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

Hui Ouyang
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

Comparing Relative Weight Reduction and Principal Component Analysis Methods Applied to Perinatal Outcome Classification ANN Models

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

M. Yagoub

Gary W. Slater
Le Doyen de la Faculté des études supérieures et postdoctorales / Dean of the Faculty of Graduate and Postdoctoral Studies

Comparing Relative Weight Reduction and Principal Component Analysis Methods Applied to Perinatal Outcome Classification ANN Models

By

Hui Ouyang

B.A.Sc. Electrical Engineering, University of Ottawa

Thesis submitted to the

Faculty of Graduate and Postdoctoral Studies

In partial fulfillment of the requirements for the degree of

Master of Applied Science in Electrical Engineering

Ottawa-Carleton Institute for 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: 978-0-494-25814-9
Our file Notre référence
ISBN: 978-0-494-25814-9

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

Acknowledgements

I would like to express sincere appreciation to my supervisor Dr. Monique Frize for her on-going guidance and support. She has given me the opportunity to be part of this very exciting and interesting research group.

I would also like to thank Dr. Robin Walker for his feedback on this research.

Many thanks go to all other MIRG members for their friendship and help.

Special thanks to my friend Lan Yang for her encouragement and entertainment.

Most importantly, I express my thanks and appreciation to my parents, my mother Huichang Situ and my father Liao Ouyang, for their love – the endless love forever.

Lastly, but in no sense the least, I am thankful to all friends and colleagues who made my stay at the university a memorable and valuable experience.

Abstract

This thesis compares relative weight reduction (RWR) and principal component analysis (PCA) variable selection methods applied to perinatal outcome classification artificial neural network (ANN) models. The two perinatal outcomes classified using ANN models are Apgar score at 5 minutes after baby birth and vaginal or cesarean section delivery type from Niday Perinatal Database. The important factors of determining the perinatal outcomes are found separately using RWR and PCA methods applied to ANN tools. It is found that RWR performs better than PCA on specificity measures and it has the advantage of keeping individual indicator data information. The PCA method performs better on sensitivity measures and it is suitable for large input data noise reduction. The Niday 2001 models are also verified in a five member committee of classifiers using Niday 2004 unseen data. The performance measures show that the Niday 2001 models are sufficient in classifying the desired outcomes.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ANN : Artificial neural network

ASE : Average squared error

CABG : Coronary artery bypass grafting

CBR : Case-based reasoning

CCR : Correct classification rate

CCVT : Committee of classifiers verification tool

CDSS : Clinical decision support system

CHEO : Children's Hospital of Eastern Ontario

CHIRPP : Canadian Hospitals Injury Reporting and Prevention Program

CNN : Canadian neonatal network

CP : Constant predictor

DECH : Doctor Everett Chalmers Hospital

ICA : Independent component analysis

ICU : Intensive care unit

KNN : K-nearest neighbor

MIRG : Medical Information-technology Research Group

MLP : Multilayer perceptron

NICU : Neonatal intensive care unit

PADS : Parent decision support system

PCA : Principal component analysis

PONI-Web : Perinatal, obstetrical, and neonatal intensive care web

PPPESO : Perinatal Partnership Program of Eastern and Southeastern Ontario

RFW	:	Research framework
ROC	:	Receiver operating characteristic
RWR	:	Relative weight reduction
SFHI	:	San Francisco Heart Institute
SNAP	:	Score for neonatal acute physiology
SNAPPE	:	Score for neonatal acute physiology with perinatal extension
SNAP-II	:	Score for neonatal acute physiology version 2
SNAPPE -II	:	Score for neonatal acute physiology with perinatal extension version 2
XML	:	Extended markup language

1 Introduction

1.1 Motivation

Information technology holds enormous potential to improve the quality, safety and efficiency of health care through electronic medical records and medical data mining. Collaboration between health care and information technology experts would certainly be valuable for the general population. Significant benefits of medical information infrastructure include better quality of care, improved process efficacy, and reduced wait-time and cost.

The challenges of the health care information technology industry are mainly the ways to collect, store, display, and analyze medical information. Clinical decision support systems encompass a range of computing software tools that provide clinical practitioners with medical data analysis to facilitate medical decision making. Data mining using computing technology in health care can help to define trends in illness, to define important indicators for future prevention, and to provide real time decision support for current activities.

Artificial neural networks (ANNs) have been shown to be accurate data mining instruments in several medical domains. The Medical Information-technology Research Group (MIRG) of University of Ottawa and Carleton University has recognized the advantage of research in this area. MIRG advocates the appropriate use of engineering methods to solve clinical problems and demonstrated the performance of artificial intelligence tools in predicting clinically significant outcomes [28]. MIRG researchers have used the ANN applications written in

MATLAB to develop various classification models to determine the most important predicting variables for clinical outcomes.

One of the medical information technology research initiatives within MIRG is to explore different data classification algorithms for medical data. MIRG members have applied various approaches such as artificial neural network, radial basis function network and Bayesian network to analyze medical data. MIRG members also have examined some statistical analysis methods such as logistic regression and Bayes' theorem to assist in predicting outcomes related to certain clinical conditions. This thesis broadens MIRG's research on applying different data variable selection methods into ANN classification models. The strength and weakness of RWR and PCA methods are studied from theoretical and practical perspectives. The analysis of the results will provide evidence of a superior hybrid method in ANN for MIRG's comprehensive clinical decision support system.

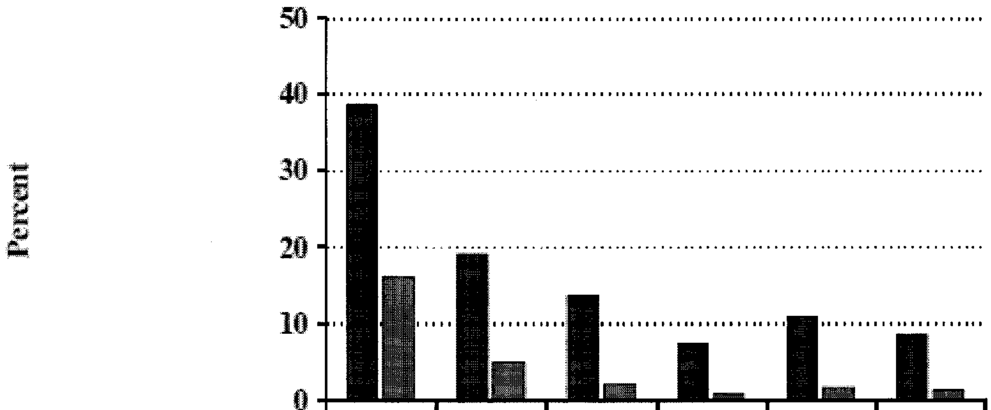
Another important research initiative within MIRG is developing ANN models for perinatal health domain. The perinatal period is traditionally defined as the period around childbirth, commonly referred to the period of five months before and one month after birth. However, it has been extended to include the preconception period, with the recognition that health status prior to pregnancy also influences maternal and newborn outcomes. Considering the importance of perinatal information extraction and usage, the Perinatal Partnership Program of Eastern and Southeastern Ontario (PPPESO) began collecting perinatal data 25 years ago.

1.2 Medical Background

PPPESO works with hospitals, health departments, community agencies, academic institutions, private practitioners, and consumers to effectively link perinatal care, education, and research. As a regional program, PPPESO works with its partners to identify issues, develop and implement solutions and produce results that improve evidence based regionalized 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 analyzing data, dissemination of information, communication, advice, facilitation and education [54].

The Niday Perinatal Database was developed under the direction of PPPESO. It consists of massive amounts of information covering many aspects of perinatal care. Two of the important indicators collected by PPPESE are of interest in this thesis: APGAR5 (Apgar score at 5 minutes after birth) and DELTYPE (vaginal or cesarean section delivery type).

Apgar score is an indication of the health of the baby at 1 and 5 minutes after the birth, with a maximum score of 10. A score of less than 7 at 5 minutes is associated with a higher risk of health and development problems for the baby. About 1 in 6 very preterm babies (< 33 weeks) experience Apgar score less than 7 at 5 minutes after birth. This vulnerable population requires early developmental screening so that timely interventions, such as physiotherapy, speech and language and occupational therapy, are in place during the opportune periods of brain development in the early years. Among term babies (37-41 weeks), only 1 in 100 experiences Apgar score less than 7 at 5 minutes [57].

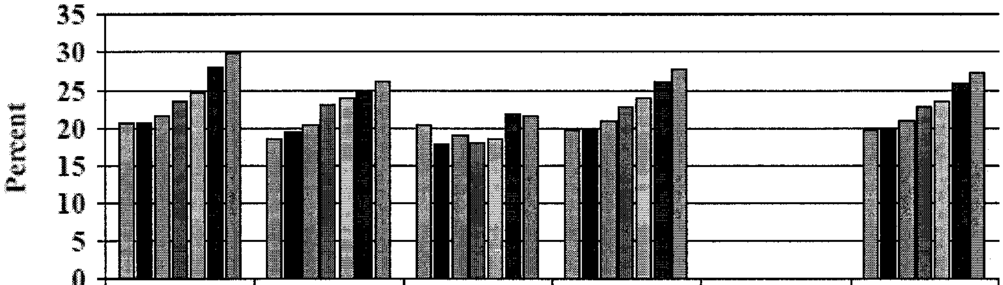


	<33 weeks	33-34 weeks	35-36 weeks	37-41 weeks	>41 weeks	Total
■ APGAR <7 at 1 minute	38.5	19.1	13.5	7.2	10.7	8.5
▨ APGAR <7 at 5 minutes	16.1	4.9	2.1	0.8	1.7	1.3

* Data provided by 64 hospitals and 9 midwifery groups across Ontario.

Figure 1-1: Proportion of babies born in Ontario* with Apgar score < 7 at birth, by gestational age, July – December, 2003 [57].

A cesarean birth is performed when labour is not progressing well, or if there is maternal or fetal compromise. A higher proportion of cesarean section operations took place in teaching hospitals, perhaps reflecting the higher risk profile of their population. In general, the province of Ontario is experiencing an increased trend in births to women of later maternal age and obstetrical interventions including cesarean birth over the last seven years [56].



	Teaching	Large Community	Small Community	All Hospitals	Home Births	All Births
▨ 1998	20.7	18.5	20.5	19.7	NA	19.7
■ 1999	20.7	19.5	17.9	19.9	NA	19.9
▨ 2000	21.5	20.5	18.9	20.9	NA	20.9
▩ 2001	23.5	22.9	18.0	22.8	NA	22.8
□ 2002	24.6	23.8	18.5	23.8	0.0	23.5
■ 2003	27.9	24.8	21.9	26.2	0.0	25.8
▨ 2004	29.8	26.2	21.6	27.7	0.0	27.3

* Eastern and Southeastern Ontario

Figure 1-2: Proportion of babies who had a cesarean birth in the region, by place of birth *, 1998 – 2004 [56].

The compelling importance of Apgar score 5 and delivery type indicators in perinatal health and the identified MIRG goal of applying ANN to health care domain lead to the study of developing artificial neural networks applied to perinatal outcome classification. The classification models can provide analytical information to clinicians when they encounter difficult decisions regarding the care of newborns and mothers. These analyses can facilitate the health care professionals to deliver better patient care throughout the health care environment, from prognosis to diagnosis to treatment care management.

1.3 Thesis Objectives

The objective of this thesis is to compare the relative weight reduction (RWR) and principal component analysis (PCA) methods applied to perinatal outcome classification ANN models. The variable selection method which provides better performance measures from Niday year 2001 database is chosen to verify the ANN models on the unseen Niday year 2004 database using committee of classifiers verification tool (CCVT). The two perinatal outcomes of interest are APGAR5 (Apgar score at 5 minutes after baby's birth) and DELTYPE (vaginal or cesarean section delivery type). In doing so, the adequacy of ANNs in classifying data information across multiple periods over a large geographical region for perinatal health care is demonstrated. Based on the results of analysis, comparisons are made to show the consistency and accuracy of the classification ability of the ANN models. These ANN models are feasible to be used by health care practitioners as a causation model to identify the importance of the perinatal variables.

In order to achieve the main objective, this thesis has the following key focuses:

- To implement and evaluate the data reduction capability of RWR and PCA variable selection methods by examining the performance measures of the ANN classification models.
- To identify the important factors for classifying perinatal outcomes APGAR5 and DELTYPE using ANN research framework (RFW) tool.
- To verify that ANN models developed from previous database from Niday year 2001 database can estimate the APGAR5 and DELTYPE perinatal outcomes on unseen Niday year 2004 database with similar data distribution.
- To study the use of constructing a multiple member committee of classifiers for ANN models utilizing the committee of classifiers verification tool (CCVT).

The long term significance of this thesis for MIRG research is to incorporate the ANN models into MIRG's web services infrastructure for perinatal, obstetrical and neonatal clinical decision support. This leads towards building a comprehensive clinical decision support system for medical information management and for medical outcomes and complications prediction in maternal and newborn health.

1.4 Thesis Outline

This dissertation is organized into five chapters as followed:

- **Chapter 1: Introduction** provides the motivation, medical background, and objectives of this thesis.
- **Chapter 2: Literature Review** describes the concepts and theories used in this thesis. There are brief comparisons on existing perinatal models on the two interested outcomes, followed by the introduction of Niday Perinatal Database. This chapter also includes the detailed studies of ANN techniques and variable selection methods, such as the relative weight reduction and principal component analysis. Then a review of the previous work by MIRG researchers is presented. The different MIRG ANN applications and development results are described in this section.
- **Chapter 3: Methodology** outlines the research methods and procedures of this thesis. The detailed data preparation steps are given along with the reasons of doing so. Some data pre-processing statistics are presented. A brief overview of the software tools used in this work is given as well. Then the break down of the meanings and manipulation

formulas of each performance measures are described in detail. A summary of research steps is listed at the end of the chapter.

- **Chapter 4: Result Analysis** presents the results of ANN experiment simulation for outcomes APGAR5 and DELTYPE, including the Niday 2001 database applied relative weight reduction method and Niday 2001 database applied principal component analysis algorithm, and Niday 2004 database using committee of classifiers verification tool.
- **Chapter 5: Conclusion** summarizes the concluding remarks, contributions and future work to build upon this thesis.

2 Literature Review

2.1 Perinatal Models in Literature

Improving maternal and child health depends on a continuum of skilled care through pregnancy, childbirth and childhood to the families. A strong multidisciplinary collaboration between healthcare professionals and the supportive participation of communities are required to ensure the effective improvements in quality of care. Examples of success have shown that the collaboration between health practitioners and scientists/engineers can help in developing systematic integrated approaches for high quality antenatal and perinatal health [3]. Several up-to-date perinatal models using different analysis techniques are discussed in this section.

The perinatal, obstetrical and neonatal periods are well connected in the maternal and child healthcare environment. Obstetrical health relates to the care of women during and after pregnancy, while neonatal health concerns the newborn during the first month after birth. The perinatal environment is situated between these two environments, which pertains to the mother and baby care of five months before and one month after birth. Apgar score and delivery type are two of the important outcomes in perinatal health studies.

2.1.1 Apgar Score

The Apgar score is a simple and repeatable method to quickly and summarily assess the health of newborns immediately after childbirth. The Apgar score [Appendix A] is determined by evaluating the newborn baby on five simple criteria (Appearance, Pulse, Grimace, Activity, and

Respiration) on a scale from zero to two and summing up the five values. The resulting Apgar score ranges from 0 to 10, and the test repeats at 1 minute, 5 minutes and 10 minutes after delivery [2]. The Apgar test result can be used individually as a health scoring system for newborns, and it is also a measuring variable in some other neonatal scoring systems.

Apgar score is a strong predictor in estimating mortality risk. Thus, to include Apgar scores over time as a risk factor in a scoring system offers enhanced mortality prediction abilities, as combining Apgar score with some other factors accounts for all difference in severity of illness adequately. For instance, several mortality risk models that are used in the neonatal intensive care unit include Apgar score at 5 minutes as a contributing variable towards the points scoring, such as SNAPPE-II. SNAP-II [59] is a modified SNAP (Score for Neonatal Acute Physiology) [58] model to assign points to each patient case according to the variable's deviation from normal physiological values (either above or below normal values) to achieve an arithmetic sum that estimates the infant's mortality risk. The six SNAP-II variables plus birth weight, small for gestational age and 5-minute Apgar score compose SNAP-Perinatal Extension-II (SNAPPE-II) [59]. Richardson's study [59] showed that the area under the receiver operating characteristic curve for the SNAPPE-II score was 0.91±0.01 with the best classification performance for the group of infants of all birth weights.

So far, there is no stand-alone predicting model for Apgar score in the medical environment. Apgar tests take into account five basic aspects of the infant at birth, and the score results rely heavily on the vital data observations involved human interpretations. Developing a systematic interpretation based on the measured data for each patient will be useful to automatically assess the risk of complication after delivery. It is valuable to separately estimate the Apgar score which indirectly predicts infant mortality.

2.1.2 Delivery Type

The decision on choosing the appropriate delivery type (vaginal or cesarean delivery) is based not only on scientific understanding but also on other factors, such as the practices of obstetricians, the preferences of mothers, and community standards. Physicians believe that a model consisting of predictor variables in determining whether a cesarean section needs to be performed on a given individual should include the following criteria [46]:

- Maternal clinical characteristics present at time of labor;
- Baby clinical characteristics;
- Non-clinical factors such as practitioner characteristics.

Logistic regression is one of the most commonly used statistical techniques in medical data analysis. This technique models the log odds of an outcome defined by the values of covariates in the model. The collected information (input variables) about the actual outcomes of patients is used to calculate regression coefficient parameters to model the results. A study [49] using multivariate stepwise logistic regression analysis indicated that the most important risk factors that influenced the decision to perform a cesarean section were the presentation of the fetus and the presence of a uterine scar, followed in descending order by placenta previa or abruption placentae, maternal disease, primiparity, low birth weight, twins, and advanced maternal age.

An artificial neural network is a powerful data modeling tool that is able to capture and represent complex input/output relationships. Another study [46] examined the important factors in affecting obstetricians' decisions to whether or not to perform cesarean section showed that artificial neural networks can be constructed to accurately predict the birthing mode decisions of clinicians. The variables kept in the final ANN model in ranked order is summarized in Table 2-1. The final model had good predictive accuracy for the birthing mode giving the classification accuracy of 83.5%, area under the ROC of 0.924, sensitivity at the information-maximizing threshold of 0.882, and specificity at the information-maximizing threshold of 0.821.

Table 2-1: Independent predictor variables in birthing mode ANN model [46].

Ranking (highest to lowest)	**Variables**
1	Mother had medical condition necessitating cesarean section
2	Gravida of mother
3	Arrest of labor
4	Cephalopelvic disproportion
5	Physician opinion of epidural use influence
6	Some degree of failed induction
7	Preterm rupture of membranes
8	Prior poor obstetrical outcome
9	Prior parity
10	Labor was induced
11	Some degree of fetal distress
12	Some degree of fetal prematurity
13	Malpresentation of baby
14	Premature rupture of membranes
15	Baby's birth weight (in grams)
16	Level of patient cooperation
17	Impact of cesarean section on room availability
18	Extent of mother's opinion about birth mode
19	Mother expected a cesarean section birth
20	Day of week
21	Timing of birth – nursing shift 11 PM to 7 AM
22	Physician judged that mother had mode expectations
23	Mother had attitude toward mode

Although some delivery type prediction models have been developed from different databases in perinatal health, no two models show the same data set for statistically significant factors. Several risk factors are either statistically found or clinically selected. Due to geographic disparity, it is a researchable approach to discover the commonality and distinctness of each database over various periods of time.

2.1.3 MIRG Models

The Medical Information-technology Research Group (MIRG) is a multidisciplinary research group whose members are investigating a number of promising information technology approaches to assist with decision making and patient management in the critical care environment. The MIRG researchers have explored the application of artificial intelligence tools to the fields of neonatal intensive care and perinatal health. Since the MIRG work is based on the confluence of engineering and medicine, the results are advantageous and beneficial to both engineering and healthcare fields. The previous works by MIRG led to the initiate of applying the feedforward back-propagation ANN to estimate the perinatal outcomes of this thesis work.

Some models developed by MIRG were shown to be reliable in predicting newborn outcomes. Ennett worked on a neural network with the weight elimination cost function and the log-sensitivity stopping criterion to develop a new neonatal mortality model that identified the most influential risk factors for predicting mortality in the neonatal intensive care unit. The Canadian Neonatal Network (CNN) database consisting 5,102 cases with 30 variables was reduced to a 13-input model. The resultant model classified the patients equally well or better than the statistically-based models in the literature [28]. Ibrahim analyzed the Perinatal Partnership Program of Eastern and Southeastern Ontario (PPPESO) database, consisting of 17,688 cases with 38 variables, to predict newborn mortality. Again, a weight elimination cost function neural network was used to reduce weights and find variables that had the most influence on the outcomes. A variable reduction technique was used whereby the input variables with the lowest relative importance was removed one at a time until the performance of the network degraded. Sixteen input variables remained [40]. For comparison, the risk factors ranked according to the impact that they had on mortality prediction for each ANN model is given in Table 2-2. The optimal ANN performance for each prediction model is shown in Table 2-3.

Table 2-2: Comparison of risk factors for mortality prediction models.

Ranking (highest to lowest)	**Neonatal Mortality Model By Ennett [15]**	**Newborn Mortality Model By Ibrahim [40]**
1	Po2/FiO2 ratio	Maternal pain relief: nitrous oxide
2	Lowest urine output	Newborn resuscitation: drugs
3	Lowest serum pH	Monitoring methods
4	Apgar score at 5 minutes	Mother inter-hospital transfer
5	Lowest platelet count	Gestational age
6	Small for gestational age status	Maternal pain relief: epidural
7	Highest sodium concentration	Newborn resuscitation: PPV
8	Highest respiratory rate	Parity
9	Highest pCO2 reading	Assisted with forceps, vacuum or none
10	Birth weight	Number of previous preterm babies
11	Lowest glucose concentration	Maternal pain relief: narcotics
12	Lowest temperature	Maternal smoking
13	Highest blood pressure	Newborn Resuscitation: chest compress
14		Apgar score at 5 minutes
15		Neonatal transfer
16		Baby's weight

Table 2-3: Comparison of performance measures for mortality prediction models.

Performance measures	**Neonatal Mortality Model By Ennett [15]**	**Newborn Mortality Model By Ibrahim [40]**
Sensitivity	38.3%	84.0%
Specificity	98.9%	98.8%
CCR	96.8%	98.8%
ASE	N/A	0.0397
ROC	0.8699	0.997

By comparing the two infant mortality models developed by MIRG researchers, Apgar score at 5 minutes, gestational age, and baby's birth weight are the common factors in ANN prediction. These factors were also found to be important in other infant outcome predictions from literature. Note that Apgar score is the only interpreted input variable from various tests. It is beneficial to apply the machine intelligence methods to assist the health practitioners in estimating the Apgar score. Meanwhile, Ennett's neonatal model includes mostly the baby's vital data after birth, while Ibrahim's newborn model contains more data on the baby's resuscitation and mother's pain relief during birth. Therefore, this thesis work focused on examining the ANN tool applied on Apgar score estimation rather than going in-depth of mortality prediction. However, more studies on perinatal mortality that combines and analyzes the overall conditions of both babies and mothers may be performed in the future.

In addition to newborn mortality prediction, Ibrahim also used the same database and ANN method to develop the predictive models for delivery type (cesarean section delivery or vaginal delivery) and Apgar5 score (low Apgar score or high Apgar score at 5 minutes) [40]. Ibrahim's optimal ANN models resulted in a classification rate higher than a CP. The results also showed that the MIRG ANN tool predicted more of the positive cases more accurately than the BIOCORE's fuzzy tool [26]. Further comparison of findings between this thesis and Ibrahim's models are discussed in Section 4.1.1 and Section 4.1.2.

