A Web-Based Perinatal Decision Support System Framework Using a Knowledge-Based-Approach to Estimate Clinical Outcomes: Neonatal Mortality and Preterm Birth in Twins Pregnancies

by

Marry Gunaratnam, B.Eng (Biomedical and Electrical Engineering)

A thesis submitted to the Faculty of Graduate and Postdoctoral Affairs

In partial fulfillment of the requirements for the degree of
Master of Applied Science in Biomedical Engineering

Ottawa – Carleton Institute for Biomedical Engineering (OCIBME)

Department of Systems and Computer Engineering

Carleton University

Ottawa, Ontario, Canada, K1S 5B6

April 2013

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: 978-0-494-94672-5

Our file Notre référence
ISBN: 978-0-494-94672-5

Abstract

There are two main contributions to knowledge presented in this thesis: (1) an improved method for predicting neonatal mortality and preterm birth in twin pregnancies, and (2) a framework to build a web-based perinatal decision support system (PEDSS) using a knowledge based approach.

Earlier identification of clinical outcomes may lead to more efficient allocation of resources. Thus, two novel prediction models using Decision Trees(DT) and Hybrid Artificial Neural Network(ANN) were evaluated. The DT prediction model had the highest performance outcome for predicting neonatal mortality (sensitivity=62.24%, specificity=99.95%, Positive Predictive Value (PPV)=72.34%, Negative Predictive Value (NPV)=99.92%) using information available within 10 minutes after birth, and preterm birth in twin pregnancies (sensitivity=79.32%, specificity=91.97%, PPV=66.85%, NPV=95.66%) before 22 weeks gestation.

The PEDSS includes three main components: the knowledge-base, a workflow engine, and a mechanism to communicate results. This tool provides prediction results within seconds and assists clinicians in the decision making process.

Acknowledgement

First and foremost, I take this opportunity to express my profound gratitude and deep regards to my thesis supervisor Dr. Monique Frize: your exemplary guidance, monitoring and encouragement are inspiring and I am very grateful to have had the opportunity to work with you. All of the help and guidance given by you from time to time shall carry me a long way in the journey of life on which I am about to embark.

A heartfelt thank you to Dr. Erika Bariciak at the Children's Hospital of Eastern Ontario (CHEO) for providing useful critiques, improvements for this research work and providing necessary information for this thesis work.

I would like to thank the current and past members of the Medical Information-Technologies Research Group, for their invaluable help throughout. I wish to express my sincere thanks to Jeff Gilchrist who always made himself available to answer my questions. Your support was instrumental in the successful completion of this thesis work

I would like to thank my fiancé Kavaskar Nadarajah for his continued support and encouragement over the course of my undergraduate and graduate studies. Further, my deepest thanks to my sister Ann Gunaratnam, whose support has always been my source of strength and inspiration.

Last but not least, I would like to express my gratitude towards my two loving parents, Anthony and Any Gunaratnam: You both gave me choices, encouraged me to persevere, think positive thoughts and taught me to never give up on my dreams. For this, I am forever appreciative.

Table of Contents

List of Tables

List of Illustrations

List of Acronyms

ANN	Artificial Neural Network
ANN RFW	Artificial Neural Network Research Framework
AGPAR	Activity, Grimace, Pulse, Appearance, and Respiration
APACHE	Acute Physiology and Chronic Health Evaluation
AUC	Area Under the Curve
CBR	Case Based Reasoning
CDSS	Clinical Decision Support System
CMS	Content Management System
CL	Cervical Length
CRIB	Clinical Risk Index for Babies
DT	Decision Tree
ECMS	Enterprise Content Management System
EMR	Electronic Medical Record
fFN	Fetal Fibronectin
FN	False Negative
FNR	False Negative Rate
FP	False Positive
FPR	False Positive Rate
GUI	Graphical User Interface
ICU	Intensive Care Unit
ID3	Iteractive Dichotomiser
JDC	Java Database Connectivity
LOS	Length of Stay in ICU
k-NN	k-Nearest Neighbor
MIRG	Medical Information-technology Research Group
MLP	Multilayer Perceptron
NICU	Neonatal Intensive Care Unit
NPV	Negative Predictive Value
ODBC	Open Database Connectivity
PBNN	Pruning Based Neural Network
PPPSEO	Perinatal Partnership Program of Eastern and Southern Ontario
PPV	Positive Predictive Value
PRAMS	Pregnancy Risk Assessment Monitoring System
PTB	Preterm Birth
ROC	Receiver Operating Curve
SNAP	Score for Neonatal Acute Physiology

SQL	Structured Query Language
TBNN	Tree Based Neural Network
TN	True Negative
TNR	True Negative Rate
TP	True Positive
TPR	True Positive Rate
URL	Uniform Resource Locator

CHAPTER 1 INTRODUCTION

1.1 MOTIVATION

1.1.1 HEALTHCARE PERSPECTIVE

Healthcare in Canada has long been depicted as a source of national pride, yet clear opportunities exist to enhance patient outcomes and thus reduce the demands on our healthcare system. In particular, preventative interventions are considered to be an important element of a modern healthcare system. Many studies have demonstrated that preventative interventions during pregnancy have led to lower mortality and morbidity rates among newborns (Lim et al., 2009).

Earlier identification of newborns at risk of neonatal mortality or women at risk of preterm[1] labour in a high risk population, would allow healthcare providers to be more efficient in the allocation of their resources and provide timely response to varying medical problems. Early diagnosis is required for effective treatment. Moving forward, leveraging the advances in technology and linking it with medical knowledge using advanced mathematical concepts would allow for the creation of sophisticated tools to generate more accurate predictions of complex outcomes in the medical field.

1.1.2 TECHNOLOGY PERSPECTIVE

The healthcare domain remains a paper-intensive and non-automated industry. A large amount of money is spent on administrative cost in healthcare, and in many cases, healthcare providers do not have a collaborative platform to share and exchange information. A

[1] Preterm: Birth of a baby of less than 37 weeks gestational age

collaborative platform would allow healthcare providers access to various clinical applications including prediction and assessment tools in order to determine appropriate care protocols and/or medication administration.

A web-based collaborative platform in a clinical setting would allow healthcare providers to collaborate in the form of sharing opinions and exchanging clinical data regardless of their geographic location (Chronaki, 1997). Further, the development of a high quality, low cost perinatal decision support system (PEDSS) to aid human decisions on a collaborative platform would improve the quality of care provided and reduce human errors.

1.2 PROBLEM STATEMENT

A number of clinical problems exist specific to perinatal care. Early and accurate diagnosis of these problems can initiate care early on and improve patient outcome. The goal of any prediction model is to successfully predict an outcome of choice while meeting clinical expectations. This thesis work includes prediction models for the identification of two distinct perinatal problems: neonatal mortality and preterm labour in twin pregnancies. Most published risk estimation models attempt to meet clinically acceptable sensitivity and specificity, however in most cases successful identification of positive cases has been difficult. Further, irrelevant variables not only add noise and complexity to the problem, but it reduces the likelihood of identifying positive cases. The primary objective of this work was to incorporate the advantages of decision trees (DTs) to create a system that can predict the two perinatal problems above at an earlier stage while maintaining high sensitivity, specificity, positive predictive and negative predictive values.

Once the fundamental objective was met, a framework for a perinatal decision support system (PEDSS) which integrated two risk assessment tools to predict the two perinatal problems identified earlier was developed. The PEDSS contains repositories for storing patient information and allows easy export of data for data mining and analysis. The combination of a decision support system with integrated assessment tool(s) to predict the risk of medical outcome(s), would initiate suitable care protocols early on and ensure effective treatment and thus improve patient care.

1.3 NEED FOR PREDICTION MODELS IN PERINATAL CARE

In order to mitigate complications associated with neonatal mortality or preterm birth in twin pregnancies, it is important to recognize important risk factors and initiate suitable care protocols in a timely manner to ensure effective treatment (Allen et al., 2002). In terms of neonatal mortality, correctly identifying a survivor from a non-survivor could lead to earlier development and execution of a targeted treatment plan which may be in the form of more medical attention allocated to the newborn, or a more aggressive intervention (i.e assign palliative care instead of restorative care for terminally ill[2] newborns) (Yu, 2009). Although it is a difficult topic to acknowledge, due to the high cost of resources required to keep a neonate within the ICU, the cost versus benefits should be analyzed, and thus the healthcare providers will require a prediction system that is clinically acceptable to aid their decisions.

Further, correctly non-invasively identifying a high risk population (i.e. twin pregnancies) at risk of preterm birth early on would prevent the need for costly and invasive screening tests (i.e. cervical length (CL) and fetal fibronectin (fFN)). This may also avoid additional stress to the

[2] Terminally ill: a disease or a condition that cannot be cured or adequately treated and that is reasonably expected to result in the death of the patient within a short period of time

mothers, which is commonly encountered with invasive testing. The physicians may also initiate interventions, treatment and allocate necessary resources early on in an attempt to prevent preterm birth. Ultimately, a prediction model that correctly identifies the risk of preterm birth in twin pregnancies early on could lead to a decrease in the overall preterm birth rates.

1.3.1 INFANT MORTALITY RATES IN CANADA

Infant mortality is an important indicator of the child's health and the country's well-being over time (The Conference Board of Canada, 2001). Infant mortality may be categorized into early neonatal mortality (death of newborn 0-7 days), late neonatal mortality (death of newborn 8-27 days) and post neonatal mortality (death of infant 28-365 days) (The Daily Motion, 2008). Between 2004 and 2005, neonatal and post-neonatal mortality rates followed different paths, where neonatal mortality rate rose from 4.0 to 4.1 deaths per 1000 live birth, and no change was observed in the post-neonatal rate, it remained at 1.3 deaths per 1000 live birth (The Daily Motion, 2008). Although Canada's overall infant mortality rate fell significantly since the 1960's, it is still higher compared to its peer countries of similar socio-economic status including Sweden, Japan, Finland and France. Canada is now tied in 2nd place with United Kingdom with the highest infant mortality rate. United States has the leading highest infant mortality rate of the developed countries (The Conference Board of Canada, 2001). For more information on infant mortality, see Appendix A.

1.3.2 PRETERM MORTALITY RATES IN CANADA

Preterm birth is an important perinatal problem in Ontario, where about 8% of babies (1 in 12) are born before term (Allen et al., 2002). In 2006-2007, 1 in 7 babies born in Canada were considered to be either preterm or too small for their gestational age (Lim et al., 2009). Preterm

birth results in a disproportionately high percentage of healthcare costs among newborns (Lim et al., 2009). The preterm birth rate in Canada has increased over the years (as with many other economically developed countries), from 6% in 1980 to 8.1% in 2006-2007. The rate varies across the provinces and territories, with the highest rate in Nunavut at 10.8%, followed by Alberta at 8.1% and the lowest rate in Prince Edward Island and New Brunswick at 7%. For more information on preterm birth, see Appendix B.

1.4 THESIS OBJECTIVE(S)

1.4.1 FIRST OBJECTIVE

Objective: Improve the Prediction of Neonatal Mortality and Preterm Birth in Twin Pregnancies

The first objective of this thesis was to construct an improved classifier system for predicting: (1) neonatal mortality non-invasively using data available at 10 minutes after birth which yields a sensitivity of 60% and specificity of 90% or greater, and (2) preterm birth in twin pregnancies non-invasively at 22 weeks gestation which yields a sensitivity of 65% and specificity of 85% or greater. These threshold values are derived from previously completed thesis work of N.Yu, C.Catley and J.Gilchrist, similar research, and in collaboration with Dr. Bariciak (Yu, 2009), (Catley, 2007), (Gilchrist, 2012). Perinatal data for this work will be obtained from the Niday (Niday Perinatal Database) and BORN (Better Outcomes Registry for Neonates) database.

The classifier system will be analyzed using two data mining techniques: decision trees (DTs) and hybrid ANN. The results of the DT and hybrid ANN approaches will be compared to past results using various performance measures to validate whether the new prediction model is superior.

1.4.2 SECOND OBJECTIVE

Objective: Design a Web-Based Content Management System Framework for Perinatal Decision Support

The second contribution of this thesis includes a conceptual framework, including the steps for the development and implementation of a web-based content management system for perinatal decision support. The system shall be low cost, easy to use and be designed to adapt to the clinician's cognitive workflow.

The Perinatal Decision Support System (PEDSS) shall consist of a knowledge-base, a workflow engine, and a mechanism to communicate the results. The knowledge-base contains a set of rules or associations related to the desired predictions; the workflow engine combines the rules in the knowledge-base with the patient data; and the communication mechanism allows entry of the patient data into the system, and delivers the results in the form of notifications, alerts or emails to the end user. This system will help physicians to inform families and to initiate preventative care, monitoring, and treatment.

1.5 THESIS OUTLINE

This thesis is organized as follows:

CHAPTER 1: Provides motivation, problem statement, and the need for a prediction model to predict neonatal mortality and preterm birth in twin pregnancies in perinatal care. The objectives are presented.

CHAPTER 2: Provides an overview of relevant medical and technological concepts related to neonatal mortality and preterm birth. An overview of a web-based content management system (CMS), and criteria for the selection of a CMS is presented. Types of neonatal decision support system are presented

CHAPTER 3: Describes the methodology used to create the prediction model for neonatal mortality and preterm birth in twin pregnancies using: (1) decision trees (DT), and (2) hybrid ANN. An overview of the proposed perinatal decision support system (PEDSS) is also discussed.

CHAPTER 4: The results for predicting neonatal mortality and preterm birth in twin pregnancies are presented. A framework of the PEDSS system is also provided.

CHAPTER 5: The concluding remarks, contributing knowledge and future work is presented.

APPENDIX A: Infant Mortality

APPENDIX B: Preterm Birth

APPENDIX C: Decision Tree Results

APPENDIX D: Hybrid ANN Results

CHAPTER 2 LITERATURE REVIEW: OVERVIEW OF RELEVANT MEDICAL AND TECHNOLOGICAL CONCEPTS

2.1 OVERVIEW OF PERFORMANCE MEASURES

Once a classification model has been developed, its predictive power must be evaluated using varying performance measures in order to determine how well the model will perform with new cases (Catley, 2007).

2.1.1 CONFUSION MATRIX (CONTINGENCY TABLE)

A confusion matrix is used to present the # of correct and incorrect predictions resulting from a model, in comparison to the actual classification from the test data. A matrix is typically *n*-by-*n*, where *n* represents the # of classes (Oracle Data Mining Concepts, 2005). Table 2.1 represents a confusion matrix for a two-class problem.

Table 2.1 Representation of a two-by-two confusion matrix

Predicted Value		Positive	Negative
	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

2.1.2 CORRECT CLASSIFICATION RATE

Accuracy is a measure of correctly predicted results by the classifier (Catley, 2007).

$$\text{Accuracy} = \frac{number\ of\ correct\ predictions}{total\ number\ of\ predictions} = \frac{TP+TN}{TP+TN+FP+FN} \quad (2.1)$$

2.1.3 ERROR RATE

Error rate is a measure of the incorrect predictions (Catley, 2007). The apparent error obtained by the prediction model using the training data is typically lower than the true error found using the validation and test data (Yu, 2009).

$$\text{Error Rate} = \frac{number\ of\ wrong\ predictions}{total\ number\ of\ predictions} = \frac{FP+FN}{TP+TN+FP+FN} \quad (2.2)$$

2.1.4 SENSITIVITY

Sensitivity refers to the true positive rate (TPR) and its ability to correctly predict a patient with a type of disease. A test with 100% sensitivity will correctly predict all patients with the disease. A test resulting in 80% sensitivity results in correctly predicting 80% of patients with the disease (True Positive), where 20% are undetected (False Negative) (Lalkhem & McCluskey, 2008).

$$\text{Sensitivity} = \text{TPR} = \frac{TP}{TP+FN} \quad (2.3)$$

2.1.5 SPECIFICITY

Specificity refers to the true negative rate (TNR) and its ability to correctly identify patients without the disease. A test with 100% specificity would correctly predict all patients without the disease. A test resulting in 80% specificity would result in correctly predicting 80% of patients without the disease (True Negative), where 20% of patients incorrectly identified (False Positives) (Lalkhem & McCluskey, 2008).

$$\text{Specificity} = \text{TNR} = \frac{TN}{TN+FP} = 1 - \text{False Positive Rate (FPR)} \quad (2.4)$$

2.1.6 LIKELIHOOD RATIO

This is a measure of how much more likely a patient who tests positive has the disease compared to a patient who has tested negative (Lalkhem & McCluskey, 2008).

$$\text{Likelihood ratio} = \frac{Sensitivity}{1-Specificity} \quad (2.5)$$

2.1.7 PREVALENCE

Prevalence is used to calculate the size of the population where the disease occurs. It is the # of occurrences of a disease within a specified period. The result is expressed as a percentage (Centre for Disease Control and Prevention, 2009).

$$\text{Prevalence} = \frac{Person\ with\ a\ given\ health\ indicator\ during\ a\ specified\ time\ period}{Population\ during\ the\ same\ time\ period} \text{ x } 100 \quad (2.6)$$

2.1.8 POSITIVE PREDICTIVE VALUE (PPV)

The positive predictive value (PPV) is the likelihood that that the positive prediction is correct, that is the probability of the actual disease given a positive prediction. The PPV is proportional to prevalence, and thus if prevalence is low, this will result in sharp decrease in PPV (Yu, 2009).

$$\text{PPV} = \frac{TP}{TP+FP} = \frac{TPR\ X\ Prevalence}{TPR\ X\ Prevalence+(1-TNR)X\ (1-Prevalence)} \quad (2.7)$$

2.1.9 NEGATIVE PREDICTIVE VALUE (NPV)

The negative predictive value (NPV) is the likelihood that the negative prediction is correct, that is the probability of a non-disease of the patient given a negative prediction. A low

prevalence results in greater assurance that the patient with a negative prediction is correctly diagnosed (Yu, 2009).

$$NPV=\frac{TN}{TN+FN}=\frac{TNR\ X\ (1-Prevalence)}{TNR\ X\ (1-Prevalence)+(1-TPR)XPrevalence} \quad (2.8)$$

2.1.10 RECEIVER OPRATING CHARACTERISTICS (ROC) CURVE

Receiver operator characteristic curve is a plot of the false positive rate (FPR) (1-specificity) on the x-axis and the true positive rate (TPR) (sensitivity) on the y-axis. The dashed line *A*, represents an ideal scenario where all points are correctly classified. Point *B*, represents higher TPR than Point *A*, with a higher FPR compared to Point *C*. Point *C* has a low FPR with a TPR approaching 45% (Yu, 2009). The dashed line D, is when the TPR is equal to the FPR, which represents a zero discrimination with AUC at 0.5 (50/50 probability, which is no better the tossing a coin). At point *E* (below the dashed line *D)* it is worse than random guessing. If the classifier's performance results in this area, this may indicate that although useful information may be contained in the classifier, the application is not correct (Yu, 2009).

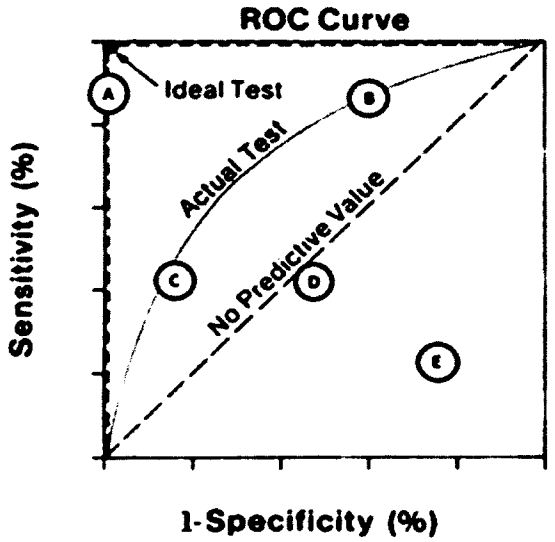

Figure 2.1 Receiver Operating Characteristic (ROC) Curve with Points of Significance Labelled (Sprawls, 2000).

