Gray et al. 1992 and SNAPP-II scoring model. Also, weight and gestational age are main factors in the CRIB risk index [Walker 2005].

Reduction	Number of input nodes	Sensitivity	Specificity	CCR
0	35	97.00%	98.82%	98.82%
1	34	90.00%	98.97%	98.94%
2	33	90.00%	99.12%	99.09%
3	32	90.00%	99.46%	99.42%
4	31	90.00%	98.66%	98.62%
5	30	95.00%	98.66%	98.64%
6	29	95.00%	98.23%	98.21%
7	28	95.00%	98.54%	98.53%
8	27	95.00%	98.71%	98.70%
9	26	90.00%	98.97%	98.94%
10	25	90.00%	99.16%	99.13%
11	24	95.00%	98.58%	98.57%
12	23	95.00%	98.10%	98.08%
13	22	95.00%	97.26%	97.27%
14	21	95.00%	98.25%	98.23%
15	20	95.00%	98.25%	98.23%
16	19	95.00%	99.35%	99.33%
17	18	95.00%	97.18%	97.19%
18	17	95.00%	98.21%	98.20%
19	**16**	**95.00%**	**98.45%**	**98.44%**
20	15	95.00%	98.41%	98.40%
21	14	95.00%	98.08%	98.07%
22	13	90.00%	99.33%	99.29%
23	12	95.00%	97.91%	97.90%
24	11	90.00%	99.24%	99.20%
25	10	90.00%	99.10%	99.07%
26	9	95.00%	97.03%	97.02%
27	8	95.00%	98.23%	98.21%
28	7	95.00%	99.03%	95.00%
29	6	90.00%	99.29%	99.26%
30	5	90.00%	99.29%	99.26%
31	4	85.00%	97.42%	97.38%
32	3	50.00%	86.77%	86.63%

Table 4.23: Variable reduction for mortality model

4.5.3 Apgar5

The initial ANN consisted of 26 input nodes in the input layer (Table 4.24). The first reduction removed *nbabies* variable, which had the lowest relative importance of 2.03% (Table 4.21). The second to fourteen reductions removed variables with relative importance less than 3.55%. Reduction fifteen removed *preterm* variable which had 3.55% relative importance ratio. The removal of *preterm* resulted in a clear degradation of the ANN performance (reduction 15). The sensitivity decreased to 31.8% for the first time (Table 4.24). Keeping *preterm,* resulted the following input set of 12 variables in the model: *gest, spinal, scalp, bf, general, weight1, smoking, pudendal, epidural, nitoxide, steroids* and *preterm.* As this was the first Apgar5 model to our knowledge, there is no previous model to compare with. Including previous preterm babies, *pretem,* and a low Apgar score, *apgar5*, risk factor was expected. The mother who had a premature baby before could face more problems in her current pregnancy and delivery than a mother who never had a premature baby. Also, having gestational age, *gest,* the most important factor was highly expected as it is a very important indicator of the baby's health condition. The preterm baby, with gestational age of less than 37 weeks, could have more medical problems than a term baby. A question about maternal pain relievers arises here, as five out of six pain relievers are in the list of predicting Apgar5. So, could the use of pain relievers seriously affect the newborn's condition? The only pain relief that is not included in the above list was *pudendal,* which had the second least relative importance (Table 4.21). This may signify a safer use of pudendal as a pain reliever, which may not affect the newborn condition as much as other drugs. It was not a surprising result to find *smoking,* even if they mother smoked very little after 20 weeks gestational age, the seventh important factor in the list. *Scalp* and *steroids* were expected to be in the list since having these done to a newborn indicates medical problems. The most surprising non-medical factor was *pf,* the intend of the mother to breast feed, and how it became the fourth important variable in the list. So, can the mother's decisions about her future baby affect the baby's health after birth? There are many questions that arise here and require more studying by future researchers. It is hoped that the list of

Apgar5's most important factors (risk factors) can help eliminate the risk factors in the future to positively affect the condition of newborns [Walker 2005].

Reduction	Number of input nodes	Sensitivity	Specificity	CCR
0	26	45.69%	96.36%	95.27%
1	25	41.38%	91.57%	90.49%
2	24	40.52%	97.08%	95.87%
3	23	44.83%	88.81%	87.86%
4	22	37.50%	97.77%	97.78%
5	21	37.07%	97.77%	96.46%
6	20	40.52%	97.20%	95.98%
7	19	38.79%	95.97%	94.74%
8	18	36.00%	96.67%	95.51%
9	17	39.66%	97.41%	96.16%
10	16	39.66%	97.41%	97.85%
11	15	38.00%	97.80%	96.57%
12	14	36.55%	97.42%	96.22%
13	13	35.54%	97.42%	96.22%
14	**12**	**36.80%**	**97.16%**	**95.94%**
15	11	31.80%	92.59%	91.55%
16	10	35.00%	95.17%	93.99%
17	9	36.50%	98.18%	96.96%
18	8	34.50%	96.87%	97.85%
19	7	34.50%	96.87%	95.64%
20	6	35.20%	97.54%	96.31%
21	5	36.50%	87.91%	87.04%
22	4	33.50%	91.29%	90.34%
23	3	32.00%	96.67%	95.40%
24	2	31.90%	97.63%	96.27%
25	1	31.89%	97.27%	95.94%

Table 4.24: Variable reduction for Apgar5 model

4.5.4 General Discussion

Combining the three risk factors, found in this section, together in an attempt to find the common risk factors that affect C/S delivery, neonatal mortality and low Apgar score resulted in the following observations:

Maternal pain relievers: there was always a direct relation between the above medical problems and maternal pain relievers. Using Nitrous oxide was common in the three models; using pudendal was the first risk factor in C/S delivery and eighth in low Apgar score. General pain relief was the second and fifth risk factor for C/S delivery and low Apgar score. Spinal was the third and second C/S and low Apgar risk factors. Epidural received the sixth and ninth positions in the mortality and low Apgar risk lists. Most of the C/S risk factors contained maternal pain relievers. Narcotics was the eleventh risk factors on mortality only. All pain relievers seemed to have an impact on the health problems under study in this thesis.

Gestational age and newborn weight were included in two of the three risk factors. Gestational age was found to be the first and fifth risk factors for low Apgar score and mortality respectively. Newborn weight was found to be one of the most important factors in predicting the three outcomes. This result was expected as most of the studies mentioned in section 2.2 identified newborn weight as a risk factor for delivery type, mortality and consequently Apgar5. Because Apgar score is a risk factor for predicting mortality, then the Apgar score risk factor could be added to the mortality risk factors in order to add more variables to the mortality predictions.

4.6 Step 6: Testing the Prediction Models

In this step, the prediction models built using the Perinatal Niday-2001 database will be tested with a data not presented in the training set. The Perinatal Updated-1999 database is used for this generalisation process. The best prediction models found in section 4.3 (4.3.1, 4.3.2 and 4.3.3) for delivery type, mortality and Apgar5 were tested first, followed by testing the three reduced models (models with only the most important variables) of section 4.5 (4.5.1, 4.5.2

and 4.5.3) for the three outcomes. The best ANN architectures used were the three-layer ANNs with 5, 5 and 13 hidden nodes for delivery type, mortality and Apgar5 respectively. As mentioned in section 3.2.6, *parity*, *delby* and *pudendal* variables of the Perinatal Niday-2001 database have no correspondent variables in the Perinatal Updated-1999 database. So, these three variables have been deleted from the testing of the best and reduced prediction models for the three outcomes using the Perinatal Updated-1999 database. The following are the results of the 1999 testing process for each outcome:

4.6.1 Delivery Type

Performance Measure	Best Model	Reduced Model
Test Sensitivity	48.88%	49.93%
Test Specificity	98.16%	97.39%
Test CCR	83.18%	82.96%
Test ASE	48.19%	51.64%
Test ROC	88.11%	84.93%
Test CP	70%	70%

Table 4.25: Testing delivery type prediction using Updated-1999 dataset

Table 4.25 shows the results of predicting delivery type using unseen dataset from the 1999 database. Comparing this table with the 2001 model from table 4.16, the sensitivity decreased by 3.93%. The constant predictor for the 1999 dataset used was 70%. The specificity and the CCR outperform the CP value, with higher Specificity than the 2001 model and less CCR of 8.05%. Testing the reduced model with the most important variables, the performance measures were stable with a sensitivity of 1.05% higher than the reduced model of the 2001 database (compared with Table 4.22).