Due to the direct correlation of neonatal infant mortality with obstetrical and perinatal care, MIRG has recently extended the research of developing ANN models in predicting preterm birth and newborn outcomes from the Perinatal Partnership Program of Eastern and Southeastern Ontario (PPPESO) database. Catley [10] achieved a maximum sensitivity of 36.6% using a very limited 8 variable obstetrical input set with a feedforward back-propagation ANN with weight elimination in predicting preterm birth. The results showed that ANNs are a potentially useful clinical decision support tool of using data readily available to the physician during obstetrical care.

2.2 Niday Perinatal Database

2.2.1 History

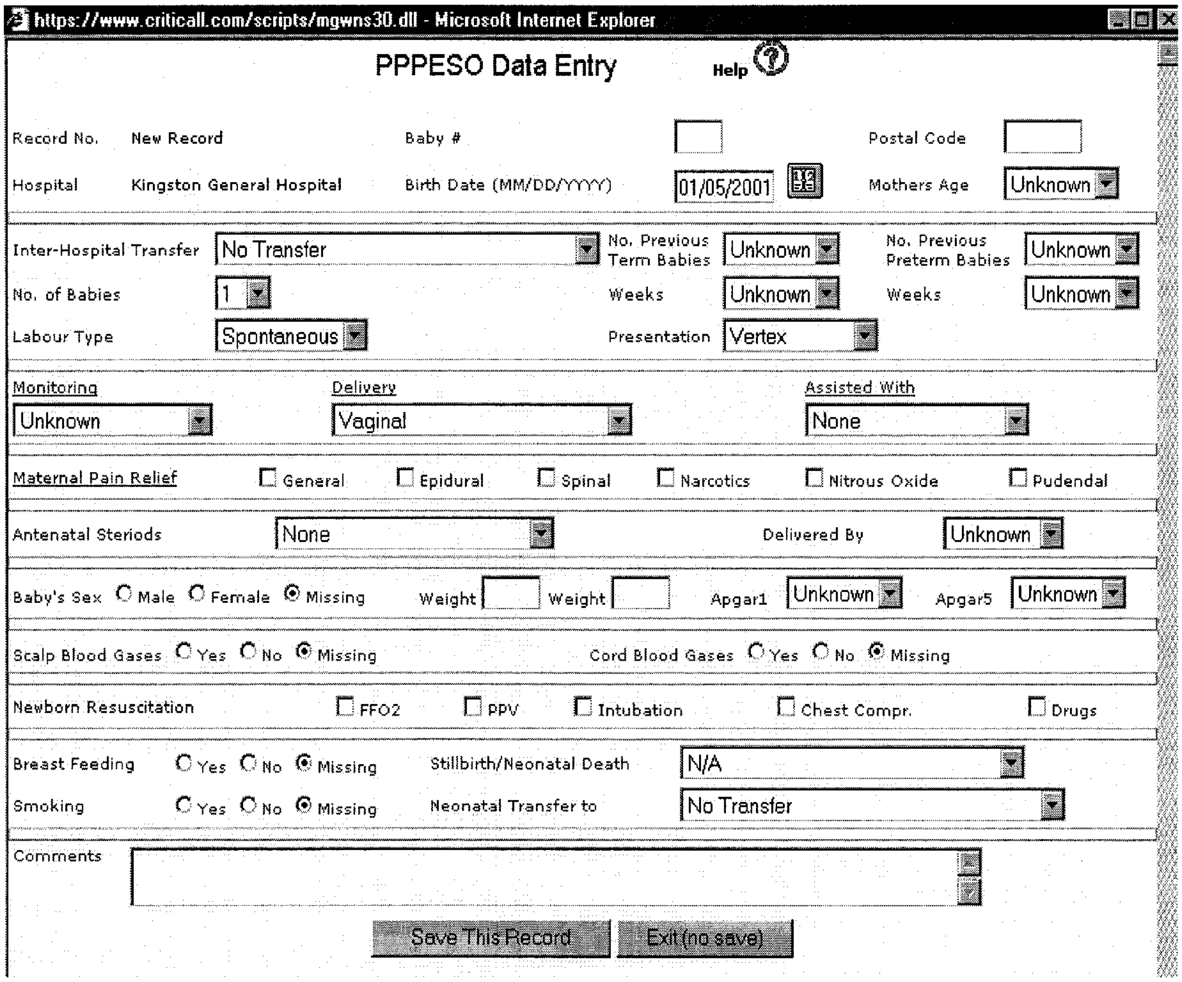
The Perinatal Partnership Program of Eastern and Southeastern Ontario (PPPESO) effectively links perinatal health care, education and research from working together with hospitals, health departments, community agencies, academic institutions, and private practitioners [54]. Under the direction of the PPPESO, the Niday Perinatal Database was implemented in 1997 with the collaboration of 37 partners of PPPESO (hospitals, health units, Health Information Partnership (HIP), community health centers, 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, which was formerly known as the Eastern and Southeastern Ontario Perinatal Database.

In January 2001, the Niday Perinatal Database was enhanced from a stand-alone computer program installed in each hospital, to the web-based CritiCall System. Instead of using a computerized data entry program to record information about each baby born in the hospital from the case room logbook, hospitals now enter data directly into the database and can generate reports independently. These changes have greatly enhanced its usefulness by improving the reporting of data and retrieval of information. The database is being used to support a growing number of planning, evaluation and research initiatives throughout the region [57].

2.2.2 Database Contents

The Niday Perinatal Database contains the profile of the mother and information about the newborn at birth. The main variables recorded by the Niday Perinatal Database reporting form are outlined below [Appendix B]. A screenshot of the electronic data entry form is shown in Figure 2-1.

- **Baby #:** the baby's record number
- **Postal Code:** the mother's complete postal code
- **Birth Date:** the baby's date of birth
- **Mother's Age:** the mother's age in years
- **Inter-Hospital Transfer:** where the mother was transferred from
- **No. Previous Term Babies:** number of previous term babies for the mother
- **No. Previous Preterm Babies:** number of previous preterm babies for the mother
- **No. of Babies:** the number of babies in this pregnancy
- **Weeks:** gestational age, the number of completed weeks
- **Labour Type:** the type of labour (Spontaneous, Induced, No Labour)
- **Presentation:** the type of presentation (Vertex, Breech, Other, Not Available)
- **Monitoring:** which monitoring methods were used (Unknown, No Monitoring, Auscultation Only, Electronic Only, Auscultation and Elctronic)
- **Delivery:** the type of delivery (Vaginal, VBAC, Primary C/S, Repeat with Trial of Labour, Repeat No Trial Labour)
- **Assisted With:** whether utilized to assist either a vaginal or cesarean birth (None, Forceps, Vacuum, Forceps and Vacuum)
- **Maternal Pain Relief:** methods of anesthesia and/or pain relief utilized for birth (General, Epidural, Spinal, Narcotics, Nitrous Oxide, Pudendal)
- **Antenatal Steroids:** whether antenatal steroids were administered
- **Delivery By:** who delivered the baby (Unknown, Physician, Midwife)
- **Baby's Sex:** the baby's gender (Male, Female, Missing)
- **Weight:** the baby's weight in grams
- **Apgar1:** the Apgar score at 1 minute
- **Apgar 5:** the Apgar score at 5 minutes
- **Scalp Blood Gases:** whether scalp blood gases were done
- **Cord Blood Gases:** whether cord blood gases were done
- **Newborn Resuscitation:** interventions that were utilized regardless of duration (FFO2, PPV, Intubation, Chest Compression, Drugs)
- **Breast Feeding:** whether the mother intends to breastfeed
- **Smoking:** whether any maternal smoking has occurred at any time after 20 weeks gestation
- **Stillbirth/Neonatal Death:** when and where the death occurred
- **Neonatal Transfer:** where the baby transferred during the first 7 days of life only
- **Comments:** additional information


https://www.criticall.com/scripts/mgwns30.dll - Microsoft Internet Explorer

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: Niday Perinatal Database entry screen [54].

A Niday Perinatal Database user guide [Appendix B] is prepared to ensure consistency of definitions for each variable among the hospitals. In addition, each hospital has received training to manage the systems data entry and reporting capabilities. At the end of each year, the data is forwarded to the PPPESO data analyst. It is checked for consistency and corrections are made with the assistance of the hospital staff.

2.3 Artificial Neural Network (ANN)

An artificial neural network (ANN) is an adaptive machine that simulates the learning activities of a human brain. It employs a massive interconnection of simple processing units called neurons to perform certain computations, such as pattern recognition, perception, and motor control. Artificial neural networks emulate biological neural networks through a learning process by acquiring knowledge through the network environment and storing the acquired knowledge through the inter-neuron connection strengths. The procedure used to perform the learning process is called a learning algorithm. The function of the learning algorithm is to modify the inter-neuron connection weights of the network in an orderly fashion to attain a desired design objective, for example, analyzing complex databases of medical information. Another computing ability of artificial neural network is to generalize the analysis to data not seen in the learning phase. A well trained network can produce reasonable outputs for inputs not encountered during learning [33]. Thus neural networks are capable to solve large scale problems and process unseen information.

2.3.1 Artificial Neuron

The artificial neuron is the basic unit of neural networks. Each neuron sends signals to other neuron via connections. The neurons accept input signals of information about the current problem which are the medical data inputs in this work. The signals that a neuron receives from the input or other neurons are weighted by multiplication and then summed to produce an overall activity level in the neuron [52]. The unit computes some function based on the sum of the synaptic weights and a bias, which can be modified during learning. The sum is then fed through an activation function to scale the output within a limited range defined by the activation function. Figure 2-2 shows the basic structure of an artificial neuron.

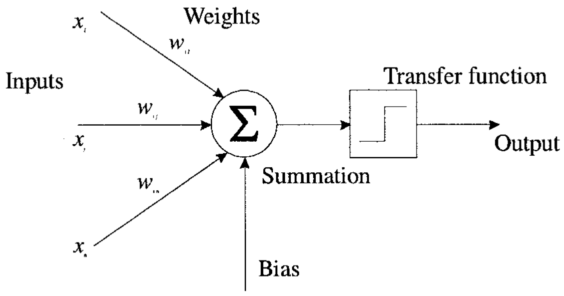


Figure 2-2: Basic structure of an artificial neuron [51].

In mathematical terms, a neuron k can be described as the following equation [33]:

$$y_k = \varphi(\sum_{i=1}^{m} w_{ki} x_i + b_k) \quad \textbf{(Equation 2-1)}$$

where x_i are the input signals; w_{ki} are the synaptic weights which weights the relevant importance of the information; b_k is the bias which acts like a weight on a connection with a constant activation of 1; $\varphi(\bullet)$ is the activation function; and y_k is the output signal of the neuron.

2.3.2 Activation Function

The activation function determines the relationship between inputs and outputs. Step function is usually used in a linear network which has no hidden layer. Nonlinear mapping function such as sigmoid transfer function is commonly used in a nonlinear network which has at least one hidden layer. The activation function used in this thesis is the hyperbolic tangent function [73]:

$$\tanh x = \frac{e^{2x}-1}{e^{2x}+1} \quad \textbf{(Equation 2-2)}$$

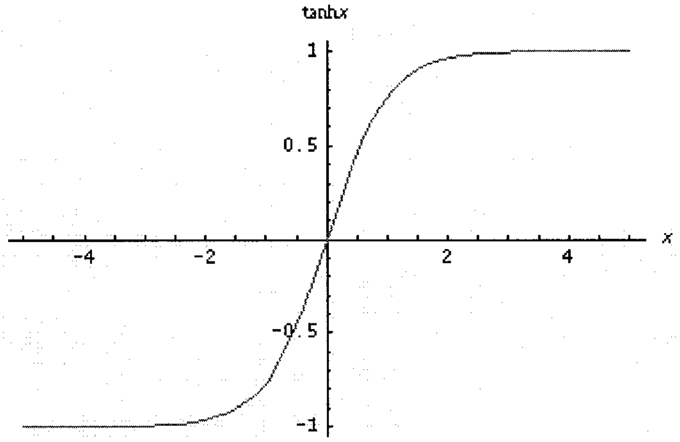


Figure 2-3: Hyperbolic tangent function [73].

The hyperbolic tangent function is a bipolar sigmoid function ranged between -1 and +1. It has a sharp sigmoidal transition at zero resulting in very few output values near zero. Providing input values scaled to have zero mean and unit variance, the output values are expected to saturate at the maximum value of +1 and minimum value of -1. The sharp transition between outputs offers a fast learning process [22]. The nonlinearity of this function results in very little change after the saturation point permits stability of simulated outputs [41].

2.3.3 Learning Method

Learning through training and subsequently being able to recognize the patterns or categorize the input information is the unique source of intelligence in an artificial neural network. There are two distinct types of learning methods: unsupervised learning and supervised learning. Both types of learning require a certain amount of training cases for the network to model the data set. The difference between the two methods is that unsupervised learning does not require output values for the training examples at the time of training, while supervised learning requires known outputs for the network to correct its errors when it compares with its output samples [69]. The back-propagation algorithm is the most common supervised learning method for artificial neural networks because of its robustness and ease of implementation [48]. As there is a sufficient number of training samples with error free input and output variables, back-propagation algorithm was chosen in this thesis work's learning process.

2.3.3.1 Back-propagation Algorithm

The back-propagation algorithm is used for the learning process of the multilayer perceptron (MLP) neural networks. This algorithm is based on the error correction learning rule. In order to reduce the network's prediction error, weights and biases need to be changed individually as a function of the error calculated from the results. Error back-propagation learning process consists of a forward pass and a backward pass through the different layers of the network. In the forward pass, input signals and its effects propagate through the network layer by layer. Actual output responses are produced during this activity. Then the synaptic weights are updated according to the error correction rule during the backward pass. The backward activity also aims to reduce the distance between the actual output response and the desired response of the network [1].

The actual output of the network is subtracted from a desired output to produce an error signal, which is then propagated backward through the network. The following equation defines the error signal at the output of neuron i at the nth training example [33]:

$$e_i(n) = d_i(n) - y_i(n) \quad \textbf{(Equation 2-3)}$$

where $d_i(n)$ is the desired response for neuron i; and $y_i(n)$ is the actual response at the output of neuron i.

The total error energy is the sum of all neurons in the output layer [33]:

$$\xi(n) = \frac{1}{2}\sum_{i \in C} e_i^2(n) \quad \textbf{(Equation 2-4)}$$

where the set C includes all the neurons in the output layer of the network.

The error signal is minimized based on the mean squared error in the training set. After each training epoch, the back-propagation is carried out using the generalized delta rule to update the weights and biases [45]. The forward and backward computations are iterated by running new epochs of training examples to the network until the stopping criterion is met.

2.3.4 Network Architectures

Depending on the learning algorithm used to train the network, neural networks are structured in three fundamentally different classes:

2.3.4.1 Single-layer Feedforward Networks

There are only two layers of neurons present in a single-layer network, that is, one input layer and one output layer. This is the simplest form of a layered network. The input source nodes are strictly fed forward to the "single" output computation nodes (the input layer of source nodes are not counted because no computation is performed) [33]. Figure 2-4 shows a single-layer neural network with one output node.

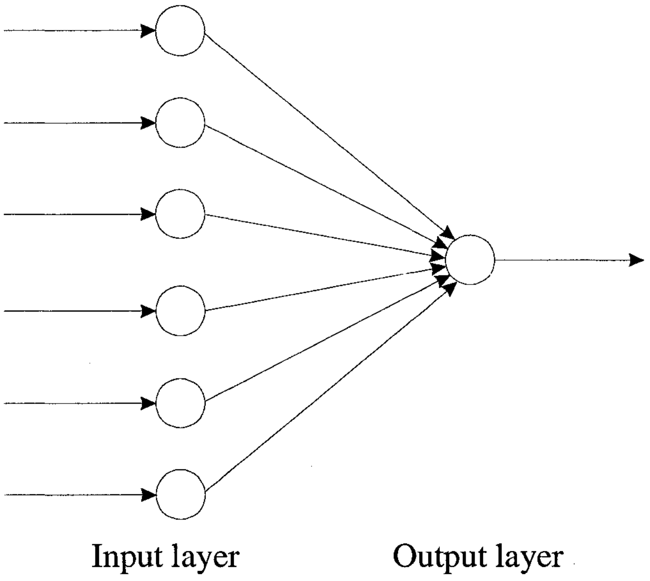


Figure 2-4: Single-layer neural network [51].

2.3.4.2 Multilayer Feedforward Networks

In a multilayer neural network, there are one or more hidden layers of neurons between the input and output layers. The hidden nodes enable the network to work with non-linear activation function and to learn complex tasks by extracting more meaningful features from the input patterns. There are two types of connections for a neural network: fully connected or partially connected. A fully connected network has every node in each layer of the network connected to every other node in the adjacent layer. Partially connected network has some of its weight connections missing from the network [33]. Based on the complexity of the domain problem and design requirements, the fully connected double-layer network is found to be sufficient in the experiment of this thesis work. Figure 2-5 is an example of a fully connected double-layer neural network with three hidden nodes and one output node.

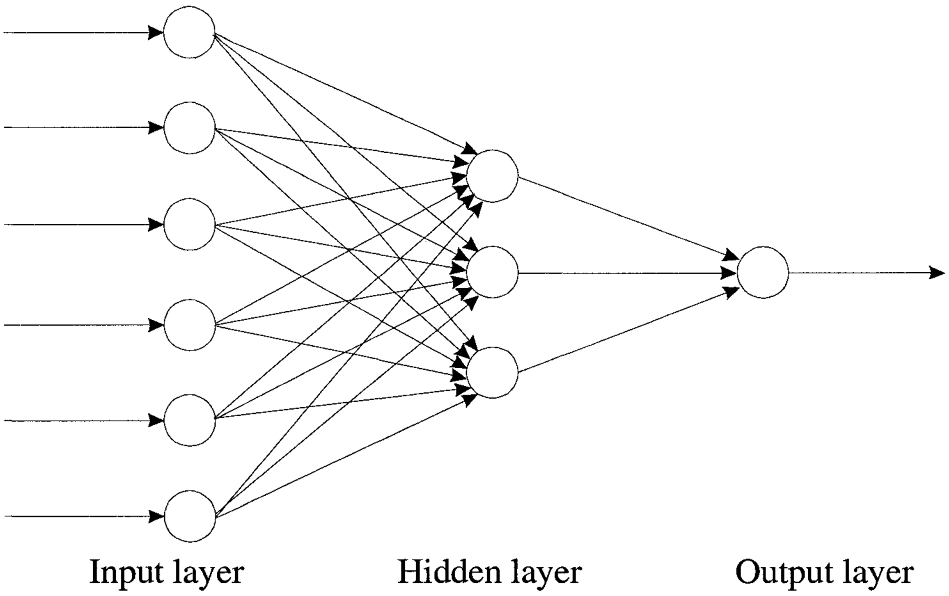


Figure 2-5: Double-layer neural network [51].

2.3.4.3 Recurrent Networks

A recurrent neural network consists of at least one feedback loop, where the neurons feeding its output signal back to its own input or the inputs of the other neurons [33]. The feedback networks can perform more complex nonlinear learning processes but also involve more complicated network settings. A feedback network is shown in Figure 2-6.

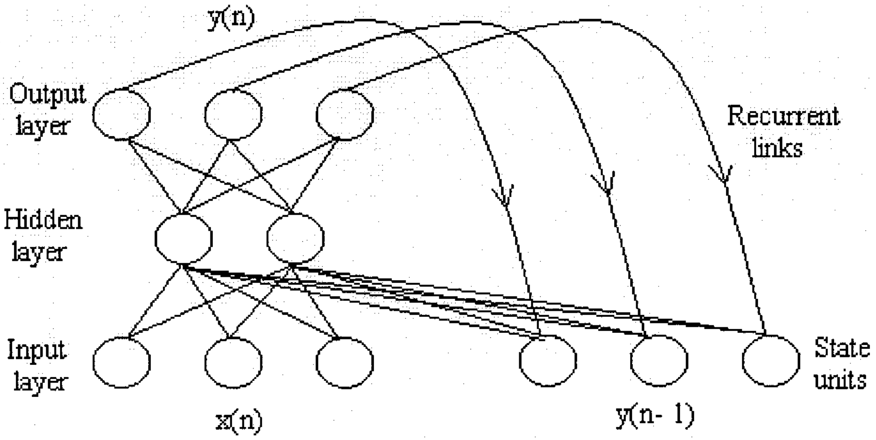


Figure 2-6: Recurrent neural network [5].

2.3.5 Network Structure

As many of the medical data sets contain nonlinearities, neural networks using non-linear algorithms can solve more complex problems and offer better results than linear methods. As mentioned in Section 2.3.4, multilayer networks can implement nonlinearly separable functions to solve non-linear problems. The primary concerns for selecting the appropriate artificial neural network structure for the specific data set are the number of hidden layers and the number of nodes per hidden layer.

2.3.5.1 Number of Hidden Layers

A multilayer neural network ideally can learn any continuous mapping to an arbitrary accuracy. Although more than one hidden layer may be beneficial for some applications, one hidden layer is sufficient in most of the problems [22]. That is because the training time becomes longer if there are more hidden layers presented in the network as there are more neurons and parameters to optimize. The majorities of medical applications use one-hidden-layer networks, as well as this thesis work.

2.3.5.2 Number of Hidden Nodes

Similar to the reasons of selecting the number of hidden layers for a neural network, the trade-off between training time and the accuracy of training also exists when choosing the number of nodes in a hidden layer. The greater number of hidden nodes in the network, the longer it takes to train the network. Too many hidden nodes cause the network to over-fit (i.e.

memorise the training data set), which decreases its generalisation performance [12]. Although fewer hidden nodes provide faster training, the cost is having fewer feature detectors resulting in under-fitting [39]. There are many methods that can be used to determine the optimal number of hidden nodes for a neural network such as: using 50% of the number of input and output nodes [53], using 75% of the number of input nodes [44], using the square root of the number of input nodes multiplied by the number of output nodes [47], or using 2n+1 hidden nodes where n is the number of input nodes [4]. In general, large size data sets which contain a large number of variables need less hidden nodes (less complex network structure) than the small size data sets. Smaller networks require more hidden nodes to provide higher stability. However, each network has its own characteristics and thus its own optimal number of hidden nodes.

2.3.6 Pruning Technique

One of the design objectives of using neural networks to solve real-world problems is to reduce the size of the network while maintaining good performance. In the case of starting with a large multilayer perceptron, network pruning is one way to "weaken or eliminate certain synaptic weights in a selective and orderly fashion" [33]. The basic ideas of the network pruning technique are regularisation and deletion of the synaptic connections from the network. The weight elimination cost function containing the weight decay procedure is one of the most popular approaches associated with the back-propagation learning algorithm, which was chosen for this thesis.

2.3.6.1 Weight Elimination Cost Function

Weight elimination cost function is an effective approach to reduce the complexity of the over sized networks. In the circumstance of back-propagation learning, the trade-off between network reliability and model complexity can be expressed as [33]:

$$\underbrace{R(w)}_{Cost} = \underbrace{\xi_s(w)}_{Error} + \underbrace{\lambda\xi_c(w)}_{Penalty} \quad \textbf{(Equation 2-5)}$$

$\xi_s(w)$ is the initial cost function which evaluates all the output neurons for all the training examples on an epoch-by-epoch basis. This error term is defined as the sum of mean squared errors which is calculated by Equation 2-5. The role of the weight decay constant, λ, is to determine the relative importance of the complexity penalty term. The complexity penalty term $\xi_c(w)$ identifies the significance of the synaptic weights which is defined by [71]:

$$\xi_c(w) = \sum_{i \in C} \frac{w_i^2 / w_0^2}{1 + w_i^2 / w_0^2} \quad \textbf{(Equation 2-6)}$$

where set C includes all the synaptic connections in the network; w_0 is the weight elimination scale factor; w_i is the weight of some synapse i in the network.