2.1.11 AREA UNDER THE CURVE

The area under the curve is a measure of the classification model's effectiveness in correctly predicting medical outcomes. The AUC values range between 0 and 1, where the dashed line *D*, in Figure 3.1 has a value of 0.5 and the dashed line *A*, has a value that is closer to 1. The AUC's index and effectiveness for discrimination is given in Table 2.2 (Yu, 2009).

Table 2.2 Area under the curve index and its effectiveness for discrimination

Min	Max	Effectiveness
0	<= 0.500	**No Discrimination (Random**
0.500	<0.700	**Poor Discrimination**
0.700	<0.800	**Acceptable Discrimination**
0.800	<0.900	**Excellent Discrimination**
0.900	1.00	**Outstanding Discrimination**

2.1.12 ENSURING CONFIDENCE IN THE MODEL

The performance measures must be maximized or minimized for varying clinical outcomes, however determining which performance measures to maximize for predicting neonatal mortality or preterm birth in twin pregnancies is a challenge. To date no guaranteed treatment plan for the conditions described exists. Thus identifying a newborn at risk of mortality or preterm birth in a twin pregnancy does not necessarily imply that the negative outcome can be avoided. However, if a newborn or a mother has been diagnosed with a complication, then this tool may be useful in indicating whether further testing is necessary (Catley, 2007). Prompt recognition of these conditions may lead to more efficient allocation of resources and facilitate antenatal monitoring to ensure a timely response to medical problems.

Table 2.3 The clinical implications of true and false positives vs. true and false negatives (Catley, 2007)

True Positives	Additional follow up such as invasive testing may be suggested to evaluate the mother's risk. Pregnancy management strategy, including attending prenatal classes may be advised to mitigate the risk.	Some mothers may prefer not to know their elevated risk due to fear, or wish to avoid additional testing. Further, knowing this may contribute to additional maternal stress.
False Positives	Additional prenatal care may be beneficial if other potential maternal and infant implications were discovered and managed as a result.	Mothers may receive unneeded additional monitoring and testing which results in added cost to the healthcare system and stress on the mother.
True Negatives	Mothers may not require additional testing and visits, which reduces the costs of healthcare. Mothers also may experience less maternal stress knowing that they are at lower risk.	None identified.
False Negatives	None identified.	Mothers may not receive the additional prenatal care required, and the pregnancy management strategy may not be modified to reflect the mothers at high risk of pregnancy.

From the analysis presented in table 2.3, it can be concluded that the greatest benefit with the least amount of negative impact is reached by maximizing the true positive rate (i.e sensitivity). However obtaining a reasonably high true negative rate (i.e specificity) is also important in order to avoid unnecessary prenatal care, testing and visits and thus decrease healthcare expenses (Catley, 2007).

2.2 NEONATAL MORTALITY

Neonatal mortality refers to death of infants aged 0-27 days. In 2007, the rate of neonatal mortality in Canada during the first month was 4.2 deaths per 1000 live births, where 3.3 deaths per 1000 live births occurred during the first week (Milan, 2011).

2.2.1 CLASSIFICATION OF NEONATAL MORTALITY

Neonatal mortality can be subdivided into early neonatal (< 7 days) and late neonatal (7-28 days) periods (Rowley et al., 2011). Neonatal mortality typically occurs as a result of surrounding events during the prenatal period and delivery.

2.2.2 FACTORS CONTRIBUTING TO NEONATAL MORTALITY

Neonatal mortality arises largely due to insufficient care during pregnancy, inappropriate management of complications during delivery, and poor hygiene during the critical hours of delivery (World Health Organization Press, 2006). During the perinatal period, factors that may lead to neonatal mortality include the presence of the following in the newborn: respiratory distress syndrome, short gestation, and low birth weight (Public Health Agency of Canada, 1996). Other risk factors associated with neonatal mortality include birth defects, disorders related to preterm birth, low birth weight and congenital anomalies (Silins et al., 1985).

2.2.2.1 BEHAVIORAL CHARACTERISTICS

Modifiable risk factors associated with an increased risk of neonatal mortality include poor maternal diet, tobacco use and alcohol use (Silins et al., 1985). It has been found that

smoking during pregnancy results in a higher still birth rate. Further, the inhalations of smoke or second hand smoke are both risk factors of neonatal mortality (Rumeau-Rouquette, 1974).

2.2.2.2 SOCIOECONOMIC CHARACTERISTICS

The socio-professional characteristics of the parents contribute to the risk of neonatal mortality. In a study, the father's occupation was used to estimate the social status of the family, and it was observed that the unskilled labour class resulted in the highest rate of neonatal mortality, whereas the intermediate class resulted in a lower rate of neonatal mortality and the executive class resulted in an even lower rate of neonatal mortality (Rumeau-Rouquette, 1974).

Furthermore, being unmarried, having a low socioeconomic status, or the use of tobacco and alcohol among adolescent mothers poses a significant risk of neonatal mortality (Silins et al., 1985). At highest risk of neonatal mortality are women under 20 years of age (Rowely et al., 2011). The neonatal mortality is significantly higher in rural areas compared to urban areas (Silins et al., 1985).

2.2.2.3 LOW BIRTH WEIGHT

Birth-weight is a good indicator of neonatal mortality. Birth weight is classified into Very Low Birth Weight (VLWB) (<1500g), Extremely Low Birth Weight (ELWB) (<1000g), and Low Birth Weight (LBW) (<2500g) (Rowely et al., 2011). In a study conducted to examine the relationship between neonatal mortality and birth weight, it was found that the leading cause of death for neonates weighing <1000g was respiratory distress syndrome, whereas the leading cause of death in neonates weighing 3000-4500g were intrauterine hypoxia or anoxia (Silins et al., 1985).

2.2.2.4 MATERNAL HEALTH CHARACTERISTICS

Maternal health characteristics related to neonatal death include uterine fibroids and tumors of the ovaries. Further, metrorrhagia during the first three months of pregnancy, or the presence of endocrine, severe cardiovascular, digestive or respiratory diseases lead to a higher rate of still births, and neonatal mortality (Rumeau-Rouquette, 1974).

2.2.3 IMPACT OF NEW TECHNOLOGIES AND NEONATAL MORTALITY

New technologies are not always advantageous, especially ones that allow for sex selection procedures and assisted reproduction to allow twins births. Twins births increase the risk of complications in both the mother and the fetus, where a large percentage of twins and nearly all triplets are born preterm. The neonatal mortality rates are higher among these groups compared to singleton pregnancy. Furthermore, gender preferences vary across the world, where most cultures prefer sons, and thus the neonatal mortality among girls is predicted to be about 1/3 higher than recorded (World Health Organization Press, 2006).

2.2.4 ACCURACY OF STANDARD SCORING SYSTEMS FOR NEONATAL MORTALITY

There are many scoring systems available for neonatal mortality prediction. These scoring systems are typically used to identify newborns at risk of morbidity in order to apply appropriate medical attention or intervention. Three of the most popular scoring systems used to date are: Clinical Risk Index for Babies (CRIB), Score for Neonatal Acute Physiology (SNAP) and AGPAR. The CRIB and SNAP models have gone through changes and have evolved into CRIB II, SNAP II and SNAP-PE. The scoring systems presented in table 2.4 are for babies

admitted to NICU; therefore the sensitivity and specificity values are high, since the data is extracted from a homogeneous population. This thesis work aims to predict neonatal mortality in a heterogeneous population using data available 10 minutes after birth.

Table 2.4 Comparison between scoring systems used in neonatal care: (A) Scoring System; (B) Year the Model was Published; (C) # of Input Variables; (D) Time to Collect; (E) Gestational Requirement; and (F) Birth Weight Requirement (Mohkam, M. et al., 2010)

AGPAR		5	1 - 30 min		
CRIB	1993	6	<12 hours	< 32 weeks	<1500g
CRIB II	2003	6	<12 hours	< 32 weeks	<1500g
SNAP	1993	34	<24 hours		
SNAP II	2001	6	<12 Hours		
SNAP – PE	1993	38	<12 hours		
SNAP – PE II	2001	6	< 12 Hours		

In 2010, a cohort study was conducted to evaluate neonatal mortality using CRIB, CRIB II, SNAP, SNAP II and SNAP PE on 404 critically ill[3] neonates, with detected mortality in 20% of neonates. The predicative accuracy was expressed as area under curve (AUC), sensitivity, specificity, PPV and NPV. The results for AUC in CRIB, CRIB II, SNAP, SNAP II and SNAP PE were 81.7, 69.8, 93.1,90.1, 91.8, respectively, the sensitivity was 87.9, 69.6, 94.4, 84.8, 89.8, respectively, the specificity was 68.5, 63.0, 86.7, 62.8, 82.4, respectively, the PPV was 92.7, 76.2, 96.5, 90.8, 96.2, respectively, the NPV was 55.6, 54.8, 80, 54.4, 62.2, respectively, and the accuracy rate was 84.4, 67.1, 92.8, 82.7 and 88.6, respectively. The study concluded that the SNAP system had the highest AUC, sensitivity, specificity, PPV and NPV values and the lowest values were found using the CRIB II system (Mohkam et al., 2010).

In another study based on 295 newborns to assess the validity of SNAP, the resulting sensitivity, specificity, PPV, and NPV were 63.0, 95.0, 72.0 and 95.0, respectively. This study

[3] **Critically ill:** A patient experiencing an acute life-threatening episode or who is believed to be in imminent danger of such an episode.

excluded neonates submitted <24 hours of admission to NICU and those shifted to NICU for observation as per specifications of SNAP (Maiya et al., 2001).

A case controlled study[4] was conducted to evaluate the validity of CRIB II on 52 survivors and 52 non-survivors. The mean gestational age was 28.5 weeks and the mean weight was 865g. The results from this study indicated a sensitivity of 0.73 and a specificity of 0.60 (Fernandez-Carrocera et al., 2011).

In a study conducted to evaluate the validity of AGPAR scores and low birth weight in 3954 newborns with 3.8% mortality rate, the findings suggested a sensitivity of 0.84 and a specificity of 0.71, with a NPV value of 0.99; however the PPV was poor at 0.15. This estimation was based on 11 variables, where the AGPAR score at 1 min, birth weight and delivery mode were found to be variables that emerged as significant (Weirich et al., 2005).

In a similar study to evaluate the validity of low birth weight and 5-min AGPAR score, the sensitivity was lower at 0.722. This study was conducted on 875 newborns with 713 survivors and 163 non-survivors using 13 variables (Weirich et al., 2005).

2.2.5 LIMITATIONS IN NEONATAL MORTALITY SCORING SYSTEMS

One of the main disadvantages to these scoring systems is the time to collect the information, with the exception of the AGPAR scoring system. In addition, although these scoring systems have been in place for over three decades, the validity of the results varies (due to different prognostic variables, timing, quality of care etc.) and the ability to predict survival is poor. Most of these systems do not meet the clinical standards set by Dr. Bariciak, and even if these systems were to meet the clinical standards, it is accompanied by a poor PPV value, and in many cases these findings produced a high sensitivity followed by a low specificity. Further,

[4] **Case Controlled Study:** A case-control study is an analytical study which compares individuals who have a specific disease ("cases") with a group of individuals without the disease ("controls")

these findings are all based on patient populations of <900. Although the SNAP scoring system did meet the clinical standards, the data can only be obtained within 24 hours from admission to NICU and required 34 variables. A data collection to include all 34 variables may be feasible for hospitals involved in R&D, but this is likely not be feasible for independent ICUs where it may not be part of their normal clinical routine. Further, the 24 hour period required for data collection is too long, especially in an environment where time is of the essence (Yu, 2009). Thus, there clearly exists the need for a scoring system that is able to predict neonatal mortality in a shorter time frame using fewer variables which meets clinical standards.

2.3 PRETERM BIRTH

Approximately 350 000 babies are born preterm in Canada each year. Despite significant advances in medical technologies, the rate of preterm birth has risen from 6.3% in 1981-1983 to 6.6% in 1991, and 7.6% in 2000. The rate of preterm births in Canada is currently ~7% (Goldenberg et al., 2008). Although advances in medical technologies have allowed for the majority of preterm babies in the current generation to survive, certain advances in medical technologies including IVF treatments or fertility drugs have been blamed for the increase in the number of preterm births (Norton, 2010).

Due to the direct correlation between preterm births and neonatal mortality, preventing preterm births is one of the main priorities within the healthcare domain. A common goal of perinatal care is to identify key factors contributing to preterm births. Previously many studies have attempted to predict women who are at risk of preterm birth, yet to date no scoring mechanism exists that has been proven to be superior to clinical judgment (Catley, 2007).

2.3.1 CLASSIFICATION OF PRETERM BIRTHS

The World Health Organization (WHO) defines preterm births as births occurring before the completion of 37 week gestational period. An ideal term birth occurs between 37-41 weeks of gestation and post term birth occurs >41 weeks gestation. The gestational period is measured from the first day of the last normal menstrual period of the mother (Blackmore et al., 2011).

Preterm births can be further categorized according to gestational age and birth weight (Table 2.5). About 70% babies born preterm are categorized as low birth weight (<2500g) (Lim et al., 2009) (Allen et al. 2002). Low birth weight is an important variable for predicting health problems and disability of the newborn (Allen et al. 2002).

Table 2.5 –Preterm birth classified according to gestational period and birth weight. (A) Sub Groups; (B) Gestational Age; (C) Typical Birth Weight; (D) Occurrence; (E) Survival; (F) Disability; (G) Cost; and (H) Length of Stay (LOS) (Goldenberg et al., 2008) (Lim et al., 2009)

(A)	(B)	(C)	(D)	(E)	(F)	(G)	(H)
[illegible]	28 weeks	<1000g	5%	>80%	15%	84 235	83.1
[illegible]	28-31 weeks	<1500g	15%	>80%	10%	43 718	42.6
[illegible]	32-33 weeks	<2250g	20%	>95%	5%	19 463	21.2
[illegible]	34-36 weeks	<2500g	60-70%	99%	1%	5 047	5.8

2.3.2 FACTORS CONTRIBUTING TO PRE-TERM BIRTHS

The pathogenesis associated with preterm birth is not clearly understood. However, certain factors may increase the risk of preterm birth (Alere's Women's & Children's Health, 2008). Defining risk factors for preterm birth is favourable for many reasons including the fact that identifying an at-risk woman may allow for earlier treatment in order to prolong the gestational period. Moreover, defining an at-risk population may lead to further studies on the identification of vital risk factors that are linked to preterm birth (Goldenberg et al., 2008).

The root cause of preterm birth is a multi-factorial problem, with the risk factors varying from a combination of the maternal demographic, socioeconomic, health, genetic, biological, pregnancy, stress, and behavioural characteristics. Due to the large number of predictor variables associated with preterm births, prediction of preterm birth in an at-risk population remains a challenging problem. Please refer to Table F.1 for a complete list of risk factor variables associated with preterm delivery derived from the BORN and Niday databases. The following section provides a brief summary of the associated risk factors of preterm births.

2.3.2.1 MATERNAL DEMOGRAPHIC & SOCIOECONOMIC CHARACTERISTICS

It has been reported that African American women are at an increased risk of preterm birth, where Afro-Caribbean women are at nearly twice the risk of preterm birth compared to other races (Goldenberg et al., 2008). Several studies have identified that low socioeconomic and educational statuses, low or high maternal age and single maternal statuses are linked to preterm birth (Goldenberg et al., 2008).

East Asian and Hispanic women are at lower risk of preterm birth. Though women from South Asia have higher rates of low birth weight due to decreased fetal growth, they are not necessarily at risk of preterm birth (Goldenberg et al., 2008).

2.3.2.2 GENETIC CHARACTERISTICS

Women who previously delivered preterm are at higher risk of another preterm birth compared to women who previously delivered term. Women with sister(s) who delivered preterm are at 80% risk of also delivering preterm (Svensson et al., 2009). Underlying disorders that lead to preterm birth include diabetes, hypertension, and obesity (Goldenberg et al., 2008).

2.3.2.3 MATERNAL HEALTH CHARACTERISTICS

The mother's health during pregnancy may be described by several indicators. It has been found that women with high or low body mass index (BMI) during pregnancy are at higher risk of adverse conditions during delivery. Women with high BMI are prone to adverse delivery outcomes including hypertensive complication and caesarean section, whereas women with low BMI are linked to preterm birth, low birth weight and increased rate of neonatal mortality (Yu, 2009). Women who have lower serum concentration of iron, folate or zinc are at higher risk of preterm birth.

Insufficient weight gain during pregnancy is also a risk factor associated with preterm birth. Thin women commonly intake lesser vitamins and minerals which may lead to decreased blood flow and increased infections. Obese women are more likely to have pre-eclampsia and diabetes or other medical conditions including glucose intolerance, hypoglycemia and gestational diabetes mellitus (GDM), which are all factors associated with preterm birth.

2.3.2.4 BIOLOGICAL CHARACTERISTICS

Several biological markers have been used to detect preterm birth, which includes amniotic fluid, urine, cervical mucus, vaginal secretions, serum or plasma, and saliva. The most powerful marker has been fetal fibronectin, which is a glycoprotein. Usually, it is absent from 24 weeks to near term, however if it is detected during routine screening in women between 24-32 weeks, this indicates that they are at increased risk of preterm birth (Goldenberg et al., 2008)

Intrauterine infection leads to 25-40% of all preterm birth due to the activation of innate immune system, although this may be an underestimation since it is difficult to detect using standard culture techniques. Bacterial vaginosis (disease that occurs due to the change in

microbial ecosystem in the vagina) detected using a vaginal pH of greater than 4.5 leads to preterm birth. Although many infections are related to preterm birth, there are also many non-genital tract infections associated with preterm birth, which include pyelonephritis and symptomatic bacteriuria, pneumonia and appendicitis. Further periodontal disease which leads to intrauterine infections is associated with preterm births. Viral infections including varicella or severe acute respiratory distress syndrome can also lead to preterm birth. Other maternal medical disorders that have led to preterm birth include thyroid disease, asthma, diabetes and hypertension (Goldenberg et al., 2008).

Decidual haemorrhage is also linked with preterm birth. This may cause bleeding in the placenta which causes thrombin formation and increased prostaglandin which leads to cervical change and rupture of the membrane (Allen et al. 2002). Uterine over distension causes the activation of the uterine muscles which may lead to preterm birth (Allen et al. 2002).

2.3.2.5 PREGNANCY CHARACTERISTICS

Women carrying twin fetuses are at higher risk of delivering prematurely. The presence of twin fetuses in the uterus limits the room available for fetal growth and development, which then leads to pregnancy related complications including uterine over distension and contractions leading to preterm birth (Goldenberg et al., 2008).

Other risk factors associated with preterm birth include vaginal bleeding caused by placental abruption or placenta previa, bleeding in the first and second trimester which is not related to placental abruption, high volume of amniotic fluid, and/or maternal abdominal fluid in second/third trimester (Goldenberg et al., 2008).