4.6.2 Mortality

Performance Measure	Best Model	Reduced Model
Test Sensitivity	87.50%	87.50%
Test Specificity	98.85%	99.16%
Test CCR	98.78%	99.08%
Test ASE	4.48%	3.40%
Test ROC	93.26%	93.45%
Test CP	99.36%	99.36%

Table 4.26: Testing mortality prediction using Updated -1999 dataset

Table 4.26 shows the results of predicting mortality using 1999 dataset with 99.36 CP. This result was similar to the best prediction results of 2001 model from table 4.17 with a sensitivity increase of 3.5%. Testing the reduced model of the most important variables, the specificity and the CCR were similar to the 2001 results (compared with Table 4.23).

4.6.3 Apgar5

Performance Measure	Best Model	Reduced Model
Test Sensitivity	35.75%	43.02%
Test Specificity	94.65%	76.85%
Test CCR	92.54%	75.64%
Test ASE	27.11%	70.76%
Test ROC	70.77%	63.33%
Test CP	96.4%	96.4%

Table 4.27: Testing Apgar5 prediction using Updated -1999 dataset

Table 4.27 shows the results of predicting Apgar5 using 1999 dataset with 96.4% CP. This result is close to the prediction results of the 2001 model from table 4.18 with a decreased sensitivity of 4.5%. Using the reduced model, the sensitivity improved compared to the 2001 model but with a cost of 18% CCR (compared with Table 4.24).

The above generalization testing of the best and reduced models for each outcome was a very important step to judge the strength of theses models.

For the 1999 database, the performance of the best and reduced models of delivery Type and mortality were very similar to each other. But, for Apgar5 the best model achieved a better specificity and CCR than the reduced model while the reduced model achieved a better sensitivity. For Apgar5, difference between the least and the most important variables in the list (Table 4.21) was only 4.37% (the largest was 6.4 % and the least 2.03%). They did not have the clear difference in the relative importance which is found in Table 4.19 for determining the relative importance of delivery type input variables. In the case of delivery type, some variables had no relative importance (~0%) while the first variable in the list had a relative importance of 19%.

5 Conclusion

The main objective of this thesis was to identify the most important factors in predicting delivery type (vaginal or caesarean), neonatal mortality (within the first 28 days of the newborn's life) and Apgar5 (Apgar score 5 minutes after birth) using the Perinatal Niday-2001 database. This objective was achieved by using feedforward backpropagation ANN and the techniques discussed in Chapter 3 to build an ANN prediction model for each outcome. The risk factors for each outcome were determined to be:

1. Delivery Type: *pudendal, general, spinal, nitoxide, term, present* and *weight1*.
2. Mortality: *nitoxide, resdrugs, monitor, momtrans, gest, epidural, resppv, parity, assisted, preterm, narcot, smoking, reschest, apgar5, neotrans and weight1*.
3. Apgar5: *gest, spinal, scalp, bf, general, weight1, smoking, pudendal, epidural, nitoxide, steroids* and *preterm*.

The above three set of variables are the minimum number of factors needed to predict each of the three outcomes without degrading the best measured performance.

The comparison of the MIRG ANN tool with the BIOCORE fuzzy logic classifier showed little difference in performance in predicting the low occurrence cases with a high sensitivity.

The two approaches predicted the three outcomes with a similar CCR but the ANN predicted the positive cases with a slightly higher sensitivity; this means a Caesarean delivery, death of a newborn, or low Apgar5 score.

The three ANN models resulted in a classification rate similar or higher than the constant predictor of each outcome. Using the log-sensitivity index as a measure of the ANN performance helped in creating the three models with the best sensitivities and specificities.

The optimal ANN architecture and performance for each outcome were found to be:

- Delivery Type: Three-layer ANN with 5 hidden nodes in the hidden layer. The measures of performance for this ANN model were: sensitivity 52.8% ± 1.3%, specificity 97.9% ± 0.2%, CCR 91.3% ± 0.4%, ASE 29.6% ± 1 %, ROC 89% ± 0.6 % and CP 85.3%.

- Mortality: Three-layer ANN with 5 hidden nodes in the hidden layer. The measures of performance were: sensitivity 84% ± 6.5%, specificity 98.8% ± 0.1%, CCR 98.8% ± 0.1%, ASE 4% ± 0.5%, ROC 99.7% ± 0.1% and CP 96.7%.

- Apgar5: Three-layer ANN with 13 hidden nodes in the hidden layer. The measures of performance for this ANN model were: sensitivity 40.2% ± 3.2%, specificity 95.1% ± 0.8%, CCR 93.8% ± 0.8%, ASE 25.9% ± 10.1%, ROC 75.4% ± 3.2% and CP 97.8%.

The generalization ability of the optimal and simplified ANNs was measured using a Perinatal Niday-1999 datasets. The generalization results using new data that bad not been presented in the training sets assessed how well the ANNs performed. The following results show that:

- Delivery Type: The measures of performance for the optimal ANN model were sensitivity 48.9%, specificity 98.2%, CCR 83.2%, ASE 48.2%, ROC 88.1% and CP 69.6%. While for the simplified ANN the measures of performance were sensitivity 49.9%, specificity 97.4%, CCR 83%, ASE 51.6%, ROC 84.9% and CP 69.6%.

- Mortality: The measures of performance for the optimal ANN model were sensitivity 87.5%, specificity 98.9%, CCR 98.8%, ASE 4.5%, ROC 93.3% and CP 69.4%. While for the simplified ANN the measures of performance were sensitivity 87.5%, specificity 99.2%, CCR 99.1%, ASE 3.4%, ROC 93.5% and CP 69.4%.

- Apgar5: The measures of performance for the optimal ANN model were sensitivity 35.7%, specificity 94.7%, CCR 92.5%, ASE 27.1%, ROC 70.8% and CP 99.4%. While for the simplified ANN the measures of performance were sensitivity 43%, specificity 76.9%, CCR 75.6%, ASE 70.8%, ROC 63.3% and CP 99.4%.

5.1 Contributions

The main contributions of this thesis are:

1. Identified the risk factors for C/S delivery, neonatal mortality and low Apgar score and the commonality analysis of these risk factors.
2. This was the first use of ANN approach to investigate Apgar5, to our knowledge, as an outcome in any perinatal or neonatal database.

Other contributions are:

3. Identified an error on applying Garson's partitioning algorithm by Goh in 1995 and reformulated Garson's algorithm in simpler three equations which follow the original idea published by Garson 1991.

4. This was the first comparison of the MIRG ANN software and a fuzzy logic software. The similar results supported the choices of ANN by MIRG for predicting medical outcomes.