The weigh decay constant (λ) is assigned a value somewhere between zero and infinite large to control how strongly the weights are penalized. Choosing a value of λ that is closer to zero will not affect the network because the network is mostly determined from the training examples of the back-propagation learning process. When λ is too large, the network relies on the constraint imposed by the complexity penalty term and reducing the efficiency of the training examples in the learning process. Since the network is sensitive to the value of λ, it is recommend to start the network with λ at zero in order to take advantage of all input variables at the beginning and then slowly increase the value. The value is increased if the weights are not diminishing or decreased if all the weights are forced to zero [72].

The weight elimination penalty term $\xi_c(w)$ "counts the number of parameters, and minimizes the sum of performance error and the number of weights by back-propagation" [72]. The weight elimination scale factor (w_0) is a user defined parameter, which enables the network to work at a preference for fewer large weights (eliminate small weights when w_0 is small) or more small weights (decay large weights when w_0 is large). When $|w_i| << w_0$, the penalty for that weight is close to zero and should be eliminated from the network as the ith weight is unreliable. For $|w_i| >> w_0$, the penalty for that weight approaches the maximum value as the weight w_i is important to the learning process. By reducing the number of connection weights and hence the model's complexity using the weight elimination cost function, the network's classification performance is expected to be improved [18].

2.3.7 Learning Parameters

In order to optimize the performance of the neural networks, there are nine driving parameters used to fine tune the network learning process. These parameters include the error ratio (err_ratio), momentum (m), learning rate (lr), learning rate increment (lr_inc), learning rate decrement (lr_dec), weight decay constant (λ), weight decay constant increment (λ_inc), weight decay constant decrement (λ_dec), and weight elimination scale factor (w_0).

These adjustable parameters change the weights and biases of the network as a function of the error between the actual and desired outputs. The following equations describe the change in weight (Δw_{ji}) and change in bias (Δb_{ji}) [33]:

$$\Delta w_{ji}(n) = m \cdot \Delta w_{ji}(n-1) + lr \cdot \delta_j(n) \cdot y_i(n) \quad \textbf{(Equation 2-7)}$$

$$\Delta b_{ji}(n) = m \cdot \Delta b_{ji}(n-1) + lr \cdot \delta_j(n) \quad \textbf{(Equation 2-8)}$$

where the local gradient $\delta_j(n)$ for neuron j is defined by [33]:

$$\delta_j(n) = e_j(n) \cdot \varphi'_j(v_j(n)) \quad \textbf{(Equation 2-9)}$$

The local gradient $\delta_j(n)$ is equal to the product of the corresponding error signal $e_j(n)$ for that neuron and the derivative $\varphi'_j(v_j(n))$ of the associated activation function.

Meanwhile, the pre-assigned error ratio value is used to adjust the momentum, learning rate and weight decay constant parameters. If the current training set error is greater than the product of the previous training set error and error ratio, the following updates are done:

$$\begin{aligned}&m = 0;\\&lr = lr \cdot lr\_dec;\\&\lambda = \lambda \cdot \lambda\_dec;\end{aligned}$$

Otherwise, if the current training set error is less than the previous training set error multiplied by the error ratio, m is unchanged, and

$$\begin{aligned}&lr = lr \cdot lr\_inc;\\&\lambda = \lambda \cdot \lambda\_inc;\end{aligned}$$

The objective of comparing the training set error with the error ratio is to have the error ratio get smaller with each epoch to improve the performance [15]. The momentum (m) parameter controls the general flow of the error changes between the previous weight change value to the new value, thereby prevents the algorithm from getting caught in local minima and allows it to continue in searching for the global minimum [24]. The learning rate (lr) determines the amount and speed of weight adjustment at each step of training. If the current-to-previous error is greater than the error ratio, the new weights and biases are rejected, and the learning rate is multiplied by the learning rate decrement factor (lr_dec) to slow down the learning. Conversely, when the current-to-previous error is less than the error ratio, the new weights and biases are retained and multiplying the learning rate by its increment factor (lr_inc) to speed up the learning process. This procedure is to ensure that the learning rate can converge to the global minimum without oscillating around it or skipping over it [13]. The weight decay constant (λ) determines how strongly the weights are penalized. When the error is greater than the pre-defined error ratio, the weight decay constant is decremented to back off the pressure on the weights to be small. On the contrary, for the smaller error, the weight decay constant is incremented to represent the small weights to ensure that they are included in further training process [15]. The weight elimination scale factor (w_0) defines the sizes of "large" and "small" weights (weight elimination or weight decay) [18]. The nine adaptive network parameters work closely to control the fashion in which the neural networks learn from one epoch to the next.

2.3.8 Network Generalisation

Generalisation is a capability expected for artificial neural networks in analysing new data. A well-trained network system can provide solutions which are appropriate and fit known situations. But it can associate good answers to previously unencountered situations that are somewhat similar to encountered ones. If the full range of possible situations is known and seen by the network, generalisation may not be an issue. However, this is non-realistic in developing a neural network model because medical databases rarely contain patient cases that cover all possible situations for network learning. Thus the interpolation is probably effective in a neural network but it still provides no guarantee for extrapolation [37]. In the case of interpolation, if the new case is similar to the cases in the training set, the predicted output is predicted based on the previous cases. In the situation of extrapolation, the new case falls in the unseen range of the training set, the extrapolated output is extracted from the model that does not necessarily represent the new pattern. The artificial neural network generalisation capability is thought to depend on various factors, such as the network structure, the operation on processing elements, learning algorithms, pattern distribution complexity, and training data density [76].

2.4 Variable Selection Methods

Variable selection involving the choice of independent variables for inclusion in the artificial neural network model is an important procedure to extract the most relevant set of variables in model development. The selection is usually taken from a set of measured potential input variables, which is essential to being able to model the system output under consideration reliably. The correct choice of model inputs is important to obtain good generalisation with finite data. When the available data set is highly dimensional, reducing the number of free parameters is also important for improving computational efficiency.

Medical data sets are generally complex and mostly associated with nonlinear processes. Consequently, the dependencies between output and input variables are difficult to measure. Artificial neural networks allow models of any complexity to be fitted between the responses (outputs) and descriptors (inputs) and these need not be specified in advance. Thus, a variable selection procedure applied to the properly trained neural network can make use of the best nonlinear model as dictated by the data itself [45]. There are a number of variable selection strategies that may be adopted, such as relative weight reduction (RWR) and principal component analysis (PCA), which were chosen for further studies in this work.

2.4.1 Relative Weight Reduction

The relative weight reduction (RWR) approach is based on the idea of developing a method to partition the neural network synaptic weights in order to determine the relative importance of each input variable in the network. In the neural network, the connection weights between neurons are the links between the inputs and outputs. The relative contributions of the input variables to the estimated output of the network depend primarily on the magnitude and direction of the connection weights [50]. Neurons with larger connection weights represent greater intensities of signal transfer, and therefore have higher importance in the process than the nodes with smaller weights. Positive connection weights enhance the intensity of the incoming signal and increase the value of the response on the neurons, whereas negative weights reduce the intensity of the incoming signal and decrease the value of the response on the neurons. Similar to PCA's dimension reduction ability, RWR can be used to remove the least important input variables depending on the least values of connection weights. Conversely, PCA components are the linear combination of all input variables, while RWR completely eliminates the less descriptive inputs.

An algorithm for calculating relative weights of multilayer perceptron neural network to include the linkage of input-hidden and hidden-output relations was proposed by Garson [31] and later modified by Goh [32]. However, the Garson-Goh algorithm has a shortcoming of using the absolute values of the connection weights when calculating variable contributions. The use of absolute values does not provide the direction of the relationship between neurons and makes it possible for a result to be misread. In the case of a positive weight and a negative weight stemming to the same node, the two values may effectively cancel each other out in the network process but would be counted twice towards the RWR calculations [50].

Applying Garson-Goh's theory of integrating the input-hidden and hidden-output weights in relative importance calculation, Rybchynski modified the computation equations for the RWR method as follows [60]:

(1) For each output neuron k, calculate the relative weight (rw_{jk}) of hidden-output connections.

$$rw_{jk} = \frac{abs(w_{jk})}{\sum_{j} abs(w_{jk})} \quad \textbf{(Equation 2-10)}$$

(2) For each hidden neuron j, calculate the relative weight (rw_{ij}) of input-hidden connections.

$$rw_{ij} = \frac{abs(w_{ij})}{\sum_{i} abs(w_{ij})} \quad \textbf{(Equation 2-11)}$$

(3) For each hidden neuron j, multiply the relative weight of input-hidden value by the relative weight of hidden-output value.

$$P_{ij} = rw_{ij} * rw_{jk} \quad \textbf{(Equation 2-12)}$$

(4) For each input neuron i, sum the relative weight products (P_{ij}) from the connected hidden nodes.

$$S_i = \sum_j P_{ij} \quad \textbf{(Equation 2-13)}$$

(5) For each input neuron i, divide the total relative weights (S_i) by the sum for all the input variables, which gives the relative importance (RI_i) to the given input.

$$RI_i = \frac{S_i}{\sum_i S_i} \quad \textbf{(Equation 2-14)}$$

From these final relative importance values, variables with the smallest relative importance can be removed until a minimum data set (only the most important indicators remained) is created. The final data set is reduced in dimensional space yet keeps the most necessary information for neural network processing. The simplicity of the RWR method to define a list of indicators and their relative importance makes it well accepted in the fields of finance and ecologies [32].

2.4.2 Principal Component Analysis

The multivariate statistical method of principal component analysis (PCA) is a very useful tool for reducing the number of variables in a data set while retaining as much information as possible. Via this approach, data represented by a set of correlated variables are transformed into a set of uncorrelated variables by a linear transformation. The new variables are linear combinations of the original variables and are ordered by reduced value. Meanwhile, the noise is simultaneously filtered while transforming the data into the new dimensions. As a result, removing the few least important variables may not result in the loss of significant information [14]. However, the reduced data set might suffer from the major disadvantage of representing a linear model, if the data is actually better fitted by some nonlinear model. To overcome this problem, PCA is used as a data processing method in the underlying model for removing cofounders and noise before the data are applied to the construction of nonlinear network models.

PCA is based on the statistical representation of the multivariate vector. Let X denote an *m*-dimensional random vector representing the environment of interest, which has zero mean. Then the covariance matrix of the random vector X is:

$$C = E[XX^T] \quad \textbf{(Equation 2-15)}$$

The eigenvectors e_i and corresponding eigenvalues λ_i of the covariance matrix can be calculated on an orthogonal basis:

$$C \cdot e_i = \lambda_i \cdot e_i, i = 1, \ldots, m \quad \textbf{(Equation 2-16)}$$

$$|C - \lambda_i \cdot I| = 0 \quad \textbf{(Equation 2-17)}$$

where λ_i are assumed to be distinct, I is the identity matrix having the same order as C, and $|\bullet|$ denotes the determinant of the matrix.

Let P_m be a matrix consisting of eigenvectors of the covariance matrix. Then to compute a new set of vector Y, whose components can be seen as the coordinates in the orthogonal base defined by the eigenvectors:

$$Y = X \cdot \mathrm{P}_m \quad \textbf{(Equation 2-18)}$$

As mentioned above, the importance of the components is ranked in the order of descending eigenvalues. The smaller value components usually suggest that the corresponding variables have little or no control on the distribution of the data set. Some of these low performance inputs can be removed from consideration in order to simplify the overall analysis.

Therefore, instead of using all the eigenvectors of the covariance matrix, the new data set Y can be represented by selecting the *n* first eigenvectors P_n as the orthogonal basis vectors:

$$Y' = X \cdot \mathrm{P}_n \quad \textbf{(Equation 2-19)}$$

This implies that the original vector has been projected from the original *m*-dimension space to a new *n*-dimension space, which is a linear transformation of the basis vectors. PCA minimises the mean square error between the data and its representation of given number of eigenvectors; meanwhile, it maximises the rate of decrease of variance [62]. Therefore, it offers a convenient way to control the trade-off between losing information and simplifying the problem. Because PCA is a non-parametric method of extracting relevant information from confusing data sets and reconstructing the original data sets from reduced vector, it is used widely in all forms of analysis, such as image compressions and face recognitions [77].

The PCA method is implemented as a data pre-processing tool using MATLAB program in this thesis. The component manipulations are automatically taken care of by the MATLAB tool. The resulting ranked components are compositions of input variables, and they are ready for further variable reductions before feeding to ANN tools. This process is described in more detail in methodology section.

2.5 Clinical Decision Support System (CDSS)

Decision support systems are a set of interrelated components, infrastructure, and process that facilitate information and knowledge aggregation, management, and dissemination for decision makers to better perform their tasks and reach their objectives [75]. Commercially available decision support systems have applications in business driven models such as banking, finance, investment, and manufacturing. In the clinical environment, decision support systems can be applied in diagnosis, prognosis, treatment selection and value clarification. The significance of using clinical decision support systems (CDSS) is to help both the physicians and patients to make better decisions on healthcare related matters.

Traditional tools for clinical decision making include statistical analysis, knowledge-based systems, and neural network models. A novel approach is the hybrid system combining the data-derived knowledge and expert-supplied knowledge in the same decision support system [38]. This approach is rapidly evolving because of fast development of information technology, increasing access to computer systems in clinical practice, and growing concern about the process and quality of medical care. Studies [30] have shown that computerized clinical decision support systems can improve clinical decision making in the aspects of practitioner performance and patient outcomes. The advantages of adopting clinical decision support system are [11]:

- Expanding a wider spectrum of medical decision from individual patients to whole populations;
- Widening the spatial support of decision making by including place of service and remote access;
- Minimizing the temporal constraints by supporting real-time and retrospective modes of decision making.

However, there are concerns and issues of using the systems such as [11], [30]:

- Patient privacy and record confidentiality;
- Physician workflow and user acceptance;
- Standardization of clinical terminology and medical data capture;
- System security, upgrade availability and compatibility with existing facilities;

- Infrastructure development, financial and regulatory issues.

Generally speaking, the use of clinical decision support system offers tremendous benefits on improving the balance between healthcare expenditure and healthcare quality (including safety, productivity, and timeliness).

As an advocate of exploring clinical decision support systems, Medical Information-technology Research Group (MIRG) of the Carleton University and the University of Ottawa is engaged in improving patient care through the development of a hybrid CDSS to support integrated ANN applications. Starting from early 90's, researchers of MIRG have developed numerous ANN models for clinical outcome estimations in both adult and neonatal intensive care unit. MIRG has also designed hybrid systems combining the data-driven tools (e.g., ANN) and evidence-based tools (e.g., case-based reasoning (CBR), k-nearest neighbor (KNN)). Meanwhile, the research group is working on designing a standard-compliant medical information infrastructure that allows for seamless integration of the different clinical decision support tools. A brief review of these works done by MIRG is given in the following section.

2.5.1 CDSS Prototypes by MIRG

As MIRG has developed more ANN models using many different databases, there is a need for adapting these clinical decision support tools in a real hospital environment. MIRG's CDSS attempt was designed in eXtended Markup Language (XML) enabled interface to standardize different medical databases and integrate various medical applications. XML was chosen as the foundation because it can enable data portability by separating the presentation of the data from the content. It is a good candidate technology to transfer self-describe data from one application to another (e.g., ANN to ANN, ANN to CBR) over communication channels (e.g., the Internet). Catley [8] deployed this XML prototype implementation for the NICU of the Children's Hospital of Eastern Ontario (CHEO) to extend their information centre to incorporate with the ANN tools designed by MIRG. The data were stored using MySQL and accessed remotely using Java Database Connectivity. The system then extracted data from the MySQL database into a new XML document and conversely parsed existing XML documents into the one database for future XML processing. This is in order that the XML database is ready to be used by MIRG's MATLAB-based ANN tools.

One of the benefits of integrating CDSS's medical context into social context is to create the maximum value for healthcare decisions involving both the physicians and patients. Yang [29] designed an "ethical" decision support system that provides support to the CHEO's NICU clinicians and parents in making difficult "ethical" decisions involved in withholding or withdrawing intensive care for infants with extremely poor prognoses. The collaborative PArent Decision Support (PADS) system consisted of four components: an evidence-based estimation block (ANN prediction), a content management block (data repository), a communication management block (message protocol and patient decision aid), and a graphical user interface. The four components were merged into two user interface platforms: a physicians' workplace and a parents' workplace. PADS allowed physicians to better predict certain outcomes in neonatal intensive care and use this information when they counsel the parents. Meanwhile, the system included the parents in the decision loop and took into consideration their values.

For the goal of offering physicians access to outcome prediction abilities in real-time and distributed decision-making environments, a unifying infrastructure for CDSSs and medical databases relating to the perinatal life cycle was proposed by Catley [9]. The proposed web service based infrastructure was referred as the Perinatal, Obstetrical, and Neonatal Intensive Care Web (PONI-Web). A web service is an interface that describes a collection of operations that are network-accessible through standardized XML-messaging. The core web services included MIRG's individual clinical decision support tools such as ANN processing, missing value replacing, and alert detection. The composite web services combined two or more core services to offer a complete system application. Figure 2-7 presents an overview of the PONI-Web infrastructure. The PONI-Web is an infrastructure that can improve perinatal, obstetrical and neonatal care through MIRG's on-going CDSS research. The perinatal outcome estimation model in this thesis can be used in the PONI-Web's data mining core service.

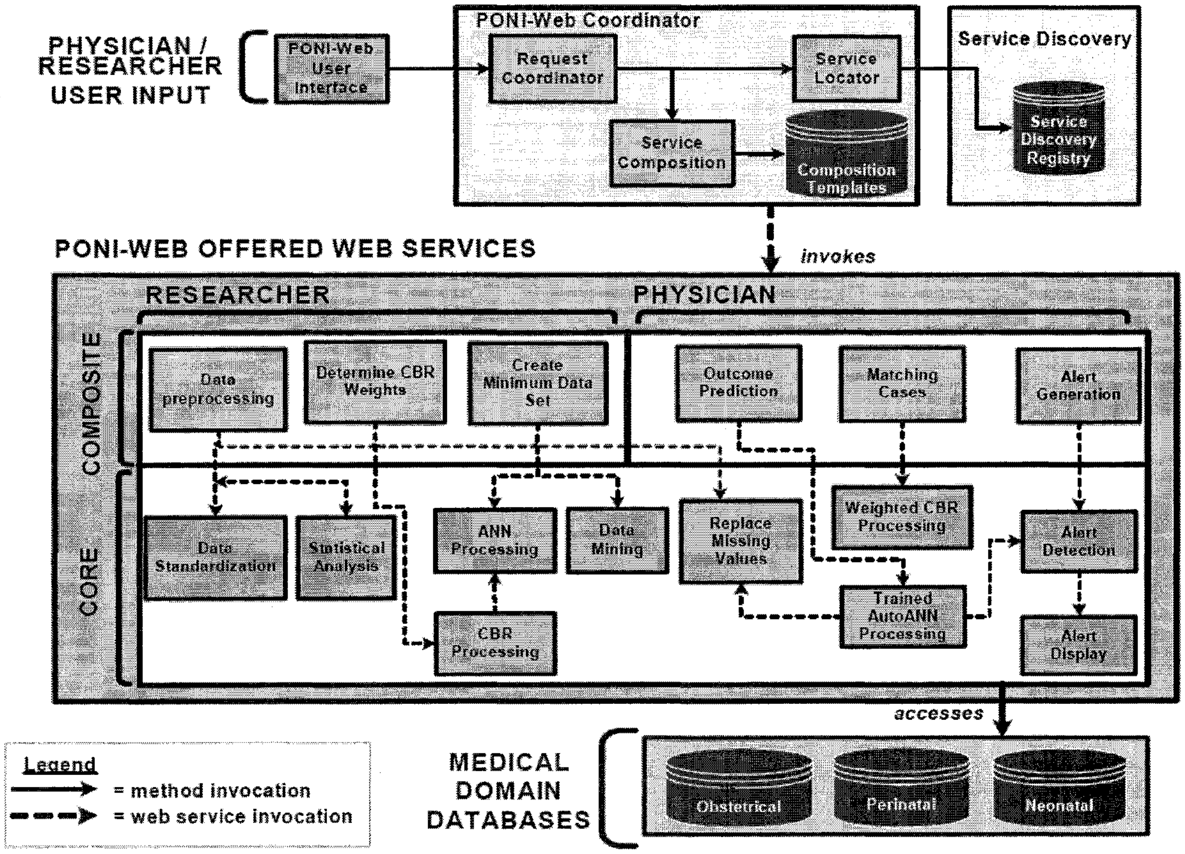


Figure 2-7: PONI-Web integration infrastructure [9].

3 Methodology

3.1 Problem Formulation

The main problem of this thesis is to compare the classification performance of RWR and PCA variable selection methods applied to ANN models for APGAR5 and DELTYPE perinatal outcomes. The method which provides better performance is chosen to verify the ANN models on unseen data.

3.2 Data Preparation

In practice, an approach of applying the raw input data directly onto the required final output values in a neural network will generally give poor results for a number of reasons. Thus it is necessary first to transform the data into some new representation before training a neural network. It is estimated that approximately 80% of a researcher's time is spent on preparing the training data for the neural network [74]. In many practical applications the choice of pre-processing will be one of the most significant factors in determining the performance of the final system. The research data used in this thesis are from the Niday Perinatal Database collected under the direction of the Perinatal Partnership Program of Eastern and Southeastern Ontario [54]. The year 2001 (Niday 2001) and year 2004 (Niday 2004) patient records are used in the simulation and result analysis for the ANN multi model outcome estimation.

3.2.1 Variable formatting

The original Niday database contained massive amount of information for each patient record collected. Since ANN required data to be used in a certain format in order to produce accurate results, the raw data were first prepared to the standard numeric ANN data format. Especially the text based data were not directly used as inputs into the ANN but were numerically coded to replace the information. The data used in this study and how they were encoded are described in Table 3-1.

Table 3-1: Niday Perinatal Database data contents.