2.3.2.6 STRESS

Women experiencing high levels of psychological or social stress are at higher risk of delivering preterm. Although the underlying mechanism causing this is unknown, a hormone known as corticotrophin has been proposed (Goldenberg et al., 2008). Furthermore, women experiencing clinical depression are also at elevated risk of delivering preterm. Although the exact link is not clear, there is a relation in the fact that women experiencing depression are often correlated with an increased risk of smoking, drug abuse and alcohol. Depressed mood is also related to a natural killer cell activity and higher plasma concentration of inflammatory cytokines and their receptors, and therefore inflammation may relate depression to preterm birth (Goldenberg et al., 2008).

It has been observed that women who work long hours or are undertaking hard physical labour are at higher risk of delivering preterm. Although the level of physical activity is not related to preterm birth, the stress that is induced from overworking is the underlying factor that leads to preterm birth (Goldenberg et al., 2008).

2.3.2.7 BEHAVIORAL CHARACTERISTICS

Tobacco, drugs (cocaine and heroin) and heavy alcohol use puts pregnant women at higher risk of delivering preterm. Smoking contains over 3000 chemicals, including nicotine and carbon monoxide which are both powerful vasoconstrictors, which lead to placental damage and decreased uteroplacental blood flow which may lead to preterm birth (Goldenberg et al., 2008).

2.3.3 EARLIER DETECTION AND PREVENTION OF PRETERM BIRTH

As much as 25% of all preterm births are associated with identifiable health problems in the mother, and thus recognizing the underlying pathway which leads to preterm births may lead to new preventative measures (Allen et al. 2002). There are certain measures which can be taken to reduce the chances of preterm birth, which include starting prenatal care early, eliminating and avoiding all modifiable risks associated with preterm birth (i.e. eliminating tobacco, drugs and alcohol use), following Canada's guide to nutrition in order to maintain a healthy diet, seeking help when symptoms and signs are observed of preterm birth, and maintaining good communication with the respective doctor/midwife regularly (Ontario's maternal newborn and early child development resource centre, 2004). Although it is not easy to tell if a women is at risk of preterm birth, a few vital signs and symptoms to look out for in pregnant women include persistent cramps, vaginal bleeding, lower back pain, a feeling of baby pushing down, contractions, and an increase in the amount of vaginal discharge (Ontario's maternal newborn and early child development resource centre, 2004).

2.3.4 ACCURACY OF STANDARD INAVSIVE TEST METHODS FOR PREDICTING PTB IN TWINS

The ability to correctly identify and predict preterm birth in a high-risk population, such as women pregnant with twins, could provide a better understanding of the risk factors leading to preterm birth. The gold standards for predicting preterm birth in women carrying twins are cervical length (CL) measurements and fetal fibronectin (fFN). Cervical length measurements are most useful between 16-32 weeks gestation. After 32 weeks, the cervix naturally shortens for women delivering term. Fetal fibronectin is a protein that is found between the maternal-fetal

interface and in low levels in vaginal secretions. This protein is normally present before 24 weeks gestation, and again after 3 weeks prior to delivery, thus in term delivery it would be after 34 weeks gestation. This test is useful in gestational interval between 24-32 weeks, where levels above >= 50 mg/mL are marked as abnormal and are linked to spontaneous preterm births (Yu, 2009).

In a cohort study to evaluate cervical length measurement for the prediction of preterm birth <34 weeks gestation in 2757 women with twins pregnancies, the resulting sensitivity and specificity were different for varying for cut off points in cervical length. The summary estimates in terms of sensitivity and specificity were 0.78 and 0.66 for 35 mm, 0.41 and 0.87 for 30 mm, 0.36 and 0.94 for 25 mm, and 0.30 and 0.94 for 20 mm, respectively (Lim, A.C et al., 2011). In another cohort study with a total of 3523 women pregnant with twins to predict preterm birth before 32 and 34 weeks at 20-24 weeks gestation with cervical length <=20 mm, the pooled sensitivity were 39% and 29%, and specificity were 96% and 97%, respectively (Conde-Agudelo, Romero and Yeo, 2010). In a study involving 147 women pregnant with twins to predict preterm birth before 35 weeks at 22-24 weeks resulted in a low sensitivity of 30% and high specificity of 88%, with PPV and NPV of 54% and 74%, respectively, for cervical length <=25 mm (Mella et al., 2009). In a similar study to evaluate the validity of cervical length measurements in women carrying twins at 27 weeks gestation, the prediction for delivery <34 weeks had a sensitivity of 77%, specificity of 86%, with a poor PPV 0.34 for cervical length <=25 mm. These results were lower at 18 weeks with sensitivity of 14.3%. At 24 weeks gestation, for cervical length <=22 mm the results were poor with a sensitivity of 28.6%. Therefore a cervical length measurement alone is not a suitable predictor of preterm birth due to the poor sensitivity prior to 24 weeks gestation (Institute Advanced Medical Education, 2013).

Table 2.6 Comparison of performance outcomes in cervical length measurements in literature

Lim, A.C et al.	<34 weeks	2757	35 mm	78%	66%
Conde-Agudelo et al.	32 weeks	3523	<=20 mm	39%	96%
Mella et al.	<35 weeks	147	<=25 mm	30%	88%
Institution of AME	<34 weeks	-	<=25 mm	77%	86%

In a study conducted on 52 pregnant women with twins to test the validity of fetal fibronectin to predict preterm birth before 37 weeks gestation, the following sensitivity, specificity, PPV, and NPV were found with respect to the following gestational age: 24-26 weeks – 0.667, 0.818, 0.778 and 0.72; 27-30 weeks – 0.857, 0.708, 0.774 and 0.809; 31-34 weeks – 0.846, 0.583, 0.688 and 0.778, respectively (Oliveira et al., 1999). In a study to predict preterm birth before 35 weeks gestation with fFN present at 28 weeks resulted in sensitivity, specificity, NPV and PPV of 50.0%, 92.0%, 62.5% and 87.3%, respectively (Wennerholm et al., 1997). In another study to predict preterm birth before 32 and 35 weeks gestation with fFN present at 28 weeks gestation, resulted in sensitivity of 28.6% and 57.1%, and specificity of 96.1% and 26.5%, respectively. The presence of fFN at 30 weeks, resulted in sensitivity of 37.5% and 62.5%, and specificity of 98.9% and 72.7%, respectively (Blickstein and Keith, 2005). In a study containing 87 twin birth, the presence of fetal fibronectin after 24 weeks testing resulted in sensitivity, specificity, PPV and NPV values of 0.71, 0.74, 0.19 and 0.97 respectively (Singer et al., 2007). In a similar study containing 48 twin births of primarily Hispanic ethnicity, the presence of fetal fibronectin between 22-24 weeks gestation to predict preterm birth before 35 weeks gestation, resulted in sensitivity, specificity, PPV and NPV values of 0.7683, 0.5833, 0.667 and 0.700, respectively (Ruiz, R.J. et al., 2004).

Table 2.7 Comparison of performance outcomes in fetal fibronectin measurements in literature

Oliveira et al.	<37 weeks	52	24-6	66.7%	81.8%
Wennerholm et al.	<35 weeks	-	28	50.0%	92.0%
Blickstein, I .& Keith, L.	<35 weeks	-	28	57.1%	26.5%
Singer, E.	<32 weeks	87	>24	71.0%	74.0%
Ruiz, R.J.	<35 weeks	48	22-24	76.83%	76.83%

2.3.5 LIMITATIONS IN INVASTIVE TESTING METHODS

The standard clinical prediction methods (CL and fFN) commonly offered to high risk patients are costly and only effective >24 weeks gestation. Further, one of the biggest challenges in past studies is trying to compare the performance outcomes with the varying cut off gestational ages for preterm including <32, <34, <35 and <37 weeks. For CL screenings, the CL cut off values and gestational timings have varying levels of risks associated. This heterogeneity makes it challenging to make a direct comparison between the various prediction accuracies in studies (Yu, 2009). A clear need exists for a non-invasive prediction tool that can predict preterm birth <24 weeks gestation in women carrying twins, with improved sensitivity while meeting clinical standards. Since predicting preterm birth in women carrying twins is controlled by several factors, this thesis will explore two data mining techniques including decision trees and hybrid ANNs to predict this complex outcome.

2.4 COLLABORATION PLATFORM FOR HEALTHCARE

Collaboration is a term used to describe two or more people, or organizations interacting and working together. To date, many collaborative systems incorporate a virtual workspace feature for users to interact. Authorized users would have transparent access to shared virtual workspace regardless of their browser, network connection or geographic location (Intapong, 2010).

Hospitals, primary care physicians, specialists, and nurses are challenged to adhere to evidence based practises and are often pressured to execute the latest clinical technological advances in order to provide safer and enhanced patient care. Furthermore, patient information and clinical data may be scattered throughout the enterprise in the form of third party databases, electronic documents, diagnostic imaging, clinical trials etc.... In order to determine appropriate care protocols, medication administration, standard operating procedures and to improve patient care, healthcare providers (i.e. physicians, nurses, pharmacists, and other clinical staff) must have a means to easily collaborate.

The evolution of the web has transformed the internet into a strategic medium for many types of organizations including healthcare. Ideally, in a clinical environment, a web-based collaborative environment would allow healthcare providers to collaborate in the form of sharing opinions, exchange clinical data, and access clinical information, irrespective of their geographic location (Chronaki et al., 1997).

2.4.1 OVERVIEW OF WEB-BASED CONTENT MANAGEMENT SYSTEMS

A Content Management System (CMS) allows collaboration within organizations and is largely composed of content, process and software. The content portion of the system includes text, images and various media. The process, is the manner in which the system is developed, and the content is published. The software serves as a bridge to the CMS, which allow site administrators to control the site's content without needing to learn extensive programming (Lurie, 2002).

In general, the site content of a CMS is stored in a database, and the content within the CMS is much easier to maintain compared to a standard web page. A CMS also supports easier management of security, allowing administrators to easily control their end users' level of permissions. Many CMSs also supports workflow integration, and thus allow automation of business processes. In the long run, a content management system would lead to reduced operational costs and increased revenue. Since the non-technical staff are able to maintain the site; this eliminates the need to hire programmers, or a development team (Lurie, 2002).

An enterprise content management system (ECMS) is designed for enterprise usage, and allows forms processing, archival and retrieval of data (converting paper based patient data to electronic form), data warehousing (storing and analyzing information), and document distribution via internet. This in return eliminates many manual error prone processes in a corporate setting (i.e. the time and resources wasted in faxing referrals) (Fujitsu Computer Products of America Inc., 2010).

2.4.2 CRITERIAS FOR SELECTING A CMS

It is important to deliver targeted and useful content to clinical users upon entering into a site; this in turn will save the healthcare providers a lot of time within a clinical setting. Selecting the best content management system tailored towards the organizational need can be a difficult task. A few decisions need to be made prior to implementing a content management system. These include determining whether the organization would benefit from open source or commercial sources, or, an enterprise content management system (ECMS) or web specific. In a large healthcare setting, it is important to choose commercially available systems in order to upkeep the performance, security, reliability and commercial grade support. Although the initial cost of open source systems are lower, the development and maintenance costs over time are much higher compared to commercial grade products. Secondly, it is important to consider that in a clinical setting, it is likely that the public will not be given access to the system, due to the sensitive patient information which it may retain. An ECMS would be ideal in a clinical setting, and in addition, an ECM would result in a reduction in manual processes, increase productivity, and can be optimized to fit the needs of the organization (Sitecore, 2009). Furthermore, a CMS should support multi-platform compatibility. Especially for larger organizations such as healthcare, a CMS should be compatible with common platforms including Windows, Mac, Linux, Solaris etc. In the event that the organization is entirely a Windows based, a CMS designed for only one OS will limit the user's functionality on the road or when working from home (Hannon Hill Corporation., 2010).

2.5 OVERVIEW OF CLINICAL DECISION SUPPORT SYSTEMS

Most Clinical Decision Support Systems (CDSS)s over the years were either designed to be a standalone system or part of a non-commercial computer-based patient record system. Recently vendors have started to include CDSS into computer based patient record systems (Berner, 2007). Studies have indicated that healthcare organizations utilizing a CDSS have led to enhanced outcome in the clinician's performance as well as better patient outcomes (Child Health Research Project Special Report, 1999). Further electronic medical records (EMRs) are designed to improve both the accessibility and legibility of information (Berner, 2009). EMRs are considered to be the foundation for providing quality healthcare. A CDSS is essential to achieve the full potential of an EMR (Berner, 2009). With over 100 EMRs to choose from, it is important that data standardization protocols be enforced. Although humans may still make sense of information if the data were not standardized, computers will not be able to (Miller, 2010).

As society moves towards a service-oriented architecture, knowledge may be represented via web services. This allows EMRs to remotely subscribe to a publicly available web service and transfer data to the knowledge repository (Miller, 2010). CDSS are designed to intelligently filter information in order to aid the decision making process in a clinical setting (Berner, 2009). CDSSs may also be used to gather large sets of data and manage medication actions effectively. In turn, using a CDSS will reduce the occurrence of practice errors, and provide higher quality of care (Peleg & Tu, 2011).

2.5.1 TARGET AREA OF CARE

CDSSs may aid physicians in assessing varying clinical issues ranging from providing accurate diagnosis to ordering medication. The general target areas of care for CDSS are presented below (Berner, 2009).

- Preventative Care – screening and disease management
- Diagnosis – search for diagnosis based on patients signs and symptoms
- Follow up Management – reminders and alerts
- Hospital Provider Efficiency – records/management of hospital beds, length of stay etc..

2.5.2 SYSTEM DESIGN

A collaborative system design will include the following subsystems:

- Communication – to handle alerts and notifications
- Knowledge discovery – rules and conditions
- Knowledge repository – large sources of patient information
- Multi-participant presentation component (Frize, 2005)

2.5.3 FACTORS LEADING TO SUCCESSFUL CDSS IMPLEMENTATION

(1) Simple, user friendly interface

(2) Automated decision support

(3) Workflow integration

(4) Timely results

(5) Continuous knowledge-base and user-interface update support (Peleg & Tu, 2011)

2.5.4 TYPES OF CLINICAL DECISION SUPPORT SYSTEMS

CDSSs may be categorized into either a knowledge-based system or non-knowledge-based system. Most systems to date are classified as knowledge-based systems, and contain rules, guidelines or compiled knowledge (commonly derived from medical literature). CDSSs may be further be classified into active or passive systems, depending on whether the system responds actively or passively to physician input. The ultimate goal of decision support systems is to support and enhance human thinking (Berner, 2007).

2.5.4.1 KNOWLEDGE-BASED SYSTEM

There are three main parts to a knowledge-based system including the knowledge-base, inference engine (or reasoning engine) and a mechanism to communicate. The knowledge-base often contains rules in the form of IF-THEN. For example, a rule designed to prevent duplicate test ordering may be in the form of IF new order was placed for a particular blood test, AND IF blood test was ordered within 24 hours, THEN alert the physician. Other forms of knowledge-bases may also include probabilistic association between drug and food interactions, signs or symptoms and diagnosis. The inference engine consists of formulas to combine the rules (or associations) with the patient data in the knowledge-base. The communication mechanism is the method for transporting the patient data into the system, and the results out of the system, to be displayed to the user, who will then make the final decision. In a standalone system, the patient data will need to be entered into the system by the end user. However, CDSSs which integrate with EMRs will auto-populate the patient data into appropriate fields from the information available in the EMR. The output of the system may be in the form of notifications, alerts or emails (Berner, 2007).

In the past, systems were built to assist the clinicians' with their decision making and were largely classified as diagnostic decision support systems. These were commonly standalone systems which would contain a large information repository, and would provide a list of potential diagnosis based on the patient's signs and symptoms (entered by the clinician or extracted from the EMR). The users of these systems were expected to be active instead of being passive, and thus interact with the system to filter important information. The interaction between the user and the system is important in order to determine how the system will be used. There are also systems built to assist primary care providers with medication order known as computerized physician order entry (CPOE). A typical input to this type of system may be the patient's lab result, and the knowledge-base may contain values and rules to output the list of potential medications as well as include various customized built-in features (i.e. an automatic alert to the physician if the toxic level of medication is too high). There are also types of CDSS that are part of CPOE, where the input of the system is the patient's current medication and new medication, and the knowledge-base would include a drug database and output the drug interaction (Berner, 2007).

2.5.4.2 NON-KNOWLEDGE-BASED SYSTEM

Non-knowledge-based decision support systems use artificial intelligence in the form of machine learning, to allow the computer to learn from past experiences, or analyze and detect patterns in clinical data. Common type of non-knowledge-based systems includes artificial neural networks (Berner, 2007).

2.5.5 DESIGN CONSIDERATION AND GOALS

There are several design elements that must be taken into consideration to produce an effective tool which provides high quality, safe and efficient patient care, which also encourages physician adoption. Successful implementation of a CDSS will be facilitated by a simple, intuitive and user-friendly interface, on an easily accessible and mobile-friendly platform. Furthermore, the system must be developed at low cost, provide timely results with automated decision support, as well as support a continuous knowledgebase and user interface updates (Peleg & Tu, 2011) (Frize & Weyand, 2010).

2.5.5.1 CLINICIAN WORKFLOW PROCESS

When designing a framework for a CDSS, a solid understanding of a typical clinician's workflow process, including the performance of both routine and complex clinical tasks in a hospital setting is required. System builders often focus on a CDSS that produces good decisions, yet researchers have shown that the ability to produce a correct diagnosis is only one part of the formula for success. It is important to recognize that systems can fail if they require that the practitioner interrupts a normal work pattern (i.e. to shift to another station to start up a non-intuitive software program that contains time consuming start-up procedures) (Musen, Shahar & Shorliffe, 2006). Thus, a seamless integration of the CDSS into the clinician's workflow routine would increase the likelihood of the CDSS being used.

One of the early steps in the system development phase would be to observe the clinicians' typical routine and determine how the CDSS should be integrated. The clinicians' workflow pattern is determined by observing their interactions with health information systems including the use of related tools and devices during their normal routine. The clinicians' may be

required to incorporate minor changes in their normal routine either prior to the adoption of CDSS or during the adaption of CDSS to optimize care (Berner, 2009).

It is also important to realize the variability that exists among clinicians, and thus a single workflow design may not meet the needs of all clinicians, since each clinician may have developed his/her own particular work style. Therefore the business workflow component must be adaptable to accommodate the needs of the clinician (Berner, 2009). A solid understanding of the clinician's workflow is required for system developers to design and implement effective workflows that encourage physician adoption.

2.5.6 OVERVIEW OF CURRENT CDSS

2.5.6.1 HEART RATE OBSERVATION SYSTEM (HeRO)

The HeRO project was led by J.Randall Moorman which focused on early detection of neonatal distress in very low birth weight (<1500g) infants using heart rate variability (Griffin et al., 2005). Heart rate variability is commonly found in healthy patients, and thus a reduction in variability could signify illness. Earlier detection would allow for earlier treatment and interventions and improve patient outcome (Gilchrist, 2012). In a study conducted with over 3000 preterm infants using the HeRO monitoring system, a 20% reduction in mortality was observed, where 1 neonate's life was saved for every 48 hours monitored (Fairchild et al., 2012).

The HeRO system is non-invasive and generates a numeric score that quantifies the prevalence of abnormal patterns in each person's heart rate. The HeRO system continuously acquires records and measurements to analyze heart rate patterns and provides a real-time display of the HeRO score. Further it supports secure remote monitoring capabilities and interfaces with any existing ECG systems (Medical Predictive Science Corporation, 2011).