5.2 Future Work

As there is a huge amount of information in the Perinatal Niday-2001 database and in the similar perinatal databases collected every year by the PPPESO, this work can continue with many new possible research questions. For example:

Perinatal PPPESO Databases:

- Design another Apgar5 model with a new threshold value of 4 between the low and normal Apgar score (7 is the threshold used in this thesis).
- Compare the risk factors of neonatal mortality with other database. Different mortality models have been tried by MIRG on NICU database and are still under study by other researcher. Comparing the risk factor of theses two MIRG works will help in identifying mortality risk factors and improving the mortality model.
- Modify the cutoff point to see if it can improve the prediction of the three outcomes. For example, changing the mortality to include newborns who die within one week after the 28 days. This will increase the number of positive cases and reduce the skewness of the database. Other modifications include deleting the "in-

between" cases. In case of mortality, newborns who die between 28 days and two month of birth would not be included.

Statistical Approach:

- Using different statistical methods (such as logistic regression) to select the most significant variables and compare this with the most influential variables found by ANNs

- Designing ANN models using the most significant variables found by logistic regression and compare the result with the most important factors and performance found in this thesis.

- Designing logistic regression models using the most important factors found by ANN and compare the results with the results of this thesis.

Appendices

Appendix A: List of Variables of the Enhanced Database Computer Entry Program

This list includes instructions to the user on how each variable entered into the Enhanced Database Computer Entry Program. Notice that each data should be collected on each birth (not pregnancy). Therefore, if a multiple gestation, then each baby is entered separately [PPPESO 2001].

Baby No.

At the beginning of each month, number each birth, beginning with the number "1".

Postal Code

Enter the mother's complete postal code (L#L#L#). No space entered.

Birth Date

Indicate the baby's date of birth (MM/DD/YYYY). The calendar can be clicked and used to select a date.

Mother's Age

Enter the mother's age in completed years using the pick list. Do not round the number (e.g. 29 years and 7 months should be entered as 29 years). Range is 12-55 years.

Inter-Hospital Transfer

Record maternal transfers here. If the woman was transferred to your hospital to give birth, indicate where she was transferred from: No transfer, Almonte, Belleville, Brockville, Cornwall, Hawkesbury, Kingston, Montfort, Napanee, Ottawa Hospital – Civic Campus, Ottawa Hospital – General Campus, Pembroke, Perth Smiths Falls, Picton, Queensway Carleton, Renfrew, Winchester, Planned Home Birth, Outside Eastern Ontario.

No. Previous Term Babies

From the Gravidity section of the Obstetrical Record Logbook, enter the number listed under "T" (ensure you take the number from the correct column). Range is 0-15.

No. of Previous Preterm Babies

From the Gravidity section of the Obstetrical Record Logbook, enter the number listed under

"P" (ensure you take the number from the correct column). Range is 0-15.

Number of Babies

Indicate the number of babies (1, 2, 3, 4, 5, 6, > 6) in this pregnancy using the pick list.

Weeks

Indicate the number of completed weeks (gestational age). The number of completed weeks is entered twice in the computer program to ensure accuracy. Do not round the number (i.e. 32 weeks and 5 days = 32 weeks) and do not enter the number of days. Range is 18-45 weeks.

Labour Type

Indicate the type of labour: Spontaneous, Induced, No Labour.

Presentation

Indicate the type of presentation (Vertex, Breech, or Other, Not Available).

Monitoring

Indicate which monitoring methods were used (at any time during this admission): Unknown, No Monitoring, Auscultation only, Electronic only, Auscultation and Electronic.

Delivery

Indicate the type of delivery: Vaginal, VBAC, Primary C/S, Repeat with Trial of Labour, Repeat No Trial Labour.

Assisted with

Indicate whether the following were utilized to assist either a vaginal or cesarean birth: None, Forceps, Vacuum, Forceps and Vacuum.

Maternal Pain Relief

Indicate all methods of anesthesia and/or pain relief utilized for this birth: General , Epidural, Spinal, Narcotics, Nitrous Oxide, Pudendal

NOTE: Even if an epidural infusion contains narcotics, select only the epidural category. Select the narcotic category if the woman receives an IV, IM, or S/C dose of narcotics.

Antenatal Steroids

Indicate whether antenatal steroids were administered: None; 1 dose < 24 hours (before the time of birth); 2 doses: Last Dose < 24 hours (before the birth); 2 doses: Last Dose => 24 hours (from the time of the last dose to the time of birth). Note: the symbol "=>" means equal or greater than on the computer program.

Delivered By

Indicate who delivered the infant: Unknown, Physician, Midwife.

Baby's Sex

Indicate whether the baby is male or female. "Missing" should only be completed if ambiguous genitalia.

Weight

Indicate the weight of the baby in grams. Do not include decimals or commas. Do not round the number. Weight is entered twice to ensure accuracy.

Apgar 1, Apgar 5

Indicate the Apgar scores at 1 and 5 minutes: range is 0 to 10. Select "unknown" only if information is missing.

Scalp Blood Gases

Indicate whether scalp blood gases were done: Yes, No, Missing.

Cord Blood Gases

Indicate whether umbilical cord blood gases were done: Yes, No, Missing.

Newborn Resuscitation

Indicate all interventions that were utilized, regardless of duration:

FF02 (ie. free flow oxygen); PPV (i.e. bag and mask ventilation); Intubation; Chest Compression, Drugs (i.e. epinephrine, sodium bicarbonate or volume expanders).

Breastfeeding

Indicate whether the woman intends to breastfeed (regardless of whether any feeding occurs while in the LDR area): Yes, No, Missing.

Smoking

Indicate whether any maternal smoking has occurred at any time after 20 weeks gestation (regardless of the duration or amount): Yes, No, Missing.

Stillbirth/Neonatal Death

Indicate when and where the death occurred (if not at your hospital) from the following options:

- N/A (Not Applicable)
- Stillbirth (i.e. > 20 wks and/or 500 g)
- Neonatal death (i.e. death which occurs within the first 28 days of life):
- < 7 days of age at birth hospital
- < 7 days of age at transfer hospital
- 7-28 days of age at birth hospital
- 7-28 days of age at transfer hospital

Neonatal Transfer

Include infants transferred during the first 7 days of life only (not back transfers). Indicate which hospital received the infant: no transfer, Almonte, Belleville, Brockville, CHEO, Cornwall, Hawkesbury, Kingston, Montfort, Napanee, Ottawa Hospital – Civic Campus, Ottawa Hospital – General Campus, Pembroke, Perth Smiths Falls, Picton, Queensway Carleton, Renfrew, Winchester, Outside Eastern Ontario.

Comments

This section should be used to write information that is important such as maternal, fetal or neonatal complications or other details or clarifications.

Appendix B: Neonatal Scoring Systems

Apgar Score

Developed in 1952 by neonatologist Dr. Virginia Apgar, the Apgar score [Apgar 1953, Apgar et al. 1958] is measured at one, five and ten minutes after birth to quickly assess a newborn's initial condition. Table B.1 shows the parameters that are assessed and scored a value of 0 to 2, summing to a maximum score of 10. A score of 7 to 10 is normal (it is rare to receive a one-minute Apgar score of 10 because newborns are usually somewhat pale from the delivery process). Infants scoring between 4 and 6 may require some minor interventions, but will likely recover quickly. A score under 4 indicates that the baby may require immediate resuscitative measures, however, there is still a possibility of good outcome. Prematurity and medication given to the mother may account for a lower Apgar score [BabyCenter 2003][Ennett 2003].