Variable	Definition	Codification
IDNUM	Birth identification number	Numbers 0 or 9999 = Missing
DOB	Date of birth	MM/DD/YYY
NBABIES	Number of babies in this pregnancy	1, 2, 3, 4, 5, 6, >6
WEIGHT1	Weight of the baby	In grams
GEST	Gestational age	Range: 18 – 45 weeks
AGE	Age of the mother	Range: 12 – 55 years
GENDER	Gender of the baby	1 = Male 2 = Female 9 = Unknown
TERM	Number of previous term babies	Range: 0 – 15
PRETERM	Number of previous preterm babies	Range: 0 – 15
PARITY	Parity	Range: 0 – 15
APGAR1	Apgar score at 1 minute	Range: 0 – 10
APGAR5	Apgar score at 5 minute	Range: 0 – 10
BF	Whether the woman intends to breastfeed	1 = Yes 2 = No 9 = Unknown
SMOKING	Whether any maternal smoking has occurred at any time after 20 weeks gestation	1 = Yes 2 = No 9 = Unknown
CORD	Whether umbilical cord blood gases were done	1 = Yes 2 = No 9 = Unknown
SCALP	Whether scalp blood gases were done	1 = Yes 2 = No 9 = Unknown
GENERAL	Maternal pain relief utilized: general	1 = Yes 2 = No 9 = Unknown
EPIDURAL	Maternal pain relief utilized: epidural	1 = Yes 2 = No 9 = Unknown

SPINAL	Maternal pain relief utilized: spinal	1 = Yes 2 = No 9 = Unknown
NARCOT	Maternal pain relief utilized: narcotics	1 = Yes 2 = No 9 = Unknown
NITOXIDE	Maternal pain relief utilized: nitrous oxide	1 = Yes 2 = No 9 = Unknown
PUDENDAL	Maternal pain relief utilized: pudendal	1 = Yes 2 = No 9 = Unknown
RESFFO2	Newborn resuscitation utilized: FFO2	1 = Yes 2 = No 9 = Unknown
RESPPV	Newborn resuscitation utilized: PPV	1 = Yes 2 = No 9 = Unknown
RESINTUB	Newborn resuscitation utilized: intubation	1 = Yes 2 = No 9 = Unknown
RESCHEST	Newborn resuscitation utilized: chest compression	1 = Yes 2 = No 9 = Unknown
RESDRUGS	Newborn resuscitation utilized: drugs	1 = Yes 2 = No 9 = Unknown
MONITOR	Fetal surveillance	1 = No monitoring 2 = Auscultation only 3 = EFM only 4 = Auscultation and EFM
LABTYPE	Labour type	1 = Spontaneous 2 = Induced 3 = No labour 4 = Unknown
PRESENT	Presentation	1 = Vertex 2 = Breech 3 = Other 4 = Not available
ASSISTED	Assisted with	1 = None 2 = Forceps 3 = Vacuum 4 = Forceps and Vacuum
DELTYPE	Delivery type	1 = Vaginal 2 = VBAC 3 = Primary C/S 4 = Repeat C/S trial of labour 5 = Repeat C/S no trial of labour
DELBY	Delivered by	1 = Physician 2 = Midwife 9 = Unknown
STEROIDS	Whether antenatal steroids were administered	1 = None 2 = 1 dose < 24 hrs 3 = 2 doses: last dose < 24 hrs 4 = 2 doses: last dose => 24 hrs
NEODEATH	When and where the neonatal	1 = N/A

	death occured	2 = Stillbirth (> 20 wks and/or 500 g) 3 = < 7 days: birth hospital 4 = < 7 days: transfer hospital 5 = 7 – 28 days: birth hospital 6 = 7 – 28 days: transfer hospital
NEOTRANS	Neonatal transferred during the first 7 days of life only	1 = No transfer 2 = Almonte General (Almonte) 3 = Quinte HC – Belleville Site (Belleville) 4 = Brockville General Hospital (Brockville) 5 = Cornwall General (Cornwall) 6 = Hawkesbury & District (Hawkesbury) 7 = Kingston General Hospital (Kingston) 8 = Hopital Montfort (Ottawa) 9 = Lennox & Addington County (Napanee) 10 = Civic – Ottawa Hospital (Ottawa) 11 = General – Ottawa Hospital (Ottawa) 12 = Pembroke General Hospital (Pembroke) 13 = Perth/Smith Falls District (Smith Falls) 14 = Quinte HC – Picton Site (Picton) 15 = Queensway – Carleton Hospital (Nepean) 16 = Renfrew Victoria Hospital (Renfrew) 17 = Winchester District (Winchester) 18 = Outside Eastern Ontario 20 = Children's Hospital of Eastern Ontario (Ottawa)
MOMTRANS	Mother inter-hospital transferred from	1 = No transfer 2 = Almonte General (Almonte) 3 = Quinte HC – Belleville Site (Belleville) 4 = Brockville General Hospital (Brockville) 5 = Cornwall General (Cornwall) 6 = Hawkesbury & District (Hawkesbury) 7 = Kingston General Hospital (Kingston) 8 = Hopital Montfort (Ottawa) 9 = Lennox & Addington County (Napanee) 10 = Civic – Ottawa Hospital (Ottawa) 11 = General – Ottawa Hospital (Ottawa) 12 = Pembroke General Hospital (Pembroke) 13 = Perth/Smith Falls District (Smith Falls) 14 = Quinte HC – Picton Site (Picton) 15 = Queensway – Carleton Hospital (Nepean) 16 = Renfrew Victoria Hospital (Renfrew) 17 = Winchester District (Winchester) 18 = Outside Eastern Ontario 20 = Children's Hospital of Eastern Ontario (Ottawa)
HOSPTYPE	Hospital type	1 = Teaching hospitals 2 = Large community hospitals 3 = Small community hospitals
Total variables	**38**	

Not all of the above variables were applicable in the ANN modeling, so that the encoded database was then filtered depending on the interested outcomes APGAR5 and DELTYPE in this study. Variables IDNUM and DOB were removed because they were important to keep track of the data but not for analysis. Other variables were deleted for medical reasons. In APGAR5 outcome analysis, the following 9 variables were removed: APGAR1, CORD, RESFO2, RESPPV, RESINTUB, RESCHEST, RESDRUGS, NEODEATH, and NEOTRANS. The APGAR1 variable measures the Apgar score at 1 minute after baby birth while APGAR5 measures the same thing after 5 minutes. The short interval of a few minutes apart in measuring the two variables made them identical in most of the cases. APGAR1 was identified as a cofounder by the medical practitioners for APGAR5. Including APGAR1 in the ANN simulation might bias the APGAR5 estimation, thus it was removed from the inputs. Other 8 variables were removed due to the reason that they were collected after APGAR5 was measured, so that they were not sufficient in determine the low Apgar score at 5 minutes. Similarly, in the case of DELTYPE output estimation, 11 variables were removed from the inputs. They are APGAR1, APGAR5, CORD, RESFFO2, RESPPV, RESINTUB, RESCHEST, RESDRUGS, ASSISTED, NEODEATH, and NEOTRANS. Again, these 11 variables took place after delivery so that they could not play role during the delivery. DELBY was identified as a cofounder because there is no cesarean section operation by midwife. The final input variables used for output APGAR5 and DELTYPE analysis are listed in Table 3-2.

Table 3-2: Niday 2001 – input variables for output APGAR5 and DELTYPE.

Output	**APGAR5**	**DELTYPE**
Input	NBABIES	NBABIES
	WEIGHT1	WEIGHT1
	GEST	GEST
	AGE	AGE
	GENDER	GENDER
	TERM	TERM
	PRETERM	PRETERM
	PARITY	PARITY
	BF	BF
	SMOKING	SMOKING
	SCALP	SCALP
	GENERAL	GENERAL
	EPIDURAL	EPIDURAL
	SPINAL	SPINAL
	NARCOT	NARCOT
	NITOXIDE	NITOXIDE
	PUDENDAL	PUDENDAL
	MONITOR	MONITOR
	LABTYPE	LABTYPE
	PRESENT	PRESENT
	ASSISTED	STEROIDS
	DELTYPE	MOMTRANS
	DELBY	HOSPTYPE
	STEROIDS	
	MOMTRANS	
	HOSPTYPE	
Total inputs	**26**	**23**

3.2.2 Missing Values

Real data often suffer from a number of deficiencies such as missing input values. A common problem is that some of the input values may be missing from the data set for some of the pattern vectors [70]. When there is too little data to discard the deficient instances, or when the proportion of deficient examples is too high, it becomes important to make full use of the information which is potentially available from the incomplete patterns. In this case, it is common to "fill in" the missing input values first. In general, missing values should be treated by integration over the corresponding variables and weighted by the appropriate distribution. For example, each missing value might be replaced by the mean of the corresponding variables over the cases for which its value is available [15]. On the other hand, if the quantity of total data available is sufficiently large, and the proportion of the patterns affected is small, then the simplest solution is to discard those patterns from the data set. Different methods are available to discard missing data such as deleting the entire variable vector with the most missing values or deleting the whole instance with any missing values. In this work, as the Niday Perinatal Database is a sufficiently large database with relatively small percentage of missing values in each variable, deleting the cases with missing data will not modify the effective data distribution. Thus the method of discarding cases with missing values was chosen. The values with empty entry and codification 9 = unknown were discarded as missing values.

The cases with missing/insufficient values were deleted from tracing each applicable input variable one after one in order to keep the maximum number of complete cases in the ANN simulation. Overall, Niday 2001 database has 1,538 cases with missing values for a total of 17,688 cases (91.3% complete cases) for output APGAR5, and 2,859 missing cases for the total of 17,688 cases (83.8% complete cases) for output DELTYPE. Table 3-3 shows the detailed breakdown of deleted cases for each input variable in Niday 2001 database. The total number of cases was reduced to 16,150 for output APGAR5 and to 14,829 for output DELTYPE.

Table 3-3: Niday 2001 – number of deleted cases for output APGAR5 and DELTYPE.

Variables	Cases deleted in APGAR5	Cases deleted in DELTYPE
NBABIES	0	0
WEIGHT1	6	6
GEST	30	30
AGE	13	13
GENDER	35	35
TERM	22	22
PRETERM	4	4
PARITY	0	0
APGAR5	58	N/A
BF	494	496
SMOKING	403	403
SCALP	291	297
GENERAL	0	0
EPIDURAL	0	0
SPINAL	0	0
NARCOT	0	0
NITOXIDE	0	0
PUDENDAL	0	0
MONITOR	0	0
LABTYPE	0	1439
PRESENT	129	57
ASSISTED	0	N/A
DELTYPE	0	0
DELBY	53	57
STEROIDS	0	0
MOMTRANS	0	0
HOSPTYPE	0	0
Total deleted	**1,538**	**2,859**

3.2.3 Data Standardization

Once a complete data set is determined, the next step of data pre-processing is to standardize the variables to have a uniform magnitude of distribution. This is often useful if different variables have typical values which differ significantly. In the case of dealing with continuous variables, data can be arranged to have similar values by applying a linear transformation. The values of each continuous variable were assumed to follow a normal distribution, so that the z-score formula as followed is applicable to be used for the data standardization [70].

$$Z_n = \frac{(X_n - \mu)}{3\sigma} \quad \textbf{(Equation 3-1)}$$

where Z_n is the standardized value of variable X for observation n; μ is the mean of variable X; σ is the standard deviation of variable X. It is easy to see that the transformed variables given by Z_n now have zero mean and unit standard deviation over the transformed data set. In the case of dealing with data taking on discrete values, it is convenient to use a 1-of-c coding for the categorical variables. Such data can simply be transformed directly into an arbitrarily artificial ordering. It is essential to standardize the variables so that the same percentage change in the weighted sum of the inputs causes a similar percentage change in the unit output [50].

Output data must be presented as Boolean (i.e., "-1" or "+1") in the data file for the ANN experiment. Normally, "-1" is chosen to represent the most frequent instance, and "+1" represent the least frequent instance. In the case of output APGAR5 for Niday 2001, high Apgar score which includes values of 7 to 10 make up 97.9% of the "-1" instance, while low Apgar score which ranges from 0 to 6 make up 2.1% of the "+1" instance.

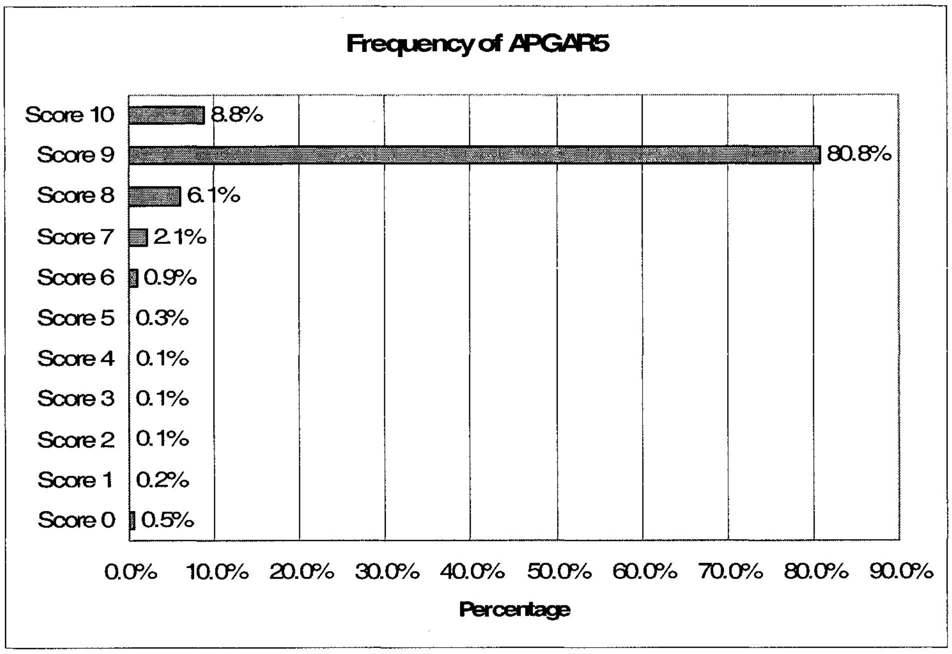


Figure 3-1: Niday 2001 – frequency analysis of output APGAR5.

In the case of output DELTYPE for Niday 2001, vaginal is the most frequent delivery type "-1", and cesarean section is the least frequent instance "+1". In Niday 2001 database, 85.4% of the cases are vaginal delivery, and 14.6% of the cases are cesarean section.

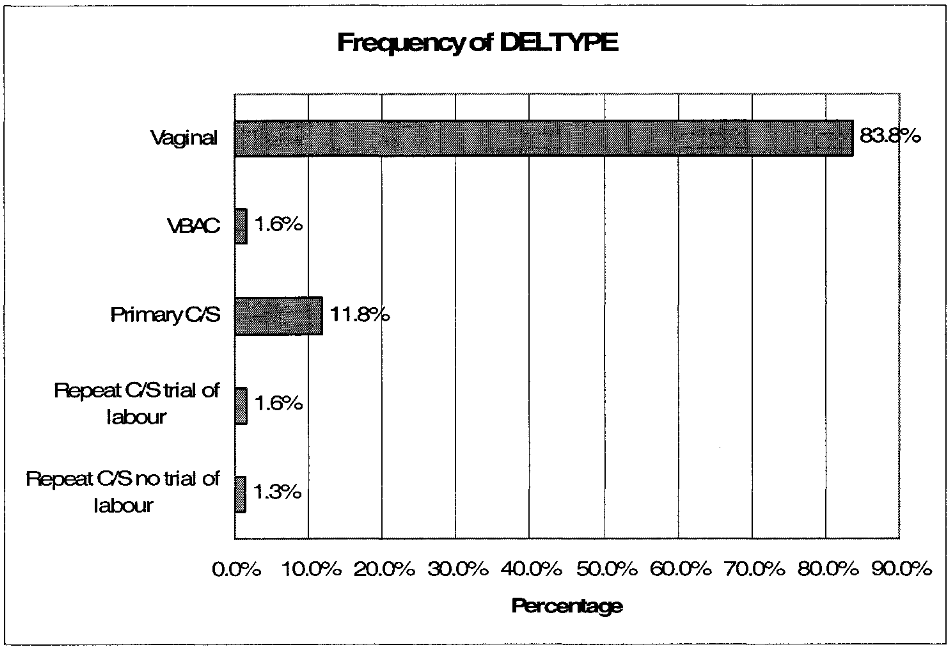


Figure 3-2: Niday 2001 – frequency analysis of output DELTYPE.

Similar frequency analysis was carried out for Niday 2004 database. Both Niday 2001 and Niday 2004 databases show that the instances of low Apgar score (values from 0 to 6) are significantly less than that of the high Apgar score. The cases of cesarean delivery are approximately 20% of the total number of cases. The output data distributions of these two databases are similar to each other, so that using the ANN model developed from Niday 2001 is feasible to predict the outcomes from Niday 2004.

Table 3-4: Niday 2001 and Niday 2004 – distribution of output instance.

Database	**Niday 2001**		**Niday 2004**	
Output	**APGAR5**	**DELTYPE**	**APGAR5**	**DELTYPE**
"-1" instance	97.9%	85.4%	98.1%	77.7%
"+1" instance	2.1%	14.6%	1.9%	22.3%
Total cases	**16,150**	**14,829**	**38,045**	**38,260**

3.2.4 Artificial Data Set

Artificial neural networks often have difficulty classifying highly skewed data, where the outcome of interest is a relatively rare event. For example, the low Apgar score value in the Niday databases is the rare outcome ("+1" instance) which only existed approximately 2% of the total cases. From a medical perspective this is a serious problem because rare outcomes are often those physicians are most interested in predicting. Previous works by MIRG members have found that artificially increasing the percentage of rare outcome to 20% in the training set improves the performance of the network [19]. Two different approaches can be used to alter the *a priori* distributions of a training set. One method is to randomly remove the more frequent outcome cases until the representation of the rare outcome becomes sufficiently high. This way will cause the loss of potentially valuable information about the more frequent cases that may determine the network's correct classification. The other method is to randomly resample the rare outcome cases from the database until the representation of the rare outcome becomes sufficiently high. The benefit of this approach is to provide the most possible case values for network learning. The second method of creating artificial data set was chosen in this work to achieve the optimal performance in estimating Apgar score when the under-represented class was artificially increased to approximately 20%.

3.2.5 Data Sub Sets

In most instances, traditional training-testing techniques prove adequate for the acceptance of a neural network system. However, the more complex and critical systems such as the ones used in health care field require a rigorous validation-verification approach [63]. Traditionally, a learning algorithm trains a neural network using one set of data. At the same time, the neural network is tested with similar data as that used in the training set to justify that the network has adequately learned the input domain. As the standard training-testing approach is insufficient to provide a reliable method for certification, the verification-validation approach is more reliable and dependable. The neural network is first trained with the training and testing data sets. The training and testing structure yields the weights and biases that would be verified on a third set of data – the verification set. Data of the verification set is never used in training of network. The verification set is presented to the network model in a single epoch using the structure to the best point in the parameter trials. If the neural network performs well on a previously unseen data set, then its generalization ability is more trusted in predicting new coming data. Otherwise, even if the verification results are good but the validation results are poor, the network is most likely over trained.

A dynamic and well verified and validated network model should be valid for predicting outcomes of cases outside the scope of the training and testing sets. Hence, the Niday databases used in this thesis for developing each ANN model is randomly selected to be placed into three data sets – training, testing and verification sets. Approximately two third of the data were used for training and testing purpose and the remaining one third of the cases were for verification purpose. Then the training and testing data were further subdivided into two third for training and one third for testing, and the verification data were subdivided into ten equal sub sets.

The following tables outline the data distribution of Niday 2001 in the ANN experiments. All the data were randomly separated into each data sub sets. Since "+1" instance for APGAR5 output has a extremely low representation in the database, an artificial data set was created to increase the low Apgar score data percentage to be about 20%. A total of 344 cases were "+1" instance. One third of 334 cases resulting 120 cases were kept for verification. One third of the remaining 224 cases resulting 75 cases were kept for testing. Two third of the rest of the cases resulting 149 cases were re-sampled 10 times to make up a total of 1,639 case for training. The re-sampled cases were randomly selected in training set. As a result of the additional 1,490 cases, Niday 2001 database were artificially increased to a total of 17,640 cases.

Table 3-5: Niday 2001 APGAR5 – distribution of training, testing and verification data sets.

APGAR5	**Training**		**Testing**		**Verification (10 sub sets)**	
"-1" instance	7,031	81.1%	3,515	97.9%	5,260	97.8%
"+1" instance	1,639	18.9%	75	2.1%	120	2.2%
Total cases	**8,670**		**3,590**		**5,380**	**17,640**

It was not necessary to manipulate artificial data set for DELTYPE output. The data were distributed according to the rules of separating one third verification, two third for training and testing where one third testing and two third training.

Table 3-6: Niday 2001 DELTYPE – distribution of training, testing and verification data sets.

DELTYPE	Training		Testing		Verification (10 sub sets)	
"-1" instance	5,628	85.4%	2,813	85.4%	4,220	85.4%
"+1" instance	966	14.6%	482	14.6%	720	14.6%
Total cases	**6,594**		**3,295**		**4,940**	**14,829**

3.3 Software Tools

MIRG researchers have been using the MATLAB Neural Network Toolbox as a base tool with many added features to develop ANN medical prediction models as mentioned in Section 2.1.3. MIRG's first generation automated ANN tool is a multilayer perceptron neural network using a feedforward back-propagation learning algorithm with weight elimination cost function [24]. Nine learning parameters are used to drive the learning: error ratio, momentum, learning rate, learning rate increment, learning rate decrement, weight decay constant, weight decay constant increment, weight decay constant decrement, and weight elimination scale factor. The parameter values are selected using a method of sensitivity analysis. The program keeps eight out of the nine parameters at default values while it searches for the value of the ninth parameter for the highest ANN performance results. The researchers have to repeat this procedure until the value resulting in the highest performance is found for each network parameter. Each parameter is searched by the program within a pre-defined range by using the divide-and-conquer algorithm [35]. The algorithm searches for the two highest network performance points, calculates their midpoint, and then selects the two new maximum points from the max-mid points and mid-min points. This is repeated recursively until the network converges upon the searched learning parameter that gives the "best" performance. In searching for the best performance of the network structure for the resulting ANN model, the "best" performance is judged by two criteria. Firstly, the network classifies specificity at greater than 80% and sensitivity at greater than 50%.

These criteria were determined as an acceptable rule for network structure optimisation by our physician partner [68]. Secondly, the network converges after reaching the highest log sensitivity index from checking the curves plotted by the ANN tool.

The second generation of MIRG's ANN tool – ANN research framework (RFW) – does not change the core ANN learning and has extended functionalities to speed up the running process and to deal with continuous outcomes [60]. The RFW [Appendix C] inherits the basic program construction of the previous tool. And one step further is that the program can automatically progress from one learning parameter to the next one, carrying forward the best parameter values. The default parameter ranges and start point values set in the program is shown in Table 3-7. The program executes the run of the nine parameters in the order listed in Table 3-7 and loops back in the same order. It sets the start point of the first parameter and runs a single divide-and-conquer run. Then the value is carried forward as the start point of that parameter for the next parameter's run if the result is better than the previous attempt. The best parameters keep carrying forward until a stagnant point is reached, that is, the current trial is no better than the previous best result. After reaching nine consecutive stagnant points, one network structure is complete and the next network structure is attempted.

Table 3-7: Learning parameter default ranges.

Parameter	**Start point**	**Minimum**	**Maximum**
Learning rate (lr)	0.0005	0.00005	0.01
Learning rate increment (lr_inc)	1.001	0.75	1.25
Learning rate decrement (lr_dec)	0.999	0.75	1.25
Weight decay constant (λ)	0.0001	0.00001	0.0001
Weight decay constant increment (λ_inc)	1.001	0.75	1.25
Weight decay constant decrement (λ_dec)	0.999	0.0001	1.25
Weight elimination scale factor (w_0)	0.01	0.0001	0.999
Momentum (m)	0.5	0.0001	0.999
Error ratio (err_ratio)	1.02	1.0001	1.5

The RFW also enhances the program to automatically attempt as many structures as from no hidden layer up to one hidden layer with $2n+1$ hidden nodes, where n is the number of input variables into the network. It is optional for the researchers to choose the full set of structure options or a sequential sub set of structures or any single structure.

In order to verify the ANN performance on unseen data, another useful tool is used in MIRG's medical data mining research. The committee of classifiers verification tool (CCVT) provides the user the option to manually set up a committee of classifiers of ANNs [60]. A committee of classifiers uses a majority rule system to "vote" for the accurately estimated outcome from an ANN [36]. Members of a committee can be selected from different prediction methods (e.g., ANN, CBR, and decision trees) and can also be chosen from the same data mining type with different data sets (e.g., multiple ANN data sets). It is important to ensure that each member has a high ability to classify correctly and operates different local minima. The CCVT program is set up with any number of weight, bias and structure combination from the RFW results. The tool calculates the predicted output of each case presented to it for each member of the committee. The overall result for each case is then determined by the committee voted from the majority. The CCVT allows for any number of data sets and/or prediction models to be tested at any time after the model is defined. This flexibility allows the researchers to verify an existing ANN model on a newly acquired database. Thus it helps to define when an ANN model should be retrained to account for changing clinical processes.

3.4 Performance Measures

There exist many statistical approaches to evaluate the performance of an artificial neural network classification system. The performance measures used in this thesis are presented in more detail in this section including constant predictor (CP), sensitivity, specificity, correct classification rate (CCR), average squared error (ASE), logarithmic sensitivity index, and area under the receiver operating characteristic (ROC) curve.