2.5.6.2 ARTEMIS

The ARTEMIS is pilot collaboration between researchers from the University of Ontario Institute of Technology (UOIT), IBM and the Hospital of Sick Children in Toronto, Canada (IBM InfoSphere, 2010). The ARTEMIS platform supports acquisition and storage of neonatal data and other clinical information data in NICU for the purpose of online real-time analytics. A series of clinical rules are executed online and in real-time to detect medical conditions including IVH, nosocomical infection, periventricular leukomalacia and pneumothorax (IBM InfoSphere, 2013).

The ARTEMIS system was first implemented at the Hospital of Sick Children in August 2006 and is still in use to date (Health Informatics Research, 2008), (IBM InfoSphere, 2013). ARTEMIS utilizes IBM's InfoSphere for the purpose of data processing. This middleware system supports real-time data processing and storage. The programming language used to support InfoSphere is IBM's Stream Processing Application Declarative Engine (SPADE) (ARTEMIS, 2008).

2.5.6.3 Realtromis

Realtromis (Real Time Risk of Mortality & Instability) is a software medical company which created an advanced predictive analytic for the purpose of continuously assessing the risk of neonatal mortality in critically ill infants in an attempt to identify high risk neonates (GlobalData, 2013). This system uses continuous physiological attributes including ECG, patient monitor data, laboratory results from organ functions, demographic, diagnosis related predictors, and other clinical parameters using the Health Level 7 (HL7) messaging standards (Gilchrist, 2012).

Their platform consists of an interface for physiological data collection, temporary data storage for analyzing the physiological data, and a mortality risk function which retrieves the data from the storage to calculate the mortality risk, and presents the outcome via a user interface. It is not clear how the mortality risk outcome is presented to the user (i.e numerical value, scoring system etc...). Further the performance result of the real-time neonatal medical system has not yet been disclosed to the public (Gilchrist, 2012).

2.6 PATTERN CLASSIFICATION METHODS

Pattern classification refers to theory and algorithms of assigning abstract objects (e.g. measurements on physical objects) into distinct categories, where these categories are typically known in advance. Methods of pattern classification have been applied for information retrieval, data mining, document image analysis and recognition, computational linguistics, forensics, biometrics and bioinformatics (Srihari, 2007). In this thesis work, we use two different pattern classification methods: Decision Trees (DTs) and Artificial Neural Networks (ANNs).

2.6.1 DECISION TREES

Decision trees are favoured in the data mining community due to the highly interpretable structure, allowing business end users and analysts to understand the models, whereas neural networks are difficult to understand and interpret (Apte et al., 2002). A decision tree consists of a root node, branch nodes and leaf nodes. The tree starts with a root node, is further split into branch nodes (each of the nodes represent a choice of various alternatives), and terminates with a leaf node which are un-split nodes (represents a decision) (Peng, 2006).

Classification of decision trees are conducted in two phases, including the tree building (top down) and tree pruning (bottom-up). Tree building is computationally intensive, and requires the tree to be recursively partitioned until all data items belong to the same class. Tree pruning is conducted to improve the prediction and classification of the algorithm and to minimize the effects of over-fitting, which may lead to misclassification errors (Anyanwu and Shiva, 2009).

There are a number of decision tree algorithms that exist including Classification and Regression Trees (CART), Iterative Dichotomiser 3 (ID3), C4.5 and C5.0. This thesis work uses C5.0 based decision tree algorithm which is an improvement over C4.5, which itself is an improvements over the earlier ID3 method

2.6.1.1 IMPROVEMENTS WITH C5.0

Decision trees created using C5.0 are smaller, use less memory and are significantly faster compared to C4.5. C5.0 algorithm also supports boosting, variable misclassification cost and winnowing. Boosting is method of creating multiple decision trees and combining it then to improve the accuracy, where each tree votes for its predicted class and the final class is determined using a vote count. C5.0 will automatically generate 10 trees using its boosting feature and will aim to minimize the expected misclassification costs. C5.0 also has the ability to winnow attributes with high dimensionality prior to constructing the tree; this is done by removing the attributes which it predicts to have minimal relevance in an attempt to generate smaller trees with higher accuracy. C5.0 also supports additional data types including case labels, dates, time and ordered discrete attributes (Rulequest Research, 2012).

2.6.1.2 DEVELOPMENT OF DECISION TREE MODEL

The classification of decision trees is dependent upon the attribute's value in the dataset. Each node in a decision tree is a representation of an attribute in a dataset, and the related branch is a representation of the values associated with the attribute (Estrabrooks, 2000). Decision trees are typically constructed using a top down greedy search algorithm. This algorithm recursively subdivides the train data based on the best classifying attribute in the train set using a statistical property known as information gain. The best classifying attribute then becomes the root node, and then this algorithm is repeated for each partition of the dataset to create sub trees. This process is continued until all of the training data is divided into subsets of the same class (Estrabrooks, 2000).

2.6.1.2.1 INFORMATION GAIN AND ENTROPY MEASUREMENTS

Information gain is a measure of how well the attribute can classify the training data to best predict the outcome. Information gain uses a measure known as entropy to calculate its result. The information gain is given below (eqn 2.9) (Estrabrooks, 2000):

$$\text{Gain (S,A)} = Entropy(\text{S}) - \sum_{v \epsilon Values(A)} \frac{S_v}{S} Entropy(S_v) \quad (2.9)$$

Where

- A – Attribute
- S - Subset
- Values(A) – All values in attribute 'A'
- S_v – Subset of examples in 'S' which have value 'v' for an attribute 'A'

The $Entropy(\text{S})$ can be defined over a collection of training data 'S' as follows (eqn 2.10) (Estrabrooks, 2000):

$$Entropy(\text{S}) = -p_{(+)} log_2 p_{(+)} - p_{(-)} log_2 p_{(-)} \quad (2.10)$$

- Where
- $p_{(+)}$- The proportion of positive examples in 'S'
- $p_{(-)}$- The proportion of negative examples in 'S'

2.6.1.3 DECISION TREE ADVANTAGES AND DISADVANTAGES

Decision algorithms are easy to interpret and comprehend. These algorithms are able to handle both metric (numerical, real values) and non-metric (categorical, descriptions) data, as well as handle missing values which are often encountered in clinical studies. The resulting Decision Trees (DTs) have comparable accuracy to other classification methods such as the ANNs. Further, little data preparation is required since the data does not need to be normalized or have missing data replaced with an imputed value. DTs can also handle large amounts of data in a short time frame compared to other pattern classification techniques and it can often be developed using common statistical techniques (Peng, 2006).

However, some drawbacks associated with decision trees is that it can over fit the data and create complex trees which may not generalize well. Further, a small change in the size of the dataset could result in a completely different tree, and thus the prediction results may not be stable when validating upon multiple datasets.

2.7.2 ARTIFICIAL NEURAL NETWORKS (ANN)

Artificial Neural Networks (ANNs) are powerful non-linear mapping structures and are especially useful for modelling relationships which are unknown. ANNs function similar to the human brain and can solve problems involving data that is complex, non-linear, imprecise and/or noisy (Jha, 2011). The human brain is a collection of more than 10 billion interconnected neurons that is able to receive, process and transmit data. The human brain also consists of a highly parallel computing structure to support computationally demanding perceptual acts and control activities (Lisbboa, 2011).

Artificial neural networks were developed as generalized mathematical models to represent the biological nervous system (Lisbboa, 2011). The ANN is trained to detect a pattern

between the inputted data and the related output value from a dataset. After training the set, the ANN can be used to predict the result of a newly inputted data (Jha, 2011). In a generalized mathematical model, the processing elements of neural networks are referred to as nodes, and the connection weights are used to represent synapses, where the non-linear characteristics exhibited by neurons are given by a transfer function. The weighted sum of the input signal, processed by the transfer function, is used to compute the impulse of the neuron. To improve the learning capability, the weights are adjusted according to the learning algorithm chosen (Lisbboa, 2011).

There are various types of ANNs including feed-forward, recurrent neural network and probabilistic network. The ANN structure used in this thesis is referred to as feed forward, back propagation multi-layer perceptron (Yu, 2009).

2.7.2.1 DEVELOPMENT OF ANN MODEL

ANNs are typically constructed using layers of units, and thus this is commonly referred to as multilayer ANNs. The first layer contains the input units, and the last layer contains the output units. The other units in the model are referred to as hidden units, and are referred to as hidden layers. Commonly there are two functions which govern the behaviours of a unit within each layer of the ANN, known as the input function and the output/activation function (Jha, 2011).

The input function is given by eq. 2.3, where the input to a node is the weighted sum of output from nodes connected to it (Jha, 2011).

$$Net_i=\Sigma_j w_{ij}x_j+\mu_j \quad (2.11)$$

- Net_i= result of the net inputs
- w_{ij}= weights connecting neuron j
- x_j= output from unit j and
- μ_j =threshold from neuron i

An activation function is then applied to each unit. The activation function is a mathematical formula that comes in many forms, and it has the largest impact on the ANN. The activation function is used to produce the output of a processing node. Activation functions typically use a non-linear function such as logistics, tanh etc... (Jha, 2011).

The ANN may be composed of various numbers of neurons, typically consisting one or two layer network, but may consist of more (Zupan, 1994). Figure 2.2 shows the difference between a one-layer and two-layer network, where the circles represent the neurons, and the lines connecting the circles represent the weights. A one layer network, has 4 neurons with each having 4 weights, thus totalling 16 weights. Each of the 4 neurons will accept an input signal and the signal coming from the bias. A two layer ANN has six neurons (2 in the first layer and 4 in the second layer), totalling 20 weights ((4x2)+ (3X4)=20).

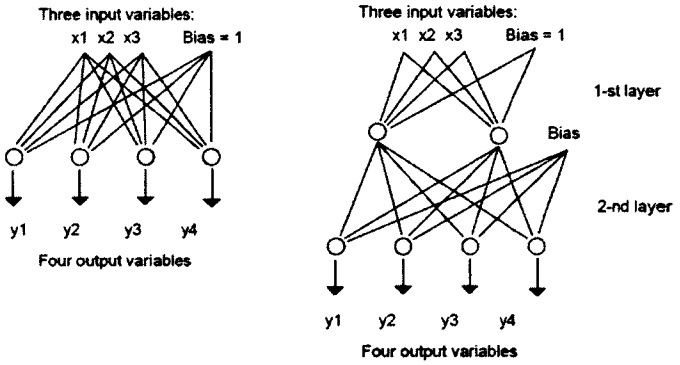

Figure 2.2 Left: One layer MLP Network, Right: Two Layer MLP Network

There are a few steps to be considered for developing a neural network as follows:

2.7.2.1.1 VARIABLE SELECTION

The input variables for modelling must be selected using an appropriate variable selection procedure (Jha, 2011). The input signals in this case are normalized between -1 and 1 (Rybchynski, 2005).

2.7.2.1.2 TRAINING, TESTING AND VALIDATION SETS

The dataset is divided into 3 distinct sets: training, testing and validation sets. The training set is used to learn patterns from the dataset; this is typically the largest set. The testing set is used to evaluate the generalized ability of the trained network. The validation set is used to check the performance of the trained network (Jha, 2011).

2.7.2.2 NEURAL NETWORK ARCHITECTURE

The neural network architecture consists of hidden layers, hidden nodes and output nodes as follows:

Number of hidden layers – the hidden layer(s) allows the network to generalize. A neural network with one hidden layer and various hidden neurons is sufficient for evaluating any continuous function. Ideally two hidden layers are used (Jha, 2011).

Number of hidden nodes – there is no formula for selecting the number of hidden nodes, however as a general rule of thumb, for a typical three layer network with n inputs and m output neurons, the hidden layer will have sqrt ($n*m$) nodes as proposed by the geometric pyramid rule (Jha, 2011). Further, a large number of hidden nodes will allow for correct learning. If too few hidden nodes are chosen, then the network may not be able to distinguish a relationship in the dataset and this may lead to poor results. Since the number of hidden nodes affects the performance of the network, choosing the correct number of hidden nodes is important (Lisbboa, 2011).

Number of output nodes – better results will be produced for neural networks with multiple outputs, compared to a network with a single output (Jha, 2011).

2.7.2.2.1 CHOOSING INITIAL WEIGHTS

There is no formula for choosing the weights. Several starting weights may be attempted to improve the network results. Typically the learning algorithm may choose the steepest descent technique, which rolls downhill in the weight space until the first valley is reached, and then this becomes the initial starting point (Oracle Data Mining Concepts, 2005).

2.7.2.2.2 CHOOSING LEARNING RATE

As each weight is modified, the learning rate controls the size of the step that is taken in the weight space. If the learning rate is too large, then the local minimum may be overstepped, and this may result in oscillations and slow convergence to lower the error rate. Conversely, if the learning rate is too low, the number of iterations required may be too large, and result in slow performance (Lisbboa, 2011).

2.7.2.3 TYPES OF NETWORKS

Two of the most widely used ANN architectures are feed forward networks and feedback/recurrent networks. The Medical Information Research Group (MIRG) has developed the ANN Research framework using MLP feed forward network with back propagation.

2.7.3.2.1 FEED FORWARD ANN

Signal flows one way from the input to the output units in a strict feed-forward direction. Although data processing may extend over multiple layers, there is no feedback (Lisbboa, 2011).

2.7.3.2.2 FEEDBACK / RECURRENT NETWORKS

Signal may travel in both directions using feedback. Feedback networks are dynamic and thus the state is continuously changing until the equilibrium is met (Jha, 2011).

2.7.2.4 THE MULTI-LAYER PERCEPTRON (MLP)

The multi-layer perceptron is a powerful system that is capable of modelling complex relationships between variables. It is the most popular neural network type due to its clear architecture and comparable simple learning algorithm (Eom, Kim, & Zhang, 2008). Multilayer feed forward neural network using back propagation algorithm architecture consists of a layered feed forward neural network, where non-linear elements are arranged in successive layers and the information flows uni-directionally from the input to output layers through the hidden layers. An MLP with one hidden layer has the ability to learn to approximate any function with high accuracy, and thus it is commonly used when little is known between the input and targets.

The weights or interconnections between the ANNs have a strength, which can be adjusted by a learning algorithm. There are three main types of learning algorithms used in neural networks known as (i) supervised learning (network constructs a model based on a known output), (ii) unsupervised learning (network constructs a model based on an unknown output), and (iii) reinforced learning (Jha, 2011).

The architecture of the MLP network is constructed by using a two-step procedure as follows (Jha, 2011):

First, the total weighted input x_j is computed using the following formula (Jha, 2011):

Where

$$X_j=\Sigma y_i W_{ij} \quad (2.12)$$

- y_i – activity level of the jth unit in the previous layer
- W_{ij} – weight of the connection between the ith and jth layer

Then typically the sigmoid function is used to represent a unit in the output layer, which determines the activity (Jha, 2011):

$$y_j = [1 + e^{-x_j}]^{-1} \quad (2.13)$$

Once the output units are determined, the error 'E' is computed as follows (Jha, 2011):

$$E = \tfrac{1}{2} \Sigma (y_j - d_j)^2 \quad (2.14)$$

- y_j – activity level of the jth unit in the top layer
- d_j – desired output of the jth unit

2.7.2.5 BACK PROPAGATION ALGORITHM

The back propagation algorithm is used to train the MLP network. In order to minimize the error in its prediction on the training set, this algorithm uses the data to adjust the network weights and threshold.

The back propagation algorithm consists of the following four steps, to determine the error changes with respect to the output activity, total inputs received, previous connection weights and activity of a unit in the previous layer (Jha, 2011):

(1) The error derivative (EA) is used to compute how fast the error changes in the activity of an output unit (Jha, 2011):

$$EA_j = \frac{\partial E}{\partial y_j} = y_i - d_j \quad (2.15)$$

(2) The error changes with respect to the total input received by an output unit (EI) is obtained by multiplying rate at which the output of a unit changes as its total input is changed as follows (Jha, 2011):

$$EI_j = \frac{\partial E}{\partial X_j} = \frac{\partial E}{\partial y_j} x \frac{\partial y_j}{\partial X_j} = EA_j y_j (1 - y_j) \text{ (2.16)}$$

(3) The error changes with respect to the weight on the connection in the output layer (EW) is obtained multiplying by the activity level of the unit from which the connection originates as follows (Jha, 2011):

$$EW_{ij} = \frac{\partial E}{\partial W_{ij}} = \frac{\partial E}{\partial X_j} x \frac{\partial X_j}{\partial W_{ij}} = EI_j y_j \text{ (2.17)}$$

(4) The error changes with respect to the activity of a unit in the previous layer (EA) is obtained by adding all of the separate effects of the output unit and multiplying the result by the weight on the connection to that output unit (Jha, 2011):

$$EA_i = \frac{\partial E}{\partial y_j} = \sum_j \frac{\partial E}{\partial X_j} x \frac{\partial X_j}{\partial y_j} = \sum_j EI_j W_{ij} \text{ (2.18)}$$

The last step is critical, since it allows for back propagation in multilayer networks. Using steps (2) and (4), it is possible to convert EAs of one layer of units into EAs for the previous layer, where this process can be applied to as many previous layers as desired. Once the EA of a unit is computed, steps (2) and (3) can be used to compute the EWs on the incoming connections (Jha, 2011).

2.7.2.6 ANN ADVANTAGES AND DISADVANTAGES

There are several advantages to ANNs. ANNs can generalize data even with noise, and does not require prior knowledge of the domain, and thus it requires less statistical training to develop. Further, ANN has the ability to implicitly detect complex relationships between dependent and independent variables and is able to detect all potential interactions with the predictor variable. This approach eliminates the tedious need to program IF-THEN rules and the need to obtain direct input from clinical experts. ANNs are also considered to be parallel networks (since the function of each neuron in each layer may be calculated in parallel). This allows the processing speed to be increased significantly by using computers with multiples processors. ANNs also have the advantage of applying different training algorithms to develop models based on the type of data to be analyzed or type of output required (Gilchrist, 2012). ANN will also improve the results every time it is trained due to its dynamic structure (Berner, 2007).

There are a few disadvantages to using ANNs. Firstly, the input data must be normalized for the ANN, and therefore the data has to be first analyzed and processed before feeding it into the ANN model. The maximum and minimum values must also be known prior to normalization. This is challenging when working in a real-time environment since the data is not known in advance, thus it will need to be calculated upon arrival (Gilchrist, 2012). The ANN does not handle missing data well, and any missing data will need to be replaced with an imputed value. Further, the training process is often time consuming. Although it is possible to determine the most contributing input, the resulting formulas and weights used are commonly not interpretable.

CHAPTER 3 METHODOLOGY FOR PREDICTING NEONATAL MORTALITY AND PRETERM BIRTH IN TWIN PREGNANCIES

3.1 INTRODUCTION

This chapter reinstates the problem in this thesis work, provides a description of the perinatal databases used, and an overview of performance measures. This chapter also discusses the methodology for the development of a prediction model for neonatal mortality and preterm birth in twin pregnancies. A framework to build a Perinatal Decision Support System (PEDSS) is also presented.

The first objective was presented and analyzed using two methods. First the development of a prediction model using the concept of knowledge discovery and decision trees is discussed, followed by a method using the hybrid ANN approach. Next, a section on achieving the second objective is presented where the development of the Perinatal Decision Support System (PEDSS) is discussed. The development of PEDSS is divided into four parts, including the following sections: platform consideration, overview of the system composition, system integration, and knowledge maintenance.

3.2 RESTATING THE PROBLEM

Fundamentally, there are two main problems of interest in this thesis work for the prediction of neonatal mortality, and preterm births in twin pregnancies:

(i) Assess various data analysis methods to improve the prediction of neonatal mortality and preterm birth in twin pregnancies

(ii) Derive a conceptual framework for a Web-Based Content Management System for Perinatal Decision Support.