Table B.1 Apgar scoring system

	0	1	2
Appearance	Pale	Blue	Pink
Pulse	Absent	<100	>100
Grimace	Absent	Grimace	Cry active
Activity	Limp	Some tone	Active
Respiration	Absent	Irregular	Regular and cry

CRIB Score

The Clinical Risk Index for Babies (CRIB) was developed retrospectively on a population of 735 infants whose birth weight was 1500g or less, or whose gestational age was less than 31 weeks [INN 1993]. The data was collected from four UK tertiary-level hospitals between 1988-1990. The CRIB score uses data collected in the first 12 hours of birth to predict in-hospital death, and excluded infants with inevitably lethal congenital malformations. Table A-2 defines the parameter ranges and their associated scores for this additive model. CRIB predicted death in the validation set with 51% sensitivity at 95% specificity. This score was developed before surfactant therapy became widely used [Ennett 2003].

Table B.2: CRIB scoring system

Risk factor		Score
Birth weight (g)	> 1350	0
	851-1350	1
	701-850	4
	≤ 700	7
Gestation (week)	> 24	0
	≤ 24	1
Congenital malformations (excluding inevitably lethal malformations)	None	0
	Not acute life-threatening	1
	Acutely life-threatening	3
Maximum base excess in first 12 h (mmol/L)	> -7.0	0
	-7.0 to –9.9	1
	-10 to –14.9	2
	≥ -15.0	3
Minimum appropriate FiO2 in first 12 h	≤ 0.40	0
	0.41-0.60	2
	0.61-0.90	3
	0.91-1.00	4
Maximum appropriate FiO2 in first 12 h	≤ 0.40	0
	0.41-0.60	1
	0.61-0.90	3
	0.91-1.00	5

NTISS Score

The Neonatal Therapeutic Intervention Scoring System (NTISS) [Gray et al. 1992] was based on the adult-ICU score called the Therapeutic Intervention Scoring System (TISS) [Cullen et al. 1974] to predict in-hospital mortality. The score was developed using data from 1643 newborns admitted to 3 NICUs between November 1989 and September 1990, and excluded moribund patients. A detailed description of the variables measured using NTISS is shown in Table B.3 [Ennett 2003].

Table B.3: NTISS scoring system

Item	Subscore
RESPIRATORY	
Supplemental oxygen	1^{a}
Surfactant administration	1
Tracheostomy care	1^{b}
Tracheostomy placement	1^{b}
Continuous positive airway pressure administration	2^{a}
Endotracheal intubation	2
Mechanical ventilation	3^{a}
Mechanical ventilation with muscle relaxation	4^{a}
High-frequency ventilation	4^{a}
Extracorporeal membrane oxygenation	4
CARDIOVASCULAR	
Indomethacin administration	1
Volume expansion (<= 15 mL/kg)	1^{c}
Vasopressor administration (1 agent)	2^{d}
Volume expansion (> 15 mL/kg)	3^{c}
Vasopressor administration (>1 agent)	3^{d}
Pacemaker on standby	3^{2}
Pacemaker used	4^{e}
Cardiopulmonary resuscitation	4
DRUG THERAPY	
Antibiotic (<=2 agents)	1^{f}
Diuretic (enteral)	1^{g}
Steroid (postnatal)	1
Anticonvulsant	1
Aminophylline	1
Other unscheduled medication	1
Antibiotic (> 2 agents)	2^{f}
Diuretic (parenteral)	2^{g}
Treatment of metabolic acidosis	3
Potassium binding resin	3
MONITORING	
Frequent vital signs	1
Cardiorespiratory	1
Phlebotomy (5-10 blood draws)	1^{h}
MONITORING (con't)	
Thermoregulated environment	1
Noninvasive oxygen	1
Arterial pressure	1
Central venous pressure	1
Urinary catheter	1
Quantitative intake and output	1
Extensive phlebotomy (>10 blood draws)	2^{h}
METABOLIC/NUTRITION	
Gavage feeding	1
Intravenous fat emulsion	1
Intravenous amino acid solution	1
Phototherapy	1
Insulin administration	2
Potassium infusion	3
TRANSFUSION	
Intravenous gamma globulin	1
Red blood cell (<=15 mL/kg)	2^{i}
Partial volume exchange	2
Red blood cell (>15 mL/kg)	3^{i}
Platelet	3
White blood cell	3
Double volume exchange	3
PROCEDURAL	
Transport of patient	2
Single chest tube in place	2^{j}
Minor operation	2^{k}
Multiple chest tubes in place	3^{j}
Thoracentesis	3
Major operation	4^{k}
Pericardiocentesis	4^{l}
Percardial tube in place	4^{l}
Dialysis	4
VASCULAR ACCESS	
Peripheral intravenous line	1
Arterial line	2
Central venous line	2

* Superscript letters represent mutually exclusive variables.

Table B.4: SNAP variable list and scoring system [Richardson et al. 1993a]

Variable		Points scored			
		0	1	3	5
Mean blood pressure	High	<66	66-80	81-100	>100
	Low	>35	30-35	20-29	<20
Heart rate	High	<180	180-200	201-250	>250
	Low	>100	80-100	40-79	<40
Respiratory rate	High	<60	60-100	>100	–
Temperature (°F)	Low	>96	95-96	92-94.9	<92
pO2	Low	>65	50-65	30-50	<30
pO2/FiO2 ratio	Low	>3.5	2.5-3.5	0.3-2.49	<0.3
pCO2	High	<50	50-65	66-90	>90
Oxygenation index	High	<0.07	0.07-0.2	0.21-0.40	>0.40
Hematocrit	High	<66	66-70	>70	–
	Low	>35	30-35	20-29	<20
White blood count	Low	>5.0	2.0-5.0	<2.0	–
Immature/total neutrophil ratio	High	<0.21	≥0.21	–	–
Absolute neutrophil count	Low	>999	500-999	<500	–
Platelet count	Low	>100	30-100	0-29	–
Blood urea nitrogen	High	<40	40-80	>80	
Creatinine	High	<1.2	1.2-2.4	2.5-4.0	>4.0
Urine output	Low	>0.9	0.5-0.9	0.1-0.49	<0.1
Indirect bilirubin	High				
- bili for birth weight>2kg		<15	15-20	>20	–
- bili/kg for birth weight<2kg		<5	5-10	>10	–
Direct bilirubin	High	<2.0	>2.0	–	–
Sodium	High	<150	150-160	160-180	>180
	Low	>130	120-130	<120	–
Potassium	High	<6.6	6.6-7.5	7.6-9.0	>9.0
	Low	>2.9	2.0-2.9	<2.0	–
Total calcium	High	<12	>12	–	–
	Low	>6.9	5.0-6.9	<5.0	–
Ionized calcium	High	<1.4	>1.4	–	–
	Low	>1.0	0.8-1.0	<0.8	–
Glucose	High	<150	150-250	>250	–
	Low	>40	30-40	<30	–
Serum bicarbonate	High	<33	>33	–	–
	Low	>15	11-15	<10	–
Serum pH	Low	>7.30	7.20-7.30	7.10-7.19	<7.10
Presence of seizures		None	Single	Multiple	–
Presence of apnea		none	Response to stimuli	No response to stimuli	Complete apnea
Stool guaiac		Negative	Positive	–	–

* Additional points scored for the Perinatal Extension (SNAPPE) are:

Birth weight ≤ 749 g	30 points
Birth weight 750-999 g	10 points
Apgar < 7 at 5 minutes	10 points
Small for gestational age (<5th percentile)	5 points