3.4.1 Two-class Classifier

A confusion matrix is a table used to represent the basic statistics of ANN analysis. Table 3-8 shows the statistical split of results for a two-class classifier [43].

Table 3-8: Confusion matrix for two-class classifier [43].

		Actual Outcome	
		Negative Instance (-1)	Positive Instance (+1)
Predicted Outcome	Negative Instance (-1)	True Negative (TN)	False Negative (FN)
	Positive Instance (+1)	False Positive (FP)	True Positive (TP)

Important terms for a two-class classifier are listed below:

- **Constant predictor (CP):** the percentage of classified instances as belonging to the class with the highest *a priori* probability.
- **Sensitivity or true positive rate:** the percentage of the actual positive instances that are correctly classified as positive cases.

$$Sensitivity = \frac{TP}{TP + FN} \quad \textbf{(Equation 3-2)}$$

- **Specificity or true negative rate:** the percentage of the actual negative instances that are correctly classified as negative cases.

$$Specificity = \frac{TN}{TN + FP} \quad \textbf{(Equation 3-3)}$$

- **Correct classification rate (CCR) or accuracy:** the percentage of the total instances that are correctly identified.

$$CCR = \frac{TN + TP}{TN + TP + FN + FP} \quad \textbf{(Equation 3-4)}$$

- **Average squared error (ASE):** the difference between each predicted output and its corresponding actual output squared and averaged over all cases.

$$ASE = \frac{(p_i - y_i)^2}{\sum_{i \in AllCases} (p_i - y_i)^2} \quad \textbf{(Equation 3-5)}$$

where p_i is the value of predicted outcome, and y_i is the value of actual outcome.

- **Logarithmic sensitivity index:** the stopping criteria to achieve optimal sensitivity and specificity classification while gives more weight to the sensitivity and therefore better predict the less frequent outcome.

$$\log\_sens\_index = -sesitivity^n * \log_{10}(1 - sensitivity * specificity) \quad \textbf{(Equation 3-6)}$$

where exponent n is a factor to the weighing towards higher sensitivity.

- **Receiver operating characteristic (ROC) curve:** the plot of sensitivity versus one minus specificity, which determines the ability of a model to discriminate between two outcomes.
- **Area under ROC curve:** the measure of the probability that the model can accurately predict the outcome.

3.4.2 Classification Ability

The CP is mostly used as a comparing scale with the CCR to determine the classification ability of a network. The CP is not an ideal measure because it can falsely give the impression that it is a better model based on a constant value of correctly classified cases, which is usually the percentage of actual negative instances (*a priori* negative). The CCR is more accurate as it takes account of the correctly classified true positive and true negative instances at the same time for every network setting. The CCR is desired to be higher than the CP to show that the network performs better than randomly "guessing" all negative values while "leaving" all positives to be classified incorrectly.

Ideally both sensitivity and specificity are desired to be high in choosing the optimal network performance. In reality, especially for medical data, predicting the less frequent value can be very difficult when the two-class classification is highly imbalance, yet the infrequent case is the one that requires more attention to. Thus, the model that predicts the rare outcome with greater success is often chosen [10]. That is, to achieve the satisfactory balance between the sensitivity and specificity while slightly favoring the sensitivity as it reflects better on the rare outcome prediction.

3.4.3 Stopping Criteria

A practically realized or well trained neural network is considered to have convergent activation dynamics, that is, every trajectory of the vector of activations of all neurons moves finally towards some equilibrium or stationary state [6]. A stopping criterion is used to decide when the network has reached the optimal and equilibrium value. The above mentioned statistics measures (i.e., highest CCR, lowest ASE, and highest logarithmic sensitivity index) can be used as stopping criteria in network optimization. The logarithmic sensitivity index, created by MIRG, allows the sensitivity and specificity performance indices to be taken consideration simultaneously, yet slightly favoring higher sensitivity. The degree of weighing on sensitivity can be adjusted by increasing the exponent n in Equation 3-5. Studies have shown that the logarithmic sensitivity index can efficiently optimize an ANN for highly skewed rare outcome data sets [20]. As the outcomes under interest in this thesis had a low occurrence in the data set, the logarithmic sensitivity index was used to stop network training after no improvement in the performance measure occurred in the subsequent 500 epochs.

3.4.4 Receiver Operating Characteristic Curve

The ROC curve is a graphical plot of the true positive rate (sensitivity) against false positive rate (1 – specificity) for a two-class classification system. The performance of a classifier is represented by a single point – the (sensitivity, 1 – specificity) pair – in the graph. A collection of the points generated by different settings of certain parameters is used to plot the curve. The ROC curve corresponds to a perfect classifier is a step function as the system has 100% sensitivity and 100% specificity. The 45-degree line corresponds to a worthless classifier as the system is a completely random predictor. The area under ROC curve provides a numeric measure as the accuracy of a classifier, i.e., the probability that the classifier will rank a randomly chosen positive instance higher than a randomly chosen negative instance [23]. Apparently it is preferred that the area under ROC is more than 0.5 (the area covered by the 45-degress line) and closer to 1.0 (the area of a unit square). Figure 3-3 provides examples of an excellent, good and worthless ROC curves.

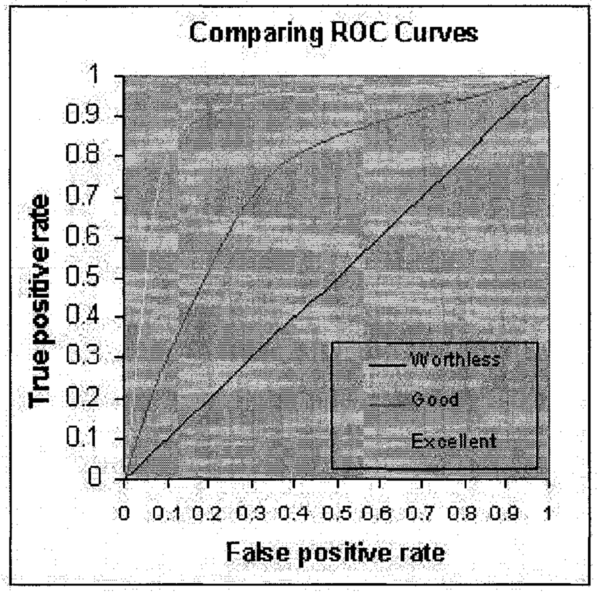


Figure 3-3: Receiver operating characteristic curves [63].

3.5 Research Steps

This section provides a summary of research steps used in this thesis.

- **Step 1: Data preparation of Niday 2001 database.**

The raw data from Niday 2001 database were carefully prepared to be a standard ANN data file format, such as encoding the text information into numeric values, removing non-applicable variables, deleting cases with missing values, normalizing input values between -1 to 1 range, categorizing output values to be "-1" and "+1" instances, artificially increasing data sets for under-represented "+1" instances, and randomly dividing data into train-test-verification sub sets. Then the clean data were saved as comma delimited text files without any file extension that is ready to be run by the MATLAB ANN software tools.

- **Step 2: Simulation of Niday 2001 applied relative weight reduction.**

The well-prepared Niday 2001 database was simulated in the automated ANN research framework which was set to attempt structures from no hidden layer up to one hidden layer of $n/2$ hidden nodes (i.e., running from 0 to 13 hidden nodes for output APGAR5, and running from 0 to 12 hidden nodes for output DELTYPE). The number of hidden nodes was chosen to be half of the number of input variables because it was found that a large size database with a large number of variables needs less complicated network structure. The best performance structure was chosen based on the measurements of sensitivity, specificity, CCR, ASE, and ROC. The selected structure was then applied the relative weight reduction (RWR) algorithm to define the relative importance of the input variables. By removing the least important variables one by one until the network performance significantly degraded, an ANN model with the most important indicators was determined. The Niday 2001 APGAR5 model and Niday 2001 DELTYPE model found in this work were compared with the other ANN models from the literatures.

- **Step 3: Simulation of Niday 2001 applied principal component analysis.**

In this step, a software tool applied principal component analysis (PCA) algorithm was developed from MATLAB. For the consistency with the existing MIRG ANN research tools, the input interface of loading data file and the interface of the experiment parameter configuration remain the same as the other tools, although the computational logic is different. Therefore, the well-prepared Niday 2001 database (except dividing into train-test-verification sub sets) was able to be used in this step. The output file of the PCA tool combined and ranked the importance of components in the order of descending eigenvalues. By removing the least important components one by one according to the ranking of eigenvalues, a new PCA database was available to be subdivided into train-test-verification sub sets. This new PCA data were then fed to the automated ANN research framework. Since the number of inputs was reduced to smaller components, the network structure from 0 to $2n+1$ hidden nodes (where n is the number of PCA components) was attempted. Again, the best performance structure was chosen based on the measurements of sensitivity, specificity, CCR, ASE, and ROC. The performance results for APGAR5 and DELTYPE were compared with the results from Step 1.

- **Step 4: Data preparation of Niday 2004 database.**

Before applying the committee of classifiers verification tool (CCVT) to the Niday 2004 database, some data pre-processing procedures were performed. Only the variables determined as the important factors from Niday 2001 models were selected, followed by the procedures of deleting cases with missing values, normalizing input values, categorizing output instances, and randomizing data. Without separating the data into sub sets, one comma delimited text file containing the necessary patient information from Niday 2004 is ready to be run in the CCVT.

- **Step 5: Verification of Niday 2004 using committee of classifiers verification tool.**

In order to verify the Niday 2001 ANN models developed from Step 1 and 3 were capable to estimate the same interested outputs APGAR5 and DELTYPE in Niday 2004 database, the committee of classifiers verification tool (CCVT) was used in this step. Before applying the tool to the Niday 2004 database, the Niday 2001 database was re-run from 0 to $2n+1$ hidden nodes only on the important factors, and then five of the best performance network structures were chosen to set up a committee of classifiers for each output. All the network elements were extracted from the five structures including weights, biases, error ratio, momentum, learning rate, learning rate increment, learning rate decrement, weight decay constant, weight decay constant increment, weight decay constant decrement, and weight elimination scale factor. The pre-processed Niday 2004 data were then simulated in the CCVT. And the performance results were verified to be sufficient or not for predicting outcomes of future data.

4 Result Analysis

4.1 Results of Niday 2001 Applied RWR

Following Step 2 mentioned in Section 3.5, the experiment results presented in this section are focused on the Niday 2001 database applied relative weight reduction (RWR) variable selection method. The carried out Niday 2001 models are compared with the risk factors found by Ibrahim in the ANN newborn outcome prediction models.

4.1.1 Niday 2001 APGAR5 Applied RWR

Since the "+1" instances (low Apgar score range from 0 to 6) were extremely under-represented (2.2% original low APGAR5) in the Niday 2001 database, 1,490 "+1" cases were duplicated and added into the training data set (18.9% artificial low APGAR5). The total number of cases was increased to 17,640 and each patient case matched with 26 input variables and 1 output APGAR5. The automated ANN RFW was set to attempt the network structures from 0 hidden nodes to 13 hidden nodes. Many trial and error experiments have proved that a less complex network structure gives a relatively better performance in a faster running time. The performance measures of each network structure are shown in Table 4-1. Note that the statistics of verification set are the average taken from the 10 verification sub sets of each attempted structure.

Table 4-1: Niday 2001 APGAR5 – performance measures of structure 0 to 13 hidden nodes.

Hidden nodes	Data sets	Log sens index	Sensitivity	Specificity	CCR	ASE	ROC
0	Train	0.18876	0.59732	0.86545	81.4764	0.54211	0.79094
	Test	0.15367	0.54667	0.87169	86.4903	0.42898	0.74563
	Verify	0.120581	0.475	0.862545	85.39033	0.4435	0.733531
1	Train	0.15205	0.57047	0.80401	75.9862	0.94366	0.69001
	Test	0.14374	0.56	0.79687	79.1922	0.80814	0.67637
	Verify	0.150708	0.533334	0.807414	80.13013	0.773315	0.678691
2	Train	0.15408	0.66443	0.62267	63.0565	0.8238	0.64022
	Test	0.1734	0.70667	0.61081	61.2813	0.74748	0.66041
	Verify	0.12387	0.583335	0.614449	61.37548	0.747508	0.603161
3	Train	0.20469	0.65101	0.79135	76.4821	0.93621	0.72592
	Test	0.18782	0.62667	0.79545	79.1922	0.82919	0.70161
	Verify	0.152183	0.533335	0.799049	79.31226	0.821988	0.674335
4	Train	0.18028	0.65772	0.71156	70.1384	1.1857	0.69401
	Test	0.23397	0.73333	0.70953	71.0028	1.1563	0.71111
	Verify	0.152298	0.591666	0.708936	70.63196	1.16727	0.646522
5	**Train**	**0.16639**	**0.5302**	**0.97042**	**88.7197**	**0.38365**	**0.80219**
	Test	**0.098048**	**0.42667**	**0.96302**	**95.1811**	**0.1921**	**0.73306**
	Verify	**0.06106**	**0.316667**	**0.964258**	**94.98142**	**0.19225**	**0.722258**
6	Train	0.18458	0.63087	0.77699	74.9366	0.98131	0.71316
	Test	0.1856	0.62667	0.7889	78.5515	0.8374	0.69848
	Verify	0.15026	0.550001	0.785361	78.01114	0.848016	0.67737
7	Train	0.093994	0.42282	0.94752	84.8328	0.5565	0.70808
	Test	0.082177	0.4	0.94225	93.0919	0.26142	0.67377
	Verify	0.059749	0.308334	0.947148	93.28997	0.248216	0.636746
8	Train	0.21443	0.63758	0.8454	80.6113	0.57163	0.81384
	Test	0.15572	0.56	0.84438	83.844	0.51689	0.73275
	Verify	0.110191	0.466666	0.846008	83.75466	0.515698	0.720111
9	Train	0.19304	0.71141	0.65311	66.4129	0.99029	0.72558
	Test	0.23508	0.77333	0.65092	65.3482	0.99916	0.74172
	Verify	0.174849	0.658334	0.659506	65.94796	0.998979	0.689417
10	Train	0.11598	0.4698	0.92291	83.7255	0.64322	0.69047
	Test	0.092853	0.42667	0.92376	91.337	0.33657	0.67595
	Verify	0.081322	0.358334	0.924715	91.20817	0.341256	0.646293
11	Train	0.18504	0.65772	0.72493	71.2226	1.0989	0.68863
	Test	0.20384	0.68	0.73314	73.2033	1.0135	0.70436
	Verify	0.130426	0.549999	0.73498	73.08551	1.010314	0.670723
12	Train	0.14309	0.61074	0.68269	66.9089	0.86925	0.71491
	Test	0.21015	0.72	0.67966	68.0501	0.83369	0.721
	Verify	0.137308	0.574999	0.680609	67.82527	0.836704	0.667245
13	Train	0.17753	0.54362	0.97227	89.1234	0.35948	0.83475
	Test	0.10636	0.44	0.97013	95.9053	0.15647	0.76613
	Verify	0.058856	0.308334	0.96882	95.40893	0.163445	0.735203

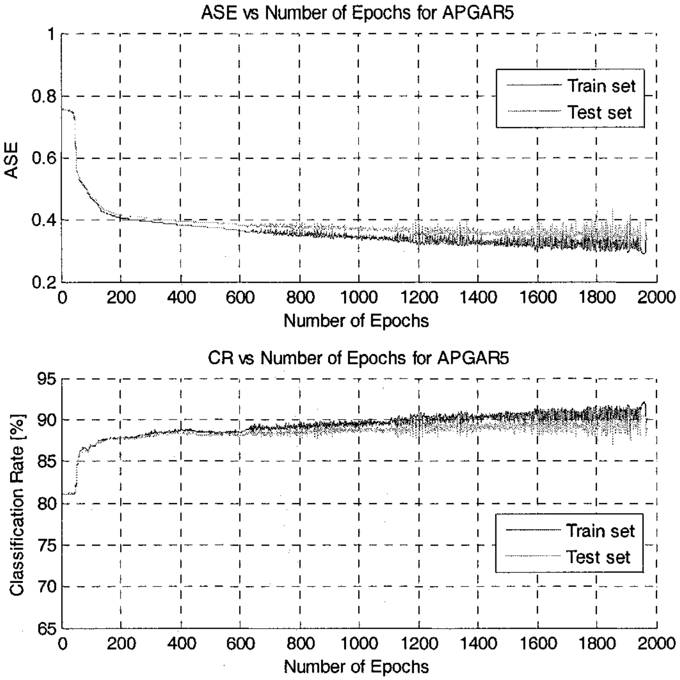


Figure 4-1: Niday 2001 APGAR5 – structure 5 ASE and CCR graphs.

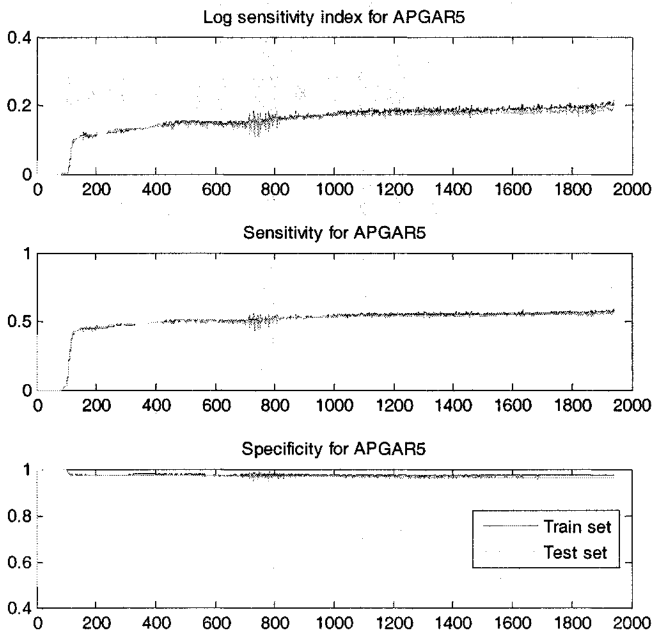


Figure 4-2: Niday 2001 APGAR5 – structure 5 log sens index, sensitivity and specificity graphs.

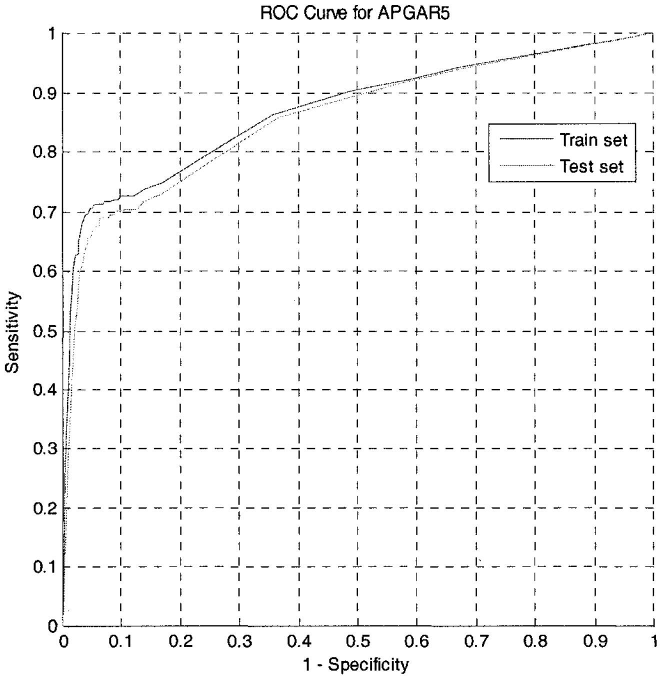


Figure 4-3: Niday 2001 APGAR5 – structure 5 ROC curve.

Overall, structure 5 had the best performance over the other structures. Both the ASE_CR and log_sensitivity_index graphs showed that the network converged, and the ROC curve was closed to a step function. Thus structure 5 was chosen to apply the relative weight reduction (RWR) method. The weights of input-to-hidden nodes and hidden-to-output nodes were extracted and calculated by RWR equations. The relative importance (RI) for the input variables are ranked from highest to lowest in Table 4-2.

Table 4-2: Niday 2001 APGAR5 – input variables relative importance ranking.

Ranking (highest to lowest)	Variables	RI
1	WEIGHT1	0.09496
2	PUDENDAL	0.07803
3	TERM	0.06563
4	LABTYPE	0.05562
5	GEST	0.05236
6	DELBY	0.05214
7	GENERAL	0.04895
8	PRETERM	0.04622
9	MONITOR	0.04616
10	SCALP	0.04540
11	NBABIES	0.04147
12	SPINAL	0.03632
13	NARCOT	0.03454
14	HOSPTYPE	0.03310
15	STEROIDS	0.03258
16	SMOKING	0.03126
17	PARITY	0.03019
18	DELTYPE	0.02570
19	MOMTRANS	0.02346
20	GENDER	0.02301
21	AGE	0.02124
22	ASSISTED	0.02001
23	PRESENT	0.01835
24	BF	0.01669
25	NITOXIDE	0.01378
26	EPIDAL	0.01283

The relative importance was used to find the minimum input variable set required to construct an ANN model to predict the APGAR5 output. The input variables with the lowest relative importance were removed one at a time until the ANN performance degraded. The 12 highest ranked input variables kept were WEIGHT1, PUDENDAL, TERM, LABTYPE, GEST, DELBY, GENERAL, PERTERM, MONITOR, SCALP, NBABIES, and SPINAL. The network performance using structure 5 are given in Table 4-3.

Table 4-3: Niday 2001 APGAR5 – structure 5 performance for 12 highest ranked variables.

Hidden Nodes	Data sets	Log sens index	Sensitivity	Specificity	CCR	ASE	ROC
5	Train	0.27877	0.65101	0.96302	90.4037	0.32674	0.84006
	Test	0.11192	0.45333	0.95647	94.5961	0.19134	0.75004
	Verify	0.068758	0.325001	0.960455	94.62825	0.198546	0.648947

MIRG member Ibrahim had done a similar work using the Niday Perinatal Database to identify the risk factors in ANN newborn outcome prediction models. The following tables compared the findings of Ibrahim's model and the Niday 2001 model developed in this work for outcome APGAR5.

Table 4-4: Comparison of important indicators for APGAR5 models.

Ranking (highest to lowest)	Ibrahim model	Niday 2001 model
1	***GEST***	***WEIGHT1***
2	***SPINAL***	***PUDENDAL***
3	***SCALP***	TERM
4	BF	LABTYPE
5	***GENERAL***	***GEST***
6	***WEIGHT1***	DELBY
7	SMOKING	***GENERAL***
8	***PUDENDAL***	***PRETERM***
9	EPIDURAL	MONITOR
10	NITOXIDE	***SCALP***
11	STEROIDS	NBABIES
12	***PRETERM***	***SPINAL***

Table 4-5: Comparison of performance for APGAR5 models.

Performance measures	Ibrahim model	Niday 2001 model
Sensitivity	40.2%	32.5%
Specificity	95.1%	96.0%
CCR	93.8%	94.6%
ASE	0.259	0.199
ROC	0.754	0.649

In general, the Niday 2001 model developed in this work is comparable with Ibrahim's model. The sensitivity and area under ROC of Ibrahim's model were marginally higher than that of Niday 2001 model, but specificity and CCR of Niday 2001 model were slightly better than that of Ibrahim's model. Though Niday 2001 model is developed using a different year of data from Ibrahim's model, the commonality and difference of network performance reveals the consequence of multiple year analysis. Carrying out multiple years ANN model development will offer a systematic overview on the one outcome estimation. The differences between variables selected by Ibrahim's model and this Niday 2001 model could be due to the fact that Ibrahim used a manual approach for determining best ANN structures, using a limited number of possibilities, while the approach here was fully automated and used a much larger range of possible structures.