Previous research on predicting neonatal mortality was obtained using knowledge-based techniques. The DT model used only 3 attributes: lowest serum pH, lowest blood pressure, and lowest heart rate to produce good preliminary results where a sensitivity of 75% and specificity of 96% was achieved. However these results were obtained from a small database with a total of n=256 patient cases with 232 survivors and 24 neonatal death, thus an above than normal mortality rate of 9%.When examining the DT model on a larger dataset with a total of n=17,364 newborn cases including 16677 survivors and 687 neonatal death and a 4% mortality rate using the same attributes, the DT model did not meet the clinical expectation, and achieved a low mean test sensitivity of 56% and a mean test specificity of 97% (Gilchrist, 2012).

Previous research on PTB prediction models focused primarily on non-knowledge-based techniques, and in particular these models were only useful for predicting the risk of preterm labour in parous[5] women. Previous studies focused on artificial intelligence based techniques using the Artificial Neural Network (ANN), and later a newer hybrid approach was introduced where it utilized a decision tree-artificial neural network model (Yu, 2009). One of the earlier models derived, utilized a hybrid method where the network parameters were extracted from a classification based ANN, and then a risk-stratification ANN was applied to estimate preterm birth. This method used only parous cases and US specific variables. The test data was processed in the system in three passes: first pass to maximize sensitivity, second pass to maximize specificity, and the third pass used a decision-tree voting algorithm to classify ambiguous cases. This achieved a sensitivity of 64%, a specificity of 84% and an ROC of 0.7795. Though these

[5] **Parous**: A medical term for a woman who has previously given birth one or more times.

results are acceptable, the main drawback to this system is that it requires 48 variables to maintain this performance and this model is not applicable in the case of nulliparous[6] women pregnant with twins (Catley, 2007).

The newer model used a hybrid approach integrating decision trees, classification ANN with weight elimination and risk stratification. The hybrid classifier used a decision tree to eliminate variables with little impact on predicting the outcome of interest; the remaining variables were processed through an artificial neural network with weight elimination (ANN-we). This method achieved a sensitivity of 65.13%, a specificity of 84.07% and an AUC of 0.8195. This method used 19 variables and included parous cases. Though nulliparous cases were analysed separately, the prediction model did not meet clinical standards. Further, this model was not applicable to the cases of nulliparous women pregnant with twins.

Both the previous and newer approach focused on primarily non-knowledge-based systems, and thus knowledge-based systems were not considered. Yet many of the systems to date use knowledge-based approach with rules, guidelines or compiled knowledge which is commonly derived from medical literature (Miller, 2010). The main drawback to non-knowledge-based systems, and especially artificial neural networks, is that the results are derived from a 'black box' system, and thus, the reliability and accountability of these systems is a concern since the system cannot be justified as to why it works the way it does. Further, this type of system is best suited when supplied with a large amount of data with known results, and is typically focused on a single disease.

While the results obtained in both the older and newer models met the expectations set by their clinical advisor, there are a few enhancements to be performed: (1) explore methods using knowledge-based approach since it is more practical in an obstetrical environment, which allows

[6] **Nulliparous**: A medical term used to describe a woman who has never given birth to a viable, or live, infant.

the obstetrical outcome to be obtained at point of care, and since the results are derived from a rule based system, the outcome is intuitive and can be easily comprehended by the clinicians; (2) change the scope from predicting only preterm labour in parous women to include a high risk population of twin pregnancies and neonatal mortality in a heterogeneous population, and (3) derive a conceptual framework to build and implement the perinatal decision support system in a healthcare setting. These modifications should ideally maintain previous performance results or preferably exceed them.

3.3 ETHICAL CLEARANCE

The Carleton University Research Ethics approved this thesis work for the duration of September 2011- April 2013. The database used in this thesis work was the Niday (Niday Perinatal Database from the Perinatal Partnership Program of Eastern and Southeastern Ontario (PPPSEO)) (2006-2007) and the BORN (Better Outcomes Registry & Network) database. An ethical clearance was signed prior to accessing the database due to the sensitive nature of the data.

3.4 THE NIDAY DATABASE

The Niday Perinatal Database ("Niday") was created in 1997 under the direction of the Perinatal Partnership Program of Eastern and Southeastern Ontario (PPPSEO) and with collaboration of all hospitals across Eastern and Southeastern Ontario (Niday Perinatal Database Project, 2008). In January 1st, 2001, database systems which were originally housed within hospitals on standalone computers throughout Ontario were migrated into a web-based system

known as CritiCall. This initiative was largely funded by the Ministry of Health and Long Term Care (MOHLTC) (Niday Perinatal Database Project, 2008).

The Niday database is a multidimensional data collection system for perinatal care providers, decision makers, educators and researchers (Dunn et al., 2011). The information available in the Niday database may be used for program management, benchmarking, quality improvement initiatives, planning, evaluation and research (Niday Perinatal Database Project, 2008). In Ontario, this is referred to as the source of data in order to evaluate outcomes, risk factors and interventions associated with perinatal care (Dunn et al., 2011).

This database has significantly evolved over time and currently serves as a joint venture with over 100 healthcare organizations in Ontario. Each of the participating organizations contributes real-time perinatal data to this web-based system (Dunn et al., 2011). The participation of the hospitals in Ontario allows for inter-hospital comparisons. This is required in order benchmark and to facilitate change based on learning from partner hospital's successes (Dunn et al., 2011).

In order to ensure the consistency of each of the variables found within the Niday database, a user guide is available which includes the variable definition and other information. Furthermore, each of the participating organizations undergo training in order to handle data entry and produce reports from the system (Niday Perinatal Database Project, 2008). The Niday database contains 90 defined patient elements. Of the total variables in the Niday database, 24 are mandatory and 66 are non-mandatory variables (Dunn et al., 2011). The province has adopted the variables in the Niday database as the minimum dataset in 2001 (Dunn et al., 2011).

3.5 THE BORN DATABASE

The BORN (Better Outcomes Registry & Network) was created on January 25th 2010 under the direction of the Ministry of Health and Long Term Care (Born Ontario, 2010). The BORN database holds the most extensive maternal-child health data in the world including information about pregnancy, birth and early childhood. All of the birthing hospitals within Ontario have committed to contributing data to the BORN registry (The Mothers Program, 2013).

The BORN's vision is to obtain the knowledge required to sustain lifelong health, and its mission is to facilitate and improve the care provided to maternal-fetal and child populations via linking the information and providers to recognize gaps in the care. The BORN database contains accurate, trusted and timely information in order to contribute to the high-performing healthcare system (The Mothers Program, 2013).

The data obtained in the BORN registry is mainly focused on maternal newborn outcomes/midwifery, congenital anomalies surveillance, newborn screening and prenatal screening. The data gathered from the BORN database may be used by health care providers and policy makers in order to improve the delivery of maternal-child care. The BORN data can also be used for research purposes, for tracking province-wide health trends, health effects and outcomes from early pregnancy and onwards (The Mothers Program, 2013).

The BORN database was streamlined from five existing maternal-fetal databases (The Mothers Program, 2013). To ensure consistency of each of the variables available in the BORN database, there is a guideline available which contains the variable and value types. The value(s) are then converted into a numeric code. The BORN database contains over 290 variables (BORN Ontario, 2013).

3.6 FIRST OBJECTIVE

Objective 1: Improve the Classification of Neonatal Mortality and Preterm Birth in Twin Pregnancies

The first objective was to improve the classification of neonatal mortality and preterm birth in twin pregnancies. Then two classification methods were used: decision trees (DT) and hybrid ANN. First the datasets were split into 10 equal sized sets using repeated 2-fold cross validation (section 3.6.1.4), then the C5.0 algorithm was run against the split dataset using the See5 application, and DTs were generated. The outputted results were in the form of a confusion matrix and included the attribute usage. Based on the attribute usage, the lowest contributing attribute was removed from the dataset to remove noise, and the C5.0 algorithm was re-run to determine whether the performance outcome improved. If the performance outcome didn't improve, the attribute was re-inserted, and the next lowest contributing attribute was removed and the process described above was repeated on a new dataset, until the optimal results were achieved.

The second method utilized a DT-ANN hybrid approach. The attributes that are best contributing (section 3.6.2.1) in a given dataset were extracted from the DT and was used to generate a table (section 3.6.2.4). Since the ANN cannot handle missing values, we used a Case Based Reasoning (CBR) tool developed by C. Cotea and S.Jiwani to fill in missing information (Cotea & Jiwani, 2003). The results of the CBR tool were stored in a .txt file. The .txt file was converted to a .csv file and repeated 2-fold cross validation was applied to split the dataset into 10 equal sized sets. The training, verification and test sets were created in preparation for running the Artificial Neural Network (ANN) (section 3.6.2.7). The ANN was run using

MATLAB (section 3.6.2.8), and the results were in the form of a confusion matrix. Results of both the DT and DT-ANN hybrid compared using various performance measures.

3.6.1 DEVELOPMENT OF A CLASSIFICATION MODEL USING THE CONCEPT OF KNOWLEDGE DISCOVERY AND DTs

Accomplishing Objective 1 involved the following to predict neonatal mortality and preterm birth in twin pregnancies using Decision Trees (DTs):

Step 1: Partitioning the Original Dataset

Step 2: Data Pre-Processing

- Handle Missing Values
- Feature Selection & Importance
- Grouping Nominal Attributes

Step 3: Preparing the Dataset

Step 4: Training and Testing the Dataset

Step 5: Feature Selection

Step 6: Software Execution and Results

Step 7: Decision Tree Ensemble Classifiers

3.6.1.1 STEP 1: PARTITIONING THE ORIGINAL DATASET

The datasets were divided into noMOD and MOD sets. The noMOD dataset(s) contained raw data, and nominal attribute(s) were not grouped. In the MOD dataset(s), all nominal attribute(s) were grouped to prevent spurious splits (section 3.6.1.2.3). The Niday database was divided into noMOD_NEONATALMORT and MOD_NEONATALMORT to predict neonatal mortality. The BORN database was divided into noMOD_TWINGEST and MOD_TWINGEST to predict preterm birth in twin pregnancies, followed by noMOD_SINGLETON and MOD_SINGLETON to predict preterm birth in singletons (Table 3.1). For both the prediction of

preterm birth in twin pregnancies and singletons, each dataset was further split for parous and nulliparous cases, where each split dataset were labelled with an extension of _parous and _nulliparous, respectively.

Table 3.1: Division of BORN and Niday database. (A) Dataset; (B) Parent Database; (C) Positive Cases; (D) Negative Cases; (E) # Grouped Attributes; (F) Total Cases

(A)	(B)	(C)	(D)	(E)	(F)
noMOD_NEONATALMORT	Niday	215	32 971	0	33186
MOD_NEONATALMORT	Niday	215	32 971	1	33186
noMOD_TWINGEST	BORN	323	1 591	0	1 914
MOD_TWINGEST	BORN	323	1 591	5	1 914
noMOD_SINGLETON	BORN	15 575	125 000	0	140 575
MOD_SINGLETON	BORN	15 575	125 000	4	140 575

The BORN database did not contain neonatal death information; therefore it was not used for predicting neonatal death. Although the Niday database did contain preterm birth information and twin gestations, this database lacked maternal health information and thus it was not used for the prediction of preterm birth in twin pregnancies.

3.6.1.2 STEP 2: DATA PRE-PROCESSING

Data pre-processing is the first step in knowledge discovery, and it is typically executed before the creation of a pattern classification model (Yu, 2009). This step is used to clean the raw data. Different error types may exist in a database such as inconsistent data, missing values, outliers etc...and thus well-defined measures must be taken to correct these errors. This may be in the form of eliminating irrelevant cases/attributes and/or imputing data as presented in section 3.6.1.2.1 (Yu, 2009).

3.6.1.2.1 HANDLING MISSING VALUES

In the presences of missing data, the C5.0 algorithm is quite robust. A value that is missing or unknown was replaced with '?' and a value that is not applicable to that particular

case was be denoted as 'N/A'. The C5.0 algorithm takes the weighted average of the predictions across all branches upon encountering missing data.

However, in the event of a large number of values missing per case or per attribute in the Niday or BORN database, for reasons such as human/technical error or participants not wishing to answer questions, the following techniques were applied to this thesis work in order to handle missing information (Yu, 2009).

Eliminate Cases: Discrete cases with more than 50% of variables missing may be deleted. The outcome of interest for neonatal mortality is neonatal death or not applicable, for preterm birth, it is term or preterm; if the outcome is missing, the case may be deleted.

Eliminate Attributes: If more than 50% of data is missing for a particular attribute and it is not a highly contributing attribute as per attribute usage outcome (see section 3.7.1.5), then the attribute itself may be deleted.

Imputation of Data: If data is missing because the information does not apply to that particular variable, then it may be corrected. For example, if the question was "Did you previously have caesarian?" and the answer was 'No', and another question was "How many caesarians have you previously had?" and the answer was blank, this may be corrected to '0'(Yu, 2009).

3.6.1.2.2 FEATURE SELECTION AND IMPORTANCE

The preliminary task with the given dataset was to determine which attribute was likely to be predictive. Firstly any attribute that was not be available prior to 10 minutes after birth or could not be obtained non-invasively in the Niday database was eliminated. Similarly, any attribute that was not be available prior to 22 weeks gestation or could not be obtained non-invasively in the BORN database was eliminated.

The noMOD_NEONATALMORT and MOD_NEONATALMORT datasets originally contained 26 attributes from the Niday database which could be obtained before 10 minutes after birth non-invasively for the prediction of neonatal mortality. The list of attributes can be seen below in Table 3.2.

Table 3.2 List of attributes used for noMOD_NEONATALMORT and MOD_NEONATALMORT datasets

1	Baby Birth Date	8	Gestation	15	Birth Weight	22	Reproductive Assistance
2	Mothers Age	9	Labour Type	16	AGPAR1	23	Aboriginal Status
3	Baby's Sex	10	Presentation	17	AGPAR5	24	First Trimester Visit
4	No. Previous Preterm Babies	11	Delivery Type	18	Intention to Breast Feed	25	AGPAR 10
5	No. Previous Term Babies	12	Previous C/S	19	Mother's Primary Language	26	Neonatal Death/Stillbirth
6	Twins Gestation	13	Forceps/Vacuum	20	Smoking		
7	Prenatal Classes	14	Previous C/S Number	21	Maternal Province		

The noMOD_TWINGEST and MOD_TWINGEST datasets originally contained 20 attributes from the BORN database which could be obtained before 22 weeks gestation non-invasively for the prediction of preterm birth in twin pregnancies. The list of attributes can be seen below in Table 3.3

Table 3.3 List of attributes used for noMOD_TWINGEST and MOD_TWINGEST datasets

1	MATAGE	-	Maternal Age	11	MATHP4	-	Diabetes Insulin Dependent
2	PPRETERM	-	Previous Preterm	12	MATHP18	-	HIV
3	PTERM	-	Previous Term	13	MATHP27	-	Lupus
4	PARITY	-	Parity	14	REPASS	-	Reproductive Assistance
5	PREVCS	-	Previous Caesarian	15	FIRSTVIS	-	First Trimester Visit
6	GENDER	-	Baby Gender	16	PRENCLAS	-	Prenatal Classes
7	INTBF	-	Intention to Breastfeed	17	LANGUAGE_up	-	Mother's Language
8	SMOKING	-	Smoking	18	MATHP_sub	-	Substance Abuse
9	MATHP0	-	None	19	MATHP_ment	-	Mental Health
10	MATHP3	-	Chronic Hypertension	20	Diagnosis	-	Preterm/Term

The noMOD_SINGLETON and MOD_SINGLETON datasets originally contained 20 attributes from the BORN database which could be obtained before 22 weeks gestation non-invasively for the prediction of preterm birth in singleton pregnancies. The list of attributes can be seen in Table 3.4.

Table 3.4 List of attributes used for noMOD_SINGLETON and MOD_SINGLETON datasets

1	MATAGE	-	Maternal Age	11	MATHP4	-	Diabetes Insulin Dependent
2	PPRETERM	-	Previous Preterm	12	MATHP18	-	HIV
3	PTERM	-	Previous Term	13	MATHP27	-	Lupus
4	PARITY	-	Parity	14	REPASS	-	Reproductive Assistance
5	PREVCS	-	Previous Caesarian	15	FIRSTVIS	-	First Trimester Visit
6	GENDER	-	Baby Gender	16	PRENCLAS	-	Prenatal Classes
7	INTBF	-	Intention to Breastfeed	17	LANGUAGE_up	-	Mother's Language
8	SMOKING	-	Smoking	18	MATHP_sub	-	Substance Abuse
9	MATHP0	-	None	19	MATHP_ment	-	Mental Health
10	MATHP3	-	Chronic Hypertension	20	Diagnosis	-	Preterm/Term

There are advantages and disadvantages of having too many or too few input attributes. Although having many attributes may lead to an improved classifier, this may also increase the complexity of the dataset and thus lead to added computational costs and noise. Further, having too many attributes may lead to spurious splits. These spurious splits will have almost no impact when evaluating the decision tree against a new dataset. Having fewer input attributes reduces the complexity of the database and improves the generalization of classifier; however it is important to not decrease the performance of the classifier. Therefore an ideal feature set must include only variables that are important (Yu, 2009).

3.6.1.2.3 GROUPING NOMINAL ATTRIBUTES

In order to achieve better results using decision trees, nominal inputs with many values was grouped. This was done to avoid spurious splits in the decision tree due to too many values. In the case of an attribute with integer values such as previous count of a condition (i.e previous # term birth, previous # preterm birth, previous # C/S), the values were grouped into 0, 1 and >2,

or >3, or >4 etc.., and steps 3-5 were executed to determine which grouping improved the performance outcome. After executing steps 3-5, it was found that grouping attribute(s) with integer values into 0, 1, 2 and >3, produced the optimized performance outcome. In extreme cases such as maternal age, where the input contains many values for each case, this was be grouped into high risk and low risk maternal age categories according to literature (i.e the world health organization considers women <18 and >40 to be high risk for preterm labour, therefore the maternal age was grouped into <18 (high risk), >=18 and <= 40 (low risk) and >40 (high risk)).

3.6.1.3 STEP 3: PREPARING THE DATASET

The c5.0 algorithm was executed using a commercially available application known as See5. The See5 application requires the dataset to be divided into 3 distinct file types as follows: names, data and test. The names file contains the attributes and values. The attributes may be divided into explicitly-defined attribute such as discrete, continuous, date, time etc. or implicitly-defined attribute which may be specified by a formula. The values may be specified by separating each identified value by a comma (Rulequest Research, 2012).

The data file includes the dataset for training the algorithm. The values of all explicitly-defined attributes are entered as a new row of data for each case. The data file may consist of one or more lines, where each value was separated by a comma, and the case entry was terminated with a period. A '?' may be used for missing values, and 'N/A' may be used when a value is not applicable for that particular case (Rulequest Research, 2012).

The test file is the same format as the data file. The test file contains unseen data for validation and was used to evaluate the classifier. (Rulequest Research, 2012).