Appendix C: Description of the Perinatal Niday-2001 Database

IDNUM	Baby's birth identification number	I	Beginning with the number 1
DOB	Baby's date of birth	T	(MM/DD/YYYY)
GENDER	Indicates baby's sex	B	1=male, 2=female
BF	Indicates weather the women intends to breastfeed	B	1=yes, 2=no
SMOKING	Indicates whether any maternal smoking has occurred at any time after 20 weeks gestation	B	1=yes, 2=no
CORD	Indicates whether umbilical cord blood gases were done	B	1=yes, 2=no
SCALP	Indicates whether scalp blood gases were done	B	1=yes, 2=no
GENERAL	Indicates whether maternal pain relief: General was used	B	1=yes, 2=no
EPIDURAL	Indicates whether maternal pain relief: Epidural was used	B	1=yes, 2=no
SPINAL	Indicates whether maternal pain relief: Spinal was used	B	1=yes, 2=no
NARCOT	Indicates whether maternal pain relief: Narcotics was used	B	1=yes, 2=no
NITOXIDE	Indicates whether maternal pain relief: Nitrous Oxide was used	B	1=yes, 2=no
PUDENDAL	Indicates whether maternal pin relief: Pudendal was used	B	1=yes, 2=no
RESFFO2	Indicates whether newborn resuscitation: FF02 was used	B	1=yes, 2=no
RESPPV	Indicates whether newborn resuscitation: PPV was used	B	1=yes, 2=no
RESINTUB	Indicates whether newborn resuscitation: Intubation was used	B	1=yes, 2=no
RESCHEST	Indicates whether newborn resuscitation: Chest Compressions was used	B	1=yes, 2=no
RESDRUGS	Indicates whether newborn resuscitation: Drugs was used	B	1=yes, 2=no

DELBY	Indicates who delivered the baby: physician or midwife	B	1=physician, 2=midwife
MOMAGE	Age of the mother	I	Range:12-55
PARITY	Parity	I	Range: 0-15
APGAR1	Apgar score at 1 min	I	Range: 0-10
MONITOR	Monitoring methods	I	1=No monitoring 2=Auscultation only 3=Electronics only 4=Auscultation and electronics
LABTYPE	Labour type	I	1=Spontaneous 2=Induced 3=No Labour 4=Not Available
PRESENT	Type of presentation	I	1=vertex 2=breech 3=other
ASSISTED	Assisted with forceps, vacuum or none	I	1=none 2=forceps 3= vacuum 4=forceps and vacuum
STEROIDS	Indicate whether antenatal steroids were administered	I	1=None 2=1 dose < 24 hours (before birth) 3=2 doses: Last Dose < 24 hours (before birth) 4=2 doses: Last Dose => 24 hours (from the time of the last dose to birth).
NEOTRANS	Neonatal transfer during the first 7 days of life	I	Indicate which hospital received the infant: no transfer, Almonte, Belleville, Brockville, CHEO, Cornwall, Hawkesbury, Kingston, Montfort, Napanee, Ottawa Hospital – Civic Campus, Ottawa Hospital – General Campus, Pembroke, Perth Smiths Falls, Picton, Queensway Carleton, Renfrew, Winchester, Outside Eastern Ontario.
MOMTRANS	Mother inter-hospital Transfer	I	If the woman was transferred to the hospital to give birth. Indicate where she was transferred from: No transfer, Almonte, Belleville, Brockville, Cornwall, Hawkesbury, Kingston, Montfort, Napanee, Ottawa Hospital – Civic Campus, Ottawa Hospital – General Campus, Pembroke, Perth Smiths Falls, Picton, Queensway Carleton, Renfrew, Winchester, Planned Home Birth, Outside Eastern Ontario.
HOSPTYPE	Hospital type	I	1=Teaching Hospitals 2=Large Community Hospitals 3=Small Community Hospitals
TERM	No. previous term babies	I	range: 0-15
PRETERM	No. previous preterm babies	I	range: 0-15
NBABIES	Number of babies in this pregnancy	I	1,2,3,4,5,6,>6

WEIGHT1	Baby's weight	I	in grams
GEST	Gestational Age	I	range: 18-45 weeks
DELTYPE	Delivery Type	I	+1= Primary C/S, repeat C/S with trial labour or repeat C/S with no trial labour -1=vaginal or VBAC
NEODEATH	Neonatal Death	I	+1=Neonatal death(i.e. death which occurs within the first 28 days of life) -1= N/A (i.e. survivor)
APGAR5	Apgar score at 5 mins	I	range: 0-10 +1=from 0-7 (low Apgar score) -1=from 7 to 10 (high Apgar score)

Table C.1: Niday-2001 variables and definitions

Appendix D: Description of the Perinatal Niday-1999 Database

IDNUM	Baby's birth identification number	I	Beginning with the number 1
YEAR	Baby's year of birth	I	Number
MONTH	Baby's month of birth	I	Range: 1-12
DAY	Baby's day of birth	I	Number
GENDER	Indicates baby's sex	B	1=male, 2=female
BF	Indicates weather the women intends to breastfeed	B	1=yes, 2=no
SMOKING	Indicates whether any maternal smoking has occurred at any time after 20 weeks gestation	B	1=yes, 2=no
CORD	Indicates whether umbilical cord blood gases were done	B	1=yes, 2=no
SCALP	Indicates whether scalp blood gases were done	B	1=yes, 2=no
GENERAL	Indicates whether maternal pain relief: General was used	B	1=yes, 2=no
EPIDURAL	Indicates whether maternal pain relief: Epidural was used	B	1=yes, 2=no
SPINAL	Indicates whether maternal pain relief: Spinal was used	B	1=yes, 2=no
NARCOT	Indicates whether maternal pain relief: Narcotics was used	B	1=yes, 2=no
NITOXIDE	Indicates whether maternal pain relief: Nitrous Oxide was used	B	1=yes, 2=no
RESFFO2	Indicates whether newborn resuscitation: FF02 was used	B	1=yes, 2=no
RESPPV	Indicates whether newborn resuscitation: PPV was used	B	1=yes, 2=no
RESINTUB	Indicates whether newborn resuscitation: Intubation was used	B	1=yes, 2=no
RESCHEST	Indicates whether newborn resuscitation: Chest Compressions was used	B	1=yes, 2=no
RESDRUGS	Indicates whether newborn resuscitation: Drugs was used	B	1=yes, 2=no
MOMAGE	Age of the mother	I	Range:12-55

APGAR1	Apgar score at 1 min	I	Range: 0-10
MONITOR	Monitoring methodes	I	1=No monitoring 2=Auscultation only 3=Electronics only 4=Auscultation and electronics
SPONTDEL	Spontaneous labour	B	T=yes, F=no
LABTYPE	Labour type	I	1=Spontaneous 2=Induced 3=No Labour 4=Not Available
PRESENT	Type of presentation	I	1=vertex 2=breech 3=other
FORCEPS	Assisted with forceps	B	T=yes, F=no
VACUUM	Assisted with vacuum	B	T=yes, F=no
STEROIDS	Indicates whether antenatal steroids were administrated	I	0=None 1=1 dose < 24 hours (before birth) 2=2 doses: Last Dose < 24 hours (before birth) 3=2 doses: Last Dose => 24 hours (from the time of the last dose to birth).
NEOTRANS	Neonatal transfer during the first 7 days of life	I	1=N/A 2=CHEO 3=Ottawa Hospital: General 4=Kingston General 5=Other Hospital 6=Ottawa Hospital: Civic
MOMTRANS	Mother inter-hospital transfer	I	1.N/A 2.Metro 3.Region 4.Outside Eastern Ontario
HOSPTYPE	Hospital type	I	1=Teaching Hospitals 2=Large Community Hospitals 3=Small Community Hospitals
TERM	No. previous term babies	I	Range: 0-15
PRETERM	No. previous preterm babies	I	Range: 0-15
NBABIES	Number of babies in this pregnancy	I	1,2,3,4,5,6,>6
WEIGHT1	Baby's weight	I	in grams
GEST	Gestational age	I	Range: 18-45 weeks
PRIMCS	Primary C/S	B	T=yes, F=no
TOL	Trial of labour	B	T=yes, F=no
VBAC	Vaginal birth after C/S	B	T=yes, F=no
REPEATCS	Repeated C/S	B	T=yes, F=no