The same 7 out of 12 variables are presented in both models regardless of the different ranking. Three of the maternal pain relievers (pudendal, general, spinal) are presented in both models, which lead to a medical question – could the use of pain relievers seriously affect the baby's condition? It is a fact that a very high percentage of women are given maternal pain relievers during labour. The safe use, dosage and timing of giving pain relievers may play a role of helping the mother to give birth, as well as affecting the newborn's condition. The significance of the 3 pain relievers found from the Niday 2001 Apgar score estimation model provides analytical evidence for a future medical study on perinatal health care. Moreover, some of the variables are appeared to be important indicators of baby's health condition as expected, such as baby's weight, gestational age, and number of babies in the pregnancy. A preterm baby with a gestational age of less than 37 weeks could have more medical problems than a term baby. The number of previous term babies and number of previous preterm babies can also indicate the chances of having preterm baby for the current birth. The labour type (spontaneous, induced or no labour) indicates the baby's activity at the moment of birth. The other factors (delivered by, fetal monitoring, and whether scalp blood gases were done) are more due to the human skills and use of technology, which are fluctuated at each birth, so that it is not surprising to see them play a role in the Apgar score estimation.

4.1.2 Niday 2001 DELTYPE Applied RWR

There were 14.6% of the cases were of "+1" cases (cesarean section delivery) in the Niday 2001 database, which need not be artificially increased for ANN training. The DELTYPE model was simulated from the database containing 14,829 cases with 23 input variables and 1 output DELTYPE. The automated ANN RFW was set to attempt the network structures from 0 hidden nodes to 12 hidden nodes. The performance measures of each network structure are shown in Table 4-6.

Table 4-6: Niday 2001 DELTYPE – performance measures of structure 0 to 12 hidden nodes.

Hidden nodes	Data sets	Log sens index	Sensitivity	Specificity	CCR	ASE	ROC
0	Train	0.33832	0.72774	0.90299	87.7313	0.37725	0.90451
	Test	0.31904	0.71162	0.90473	87.648	0.37796	0.89966
	Verify	0.299301	0.695834	0.896208	86.70042	0.403533	0.885314
1	Train	0.33294	0.72567	0.8989	87.3521	0.35641	0.88102
	Test	0.31795	0.71162	0.90295	87.4962	0.3562	0.86966
	Verify	0.281659	0.681945	0.890995	86.05263	0.389799	0.853951
2	Train	0.33971	0.73085	0.89908	87.4431	0.34932	0.8822
	Test	0.32252	0.71784	0.89797	87.1624	0.35494	0.87532
	Verify	0.305934	0.702779	0.892416	86.47773	0.382971	0.856909
3	**Train**	**0.19894**	**0.56625**	**0.97957**	**91.9017**	**0.26521**	**0.90609**
	Test	**0.20053**	**0.56846**	**0.97831**	**91.8361**	**0.26678**	**0.90492**
	Verify	**0.195216**	**0.556944**	**0.976537**	**91.53847**	**0.283576**	**0.887227**
4	Train	0.32499	0.70911	0.91933	88.8535	0.31918	0.90604
	Test	0.30166	0.68465	0.93103	89.4992	0.32206	0.89548
	Verify	0.282629	0.672223	0.912084	87.71255	0.351731	0.878132
5	Train	0.24089	0.61491	0.96642	91.4923	0.27669	0.90485
	Test	0.21134	0.58299	0.97085	91.4112	0.27721	0.90689
	Verify	0.213169	0.581944	0.960663	90.54655	0.299302	0.890527
6	Train	0.30842	0.69255	0.92608	89.1871	0.33603	0.90838
	Test	0.27117	0.65768	0.9321	89.1958	0.34029	0.90293
	Verify	0.267447	0.654166	0.918956	88.03643	0.368014	0.889587
7	Train	0.30928	0.72153	0.8694	84.774	0.40669	0.87169
	Test	0.29572	0.70954	0.86953	84.6131	0.41545	0.86965
	Verify	0.273079	0.683334	0.867063	84.02835	0.432654	0.857259
8	Train	0.33232	0.70911	0.93088	89.8392	0.30541	0.91445
	Test	0.29238	0.67635	0.9321	89.4689	0.33571	0.8938
	Verify	0.276664	0.663889	0.916586	87.97571	0.369357	0.878927
9	Train	0.39277	0.78261	0.87544	86.1844	0.38394	0.91682
	Test	0.35423	0.75519	0.87451	85.7056	0.40188	0.90261
	Verify	0.326237	0.733334	0.866114	84.67612	0.424579	0.893287
10	Train	0.22159	0.5911	0.97814	92.1444	0.25962	0.91113
	Test	0.19661	0.56432	0.9776	91.7147	0.27213	0.9044
	Verify	0.199355	0.56389	0.970377	91.11336	0.288049	0.894791
11	Train	0.36998	0.73706	0.92964	90.1426	0.30444	0.91636
	Test	0.27631	0.6639	0.92855	88.9833	0.34481	0.88981
	Verify	0.285912	0.668056	0.917771	88.13766	0.369159	0.879831
12	Train	0.38461	0.76294	0.90014	88.0042	0.33644	0.90802
	Test	0.33598	0.72614	0.9026	87.6783	0.36518	0.8933
	Verify	0.304897	0.702779	0.882703	85.64777	0.400164	0.882079

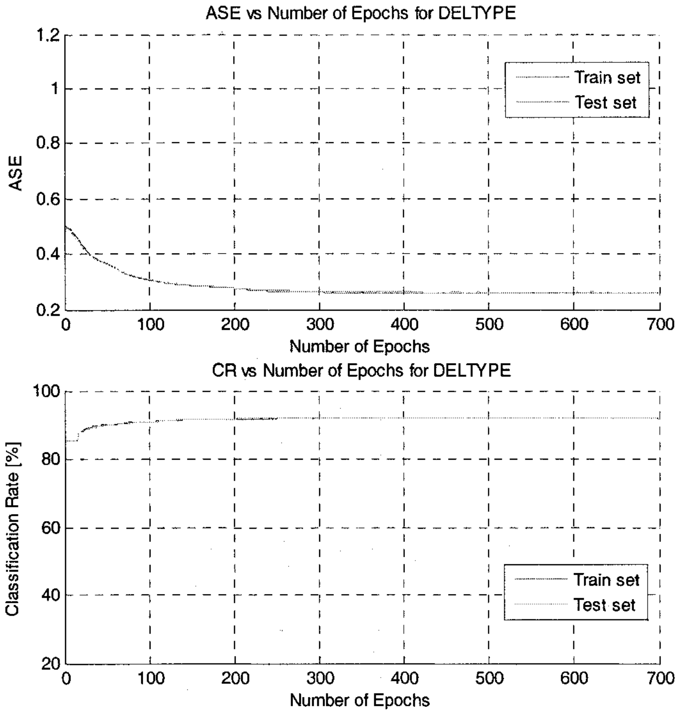


Figure 4-4: Niday 2001 DELTYPE – structure 3 ASE and CCR graphs.

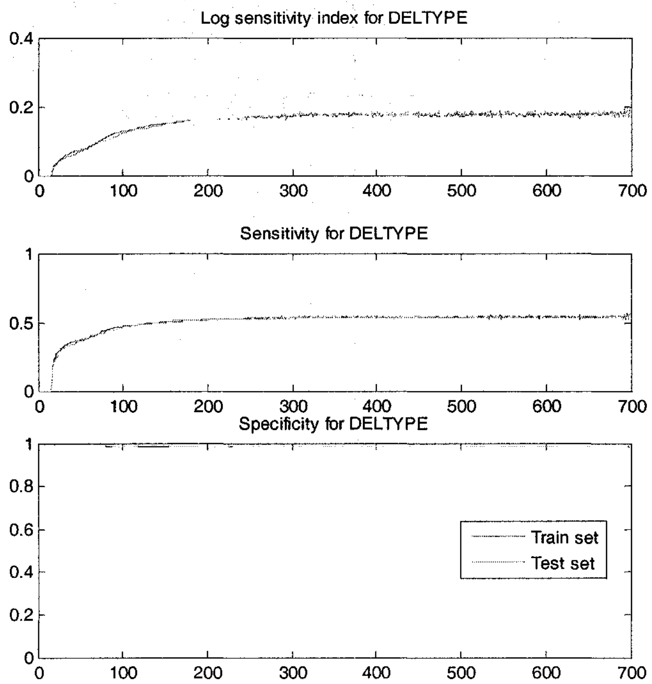


Figure 4-5: Niday 2001 DELTYPE – structure 3 log sens index, sensitivity and specificity graphs.

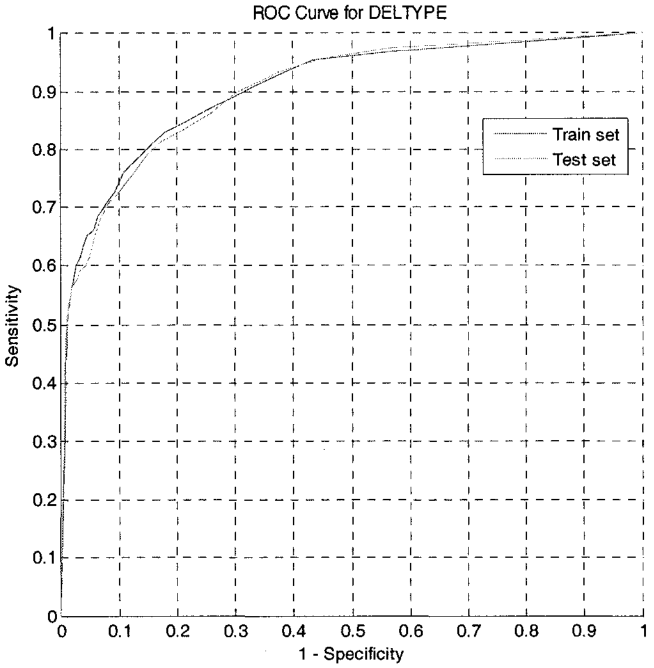


Figure 4-6: Niday 2001 DELTYPE – structure 3 ROC curve.

Structure 3 was chosen to apply the RWR method for minimum variable selection for the following reasons. Structure 3 had the second highest CCR and the second lowest ASE over the 12 structures ran. Although the performance of structure 10 should yield ideally the best results, its ASE_CR and log_sensitivity_index graphs showed that the network had not yet converged at 2,500 epochs. And its ROC curve performed poorly as closed to a 45 degree line, that is, it was more of a random guesser at the train-test phase. Even though other structures such as structure 11 and 12 performed relatively well, structure 3 was still chosen as it gave a simpler network set up which would significantly affect the experiment running time. The weights of input-to-hidden nodes and hidden-to-output nodes of structure 3 were computed in RWR equations. The ranking of the variable RI is showed in Table 4-7.

Table 4-7: Niday 2001 DELTYPE – input variables relative importance ranking.

Ranking (highest to lowest)	Variables	RI
1	GENERAL	0.16167
2	SPINAL	0.13235
3	TERM	0.12239
4	PUDENDAL	0.09801
5	NITOXIDE	0.06645
6	AGE	0.05565
7	PRESENT	0.05348
8	WEIGHT1	0.04587
9	GEST	0.03274
10	PRETERM	0.03265
11	NBABIES	0.03188
12	EPIDURAL	0.03134
13	PARITY	0.03000
14	SCALP	0.01918
15	MOMTRANS	0.01887
16	HOSPTYPE	0.01238
17	BF	0.01005
18	MONITOR	0.00985
19	STEROIDS	0.00912
20	LABTYPE	0.00903
21	NARCOT	0.00799
22	SMOKING	0.00628
23	GENDER	0.00277

After the process of removing the lowest ranked variables one at a time until the ANN performance degraded, 10 factors were kept to construct a DELTYPE model, including GENERAL, SPINAL, TERM, PUDENDAL, NITOXIDE, AGE, PRESENT, WEIGHT1, GEST and PRETERM. Table 4-8 is the performance measures of the re-run of structure 3 on the 10 highest ranked variables.

Table 4-8: Niday 2001 DELTYPE – structure 3 performance for 10 highest ranked variables.

Hidden nodes	Data sets	Log sens index	Sensitivity	Specificity	CCR	ASE	ROC
3	Train	0.17847	0.54141	0.98241	91.7804	0.27797	0.88255
	Test	0.18243	0.54564	0.984	91.9879	0.27879	0.87808
	Verify	0.186805	0.545833	0.979618	91.63967	0.292761	0.864448

Again, to examine how well the Niday 2001 DELTYPE model predicts the outcome, it was compared with the model Ibrahim developed before.

Table 4-9: Comparison of important indicators for DELTYPE models.

Ranking (highest to lowest)	Ibrahim model	Niday 2001 model
1	***PUDENDAL***	***GENERAL***
2	***GENERAL***	***SPINAL***
3	***SPINAL***	***TERM***
4	***NITOXIDE***	***PUDENDAL***
5	***TERM***	***NITOXIDE***
6	***PRESENT***	AGE
7	***WEIGHT1***	***PRESENT***
8		***WEIGHT1***
9		GEST
10		PRETERM

Table 4-10: Comparison of performance for DELTYPE models.

Performance measures	Ibrahim model	Niday 2001 model
Sensitivity	52.8%	54.6%
Specificity	97.9%	98.0%
CCR	91.3%	91.6%
ASE	0.296	0.293
ROC	0.890	0.864

Observed from the above tables, the Niday 2001 model performed as well as Ibrahim's model. The sensitivity, specificity, and CCR of Niday 2001 model were marginally higher than those of Ibrahim's model, but the ASE and area under ROC of Niday 2001 model were marginally lower than that of Ibrahim's. In fact, all seven risk factors identified in Ibrahim's model are presented in the Niday 2001 model regardless of the order of ranking. There are three additional factors shown as important variables in the Niday 2001 model.

Similar to the Niday 2001 APGAR5 model, 3 of the maternal pain relievers – spinal, general, and pudendal – plus another pain reliever, nitrous oxide, are presented in the Niday 2001 DELTYPE model. This confirms the urgency of a medical research on the impact of using pain relievers to perinatal health care. More experiments can be run in the future to observe the relationship of all different sorts of pain relievers and various perinatal outcomes. The factors expected to be presented in the model are the presentation of the baby, gestational age, and the weight of the baby. The position of the baby during labour provides important information to the doctor to decide if the mother needs a cesarean section operation. For medical reasons, having a baby with a breech position during labour increases the chance of cesarean delivery. The maturity, weight and size of the baby can also lead to a cesarean section operation from the doctor's suggestions. However, the age of the mother was not expected to be an important factor in the delivery type prediction. Yet it is found that it's one of the top 10 variables in the Niday 2001 DELTYPE model.

4.2 Results of Niday 2001 Applied PCA

Research Step 3 in Section 3.5 gave procedures of running ANN model development on the Niday 2001 database applied principal component analysis (PCA) variable selection method. The results are compared with the network performance found in the previous Step 2.

4.2.1 Niday 2001 APGAR5 Applied PCA

Although the "+1" APGAR5 cases were extremely under-represented in the Niday 2001 database, it was not advisable to artificially increase the skewed data when applying PCA method. The true data distribution must be imported to the PCA tool for a full data transformation. The 26 input variables were transformed to a set of combined variables. The mapping function in the PCA program exported a newly transformed database. According to the calculated eigenvalues, the proportion of every variable by PCA is ranked in Table 4-11.

Table 4-11: Niday 2001 APGAR5 – variable principal component analysis ranking.

Ranking (highest to lowest)	Proportion of variables (percentage)
1	19.54
2	15.48
3	10.73
4	7.74
5	6.52
6	6.00
7	5.99
8	5.15
9	5.06
10	3.58
11	3.29
12	2.89
13	2.66
14	1.93
15	1.72
16	1.24
17	0.21
18	0.08
19	0.05
20	0.06
21	0.04
22	0.02
23	0.02
24	0.00
25	0.00
26	0.00

The last 3 components are almost of no importance since they count for 0% in the new data transformation. The components ranked from 17th to 23rd are each less than 0.3% of importance. By removing the 17th components onward, the proportion of loss of importance is about 0.5% in total, which is negligibly small in a large database. Different selections of components were tried and it was found that removing the last 10 components gave relatively stable network results. The newly transformed database was separated into train-test-verification sets following the one-third and two-third rules. A total of 16,150 cases with 16 components and 1 output APGAR5 were then simulated in ANN RFW tool given the best network performance at 8 hidden nodes.

Table 4-12: Niday 2001 APGAR5 – structure 8 performance for 16 components.

Hidden nodes	Data sets	Log sens index	Sensitivity	Specificity	CCR	ASE	ROC
8	Train	0.17945	0.6443	0.73475	71.7647	1.0748	0.71822
	Test	0.17207	0.64	0.72119	71.9499	1.1735	0.70554
	Verify	0.121329	0.533333	0.729849	72.54647	1.13153	0.645206

In comparison of the two ANN models developed using different variable selection methods, the RWR method out-performed the PCA method in specificity and in Correct Classification Rate. However the PCA approach provided a higher sensitivity.

Table 4-13: Comparison of RWR and PCA for output APGAR5.

Performance measures	RWR method	PCA method
Number of inputs	12	16
Number of hidden nodes	5	8
Sensitivity	32.5%	53.3%
Specificity	96.0%	73.0%
CCR	94.6%	72.5%
ASE	0.199	1.131
ROC	0.649	0.645

The APGAR5 model developed from RWR method takes a lower number of input variables and hidden nodes to get better CCR results. As studied from the theory of PCA method, it simply uses linear transformation to filter out the noise while manipulating the important mathematical relationship among components. The less than favorable results from PCA analysis imply that the variables of the perinatal data have no linear correlation. The filtering effect of the PCA tool might actually have removed some important information. Moreover, applying PCA method has not saved time or work to pre-process the data in running the ANN RFW simulation. However, this approach may provide some improvement in the sensitivity (correctly classifying the true positive cases. This should be explored in future work.

4.2.2 Niday 2001 DELTYPE Applied PCA

This section repeated procedures of applying PCA variable selection method to predict DELTYPE output in Niday 2001 database. The 24 input variables were transformed to a new set of components ranked by the proportion of importance according to eigenvalues.

Table 4-14: Niday 2001 DELTYPE – variable principal component analysis ranking.

Ranking (highest to lowest)	Proportion of variables (percentage)
1	31.36
2	24.11
3	14.34
4	11.65
5	5.98
6	3.56
7	3.07
8	2.68
9	1.96
10	1.06
11	0.14
12	0.04
13	0.02
14	0.01
15	0.01
16	0.01
17	0.00
18	0.00
19	0.00
20	0.00
21	0.00
22	0.00
23	0.00

Observing from bottom up, the 17th to 23rd components are calculated as 0% proportion in the new data transformation; then the 11th to 16th components are each less than 0.2% of proportion of the new data sets. Removing components 11th to 23rd loses approximately 0.3% of the information from the original database, which is negligibly small in a large database. Then a total of 14,829 cases with 10 components and 1 output DELTYPE were simulated in the ANN RFW tool. Network structure 5 gave the best performance measures as shown below.

Table 4-15: Niday 2001 DELTYPE – structure 6 performance for 11 components.

Hidden nodes	Data sets	Log sens index	Sensitivity	Specificity	CCR	ASE	ROC
5	Train	0.34107	0.72774	0.90707	88.0801	0.33562	0.90446
	Test	0.32064	0.70954	0.91148	88.1942	0.34577	0.89689
	Verify	0.312533	0.705557	0.894551	86.70042	0.376762	0.886711

Table 4-16 shows the comparison of estimation ability of the ANN model for the same outcome DELTYPE but using different RWR and PCA variable selection methods.

Table 4-16: Comparison of RWR and PCA for output DELTYPE.

Performance measures	RWR method	PCA method
Number of inputs	10	10
Number of hidden nodes	3	5
Sensitivity	54.6%	70.6%
Specificity	98.0%	89.5%
CCR	91.6%	86.7%
ASE	0.293	0.377
ROC	0.864	0.886

Again, the RWR method presents better specificity measures than the PCA method, while PCA is better at sensitivity measurements than RWR. As previously mentioned, this implies that the relationship of the perinatal data must be non-linearly distributed, so that the preliminary transformation of PCA is not necessary to capture certain information. On the other hand, sensitivity is the percentage of actual positive instance that are correctly classified. PCA models have a better ability to predict the "+1" cases, which are the least frequent instance. The least frequent instance is often the cases most interest of the doctors. PCA has its advantage of being more "sensitive" in filtering the noise information and predicting the rare outcomes.

In order to take the advantages of both RWR (high specificity) and PCA (high sensitivity) methods, it is recommended that future work explore the design of a hybrid system predicting specificity from the RWR approach and predicting sensitivity using the PCA approach. The classified outputs from each method would become the "new" inputs of the hybrid system. From giving selection on the preferred performance measures, the "new" inputs can be classified by running in another learning network such as decision tree.

4.3 Results of Niday 2004 Using CCVT

In order to verify if the ANN model developed from Niday 2001 database can also be used to estimate Niday 2004 APGAR5 and DELTYPE outputs. The same input variables selected from Niday 2001 model was chosen in Niday 2004 database, and only the complete cases were used for the committee of classifiers verification tool (CCVT). Table 4-17 and Table 4-18 give the detailed breakdown of deleted cases for each input variable in Niday 2004 database. Niday 2004 has the total number of complete cases 38,045 for output APGAR5 and complete cases 38,260 for DELTYPE output. Note that in Niday 2004 database, the maternal pain relievers were grouped under one variable column. After the incomplete cases for pain reliever variable were removed, the variables PUDENDAL, GENERAL, SPINAL and NITOXIDE were manually separated for ANN CCVT simulation.

Table 4-17: Niday 2004 – number of deleted cases for output APGAR5.

Variables	**Cases deleted in APGAR5**
APGAR5	177
PAINRELIEV	9,314
WEIGHT1	115
TERM	86
LABTYPE	11
GEST	69
DELBY	342
PRETERM	555
MONITOR	473
SCALP	0
NBABIES	0
Total deleted	**11,142**

Table 4-18: Niday 2004 – number of deleted cases for output DELTYPE.

Variables	Cases deleted in DELTYPE
DELTYPE	19
PAINRELIEV	9,223
WEIGHT1	115
GEST	120
AGE	56
TERM	110
PRESENT	949
DELBY	335
Total deleted	**10,927**

The Niday 2001 APGAR5 RWR model (17,640 cases) was then re-run using 12 important factors from 0 to 25 $(2n+1)$ hidden nodes, and the Niday 2001 DELTYPE RWR model (14,829 cases) was re-run using 10 important factors from 0 to 21 $(2n+1)$ hidden nodes. This step was necessary for selecting the five best structures to construct a committee of classifiers to verify Niday 2004 data. Table 4-19 and Table 4-20 are the performance measures of only verification sets on important factors from the Niday 2001 models. The highlighted structures from the following tables are selected to extract the network elements to construct the five members of committees. For output APGAR5, structure 4, 14, 18, 19, and 22 are selected. For output DELTYPE, structure 3, 4, 5, 16, and 20 are selected. These structures are selected because they have relatively "best" specificity and sensitivity performance and converged to stable at log sensitivity index curves.

Table 4-19: Niday 2001 APGAR5 – performance measures of verification sets on 12 factors.