3.6.1.4 STEP 4: REPEATED 2-FOLD CROSS VALIDATION

In order to reduce the bias on the dataset and randomize, a repeated 2-fold cross validation method was applied to the dataset using three bash scripts (create_5by2_folds.sh, duplicate_names.sh and run_dt_5by2.sh) developed by J.Gilchrist (Gilchrist, 2012). The advantage of the repeated 2-fold cross validation method is that all data points can be used for both training and testing the data; therefore important attributes that may lead to better estimations are not left out. In this method, the dataset was first split into positive and negative cases (figure 3.1- step1). This set was then further randomly split into equal sized training and test sets. The first positive and negative sets (figure 3.1- step 2 (set A)) were merged for training, and the second positive and negative sets (figure 3.1- step 2 (set B)) were merged for testing. This allows there to be equal number of positive and negative cases in the training and testing sets. The dataset in the merged train and test set was then reversed such that the test set was used for training and training set was used for testing (figure 3.1- step 3). This process was repeated 5 times to create 10 separate training and test sets (Gilchrist, 2012).

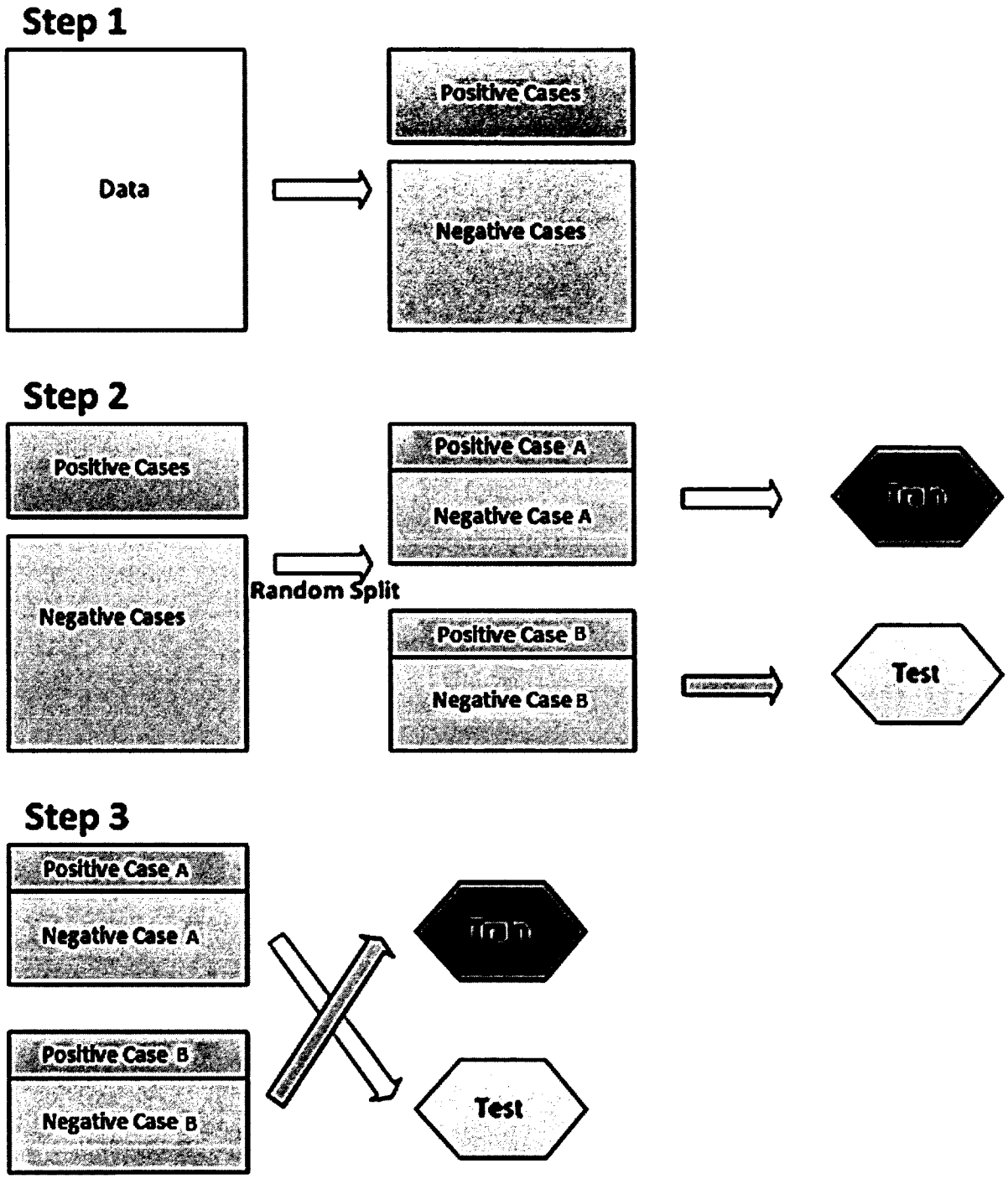

Figure 3.1 Steps to randomize a dataset using repeated 2-fold cross validation technique (Gilchrist, 2012).

3.6.1.5 STEP 5: SOFTWARE EXECUTION AND RESULTS

The results of the execution are stored in an .out text file. This file includes a list of important attributes to be used to construct the classifier, with the usage of the attributes represented by percentage and listed in order of highest importance. This file also includes ten decision trees and a confusion matrix. The confusion matrix provides additional details on the correct and incorrect classifications, and was used to evaluate various performance measures, see section 2.1 (Rulequest Research, 2012).

3.6.1.6 STEP 6: REMOVE VARIABLE OF LEAST RELEVANCE

Based on the results of the execution, variable(s) of poor usage as per the attribute usage list were removed and steps 3-5 were repeated to see if better performance outcomes could be achieved. If the overall prediction accuracy improved, these variable(s) were removed from the dataset, otherwise it was re-inserted.

3.6.2 DEVELOPMENT OF A CLASSIFICATION MODEL USING THE CONCEPT OF HYBRID ANN

The hybrid ANN architecture includes both knowledge-based components (decision tree algorithm) and non-knowledge-based components (Case Based Reasoning (CBR) tool and the ANN component). In a hybrid system, a decision tree algorithm (C5.0) was used to extract important attributes. A database was then created, and the ANN weights are adjusted according to the importance of the attributes. Then the CBR tool was run against the database for the purpose of case matching, case retrieval and reasoning. The ANN was used to train the network to learn general domain knowledge, and thus the trained network may be used as a source of general knowledge (Chen & Burrel, 2011).

One of the hybrid structures under consideration was the decision tree combined with a neural network known as the tree based neural network (TBNN). Although this method may improve the classification accuracy, it would be difficult to implement this on the medical information and technology research group's (MIRG) ANN Research Framework (ANN-RFW) since it has one hidden layer, and thus direct mapping of this algorithm would be complex and computationally expensive.

The hybrid algorithm used in this thesis work was pruning based neural networks (PBNN). This method integrates well into the ANN-RFW and uses a slightly different approach, where instead of directly mapping variables and rules to the ANN, decision trees are used to construct binary patterns which are then fed into the ANN (Yu, 2009).

Accomplishing Objective 1 to predict neonatal mortality and preterm birth in twin pregnancies using hybrid Artificial Neural Network (ANN):

Step 1: ANN Feature Selection
Step 2: Normalization / Standardization
Step 3: Incorporating a K-Nearest Neighbor (KNN) Case-Based Reasoning (CBR) Tool to Impute Data
Step 4: Database Format Requirements
Step 5: Execution of the KNN CBR Tool
Step 6: Configuring ANN
Step 7: Preparing the Dataset for ANN
Step 8: Running ANN & Result

3.6.2.1 STEP 1: ANN FEATURE SELECTION

In order to select variables of importance, section 3.6.1.1-6 were applied. The features that remained in the resulting DTs were then selected as the features to be used in the ANN classification model.

3.6.2.2 STEP 2: NORMALIZATION/STANDARDIZATION

In order to account for the large range differences that may occur between values, and to increase the efficiency of the ANN, the data inputs were normalized (Durai & Saro, 2006). Normalization is used to minimize the bias of one feature over another. Studies suggest that neural networks perform best when the values range between -1 and 1. Further, if the values were not normalized, the ANN may place importance on higher values (Durai & Saro, 2006). Normalization may speed up the training process by initiating the training process of each feature within the same range of values (Jayalakshmi & Dr. Santhakumaran, 2011). Various techniques can be used to normalize the data. In this study, for variables which were in two categorical format, the most frequent cases were set to -1, and for nominal cases the method used for

normalizing the data was derived from the MIRG ANN Guide, which suggests a modified Z-score transformation equation as follows (Rybchynski, 2005).:

$$x_i^n = \frac{x_i^n - \mu_i}{3\sigma_i} \quad (3.1)$$

where: x_i^n = Normalized value

x_i^n= Value

μ_i= Average

σ_i= Standard deviation

3.6.2.3 STEP 3: INCORPORATING A K-NEAREST NEIGHBOR (KNN) CASE-BASED REASONING (CBR) TOOL TO IMPUTE DATA

The MIRG team had developed and validated a K-Nearest Neighbor (KNN) case-based reasoning (CBR) tool for a Neonatal Intensive Care Unit (NICU) (Cotea & Jiwani, 2003). The CBR system was used to fill in missing data values from the knowledge retained from similar previously encountered cases (Cotea & Jiwani, 2003). The CBR methodology includes the following 4 steps (Cotea & Jiwani, 2003).

1. **Retrieve** – Retrieve similar past case(s) and the result(s)
2. **Reuse** - Based on previous cases, adapt a solution for the current case
3. **Revise** - Adjust any differences between the current case and retrieved case
4. **Retain** – Store the solution in the current case and re-use for future problems

3.6.2.4 STEP 4: DATABASE FORMAT REQUIREMENTS

The CBR tool runs on a Microsoft Access based database. The database must include four tables labelled: match, query, weights and index. The descriptions of these tables are given below (Cotea & Jiwani, 2003):

Match – contains the portion of the dataset with known values

Query – contains the remainder of the dataset with each row missing one or more value(s)

Index – contains the minimum and maximum value for each attribute

Weights – contains the weight associated with each attribute

The database also requires that an attribute labeled 'Caselink' to be set as the last column consisting of unique identifiers for each case (Cotea & Jiwani, 2003).

3.6.2.5 STEP 5: EXECUTION OF KNN CBR TOOL

The CBR system was used to fill in missing values and validate inconsistent predicted outcomes. The KNN algorithm was used to retrieve the closest matches from the training sample for each missing value and to replace the missing value by accounting for the weights assigned to each input variable (Catley, 2007). This application was developed in Java 1.4.1 by C.Cotea and S.Jiwani, and the connection to the database was made available using Java Database Connectivity (JDC) interface. The Graphical User Interace (GUI) was designed using the Swing API (Cotea & Jiwani, 2003).

The following steps were taken to run the KNN CBR Tool:

Step 1: Connected to the database via open database connectivity (ODBC) by running the following command C:\CBRS\run_cbrs.bat;
Step 2: Entered 'CBRS' for data source name
Step 3: '# of Primary Keys' was set to 1

Step 4: Clicked on 'Set Connection' button, if it worked properly the data would appear under "Patient ID"
Step 5: Set the 'k' value to the desired number of matches (default 10) in the '# of matches' field.
Step 6: Set the weights (ranging from 0-100) based on the results of the attribute usage list (section 3.6.1.6) in the 'Weights' table.
Step 7: Clicked on the 'Build Index' button.
Step 8: Clicked on the 'Match' Button.

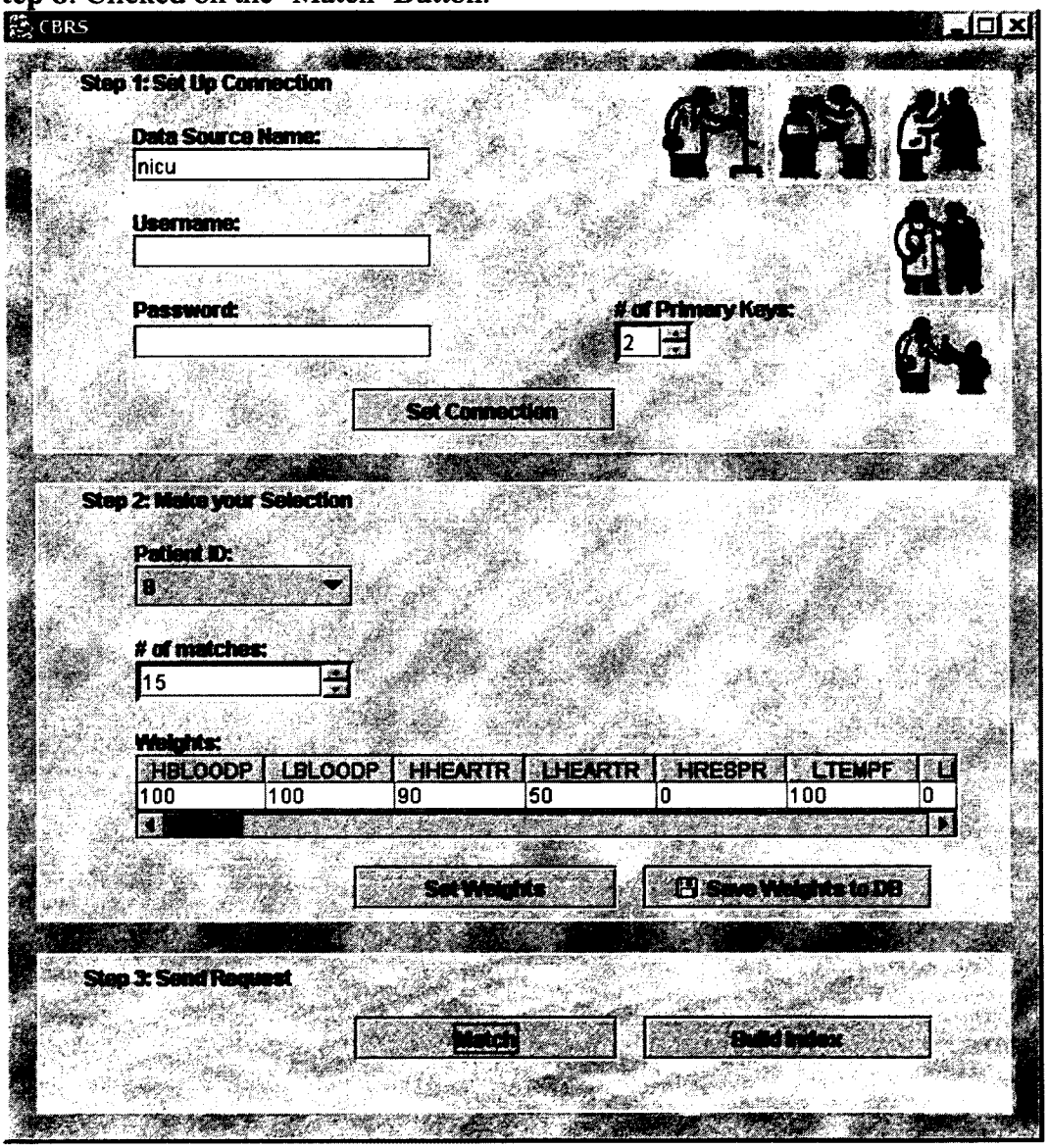

Figure 3.2 Screenshot of the CBR Interface

In a large dataset, this tool may require up to 3 days to process the information on a standard computer. The results of this tool are stored in a text (Filled_results.txt) file.

3.6.2.6 STEP 6: CONFIGURING ANN

The 'ANN_Research_Framework_1.3.doc' includes steps on setting up the MATLAB files. The following files were edited.

- main_autoANN_Sequential.m
- getData.m

3.6.2.7 STEP 7: PREPARING THE DATASET FOR ANN

The resulting dataset from the CBR tool was imported in Excel and split into positive and negative cases. A text file was then created for each case. Next, to randomize dataset, a repeated 2-fold cross validation method was applied using a bash script written by J.Gilchrist (section 3.6.1.4) (Gilchrist, 2012). The resulting .data and .test files from the script execution were then copied to a separate directory. Next, to create the train, test and validation files, two additional bash scripts (create_folds_ANN.sh and create_ANN_folders.sh) written by J. Gilchrist were used. These scripts produced 10 datasets, where each dataset was categorized into three distinct files including: training, testing and validation. The train and test sets are used to generate the structure. The validation set containing unseen data was used for confirming the model. The validation set contains unseen data. If the classifier generalizes well on new cases using the validation set, then this indicates a satisfactory performance. However, if poor performance was observed, then this may indicate that the network has been over-trained (Yu, 2009).

3.6.2.8 STEP 8: RUNNING ANN & RESULTS

To run the ANN, the following steps were executed:

1. Load MATLAB 2010a
2. Double click on the folder to process
3. Run ANN code in MATLAB by dragging the file "main_autoANN_Sequential.m" to the command window.

The results are stored in a log file and the performance outcomes were calculated.

3.7 SECOND OBJECTIVE

The second objective of the thesis was to construct a framework for a Perinatal Decision Support System (PEDSS) using a web-based content management system. Once the fundamental components were established, the integration and extension onto a web-based collaborative platform was considered. A detailed description of a conceptual framework to build a PEDSS using a knowledge-based approach is to follow.

3.7.1 SYSTEM ARCHITECTURAL CONSIDERATION

A PEDSS may be deployed onto varying platforms such as internet-based, local computer, networked with EMR or a handheld device. Further, varying techniques may be used as follows: (i) system may be integrated into an EMR, (ii) based on knowledge extracted from a central repository, or (iii) the entire system may be housed externally and accessed remotely, and thus not integrated to an EMR. Although the PEDSS may be built using any of the above described computational architecture, the final design will be based upon the clinical systems in place already, workflow, security, budget etc... (Berner, 2009).

For the purpose of this thesis, our PEDSS framework focuses on a web-based, mobile supported platform, where the entire system is housed externally and is accessed remotely. The PEDSS may be accessed via the internet by authenticated users on any computer or hand held device.

3.7.2 DEVELOPMENT OF PERINATAL DECISION SUPPORT SYSTEM

The web-based PEDSS introduced in this thesis is to serve as diagnostic tool to predict neonatal mortality and preterm labour in twin pregnancies based on patient's signs and symptoms. Ideally, a web-based collaborative system in a clinical environment would allow healthcare personnel to share opinions, exchange clinical data, and access clinical information regardless of their geographic location. Furthermore, the development of a high quality, low cost PEDSS to augment human decision-making on a collaborative platform will provide intelligently filtered information to the clinicians, staff and patients. This in turn will improve the quality of care provided and may reduce human errors.

The PEDSS was to include the following: a large knowledge-base of perinatal information extracted from the BORN database, a workflow to combine the knowledge with patient specific information, a communication mechanism to allow the clinicians to enter in patient information, and an interface to provide relevant results back to the clinician (Berner, 2009).

3.7.2.1 STEP 1: PLATFORM CONSIDERATION

The PEDSS shall be implemented on a user friendly, low cost platform that is mobile supported. The platform must adhere to confidentiality and privacy requirements as per national standards. An ideal platform shall support and meet the needs of two distinct audiences, which include healthcare providers, and the MIRG research group. The healthcare providers' require a system that is user-friendly, minimally disrupts their normal workflow and produces clinically satisfactory results (Catley, 2007). Importantly, healthcare providers will not use a system that does not guarantee protection and security of the data that is captured and recorded (Chakravarti, & Battacharyya, 2006). The MIRG research group will be more focused on a flexible framework

design, with the ability to access and customize each system component to continuously improve the system performance and reliability over time.

3.7.2.2 STEP 2: DEFINING SYSTEM COMPOSITION

3.7.2.2.1 SYSTEM OVERVIEW

The proposed PEDSS framework employs a web-based modular design that allows for extensibility and scalability; it is designed for the healthcare industry where retrieval and processing of real-time patient information is crucial. The solution architecture of this web-based system permits for 'anytime and anywhere' access utilizing an asynchronous data-driven design to allow for real-time information flow. Authenticated users may access the PEDSS from a laptop/desktop or a hand held device such as a mobile phone, PDA etc. This framework can be viewed as a 6-layer encapsulated system as presented in Figure 3.3.