BIRTHMOD	Birth mode	B	1=Vaginal 2=C/S delivery
NEODEATH	Neonatal death	I	+1=Neonatal death(i.e. death which occurs within the first 28 days of life) -1= N/A (i.e. survivor)
APGAR5	Apgar score at 5 mins	I	Range: 0-10 +1=from 0-7 (low Apgar score) -1=from 7 to 10 (high Apgar score)

Table D.1: Niday-1999 variables and definitions

Appendix E: ROC Curves of Best-Performing Networks

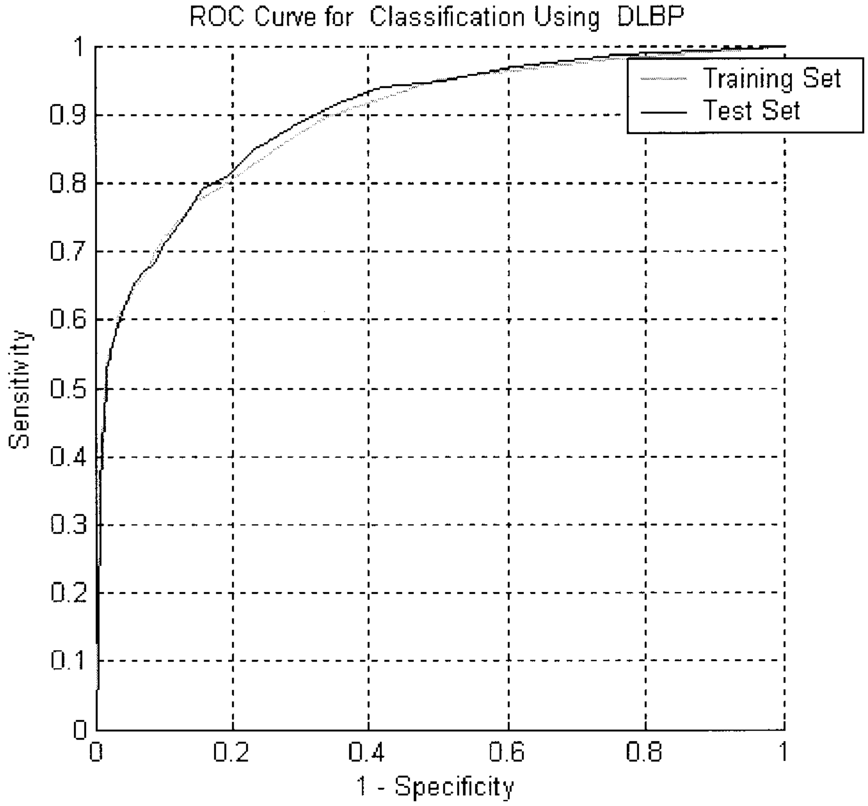


Figure E.1: The ROC curve of the best-performing network for delivery type prediction

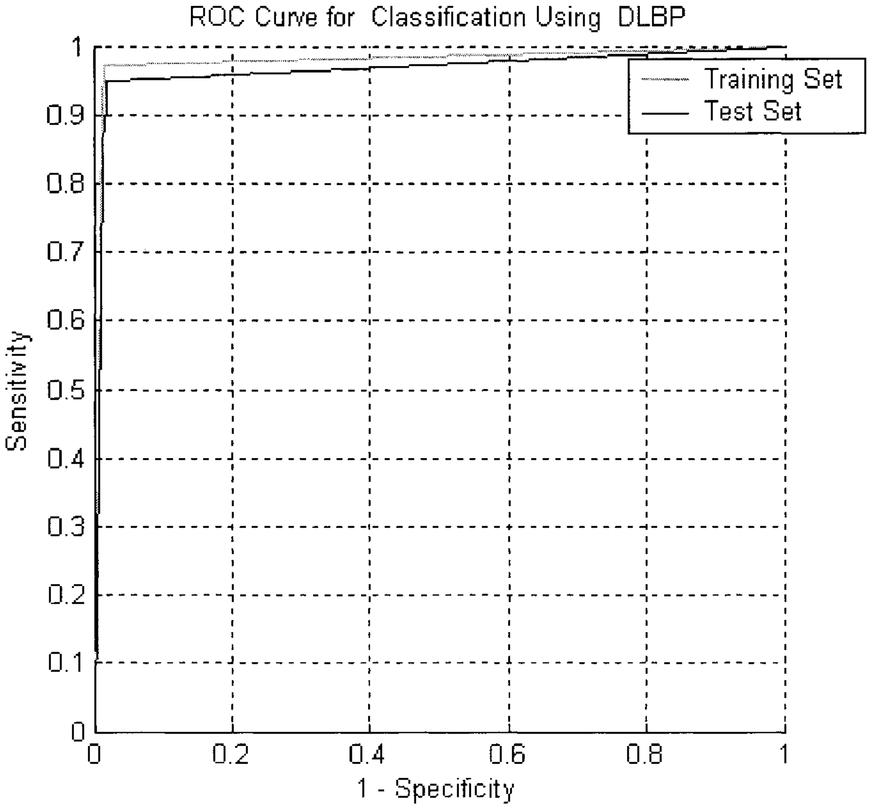


Figure E.2: The ROC curve of the best-performing network for mortality prediction

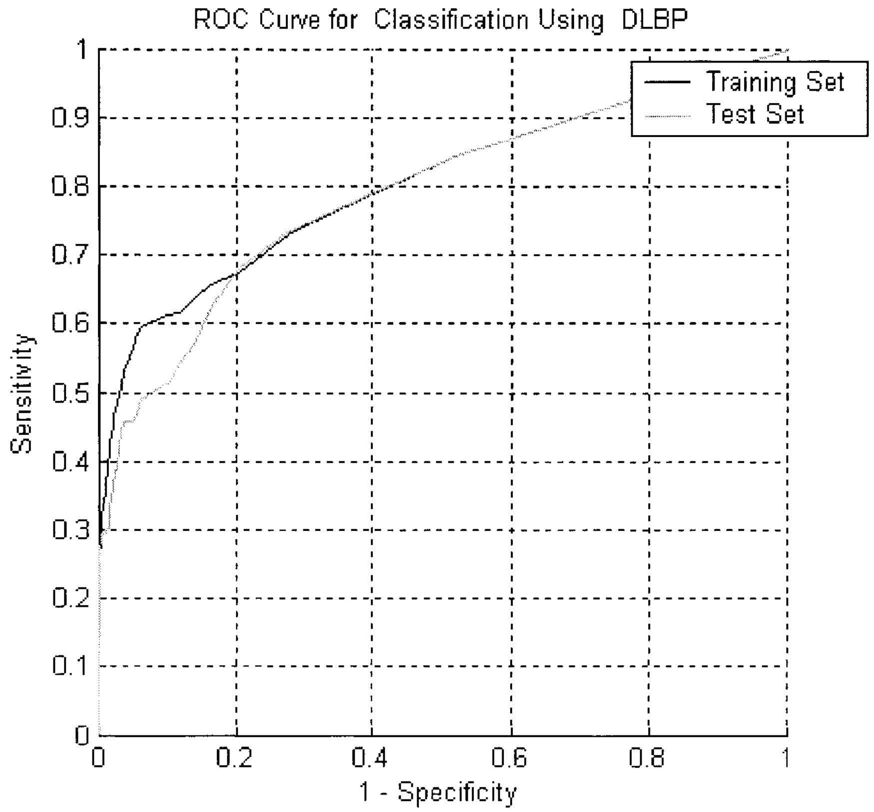


Figure E.3: The ROC curve of the best-performing network for Apgar5 prediction

References

[Ambalavanan & Carlo 2001]
Ambalavanan N, Carlo WA, "Comparison of the prediction of extremely low birth weight neonatal mortality by regression analysis and by neural networks," *Early Human Development*, vol. 65, pp. 123-137, 2001.