Hidden nodes	Data sets	Log sens index	Sensitivity	Specificity	CCR	ASE	ROC
0	Verify	0.140447	0.583333	0.691062	68.86617	1.24525	0.636533
1	Verify	0.153255	0.533335	0.828708	82.21189	0.567823	0.702464
2	Verify	0.173226	0.708333	0.590304	59.29368	1.35383	0.619417
3	Verify	0.141988	0.549999	0.783269	77.80671	0.85977	0.672386
4	**Verify**	**0.068676**	**0.325**	**0.971294**	**95.68773**	**0.159136**	**0.703716**
5	Verify	0.117913	0.458334	0.898098	88.82901	0.432092	0.690748
6	Verify	0.109897	0.475	0.808744	80.13011	0.708657	0.672892
7	Verify	0.122843	0.474998	0.863307	85.46469	0.464126	0.690819
8	Verify	0.157512	0.608332	0.707985	70.57621	0.932739	0.705275
9	Verify	0.15759	0.608333	0.701332	69.92565	1.026697	0.677249
10	Verify	0.11978	0.574999	0.646768	64.51671	1.15865	0.660345
11	Verify	0.128414	0.575	0.673575	67.13754	1.2997	0.630449
12	Verify	0.141918	0.650002	0.588403	58.97769	1.64091	0.619201
13	Verify	0.166147	0.641667	0.651141	65.09294	1.061337	0.686604
14	**Verify**	**0.051103**	**0.283334**	**0.976806**	**96.13383**	**0.129609**	**0.661787**
15	Verify	0.152527	0.55	0.78536	78.01116	0.76499	0.689116
16	Verify	0.153159	0.600002	0.704563	70.22306	0.954033	0.701932
17	Verify	0.153639	0.575	0.738972	73.5316	0.829526	0.689528
18	**Verify**	**0.04799**	**0.283333**	**0.981748**	**96.61711**	**0.13982**	**0.725437**
19	**Verify**	**0.145457**	**0.508332**	**0.898098**	**88.94053**	**0.434719**	**0.703557**
20	Verify	0.165295	0.566667	0.809126	80.37173	0.702506	0.706742
21	Verify	0.125283	0.566668	0.672244	66.98884	1.22218	0.632851
22	**Verify**	**0.06057**	**0.308334**	**0.979469**	**96.4498**	**0.142812**	**0.757517**
23	Verify	0.163817	0.683333	0.60019	60.20447	1.17649	0.696966
24	Verify	0.205053	0.741667	0.61749	62.02603	0.997273	0.716849
25	Verify	0.121329	0.533333	0.729849	72.54647	1.13153	0.645206

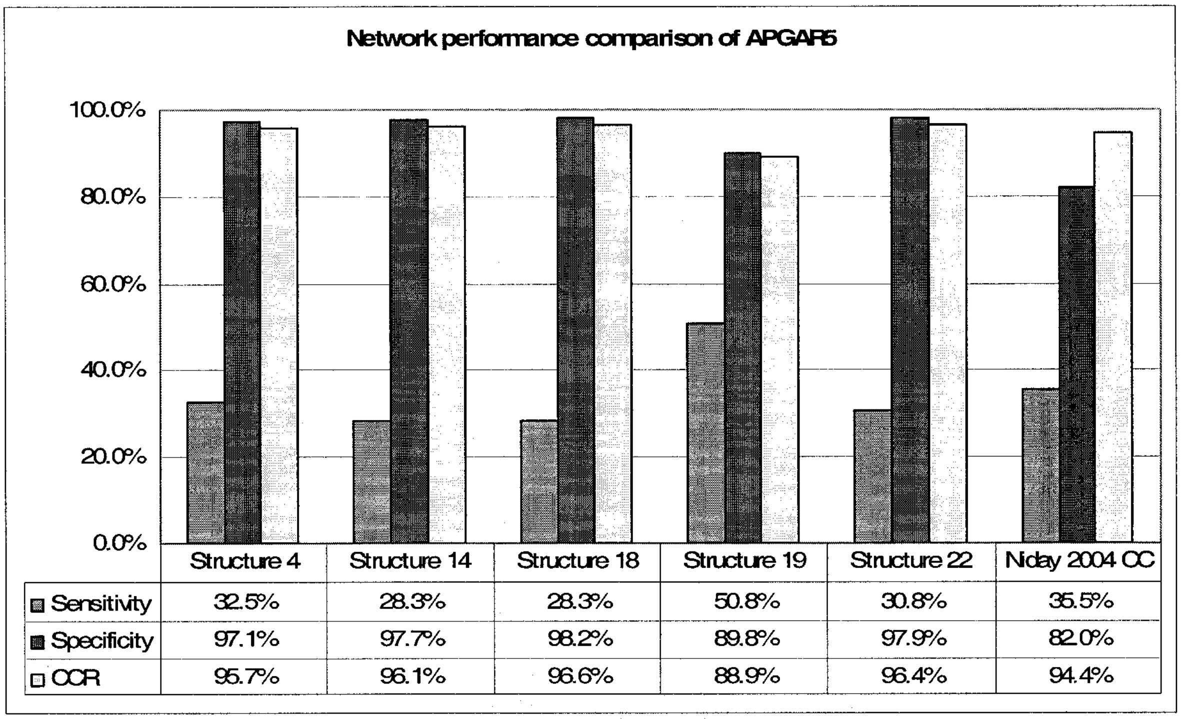
Table 4-20: Niday 2001 DELTYPE – performance measures of verification sets on 10 factors.

Hidden nodes	Data sets	Log sens index	Sensitivity	Specificity	CCR	ASE	ROC
0	Verify	0.205459	0.572222	0.963507	90.64777	0.316887	0.846156
1	Verify	0.282546	0.699999	0.855687	83.2996	0.468482	0.836182
2	Verify	0.285234	0.740278	0.783888	77.75302	0.502681	0.84035
3	**Verify**	**0.186805**	**0.545833**	**0.979618**	**91.63967**	**0.292761**	**0.864448**
4	**Verify**	**0.177034**	**0.5375**	**0.973219**	**90.97165**	**0.354652**	**0.823712**
5	**Verify**	**0.165262**	**0.519443**	**0.98341**	**91.57894**	**0.290568**	**0.862421**
6	Verify	0.200894	0.570835	0.957348	90.1012	0.334192	0.838134
7	Verify	0.313112	0.737499	0.833886	81.98382	0.45932	0.859387
8	Verify	0.257565	0.649999	0.91469	87.61134	0.370345	0.854156
9	Verify	0.276019	0.683334	0.878674	85.02025	0.396153	0.846551
10	Verify	0.25599	0.652777	0.904976	86.82186	0.420287	0.835027
11	Verify	0.27346	0.683334	0.872039	84.45344	0.463713	0.832523
12	Verify	0.264765	0.672223	0.874171	84.47369	0.395264	0.838388
13	Verify	0.202093	0.575001	0.951187	89.63563	0.351961	0.815549
14	Verify	0.265663	0.659722	0.908765	87.24698	0.389675	0.86325
15	Verify	0.246366	0.700001	0.785782	77.32794	0.606387	0.820241
16	**Verify**	**0.21665**	**0.588889**	**0.958296**	**90.44535**	**0.328582**	**0.834618**
17	Verify	0.222416	0.604169	0.928199	88.09718	0.396587	0.826799
18	Verify	0.272579	0.677778	0.882937	85.30365	0.402519	0.846772
19	Verify	0.28637	0.705555	0.853554	83.19838	0.498359	0.837892
20	**Verify**	**0.198478**	**0.563889**	**0.964454**	**90.60728**	**0.329285**	**0.824458**
21	Verify	0.317138	0.763888	0.796919	79.21052	0.609423	0.853366

The following table and figures provide comparison of the network performance of RWR models on Niday 2001 and Niday 2004 databases. For both APGAR5 and DELTYPE outcome estimations, Niday 2004 committee of classifiers models perform as good as Niday 2001 RWR models. Both Niday 2004 CC models correctly classified the output over 90% of the time.

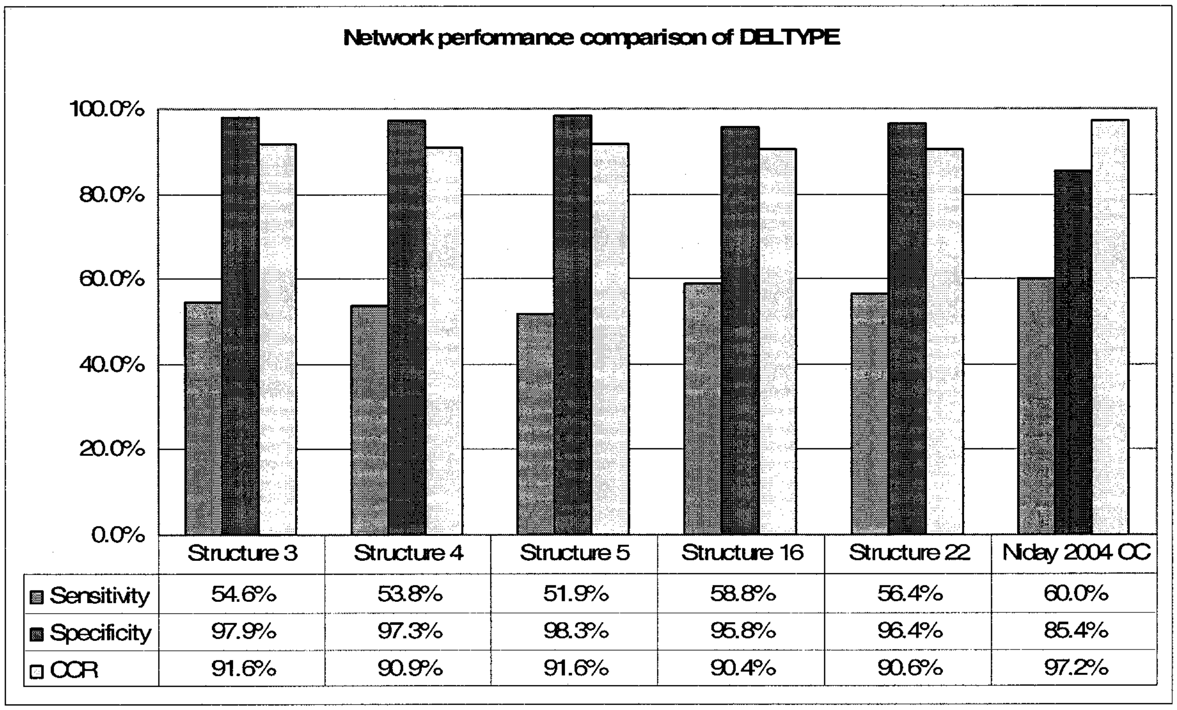
Table 4-21: Niday 2001 and Niday 2004 – best network performance.

Output	**APGAR5**		**DELTYPE**	
Database	**Niday 2001**	**Niday 2004**	**Niday 2001**	**Niday 2004**
Structure	18 hidden nodes	5 member CC	5 hidden nodes	5 member CC
Sensitivity	28.3%	35.5%	51.9%	60.0%
Specificity	98.2%	82.0%	98.3%	85.4%
CCR	96.6%	94.4%	91.6%	97.2%



	Structure 4	Structure 14	Structure 18	Structure 19	Structure 22	Nday 2004 CC
▨ Sensitivity	32.5%	28.3%	28.3%	50.8%	30.8%	35.5%
■ Specificity	97.1%	97.7%	98.2%	89.8%	97.9%	82.0%
☐ CCR	95.7%	96.1%	96.6%	88.9%	96.4%	94.4%

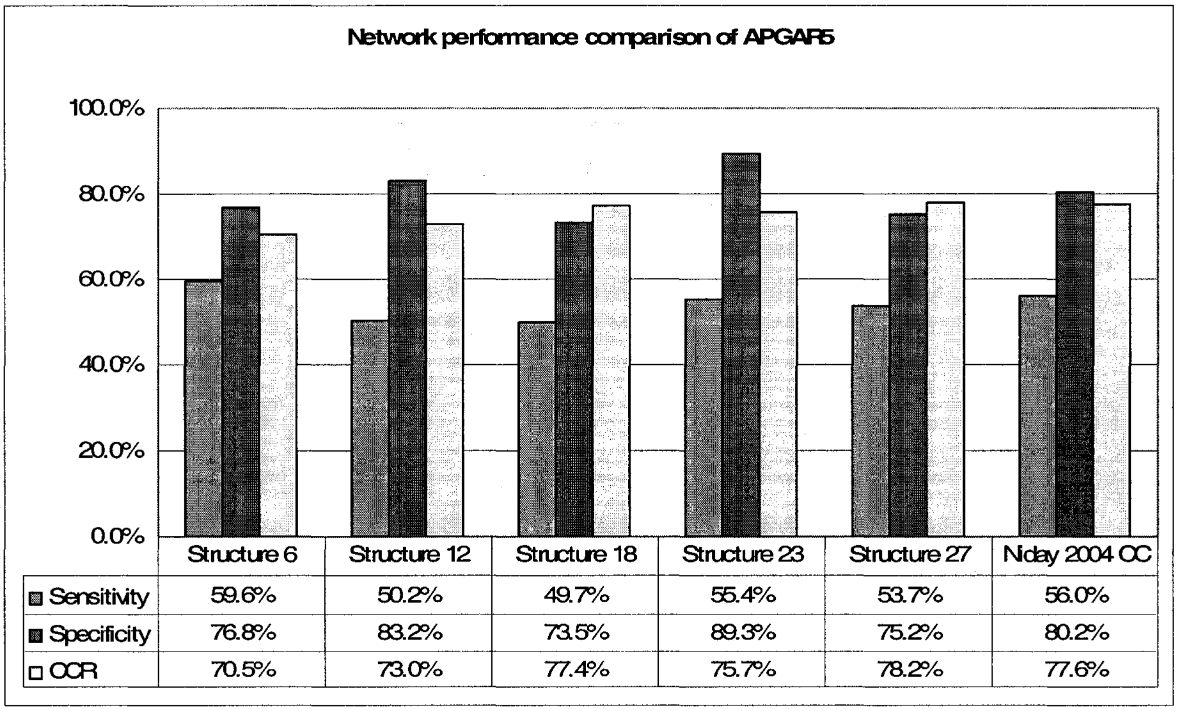
Figure 4-7: APGAR5 – 5 member committees of classifiers RWR performance measures.



	Structure 3	Structure 4	Structure 5	Structure 16	Structure 22	Nday 2004 CC
▨ Sensitivity	54.6%	53.8%	51.9%	58.8%	56.4%	60.0%
■ Specificity	97.9%	97.3%	98.3%	95.8%	96.4%	85.4%
☐ CCR	91.6%	90.9%	91.6%	90.4%	90.6%	97.2%

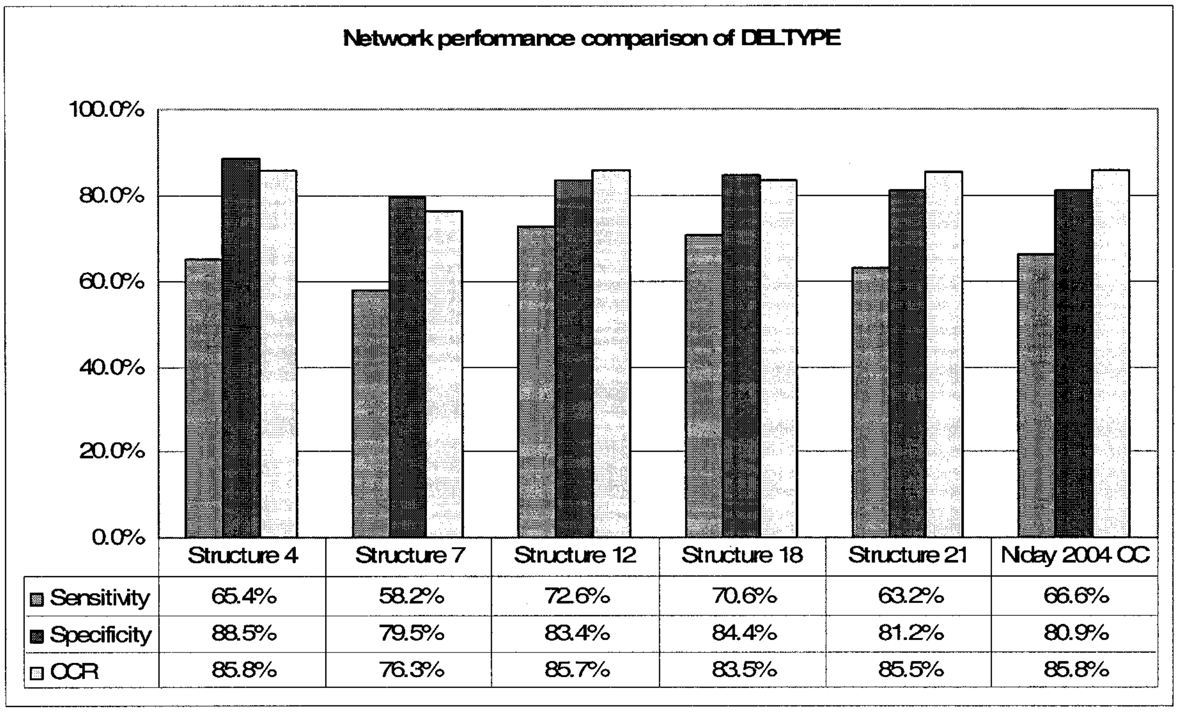
Figure 4-8: DELTYPE – 5 member committees of classifiers RWR performance measures.

The Niday 2001 APGAR5 PCA model (17,640 cases) was then re-run using 16 important components from 0 to 33 $(2n+1)$ hidden nodes, and the Niday 2001 DELTYPE PCA model (14,829 cases) was re-run using 10 important components from 0 to 21 $(2n+1)$ hidden nodes. This step was necessary for selecting the five best structures to construct a committee of classifiers to verify Niday 2004 data. For output APGAR5, structure 6, 12, 18, 23, and 27 are selected. For output DELTYPE, structure 4, 7, 12, 18, and 21 are selected. The following figures provide comparison of the network performance of PCA models on Niday 2001 and Niday 2004 databases. For both APGAR5 and DELTYPE outcome estimations, Niday 2004 committee of classifiers models perform as good as Niday 2001 PCA models.



	Structure 6	Structure 12	Structure 18	Structure 23	Structure 27	Niday 2004 CC
▨ Sensitivity	59.6%	50.2%	49.7%	55.4%	53.7%	56.0%
■ Specificity	76.8%	83.2%	73.5%	89.3%	75.2%	80.2%
□ OCR	70.5%	73.0%	77.4%	75.7%	78.2%	77.6%

Figure 4-9: APGAR5 – 5 member committees of classifiers PCA performance measures.



	Structure 4	Structure 7	Structure 12	Structure 18	Structure 21	Niday 2004 CC
▨ Sensitivity	65.4%	58.2%	72.6%	70.6%	63.2%	66.6%
■ Specificity	88.5%	79.5%	83.4%	84.4%	81.2%	80.9%
□ OCR	85.8%	76.3%	85.7%	83.5%	85.5%	85.8%

Figure 4-10: DELTYPE – 5 member committees of classifiers PCA performance measures.

The above results show that the committee of classifiers verification tool constructed from Niday 2001 models is able to classify Niday 2004 APGAR5 and DELTYPE outcomes with an acceptable percentage of accuracy. This means that a committee of selected well-trained artificial neural networks is capable of predicting the expected outcome from some completely unseen input data. The experiments done here applied a new database with similar data distribution but are 4 years apart to an ANN model developed from an old database. The knowledge of predicting the outcome is preserved by the ANN model and transferred in predicting the outcomes from new but similar information.

5 Conclusion

5.1 Concluding Remarks

The main objective of this thesis was to compare the performance of relative weight reduction (RWR) and principal component analysis (PCA) variable selection methods to discover the superior classification models for perinatal outcome Apgar score at 5 minutes and delivery type (vaginal or cesarean section). The ANN models developed in this thesis were compared with the models found in the literature and verified by the unseen data. The ANN model using the PCA approach increases the sensitivity of the network performance on both outcomes (APGAR5 and DELTYPE) by 10-15 %. The method was validated by applying it to unseen data (Niday 2004 database), which is another contribution of this research work.

It was found that the specificity performance of the RWR method applied ANN models is better than that of PCA method, while PCA performs better than RWR on sensitivity measurements. The PCA method is acknowledged to be useful in data transformation. This implies that the perinatal data has complicated non linear relationships, which requires using a one hidden layer non linear ANN structure after PCA pre-processing for classification. Both the PCA and RWR methods could reduce the numbers of input variables for ANN training. However, the RWR method had the advantage of retaining each individual important variable, so that it provides more descriptive information for input data, while the PCA method was able to reduce input information "noise". Since RWR and PCA methods each have their advantages and weaknesses in variable selection for developing ANN models, it is recommended that MIRG consider constructing a hybrid system of applying both methods with different preferred weights on the performance measures in future work.

The factors found by the back-propagation feedforward network with one hidden layer using weight elimination cost function applied relative weight reduction method are:

- APGAR5: baby's weight, maternal pain reliever "pudendal", number of previous term babies, labour type, gestational age, delivered by, maternal pain reliever "general", number of previous preterm babies, fetal surveillance, whether scalp blood gases were done, number of babies in this pregnancy, and maternal pain reliever "spinal".
- DELTYPE: maternal pain reliever "general", maternal pain reliever "spinal", number of previous term babies, maternal pain reliever "pudendal", maternal pain reliever "nitrous oxide", age of the mother, presentation of the baby, baby's weight, gestational age, and number of previous preterm babies.

The development of the ANN model involved finding optimal network performance. Although the process of running network structures in finding optimal network learning parameters were automated by the ANN RFW tool, the evaluation of performance measures were still time consuming and labor intensive. The optimal network structures used in CCVT tool and performance for each outcome are:

- APGAR5: Five members of committees were constructed by 4, 14, 18, 19 and 22 hidden nodes of Niday 2001 RWR model. This produces a sensitivity of 35.5%, specificity of 82.0% and correct classification rate of 94.4% on the Niday 2004 database. Five members of committees were constructed by 6, 12, 18, 23 and 27 hidden nodes of Niday 2001 PCA model. This produces a sensitivity of 56.0%, specificity of 80.2% and correct classification rate of 77.6% on the Niday 2004 database.

- DELTYPE: Five members of committees were constructed by 3, 4, 5, 16 and 20 hidden nodes of Niday 2001 RWR model. This produces a sensitivity of 60.0%, specificity of 85.4% and correct classification rate of 97.2% on the Niday 2004 database. Five members of committees were constructed by 4, 7, 12, 18 and 21 hidden nodes of Niday 2001 PCA model. This produces a sensitivity of 66.6%, specificity of 80.9% and correct classification rate of 85.8% on the Niday 2004 database.

5.2 Contributions

The summary of contributions to knowledge made by this thesis is outlined:

- Demonstrated that RWR method gives high specificity ANN network performance while PCA method gives high sensitivity measures.
- Showed that the RWR method can be used to select the important network factors more effectively than the PCA method for Niday Perinatal Database.
- Developed ANN models for APGAR5 and DELTYPE perinatal outcomes, and identified important factors for outcome classification from Niday Perinatal Database.
- Confirmed that the automated ANN RFW tool is suitable for ANN model development, and hidden layer performance is better for complex non linear medical data analysis.
- Verified ANN models that are developed from a previous database are appropriate on unseen data for perinatal outcome classification using committee of classifiers verification.

- Validated that the CCVT tool is a good approach for ANN model verification, as the comparison results show that committee of classifiers performs well for complex medical data network structure constructions.

5.3 Future Work

The decrease in child mortality and improvement in maternal health was identified by the United Nations in the year 2000 as the fourth and the fifth millennium development goals to be achieved by the year 2015. In the long term, it is hoped that the current work can provide useful tools for developing countries in achieving these goals. The Niday Perinatal Database contains a huge amount of perinatal care information which is collected every year by PPPESO. Using this wealth of information, numerous extensive works can be carried out to continue and improve this research, for example:

- Explore application of PCA method to other MIRG databases to examine if it improves sensitivity measurements.
- Discover other viable variable selection method such as independent component analysis (ICA) applied to ANN model developments and contrast the network performance with the ones found in this work.
- Construct a hybrid system applied various variable selection methods for achieving the overall high network performance.
- Strive to identify the important factors in classifying the multitude of perinatal outcomes using the Niday Perinatal Database.
- Consolidate the perinatal data with obstetrical and neonatal data to formulate additional comprehensive outcome analysis throughout the maternal and newborn period.