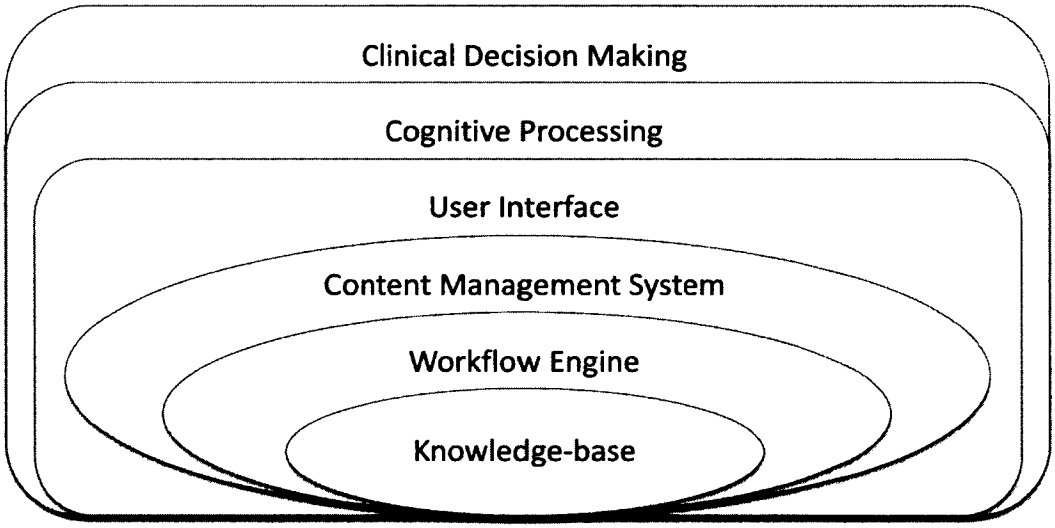

Figure 3.3: Integration of Technical and Clinical Processes (Frize, Yang, Walker, & O'Conner, 2005).

3.7.2.2.2 SYSTEM HARDWARE / SOFTWARE / COMMUNICATION ARCHITECTURE

The web-based PEDSS architecture is highly scalable and can easily be extended to many new users at low cost The PEDSS system architecture consists of the components as outlined in table 3.5 and figure 3.4.

Table 3.5 Components for a web-based PEDSS

Authentication Server	Authenticates users into the system
Content Management System	The heart of the system used to display, search, and process the data, based upon the user request
Workflow Engine	Required to automate alerts, warning and actions
External Data Source	A repository of the patient, or user information
Directories	A database of user information, etc.
Other Web Servers	Other servers required to operate the PEDSS
ASP.Net, XML, HTML	The interface presented to the user

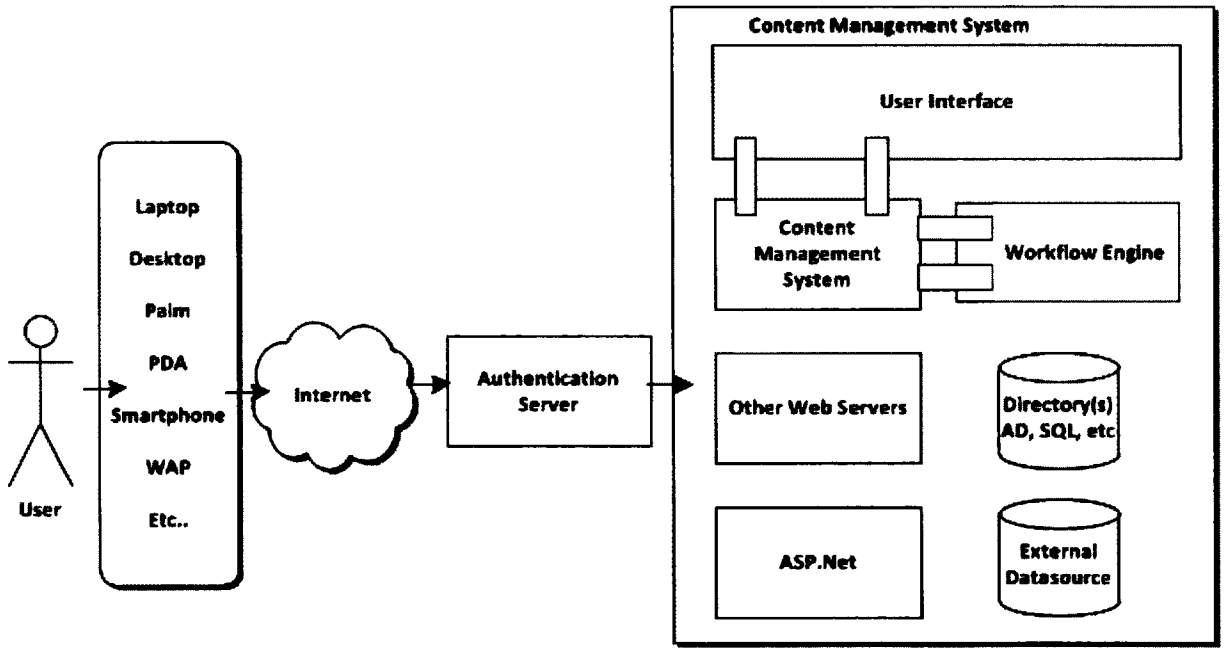

Figure 3.4 System Architecture of Web-based PEDSS

Details of the main system components are further discussed as follows:

- Authentication Server

The authentication server is used to verify that the user accessing the PEDSS has the appropriate privileges to view content on the page. The end user typically will need to provide a username and password; this information is checked to see if it matches the credentials stored in the system directory. This step is required to comply with the data security and confidentiality requirements to ensure that the transmission of data is in keeping with current privacy legislation. This is especially important for medical applications.

- Content Management System (CMS)

The content management system is considered to be the heart of the PEDSS and it includes all of the engines and support modules. A web-based content management system is required to host the PEDSS. The interface that is developed is published to the content management system. The CMS receives the inputted information from the clinician and auto starts the workflow engine for processing, the results are then displayed back to the clinician. The CMS is also composed of several databases which may be physically located over a large geographic area (Chakravarti, & Battacharyya, 2006).

The CMS supports advanced features including audience targeting, where filters can be set up based on the user logged in, so that varying levels of user-groups are presented with filtered views on what is relevant to them (Figure 3.5). The CMS also includes a search module which consists of an intelligent filter, thus only information applicable to the user will be displayed based on the search criteria.

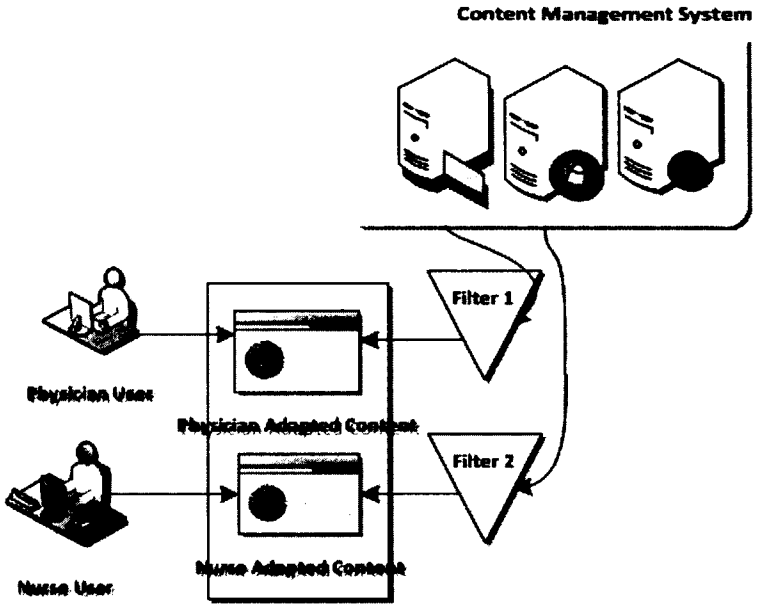

Figure 3.5 Custom Views based on Audience Targeting

- Workflow Engine

A workflow engine is a new class of software which allows the translation of code into a graphical model. This is required to automate the knowledge-base (Rasmussen et al., 2011). The workflow can incorporate various actions including sending out timed alerts, warnings and recommendations. The workflow engine is used to process the inputted patient data, as well as send alerts, diagnostic results, and/or recommendation to the clinician depending upon the requirements (Chakravarti, & Battacharyya, 2006). The workflow alerts and warnings are typically rule based and predetermined, based on the decision tree results obtained using the See5 program (C5.0). The recommendations are derived from evidence based medicine outcome, as well as information compiled from historical data stored in a clinical data warehouse (Chakravarti & Battacharyya, 2006). The workflow is customizable to the users' needs, as this will ultimately lead to increased efficiency, ease of use, and usefulness.

- External Data Source:

The external data sources will contain various SQL based databases, including a database containing the patient information extracted from an EMR or HER, and a database containing the users' information. A connection from the content management system to the external data source will be established for authenticated users to store and retrieve patient information.

3.7.2.3 STEP 3: SYSTEM INTEGRATION

The Perinatal Decision Support System (PEDSS) is composed of many discrete subsystems and components including the knowledge-base, content management system and workflows, which must be integrated physically and functionally in order to serve as a coordinated system. The entire PEDSS can be separated into three layers: Layer 1: Knowledge Discovery and Data Mining, Layer 2: Perinatal Decision Support System (PEDSS) and Layer 3: Knowledge Integration: Semantic Web Services for Healthcare. The relationship and linkages between each of these systems is shown in Figure 3.6. Each of the layers are further discussed as follows:

Layer 1: Knowledge Discovery and Data Mining

In the first layer, a decision tree was derived using the concept of knowledge discovery and data mining. The knowledge discovery process is an interactive and iterative process where most of the steps are manually processed by the user. For our framework, the data used for predicting neonatal mortality was extracted from the Niday perinatal database and the data for predicting preterm birth in twin pregnancies was extracted from the BORN database. The algorithm used to produce the DTs was C5.0, using the commercial See5 program (available in the MIRG laboratory). Each of the derived trees was evaluated using varying performance measures outlined in section 2.1. Once satisfactory results were obtained, the resulting decision

tree was presented to Dr.Bariciak. The practitioner's tacit domain knowledge defines the PEDSS composition template.

Layer 2: Clinical Decision Support System

The second layer is where the PEDSS was developed and deployed. The system is to be hosted on a web-based content management system with the architecture set up as outlined in section 3.7.2. A workflow engine shall be used to convert the decision tree results into a graphical model in the form of logical statements (i.e. IF 'Twins Gestation' AND 'Smoking THEN 'Preterm'). The workflow can also incorporate various actions including sending out timed alerts, warnings and recommendations based on requirements set in CDS composition template. This workflow shall also enable data connections to retrieve information / send captured data to SQL based databases via web services. Once the workflow model has been built, it is published to the content management system.

The graphical user interface is developed using asp.net. This interface is then browser enabled, such that the content may be viewed using only an internet browser, and thus preventing the need of users having to install additional add-ons or software to run the PEDSS. The developed .asp form was then published to the content management system.

Layer 3: Knowledge Integration: Semantic Web Services for Healthcare

The last layer ensures that authenticated users only view content relevant to them. Once the PEDSS is developed and deployed, the system shall be filtered based on the various user groups which exist within the system. There are 5 user groups under consideration: physicians, nurses, MIRG Researchers, system administrators, and visitors/auditors. The physicians shall have the ability to fill out assessments electronically, retrieve/view patient information and results, as well as have access to any relevant announcements. The nurses shall have the ability

to fill out assessments electronically, submit patient information, and have access to any relevant announcements. The MIRG research group shall have access to de-identified data, and view system components. The system administrator shall have full access to the system, including the ability to set roles, modify workflows and control the design privileges. Visitors/auditors shall have restricted read capability on the system. The system shall be set up and configured to present targeted views to the groups of users identified above, and thus only information relevant to the user shall be filtered and displayed.

Only authenticated users may access the system. The system shall include an authentication server in order to validate the user. The user would provide a username and password; this information is checked to see if it matches the credentials stored in the system directory. If the user's credentials match, then filtered content shall be displayed to the user based on their role stored in the system directory.

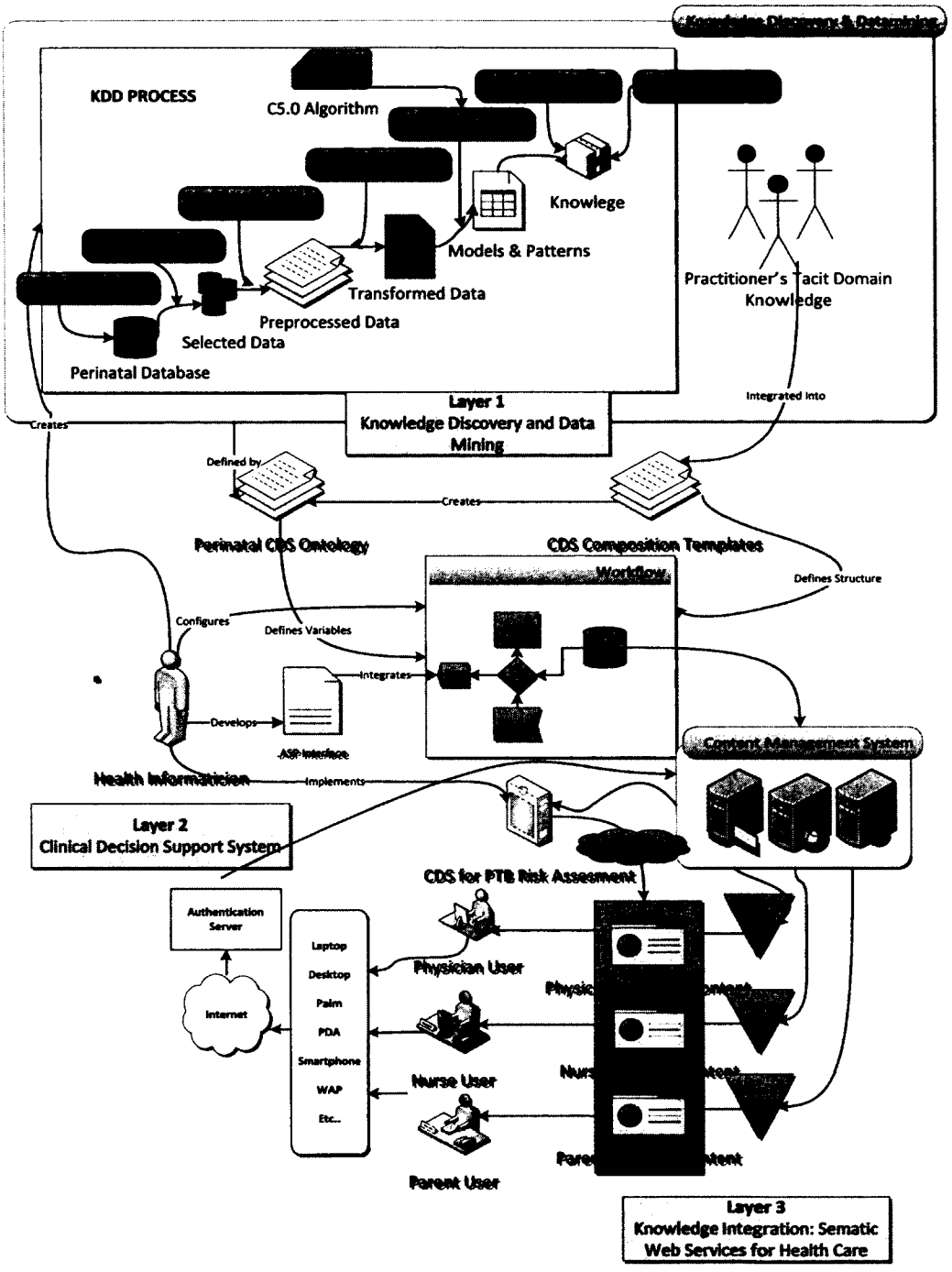

Figure 3.6 Conceptualization of the Perinatal Decision Support System

3.7.2.4 STEP 4: KNOWLEDGE MAINTENANCE

There are two main areas of knowledge maintenance that are challenging. The first challenge is to keep up to date the amount of the information within the perinatal database. As more information is recorded each year, it is important to update the system with the current information. As new medical devices emerge, if this affects the mortality rate on newborns or has an impact on delivering preterm for women pregnant with twins, then the new information must be captured, analyzed and incorporated into the system. For example, assisted pregnancy tools have resulted in twin pregnancies, where twin pregnancies often lead to preterm birth, and therefore this is a good indicator of preterm birth and must therefore be appropriately captured and recorded into the PEDSS (Berner, 2009).

Another other area of challenge that is partially related to the first problem, is maintaining the knowledge embedded within the PEDSS. As medical knowledge expands, it is important to take into account new drugs and diagnosis that are continually discovered. Therefore the PEDSS must be flexible and extensible in order to incorporate new updates without requiring significant resources and time (Berner, 2009).

There are a few potential solutions available for maintaining knowledge in a CMS. One solution would be an in-house knowledge management process; this however will require significant resources at each of the local sites. An alternate solution, and the solution developed for this thesis, was to develop a web-based repository using service oriented architecture (SOA). SOA allows for one central site to remotely maintain the knowledge. This approach is low cost, since only a few resources are required (Berner, 2009).

CHAPTER 4 RESULTS AND DISCUSSION

4.1 NEONATAL MORTALITY DECISION TREE & HYBRID ANN RESULTS

The Niday database was split into noMOD_NEONATALMORT and MOD_NEONATALMORT to predict neonatal mortality. First, a C5.0 based DT method was applied using repeated 2-fold cross validation method to generate ten test and train datasets (Section 3.6.1.4). The attributes used for each tree varied, since each DT will choose a different value to split on, and thus different folds will result in a variation of DTs based on the different subsets of the available features. Next, a hybrid ANN method was applied to the dataset (Section 3.6.1.2.1-8). The average test sensitivity and specificity exceeded 60% and 90%, respectively, for both the MOD_NEONATALMORT DT set and MOD_NEONATALMORT Hybrid ANN set.

4.1.1 NEONATALMORT DATASET

A total of 32 760 cases were evaluated using 13 attributes, where 32 695 cases were survivors and 65 resulted in neonatal death. The attributes used were Mother's Age, Baby Gender, Gestation, Labour, Delivery, Previous Cesarian, Birth Weight, AGPAR 1, AGPAR 5, AGPAR 10, Intention to Breast Feed, Smoking, and Diagnosis (for description please see Table 3.2). The prevalence of the dataset was 0.65%. This is very close to the prevalence of neonatal mortality rate of 0.40% in Canada in 2010 (UNICEF, 2010).

4.1.1 .1 noMOD_NEONATALMORT DT RESULTS

Out of the 13 input attributes, the attribute that was common in all datasets was AGPAR 5. The resulting test sensitivity was 57.17±6.84% and the test specificity was 99.98±0.01%, where during training it was slightly higher with sensitivity at 61.67% and specificity at. 99.99%. The highest performing decision tree was DT3(b) with test sensitivity of 66.67% and test specificity of 99.94%. The correct classification rate or accuracy remained high at 99.91%. The model produced excellent results to predict survival with a negative predictive value (NPV) of 99.92±0.01%, and produced satisfying results to predict neonatal death with a positive predictive value (PPV) of 87.50±8.09%. Overall the noMOD_NEONATALMORT DT model did not meet the clinical expectations set by Dr. Bariciak, since the average test sensitivity is below clinical standards of 60%.