[Anderson 1995]
Anderson, JA, *An Introduction to Neural Networks*. Cambridge: MIT Press, MA 1995.

[Apgar 1952]
http://www.alexian.org/progserv/babies/thirdtri/apgarscore.html

[Apgar 1953]
Apgar V, "A proposal for a new method of evaluation of the newborn infant," *Curr Res Anesth Analg*, vol. 32, pp.260-267, 1953.

[Apgar et al. 1958]
Apgar V, Holaday DA, James LS, Weisbrot IM, Berrien C, "Evaluation of the newborn infant: second report," *JAMA*, vol. 168, pp. 1985-1988, 1958.

[BabyCenter 2003]
BabyCenter, "Labor and delivery: The Apgar score."
http://www.babycenter.com/refcap/3074.html.

[Baxt & Skora 1996]
Baxt WG, Skora J, "Prospective validation of artificial neural network trained to identify acute myocardial infraction," *Lancet*, vol. 347, pp. 12-15, 1996.

[Blum 1992]
Blum A, Neural *Networks in C++: An Objective Oriented Framework for Building Connectionist Systems*. New York: John Wiley & Sons, 1992.

[Burns et al. 1995]
Burns LR, et al, "The effect of physician factors on the caesarean section decision," *Med Care*, vol. 33, pp. 365–382, 1995.

[Buskard et al. 1994]
T. Buskard, M. Stevenson, M. Frize, and F. Solven, "Estimation of ventilation, length of stay and mortality using artificial neural networks," *Proceedings of the 1994 Canadian Conference on Electrical and Computer Engineering (CCECE '94)*, pp. 726-729, Halifax, Nova Scotia, September, 1994.

[Charette 2000]

Charette E, "Automation of an ANN used to estimate medical outcome," *Project in computer science*, School of Information Technology and Engineering, University of Ottawa, 2000.

[Chien et al. 2002]

Chien LY, Whyte R, Thiessen P, Walker R, Brabyn D, Lee SK, "The Canadian Neonatal Network. SNAP-II predicts severe intraventricular hemorrhage and chronic lung disease in the neonatal intensive care unit," *J Perinatol*, vol. 22, pp. 26-30, 2002.

[Cullen et al. 1974]

Cullen DJ, Civetta JM, Briggs BA, Ferrara LC, "Therapeutic intervention scoring system: a method for quantitative comparison of patient care," *Crit Care Med*, vol. 2, no. 2, p.p 57-60, Mar-Apr 1974 .

[Demuth & Beale 2001]

Demuth H, Beale M, *Neural Network Toolbox User's Guide, Version 4*. Natick: The MathWorks Inc, MA 2001.

[Ennett 1999]

C.Ennett, "Coronary surgery mortality prediction using artificial neural networks," *Master's Thesis*, University of Ottawa, 1999.

[Ennett et al. 2001]

Ennett CM, Frize M, Walker CR, "Influence of missing values on artificial neural network performance," *Medinfo*, vol.10, no. Pt 1, pp. 449-453, 2001.

[Ennett et al. 2002]

Ennett CM, Frize M, Scales N, "Logarithmic-sensitivity index as a stopping criterion for neural networks," *Proc IEEE EMBS-BMES*, Texas, 2002.

[Ennett 2003]

Ennett CM, "Imputation of missing values by integrating artificial neural networks and case-based reasoning," *PhD thesis*, Electrical Engineering, Dept. of Systems and Computer Engineering, Carleton University, Ottawa, ON, May 2003.

[Ennett et al. 2003]

Ennett CM, Frize M, Scales N, "Evaluation of the logarithmic-sensitivity index as a neural network stopping criterion for rare outcomes," Proc IEEE ITAB, 2003.

[Ennett et al. 2004]

Ennett CM, Frize M., Charette E, "Improvement and automation of ANNs to estimate medical outcomes," *Medical Engineering &Physics,* vol. 26, pp. 321-328, 2004.

[Fausett 1994]

Fausett L, *Fundamentals of Neural Network*. Englewood Cliffs (NJ): Prentice-Hall, 1994.

[Frize et al. 1995]

Frize M., Solven F., Stevenson M., Nickerson B., Buskard T., Taylor K.,"Computer-Assisted Decision Support Systems for Patient Management in an Intensive Care Unit," *Medinfo*, pp. 1009-1012, Vancouver, July 1995.

[Frize et al. 1996]

M. Frize, F. Solven, Stevenson M, Nickerson BG, McGowan H., "Information Technologies Approach and Development for Various medical Applications," *Proceedings of the Canadian Conference on Electrical and Computer Engineering (CCECE '96)*, pp.365-368, 1996.

[Frize et al. 1997]
M. Frize, Trigg H, F. Solven, Stevenson M, Nickerson BG, "Decision-Support Systems Designed for Critical Care," *Proceedings of the 1997 AMIA Fall Symposium*, pp. 855,1997.

[Frize et al. 2000]
Frize M, Enter CM and Charette E, "Automated Optimisation of the Performance of Artificial Neural Networks to Estimate Medical outcomes, " *Proc. 3rd ITAB Conference (Information Technology Applications in Biomedicine) and ITIS (International Telemedical Information Society)*, pp. 168-173, Washington, Nov 2001.

[Frize et al. 2001a]
Frize M, Ennett CM, Stevenson M., Trigg HCE, "Clinical decision- support systems for intensive care units using ANNs," *Med Eng Physics,* vol. 23, no. 3, pp. 217-25, 2001.

[Frize et al. 2001b]
Frize M, Ennett CM, Hebert P, "Improving the efficiency and effectiveness of artificial neural networks as decision-support systems," *Medinfo*, vol. 10, no. Pt 1, pp. 574, 2001.

[Frize et al. 2004]

Frize M, Ibrahim D, Seker H, Walker RC, Odetayo MO, Petrovic D, Naguib RNG, "Predicting clinical outcomes for newborns using two artificial intelligence approaches," *Proc. of the 26th Annual IEEE EMBS 2004 Conference*, San Francisco, CA, USA, pp. 3202-3205, September, 2004.

[Garson 1991]
Garson, G.D., "Interpreting neural-network connection weights," *Artif. Intell. Expert*, vol. 6, pp. 47–51, 1991.

[Goh 1995]
Goh, A.T.C. "Back-propagation neural networks for modeling complex systems," *Artif. Intell. Eng.*, vol. 9, pp. 143–151, 1995.

[Gray et al. 1992]
Gray JE, Richardson DK, McCormick MC, Workmann-Daniels K, Goldmann DA, "Neonatal therapeutic intervention scoring system: a therapy-based severity-of-illness index," *Pediatrics* vol. 90, no.4, pp. 561-7, Oct 1992.