- Apply probability studies on Apgar score outcome estimations such that the outcome can be classified to the closest representation score that accurately associate with the actual measuring point.
- Develop automated tools to pre-process the raw data from the database to interface with the ANN tools. This will eliminate the rigorous data preparation for ANN analysis.
- Integrate the database, the database pre-process tools and the ANN tools into a user friendly clinical decision support system. This will provide real-time medical outcome estimations for on-site clinical trial.

References

[1] J.A. Anderson, *An Introduction to Neural Networks*, Cambridge: MIT Press, 1995.

[2] V. Apgar, "A proposal for a new method of evaluation of the newborn infant," in *Current Research in Anesthesia and Analgesia*, vol. 32, 1953, pp. 260-267.

[3] A. Bacci, and G.P. Chiaffoni, "Trends in perinatal care in developing countries," in *Seminars in Fetal and Neonatal Medicine*, vol. 11, no. 1, 2006, pp. 1-2.

[4] V. Beiu, "A novel highly reliable low-power nano architecture when von Neumann augments Kologorov," in *Proc. 15th IEEE Int. Conf. on Application-Specific Systems, Architectures and Processors*, 2004, pp. 167-177.

[5] R. Bowles, "The structure of neural networks," in *An Idiot's Guide to Neural Networks*, accessed 19 June 2006, available from <http://richardbowles.tripod.com/neural/neural.htm>.

[6] P.J. Braspenning, "Neural Cognodynamics," in *Lecture Notes of Computer Science: Artificial Neural Networks*, Heidelberg: Springer-Verlag, 1991.

[7] T. Buskard, M. Stevenson, M. Frize, and F.G. Solven, "Estimation of ventilation, length of stay and mortality using artificial neural networks," in *Proc. 1994 Canadian Conf. on Electrical and Computer Engineering*, vol. 2, 1994, pp. 726-729.

[8] C. Catley, and M. Frize, "A prototype XML-based implementation of an integrated 'intelligent' neonatal intensive care unit," in *Proc. 4th Int. IEEE EMBS Special Topic Conf. on Information Technology Applications in Biomedicine*, 2003, pp. 322-325.

[9] C. Catley, M. Frize, D.C. Petriu, C.R. Walker, and L. Yang, "Towards a web services infrastructure for perinatal, obstetrical, and neonatal clinical decision support," in *Proc. 26th IEEE EMBS Annual Int. Conf.*, vol. 2, 2004, pp. 3334-3337.

[10] C. Catley, M. Frize, R.C. Walker, and D.C. Petriu, "Predicting preterm birth using artificial neural networks," in *Proc. 18th IEEE Symp. on Computer-Based Medical Systems*, 2005, pp. 103-108.

[11] C.G. Courtright, R.S. Crawford, and D.M. Klubert, "Criteria for developing clinical decision support systems," in *Proc. 14th IEEE Symp. on Computer-Based Medical Systems*, 2001, pp. 270-275.

[12] J. Dayhoff, *Neural Networks Architectures: An Introduction*, New York: Van Nostrand Reinhold, 1990.

[13] H. Demuth, and M. Beale, *Neural Network Toolbox User's Guide*, version 3.0, Natick: The MathWorks Inc., 1997.

[14] R.O. Duda, P.E. Hart, and D.G. Stork, *Pattern Classification*, 2nd ed., New York: Jonh Wiley & Sons, 2001.

[15] C.M. Ennett, *Imputation of Missing Values by Integrating Artificial Neural Networks and Case-Based Reasoning*, Ph.D. thesis, Dept. of Systems and Computer Engineering, Carleton University, Ottawa, Ontario, 2003.

[16] C.M. Ennett, and M. Frize, "Selective sampling to overcome skewed *a priori* probabilities with neural networks," in *Proc. American Medical Informatics Association Symposium*, 2000, pp. 225-229.

[17] C.M. Ennett, and M. Frize, "Validation of a hybrid approach for imputing missing data," in *Proc. 25th IEEE EMBS Annual Int. Conf.*, vol. 2, 2003, pp. 1268-1271.

[18] C.M. Ennett, and M. Frize, "Weight-elimination neural networks applied to coronary surgery mortality prediction," in *IEEE Trans. on Information Technologies in Biomedicine*, vol. 7, no. 2, 2003, pp. 86-92.

[19] C.M. Ennett, M. Frize, and E. Charette, "Improvement and automation of artificial neural networks to estimate medical outcomes," in *Medical Engineering and Physics*, no. 26, 2004, pp. 321-328.

[20] C.M. Ennett, M. Frize, and N. Scales, "Evaluation of the logarithmic-sensitivity index as a neural network stopping criterion for rare outcomes," in *Proc. 4th Int. IEEE EMBS Special Topic Conf. on Information Technology Applications in Biomedicine*, 2003, pp. 207-210.

[21] Y. Erdebil, *Modeling Severe ATV Injuries Using Artificial Neural Networks*, M.A.Sc. Elec. Eng. thesis, School of Information Technology and Engineering, University of Ottawa, Ottawa, Ontario, 2005.

[22] L. Fausett, *Fundamentals of Neural Networks: Architectures, Algorithms, and Applications*, New Jersey: Prentice Hall, 1994.

[23] T. Fawcett, "ROC graphs: notes and practical considerations for data mining researches," *Intelligent Enterprise Technologies Laboratory*, 2003.

[24] M. Frize, C.M. Ennett, and E. Charette, "Automated optimization of neural networks in estimating medical outcomes," in *Proc. IEEE EMBS Int. Conf. on Information Technology Applications in Biomedicine*, 2000, pp. 168-173.

[25] M. Frize, C.M. Ennett, M. Stevenson, H.C.E. Trigg, "Clinical decision support systems for intensive care units: using artificial neural networks," in *Medical Engineering and Physics*, vol. 23, 2001, pp. 217-225.

[26] M. Frize, D. Ibrahim, H. Seker, R. Walker, M.O. Odetayo, D. Petrovic, R.N.G. Naguib, "Predicting clinical outcomes for newborns using two artificial intelligence approaches," in *Proc. 26th IEEE EMBS Annual Int. Conf.*, vol. 2, 2004, pp. 3202-3205.

[27] M. Frize, F.G. Solven, M. Stevenson, B.G. Nickerson, T. Buskard, and K. Taylor, "Computer-assisted decision support systems for patient management in an intensive care unit," in *Medinfo*, 8 Pt 2, 1995, pp. 1009-1012.

[28] M. Frize, R.C. Walker, and C.M. Ennett, "Development of an evidence-based ethical decision-making tool for neonatal intensive care medicine," in *Proc. 25th IEEE EMBS Annual Int. Conf.*, vol. 2, 2003, pp. 1260-1263.

[29] M. Frize, L. Yang, R.C. Walker, and A.M. O'Connor, "Conceptual framework of knowledge management for ethical decision-making support in neonatal intensive care," in *IEEE Trans. on Information Technology in Biomedicine*, vol. 9, no. 2, 2005, pp. 205-215.

[30] A.X. Garg, N.K.J. Adhikari, H. McDonald, M.P. Rosas-Arellano, P.J. Devereaux, J. Beyene, J. Sam, and R.B. Haynew, "Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review," in *Journal of the American Medical Association*, vol. 293, no. 10, 2005, pp. 1223-1238.

[31] G.D. Garson, "Interpreting neural-network connection weights," in *Artificial Intelligence Expert*, vol. 6, 1991, pp. 47-51.

[32] A.T.C. Goh, "Back-propagation neural networks for modeling complex systems," in *Artificial Intelligence in Engineering*, vol. 9, pp. 143-151.

[33] S. Haykin, *Neural Networks: A Comprehensive Foundation*, 2nd ed., New York: Prentice Hall, 1999.

[34] P.S. Heckerling, B.S. Gerber, T.G. Tape, and R.S. Wigton, "Entering the black box of neural networks: a descriptive study of clinical variables predicting community-acquired pneumonia," in *Methods of Information in Medicine*, vol. 42, 2003, pp. 287-296.

[35] N.J. Higham, "The accuracy of floating point summation," in *SIAM J. Scientific Computing*, vol. 14, no. 4, 1993, pp. 783-799.

[36] Y. Hu, J. Park, and T. Knoblock, "Committee pattern classifiers," in *1997 IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, vol. 4, 1997, pp. 3389.

[37] P.T.W. Hudson, and E.O. Postma, "Choosing and using a neural net," in *Lecture Notes of Computer Science: Artificial Neural Networks*, Heidelberg: Springer-Verlag, 1991.

[38] D.L. Hudson, and M.E. Cohen, "Using the internet to assist clinical decision making," in *Proc. 2000 IEEE EMBS Int. Conf. on Information Technology Applications in Biomedicine*, 2000, pp. 117-122.

[39] M.S. Hung, M.Y. Hu, M.S. Shanker, and B.E. Patuwo, "Estimating posterior probabilities in classification problems with neural networks," in *International Journal of Computational Intelligence and Organizations*, vol. 1, no. 1, 1996, pp. 49-60.

[40] D. Ibrahim-Swailum, *Identifying Risk Factors for Newborn Outcomes Using Artificial Neural Networks*, M.A.Sc. Elec. Eng. thesis, School of Information Technology and Engineering, University of Ottawa, Ottawa, Ontario, 2005.

[41] D. Itchhaporia, P.B. Snow, R.J. Almassy, and W.J. Oetgen, "Artificial neural networks: current status in cardiovascular medicine," in *Journals of American College of Cardiology*, vol. 28, no. 2, 1996, pp. 515-521.

[42] J.P.S. Koay, *Quantitative Analysis of Infrared Images for Early Breast Cancer Detection*, M.A.Sc. thesis, Dept. of Systems and Computer Engineering, Carleton University, Ottawa, Ontario, 2004.

[43] R. Kohavi, and F. Provost, "Glossary of terms: special issue on applications of machine learning and the knowledge discovery process," in *Machine Learning*, vol. 30, 1998, pp. 271-274.

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

[45] D.J. Livingstone, D.T. Manallack, and I.V. Tetko, "Data modeling with neural networks: advantages and limitations," in *Journals of Computer-Aided Molecular Design*, vol. 11, no. 2, 1997, pp. 135-142.

[46] M. MacDowell, E. Somoza, K. Rothe, R. Fry, K. Brady, and A. Bocklet, "Understanding birthing mode decision making using artificial neural networks," in *Medical Decision Making*, vol. 21, no. 6, 2001, pp. 433-443.

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

[48] L. Medsker, and J. Liebowitz, *Design and Development of Expert Systems and Neural Networks*, New York: Macmillan, 1994.

[49] S. Mor-Yosef, A. Samueloff, B. Modan, D. Navot, and J.G. Schenker, "Ranking the risk factors for cesarean: logistic regression analysis of a nationwide study," in *Obstetrics & Gynecology*, vol. 75, no. 6, 1990, pp. 944-947.

[50] J.D. Olden, and D.A. Jackson, "Illuminating the 'black box': a randomization approach for understanding variable contributions in artificial neural networks," in *Ecological Modelling*, vol. 154, 2002, pp. 135-150.

[51] A. Pan, "MIRG ANN Guide," NSERC Summer Student Report, Medical Information-technology Research Group, University of Ottawa, Ottawa, Ontario, 2002.

[52] W. Penny, and D. Frost, "Neural networks in clinical medicine," in *Medical Decision Making*, vol. 16, 1996, pp. 386-398.

[53] S. Piramuthu, M. Shaw, and J. Gentry, "A classification approach using multi-layer neural networks," in *Decision Support Systems*, vol. 11, no. 5, 1994, pp. 509-525.

[54] *PPPESO – Perinatal Partnership Program of Eastern and Southeastern Ontario*, accessed 2 March 2006, available from <http://www.pppeso.on.ca>.

[55] Perinatal Partnership Program of Eastern and Southeastern Ontario, *Annual Perinatal Statistical Report 2001*, 2002.

[56] Perinatal Partnership Program of Eastern and Southeastern Ontario, *Annual Perinatal Statistical Report 2004*, 2005.

[57] Perinatal Partnership Program of Eastern and Southeastern Ontario, *Perinatal Services in Ontario: How Are We Doing?* 2005.

[58] D.K. Richardson, J.E. Gray, M.C. McCormick, K. Workman, and D.A. Goldmann, "Score for neonatal acute physiology: a physiologic severity index for neonatal intensive care," in *Pediatrics*, vol. 91, no. 3, 1993, pp. 617-623.

[59] D.K. Richardson, J.D. Corcoran, G.J. Escobar, and S.K. Lee, "SNAP-II and SNAPPE-II: Simplified newborn illness severity and mortality risk scores," in *Journal of Pediatrics*, vol. 138, no. 1, 2001, pp. 92-100.

[60] D. Rybchynski, *Design of an Artificial Neural Network Research Framework to Enhance the Development of Clinical Prediction Models*, M.A.Sc. Elec. Eng. thesis, School of Information Technology and Engineering, University of Ottawa, Ottawa, Ontario, 2005.

[61] Y. Shi, *Development of a Model for Prediction of Repeat Injuries in Injured Canadian Using Artificial Neural Networks*, M.ISS thesis, Dept. of Systems and Computer Engineering, Carleton University, Ottawa, Ontario, 2004.

[62] J. Shlens, *A Tutorial on Principal Component Analysis*, version 2, La Jolla: University of California San Diego, 2005.

[63] B.J. Taylor, and M.A. Darrah, "Verification and validation of neural networks: a sampling of research in progress," in *Proceedings of the SPIE*, vol. 5103, 2003, pp. 8-16.

[64] T.G. Tape, "The area under an ROC curve," in *Interpreting Diagnostic Tests*, accessed 10 July 2006, available from <http://gim.unmc.edu/dxtests/Default.htm>.

[65] Y. Tong, *Developing ANN Approaches to Estimate Neonatal ICU Outcomes*, M.A.Sc. Elec. Eng. thesis, School of Information Technology and Engineering, University of Ottawa, Ottawa, Ontario, 2000.

[66] Y. Tong, M. Frize, and R. Walker, "Extending ventilation duration estimates approach from adult to neonatal intensive care patients using artificial neural networks," in *IEEE Trans. on Information Technology in Biomedicine*, vol. 6, no. 2, 2002, pp. 188-191.

[67] H.C.E. Trigg, *An Investigation of Methods to Enhance the Performance of Artificial Neural Networks Used to Estimate ICU Outcomes*, M. Eng. thesis, Dept. of Electrical Engineering, University of New Brunswick, Fredericton, New Brunswick, 1997.

[68] R. Walker, Personal communication of discussions via e-mails and meetings, Ottawa, Ontario, 2005.

[69] S. Walczak, and N. Cerpa, "Heuristic principles for the design of artificial neural networks," in *Information and Software Technology*, vol. 41, no. 2, 1999, pp. 107-117.

[70] A.R. Webb, *Statistical Pattern Recognition*, New Jersey: Wiley, 2002.

[71] A.S. Weigend, D.E. Rumelhart, and B.A. Huberman, "Back-propagation, weight-elimination and time series prediction," in *Proc. 1990 Connectionist Models Summer School*, 1990, pp. 105-116.

[72] A.S. Weigend, D.E. Rumelhart, and B.A. Huberman, "Generalization by weight-elimination with application to forecasting," in *Advances in Neural Information Processing Systems*, vol. 3, 1991, pp. 875-882.

[73] E.W. Weisstein, "Hyperbolic tangent," in *MathWorld – A Wolfram Web Resource*, accessed 7 October 2005, available from <http://mathworld.wolfram.com/HyperbolicTangent.html>.

[74] K. Yale, "Preparing the right data diet for training neural networks," *in IEEE Spectrum*, March 1997, pp. 64-66.

[75] L. Yang, *Pilot Usability Study of UI Prototype for Collaborative Adaptative Decision Support in Neonatal Intensive Care Unit*, M.A.Sc. Elec. Eng. thesis, School of Information Technology and Engineering, University of Ottawa, Ottawa, Ontario, 2005.

[76] S. Yoshimoto, "A study on artificial neural network generalization capability," in *Proc. International Joint Conference on Neural Networks*, vol. 3, 1990, pp. 689-694.

[77] J. Zhang, Y. Yan, and M. Lades, "Face recognition: eigenface, elastic matching, and neural nets," in *Proc. of the IEEE*, vol. 85, no. 9, 1997, pp. 1423-1435.

Appendices

Appendix A: Apgar Score

The Apgar[1] score was devised in 1952 by Dr. Virginia Apgar as a simple and repeatable method to quickly and summarily assess the health of newborn children immediately after childbirth. The Apgar score is determined by evaluating the newborn baby on five simple criteria on a scale from zero to two and summing up the five values thus obtained. The resulting Apgar score ranges from 0 to 10.

The five criteria of the Apgar score, as listed in the following table:

Table A-1: Apgar score criteria.

Acronym	Measurement	Score 0	Score 1	Score 2
Appearance	Skin color	Blue all over	Blue at extremities	Normal
Pulse	Heart rate	absent	<100	>100
Grimace	Reflex irritability	No response to stimulation	Grimace/feeble cry when stimulated	Sneeze/cough/pull away when stimulated
Activity	Muscle tone	None	Some flexion	Active Movement
Respiration	Respiration	Absent	Weak or irregular	Strong

The test is generally done at one and five minutes after birth, and may be repeated later if the score is, and remains, low. Scores below 3 are generally regarded as critically low, with 4 to 7 fairly low and over 7 generally normal. Low scores at the one minute test may require medical attention, but are not an indication of longer term problems, particularly if there is an improvement by the stage of the five minute test. If the Apgar score remains below 3 at later times such as 10, 15, or 30 minutes, there is a risk that the child will suffer longer term neurological damage. There is also a small but significant increase in the risk of cerebral palsy.

---

[1] From Wikkipedia - the free encyclopedia, <http://en.wikipedia.org/wiki/Apgar>, accessed: 27 February 2006.

However, the purpose of the Apgar test is to determine quickly whether a newborn needs immediate medical care; it was not designed to make long-term predictions on a child's health.

Some ten years after the initial publication, the acronym APGAR was coined in the US as a mnemonic learning aid: Appearance (skin color), Pulse (heart rate), Grimace (reflex irritability), Activity (muscle tone), and Respiration.

Appendix B: Niday Perinatal Database User Guide

(Note: This list[2] includes only those variables entered into the Enhanced Database Computer Entry Program).

Data should be collected on each birth (not pregnancy). Therefore, if a multiple gestation, then each baby is entered separately.

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

---

[2] From PPPESO - Perinatal Partnership Program of Eastern and Southeastern Ontario, <http://www.pppeso.on.ca>, accessed 2 March 2006.

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 C: ANN RFW Design Diagram

A high level flow chart of the operation of the ANN research framework[3] (RFW) is shown in the Figure C-2. A legend for the flow chart symbols is given in Figure C-1.

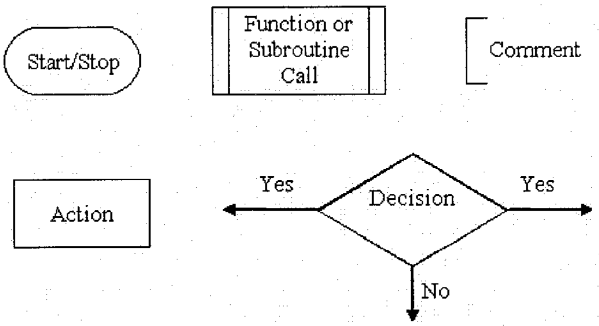


Figure C-1: Legend for flow chart symbols.

The main_autoANN_Sequential.m is the core MATLAB program of the MIRG ANN tool, which contains and calls the sequential functions during the ANN experiments.

---

[3] D. Rybchynski, *Design of an Artificial Neural Network Research Framework to Enhance the Development of Clinical Prediction Models*, M.A.Sc. Elec. Eng. thesis, School of Information Technology and Engineering, University of Ottawa, Ottawa, Ontario, 2005.

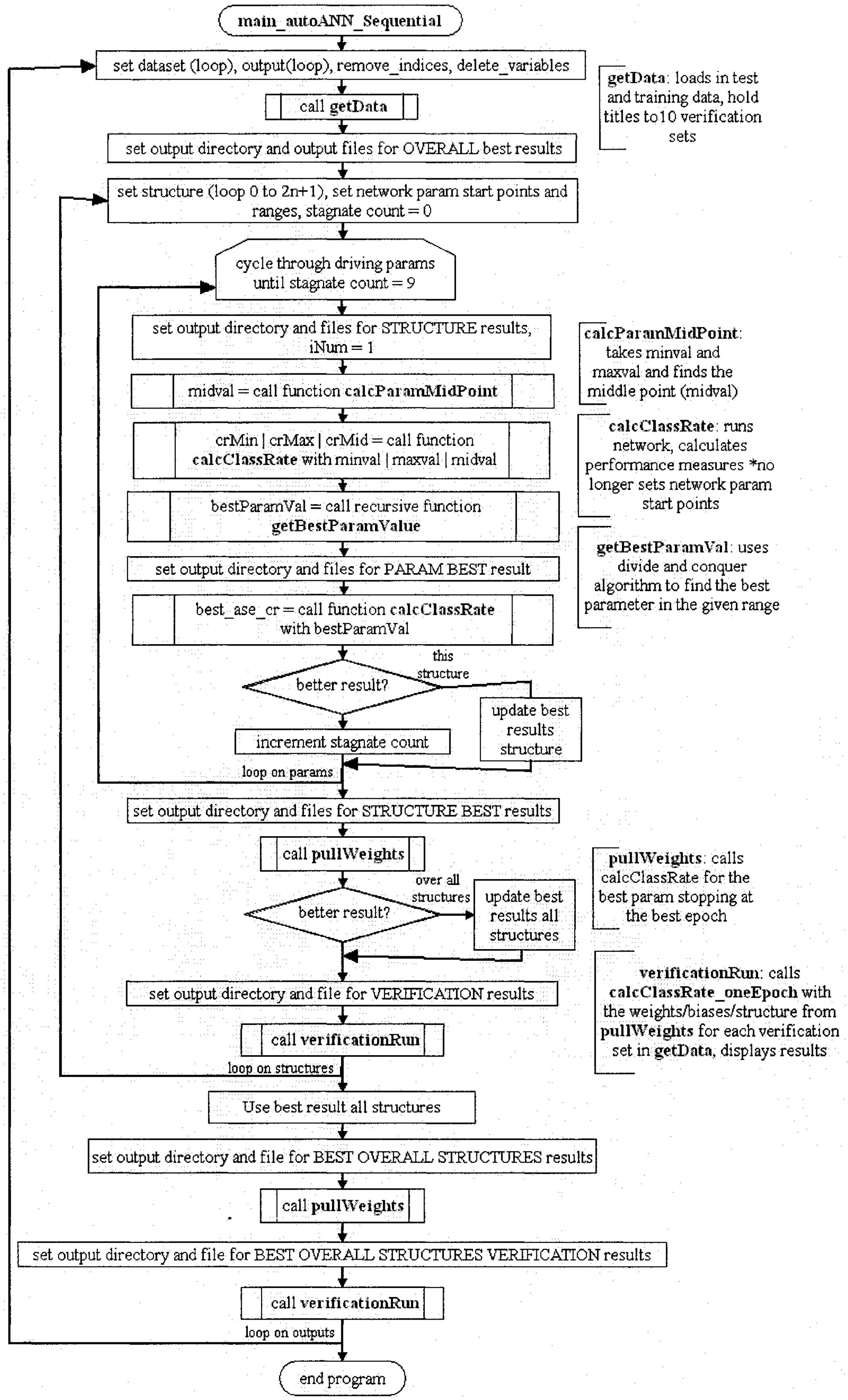


Figure C-2: Flow chart of ANN RFW – main_autoANN_Sequential.m.