[Ibrahim 2002]
Ibrahim D, "Mortality prediction using logarithmic-sensitivity index as a performance measure for ANN," *Project in Information System Science*, School of Mathematics and Statistics, Carleton University, 2002.

[INN 1993]
International Neonatal Network, "The CRIB (clinical risk index for babies) score: a tool for assessing initial neonatal risk and comparing performance of neonatal intensive care units", *Lancet* vol. 342, pp.193-198, 1993.

[James et al. 2000]

James K.S., Aitken Ian, Subramanian S.V, "Neonatal mortality in India: emerging paradoxes," *Harvard School of Public Health*, vol. 10, 2000.

[Krogh & Hertz 1992]
Krogh A, Hertz JA. "A simple weight decay can improve generalization," *In: Lippmann RP, Moody J, Touretzky, eds. Advances in Neural Information Processing Systems 4 (NIPS'91). Morgan Kaufmann: San Matteo, CA,* pp. 950-957, 1992.

[Kuncheva 2000]
Kuncheva LI, *Fuzzy Classifier Design*. Heidelberg: Springer-Verlag, 2000.

[Lenard 1995]
M.J. Lenard, P. Alam, G.R. Madey, "The application of neural networks and a qualitative response model to the auditor's going concern uncertainty decision," *Decision Sciences*, vol 26, no.2, pp. 209-227,1995.

[Livingstone & Manallack 1993]
Livingstone DJ, Manallack DT, "Statistics using neural networks chance effects," *J Med Chem.* vol. 36, pp. 295-97,1993.

[McDowell et al. 2001]
McDowell M., Somoza E., Rothe K., Fry R., Brady K. and Bocklet A, "Understanding birthing mode decision making using artificial neural networks," *Med Decis Making,* vol. 21, pp. 433-443, 2001.

[Masters 1993]
Masters T, *Practical Neural Network Recipes in C++*. San Diego, CA: Academic Press, 1993.

[MATLAB]
MATLAB built in help disk, release 13.

[Olden & Jackson 2002]
Olden J., Jackson D., " Illuminating the 'black box': a randomization approach for understanding variable contributions in artificial neural networks," Department of Zoology, University of Toronto, *Ecological Modelling*, 154:135–150, 2002.

[Penny & Frost 1996]
Penny W, Frost D, "Neutral networks in clinical medicines," *Med Decis Making*, vol. 16, pp. 386-398, 1996.

[PPPESO]
The mission of the Perinatal Partnership Program of Eastern Ontario. www.pppeso.on.ca

[PPPESO 2001]
The PPPESO data entry form and list of variables used to enter the data into the Perinatal Niday-2001 database. PPPESO, 2001.

[PPPESO 2002]
Annual perinatal statistical report 2001.PPPESO June, 2002.

[PPPESO 2005]
PPPESO strategic plan 2001-2005. www.pppeso.on.ca/english/strategicplan_2005.pdf

[Richardson et al. 1993]

Richardson DK, Gray JE, McCormick MC, Workmann K, Goldmann DA, "Score for neonatal acute physiology: a physiologic severity index for neonatal intensive care," *Pediatrics*, vol. 91, no. 3, pp. 617-623, Mar [1993a].

[Rumelhart 1986]
Rumelhart DE, Hinton GE, Williams RJ, "Learning representations by backpropagating errors," *Nature,* vol. 323, issue B, pp.533-536, 1986.

[Sargent 2001]
Sargent DJ, "Comparison of artificial neural networks with other statistical approaches – Results from medical data sets," *Cancer,* vol. 91, no. 8, pp. 1636-1642, 2001.

[Sarle 2005]
Sarle W, "Frequently Asked Question (FAQ) List," March 2005. URL: ftp://ftp.sas.com/pub/neural/.

[Scales 2001]
Scales N, " Mortality prediction in an ICU and coronary surgery database using neural networks," *Term project*, Department of Systems and Computer Engineering, Carleton University, Dec 2001.

[Seker 2003]
Seker H, Odetayo MO, Petrovic D, Naguib RNG, " A fuzzy logic based-method for prognostic decision making in breast and prostate cancers," *IEEE Transactions on Information Technology in Biomedicine*, vol. 7, no. 2, pp. 114-122, 2003.

[Stevens et al. 1994]
Stevens SM, Richardson DK, Gray JE, Goldmann DA, McCormick MC, "Estimating neonatal mortality risk: an analysis of clinicians' judgements," *Pediatrics* vol. 93, no. 6, pp. 945-950, 1994.

[Tong 1983]
Tong H, *Threshold Models in Nonlinear Time Series Analysis.* Springer, 1983.

[Tong et al. 2002]
Tong Y., Frize M., and Walker R, "Extending successful predictions of ventilation needs using artificial neural networks from adult to neonatal ICU patients," *IEEE Transaction on Information Technology in Biomedicine*, vol. 6, no. 2, pp. 188-191, June 2002.

[Trigg 1997]
Trigg HCE, "An investigation of methods to enhance the performance of artificial neural networks used to estimate ICU outcomes," *Master's thesis*, Department of Electrical Engineering, University of New Brunswick, Fredericton, NB, Jan 1997.

[Trigg et al. 1997]
Trigg H., Stevenson M., Frize M, "Estimating ventilation requirements in critical care medicine," *Proceedings of the World Congress on Biomedical Engineering and Physics*," pp. 14-19, Nice, France, September 1997.

[Yosef 1990]
Mor-Yosef S, Samueloff A, Modan B, Navot D, Schenker JG, "Ranking the risk factors for cesarean : logistic regression analysis of a nationwide study," *Obstetrics & Gynecology,* vol. 75, pp. 944-947, 1990, http://medinfo.psu.ac.th/smj2/181/smj1813.html.

[Walker et al. 1999]
Walker R, Frize M, Tong Y, "Data analysis using artificial neural networks and computer-aided decision-making in the neonatal intensive care unit," *Canadian Paediatric Society*, Annual Meeting, Paediatrics and Child Health, Winnipeg, June 1999.

[Walker et al. 2003]
Walker RC, Ennett CM, Frize M, "Use of artificial neural networks to identify the contribution of different physiologic variables to neonatal mortality in newborn infants admitted to neonatal intensive care units," *Pediatr Res,* vol. 53, pp. 444A, 2003.

[Walker 2005]
Walker RC, Chief Neonatologist, Children's Hospital of Eastern Ontario (CHEO), Ottawa Hospital – General Campus, Ottawa, ON, *Personal communication*, August 2005.

[Weigend et al. 1990]
Weigend AS, Rumelhart DE, Huberman BA, "Back-propagation, weight-elimination and time series prediction," *Proc 1990 Connectionist Models Summer School,* Morgan Kaufmann: San Mateo, pp. 105-116, 1990.

[Weigend et al. 1991]
Weigend AS, Rumelhart DE, Huberman BA. "Generalization by weight-elimination with application to forecasting. In RP Lippmann, J Moody, DS Touretzky, eds. Advances in Neural Information Processing Systems (NIPS'90), Volume 3. Morgan Kaufmann: San Mateo. 1991;3:875-882.

[Zernikow et al. 1998]
Zernikow B, Holtmannspoetter K, Michel E, Pielemeier W, Hornschuh F, Westermann A, Hennecke KH, "Artificial neural network for risk assessment in preterm neonates," *Arch Dis Child Fetal Neonatal,* vol. 79, pp. F129-F134, 1998.