Table 4.1: noMOD_NEONATALMORT decision tree results

Train	Sensitivity	0.7273	0.5938	0.6250	0.5313	0.5455
	Specificity	0.9998	0.9999	0.9999	0.9999	0.9999
	PPV	0.8571	0.9048	0.9524	0.9444	0.9474
	NPV	0.9994	0.9992	0.9993	0.9991	0.9991
Test	Sensitivity	0.6250	0.6061	0.5000	0.5758	0.5625
	Specificity	0.9996	0.9998	0.9998	0.9999	0.9999
	PPV	0.7692	0.8696	0.8421	0.9500	0.9474
	NPV	0.9993	0.9992	0.9990	0.9991	0.9991
Train	Sensitivity	0.6875	0.6061	0.5313	0.7879	0.5313
	Specificity	0.9998	0.9999	0.9999	0.9999	0.9999
	PPV	0.8800	0.9524	0.9444	0.9286	0.8947
	NPV	0.9994	0.9992	0.9991	0.9996	0.9991
Test	Sensitivity	0.6667	0.5313	0.5758	0.4375	0.6364
	Specificity	0.9994	0.9998	0.9999	0.9998	0.9998
	PPV	0.7097	0.8095	0.9500	0.8235	0.8750
	NPV	0.9993	0.9991	0.9991	0.9989	0.9993

4.1.1.2 MOD_NEONATALMORT DT RESULTS

In this set, attribute(s) with nominal values were grouped, which included Mother's Age. Out of the 13 input attributes, the attribute that was common in all datasets was AGPAR 5. The resulting test sensitivity was 62.24±3.28% and the test specificity was 99.95±0.03%, where during training it was slightly higher with sensitivity of 71.11% and specificity of. 99.96%. The highest performing decision tree was DT5(a) with test sensitivity of 68.75% and test specificity of 99.98%. The correct classification rate or accuracy remained high at 99.98%. The model produced excellent results to predict survival with a NPV of 99.93±0.01%, and produced satisfying results to predict neonatal death with a PPV of 72.35±13.91%. Overall this model exceeds the clinical expectations set by Dr. Bariciak, since the average test sensitivity and specificity exceeds the clinical standards of 60% and 90%, respectively.

Table 4.2: MOD_NEONATALMORT DT results

Train	Sensitivity	0.6970	0.5625	0.8929	0.8929	0.6061
	Specificity	0.9998	0.9998	0.9995	0.9995	0.9996
	PPV	0.8519	0.8182	0.7576	0.7576	0.7407
	NPV	0.9994	0.9991	0.9998	0.9998	0.9992
Test	Sensitivity	0.6250	0.6061	0.6250	0.6250	0.5938
	Specificity	0.9999	0.9997	0.9994	0.9994	0.9990
	PPV	0.9091	0.8000	0.6897	0.6897	0.5429
	NPV	0.9993	0.9992	0.9993	0.9993	0.9992
Train	Sensitivity	0.7188	0.6061	0.7500	0.6667	0.7188
	Specificity	0.9998	0.9994	0.9998	0.9998	0.9994
	PPV	0.8519	0.6897	0.8571	0.8462	0.7188
	NPV	0.9994	0.9992	0.9995	0.9993	0.9994
Test	Sensitivity	0.6563	0.5938	0.5758	0.6875	0.6364
	Specificity	0.9995	0.9991	0.9999	0.9998	0.9990
	PPV	0.7241	0.5758	0.9048	0.8462	0.5526
	NPV	0.9993	0.9992	0.9991	0.9994	0.9993

4.1.1.3 COMPARING BEST PERFORMING nOMOD_NEONATALMORT VS. MOD_NEONATALMORT DT

The best performing DT for noMOD_NEONATALMORT and MOD_NEONATALMORT prediction is presented in table 4.3. The best performing set for noMOD_NEONATALMORT DT showed higher sensitivity, specificity, PPV and NPV compared to the MOD_NEONATALMORT DT set. The highest increase was noted in the PPV value with an increase of 0.1365. The sensitivity, specificity and NPV values resulted in only slight increase of 0.0208, 0.0004, 0.0001 respectively, compared to the MOD_NEONATALMORT DT set. When evaluating the best performing DT set for noMOD_NEONATALMORT and MOD_NEONATALMORT, both sets verify well and are clinically acceptable.

Table 4.3: Best performing noMOD_NEONATALMORT vs MOD_NEONATALMORT DT set

Train	Sensitivity	0.6667	0.6875
	Specificity	0.9998	0.9998
	PPV	0.8462	0.8800
	NPV	0.9993	0.9994
Test	Sensitivity	0.6875	0.6667
	Specificity	0.9998	0.9994
	PPV	0.8462	0.7097
	NPV	0.9994	0.9993

Although the best performing noMOD_NEONATALMORT DT set exceeds the performance MOD_NEONATALMORT DT set, the average test sensitivity of the noMOD_NEONATALMORT DT, did not meet clinical expectations. The average test sensitivity of the noMOD_NEONATALMORT DT set was 57.17±6.84% which is below clinical standards, even though the specificity remained high at 99.98±0.01%. Thus the only clinically acceptable model is the MOD_NEONATALMORT DT set with an average test sensitivity of 62.24±3.28% and specificity of 99.95±0.03%.

4.1.1.4 COMPARING PPV AND NPV OF nOMOD_NEONATALMORT VS. MOD_NEONATALMORT

The average test PPV and NPV values for noMOD_NEONATALMORT and MOD_NEONATALMORT DT sets are presented in Table 4.4. The NPV results for both the noMOD_NEONATALMORT and MOD_NEONATALMORT DT were extremely high and stable exceeding 99.92%. The PPV values for both the noMOD_NEONATALMORT DT and MOD_NEONATALMORT DT set are both much higher than previous work of N.Yu of 0.34, where the MOD_NEONATALMORT DT set produced the highest average PPV of 0.85 (Yu, 2009). The train PPVs were higher than the test PPVs as expected. This could be attributed to fact that grouping nominal values, resulted in a simpler network allowing for better generalization of newer datasets. The smaller PPV in previous work indicates that many of the positive results from the previous prediction models were false positives. Therefore any positive results in the previous work would require a follow up in the form of a more reliable assessment to determine if the neonate will survive.

MOD_NEONATALMORT DT			
Train	PPV	0.7889	0.9206
	NPV	0.9994	0.9992
Test	PPV	0.7235	0.8546
	NPV	0.9993	0.9992

4.1.1.5 noMOD_NEONATALMORT HYBRID ANN RESULTS

The resulting average test sensitivity and test specificity for the hybrid ANN models were 21.85±12.97% and 86.55±10.83%, respectively. During training, sensitivity was slightly higher at 31.61% and specificity was lower at 87.90%. The highest performing node was Fold 5(b) with

test sensitivity of 22.63% and test specificity of 54.05%. The model produced excellent results to predict survival with a NPV of 89.84±0.51%, and produced poor results to predict neonatal death with a PPV of 18.14±5.72%. The ROC was 55.24% resulting in poor discrimination. Overall the noMOD_NEONATALMORT hybrid ANN model did not meet the clinical expectations set by Dr. Bariciak, since the average test sensitivity is below clinical standards of 60% and resulted in poor discrimination.

Table 4.5: noMOD_NEONATALMORT hybrid ANN results

Train	Sensitivity	0.2533	0.7445	0.1770	0.1891	0.1747
	Specificity	0.7690	0.2366	0.8977	0.8977	0.8790
	PPV	0.1267	0.1079	0.1868	0.1902	0.1583
	NPV	0.8861	0.8818	0.8914	0.8971	0.8910
Test	Sensitivity	0.0245	0.57356	0.1912	0.2288	0.1862
	Specificity	0.9039	0.54053	0.9020	0.9011	0.9001
	PPV	0.0251	0.13412	0.2011	0.2315	0.1947
	NPV	0.9018	0.91084	0.8963	0.8997	0.8951
Train	Sensitivity	0.1799	0.3745	0.6707	0.1716	0.2263
	Specificity	0.8885	0.6698	0.4127	0.8723	0.8406
	PPV	0.1691	0.1237	0.1310	0.1450	0.1560
	NPV	0.8958	0.8959	0.9047	0.8930	0.8930
Test	Sensitivity	0.1985	0.1875	0.2053	0.1934	0.1962
	Specificity	0.9008	0.9001	0.9016	0.9027	0.9021
	PPV	0.2064	0.2024	0.2071	0.2061	0.2058
	NPV	0.8964	0.8912	0.9007	0.8955	0.8967

4.1.1.6 MOD_NEONATALMORT HYBRID ANN RESULTS

The resulting test sensitivity was 60.74±0.59% and the test specificity was 91.30±0.65%, where during training sensitivity was slightly higher at 65.44% and specificity was lower at 80.607%. The highest performing node was Fold 2(b) with test sensitivity of 62.83% and test specificity of 84.99%. The correct classification rate or accuracy 83.33%. The model produced excellent results to predict survival with a NPV of 86.83±0.33%, and produced

good results to predict neonatal death with a PPV of 71.14±1.64%. The ROC was 77.09% resulting in acceptable discrimination. Overall the MOD_NEONATALMORT hybrid ANN model met the clinical expectations set by Dr. Bariciak, since the average test sensitivity and specificity is above clinical standards of 60% and 90%, respectively.

Table 4.6: MOD_NEONATALMORT Hybrid ANN result

Train	Sensitivity	0.6265	0.7166	0.5979	0.6283	0.5705
	Specificity	0.8438	0.7128	0.8719	0.8499	0.9137
	PPV	0.5936	0.4781	0.6337	0.6014	0.7070
	NPV	0.8612	0.8726	0.8540	0.8638	0.8535
Test	Sensitivity	0.6193	0.5960	0.6115	0.6058	0.6131
	Specificity	0.9103	0.9172	0.9028	0.9153	0.9215
	PPV	0.7046	0.7212	0.6843	0.7203	0.7314
	NPV	0.8738	0.8634	0.8708	0.8658	0.8722
Train	Sensitivity	0.6889	0.6734	0.7252	0.7387	0.5775
	Specificity	0.7538	0.7665	0.7068	0.6741	0.9139
	PPV	0.5060	0.5146	0.4736	0.4545	0.7093
	NPV	0.8687	0.8646	0.8761	0.8753	0.8561
Test	Sensitivity	0.6034	0.6068	0.6057	0.6069	0.6056
	Specificity	0.9004	0.9201	0.9123	0.9145	0.9156
	PPV	0.6835	0.7326	0.7042	0.7129	0.7185
	NPV	0.8643	0.8664	0.8704	0.8692	0.8671

4.1.1.7 COMPARING BEST PERFORMING noMOD_NEONATALMORT VS. MOD_NEONATALMORT HYBRID ANN

The best performing hybrid ANN for noMOD_NEONATALMORT and MOD_NEONATALMORT prediction is presented in Table 4.7. The best performing set for MOD_NEONATALMORT hybrid ANN showed higher sensitivity, specificity, and PPV compared to the noMOD_NEONATALMORT set. The highest increase was noted in the PPV value with an increase of 0.4454, whereas the sensitivity and specificity increased by of 0.4020 and 0.0093 respectively, compared to the noMOD_NEONATALMORT hybrid ANN set.

Table 4.7: Best performing noMOD_NEONATALMORT vs MOD_NEONATALMORT hybrid ANN sets

Train	Sensitivity	0.2263	0.6283	
	Specificity	0.8406	0.8499	
	PPV	0.1560	0.6014	
	NPV	0.8930	0.8638	
Test	Sensitivity	0.1962	0.6058	
	Specificity	0.9021	0.9153	
	PPV	0.2058	0.7203	
	NPV	0.8967	0.8658	

When evaluating the best performing hybrid ANN set for noMOD_NEONATALMORT and MOD_NEONATALMORT, the best performing noMOD_NEONATALMORT does not meet clinical expectations since both the sensitivity and specificity are low.

Further the average test sensitivity of the noMOD_NEONATALMORT set was 21.85±12.97% which is below clinical standards, even though the specificity remained high at 86.55±10.83%. Thus the only clinically acceptable model is the MOD_NEONATALMORT set with an average test sensitivity 60.74±0.59% and test specificity of 91.30±0.65%

4.1.1.8 COMPARING PPV AND NPV OF noMOD_NEONATALMORT VS. MOD_NEONATALMORT HYBRID ANN

The average test PPV and NPV values for noMOD_NEONATALMORT and MOD_NEONATALMORT hybrid ANN sets are presented in Table 4.8. The NPV results for both the noMOD_NEONATALMORT and MOD_NEONATALMORT were high and stable exceeding 86.83%. The PPV of 0.7114 in the MOD_NEONATALMORT hybrid ANN set was higher than previous work of N.Yu of 0.34; however the PPV of the noMOD_NEONATALMORT hybrid ANN set was much lower at 0.18 (Yu, 2009).

Table 4.8: Comparing PPV and NPV of noMOD_NEONATALMORT vs MOD_NEONATALMORT

Train	PPV	0.1495	0.5672
	NPV	0.8930	0.8646
Test	PPV	0.1814	0.7114
	NPV	0.8984	0.8683

4.1.1.9 PREDICTING NEONATAL MORTALITY DISCUSSION

The results of the pre-processed MOD_NEONATALMORT DT dataset with nominal values grouped using C5.0 algorithm exceeded the performance of DT-ANN hybrid method with a higher test sensitivity and test specificity of 62.24±3.28% and 99.95±0.03%, respectively. However both the DT and hybrid ANN MOD_NEONATALMORT met the clinical expectations set, with sensitivity and specificity exceeding 60% and 90%, respectively.

Further a high sensitivity and specificity, with a high PPV and NPV was achieved in both the DT and Hybrid ANN MOD_NEONATALMORT. The PPV refers to the likelihood that the positive prediction is correct and that the actual disease exists given a positive prediction (section 2.1.8). The NPV refers to the likelihood that the negative prediction is correct that the non-disease of the patient is correct given the negative prediction (section 2.1.9). These values are used as direct diagnosis accuracy of a classifier which serves as a comprehensive test for physicians (Yu, 2009). There are a few similar studies available in literature using AGPAR 5 as the main attribute, and although these studies often had a high sensitivity, these results were accompanied by a low specificity and in many cases a poor PPV.

To date, and to our knowledge, this is the first model to predict neonatal mortality non-invasively in a heterogeneous population within the first 10 minutes of birth, or later with clinically satisfactory results. This model was generated using a large dataset obtained from the Niday database where originally it contained 65 536 newborn cases, after removing 32 374 cases

due to data clean up procedures as outlined in section 3.6.1.2, there were 32 776 newborn cases remaining, with 32 695 survivors and 65 neonatal death. This is the largest dataset which produced a clinically acceptable model to date. This set also contains a more correct prevalence of 0.65%, in comparison with the neonatal mortality in Canada which is recorded to be 0.40% (UNICEF, 2010). All previous studies conducted consisted of a much smaller dataset, with artificially high prevalence. A dataset that is too small may lead to erroneous conclusions upon evaluating the reliability of the tree on new data. In addition, a dataset containing a higher than normal prevalence may bias the resulting decision tree and produce better performance results. However, the implications of the previous study results upon validating it on a larger dataset with normal prevalence are unclear.

Further, previous studies focused on newborns after being admitted to NICU, and attempted to predict the risk on neonatal mortality before the first 48 hours after admission. This model is the first to predict the risk of neonatal mortality within the first 10 minutes of birth in a heterogeneous population non-invasively using only 13 attributes, which resulted in clinically acceptable results. In order to not count cases where the neonatal death occurred already before 10 minutes, cases with AGPAR 10 score <1 were removed from the dataset.

The MOD_NEONMORT set on average consisted of a mean test sensitivity of 62.24$\pm$3.28% and 13 attributes in total. Further, the highest performing decision tree used only 5 attributes which included AGPAR 5, AGPAR 1, Birth Weight, Gestation, and Delivery, and had a high test sensitivity and a test specificity of 66.67% and 99.94%, respectively, with PPV and NPV exceeding clinical expectations of 0.863 and 0.998 respectively.

4.2 TWIN GESTATION DECISION TREE RESULTS

The BORN database was split into noMOD_TWINGEST DT and MOD_TWINGEST DT sets to predict preterm birth in twin gestation. A C5.0 based DT method was applied using repeated 2-fold cross validation method to generate ten test and train datasets (Section 3.6.1.4). The average test sensitivity and specificity exceeded 65% and 85%, respectively, for the MOD_TWINGEST DT set.

4.2.2 TWINGEST DATASET

A total of 1914 test cases were evaluated including 1591 term and 323 preterm births. A total of 20 attributes were used which included MATAGE, PPRETERM, PTERM, PARITY, PREVCS, GENDER, INTBF, SMOKING, MATHP0, MATHP3, MATHP4, MATHP18, MATHP27, REPASS, FIRSTVIS, PRENCLAS, LANGUAGE_up, MATHP_sub, and Diagnosis (for description please see Table 3.3). The prevalence of the dataset was 16.87%.

4.2.2.1 noMOD_TWINGEST DT RESULTS

Out of the 20 input attributes, the C5.0 algorithm used 1-4 attribute(s) to generate the DTs. The attribute that was in common included LANGUAGE, and it also remained to be the most contributing factor in predicting preterm birth in twin pregnancies. The resulting average test sensitivity was 15.43±3.01% and the average test specificity was 99.31±0.29%. The sensitivity and specificity during training was slightly higher at sensitivity of 15.85% and specificity of 99.37%, respectively. The correct classification rate or accuracy was 85.15±0.53%. The PPV and NPV were 81.94±6.50% and 85.26±0.44%, respectively. Due to the poor sensitivity, this prediction model does not meet the clinical standards set by Dr. Bariciak.

Table 4.9: noMOD_TWINGEST decision tree results

Train	Sensitivity	0.1790	0.1242	0.1667	0.1801	0.1296
	Specificity	0.9950	0.9925	0.9950	0.9925	0.9899
	PPV	0.8788	0.7692	0.8710	0.8286	0.7241
	NPV	0.8562	0.8484	0.8544	0.8567	0.8482
Test	Sensitivity	0.1242	0.1790	0.1988	0.1235	0.1739
	Specificity	0.9925	0.9950	0.9887	0.9950	0.9975
	PPV	0.7692	0.8788	0.7805	0.8333	0.9333
	NPV	0.8484	0.8562	0.8590	0.8480	0.8564
Train	Sensitivity	0.1739	0.1852	0.1429	0.1543	0.1491
	Specificity	0.9975	0.9937	0.9937	0.9962	0.9912
	PPV	0.9333	0.8571	0.8214	0.8929	0.7742
	NPV	0.8564	0.8570	0.8513	0.8527	0.8519
Test	Sensitivity	0.1296	0.1180	0.1914	0.1491	0.1543
	Specificity	0.9899	0.9937	0.9912	0.9912	0.9962
	PPV	0.7241	0.7917	0.8158	0.7742	0.8929
	NPV	0.8482	0.8476	0.8576	0.8519	0.8527

In order evaluate the effects on the prediction model for the case of parous vs. nulliparous women, the noMOD_TWINGEST DT was further split into noMOD_TWINGEST_parous and noMOD_TWINGEST_nulliparous, respectively. The sensitivity of the noMOD_TWINGEST_parous set was lower compared to the noMOD_TWINGEST set with an average test sensitivity of 9.46±6.75%, whereas the test sensitivity of the noMOD_TWINGEST_nulliparous increased to 25.56±2.65%, the test specificity values remained stable at 99.31±0.51% and 97.61±0.50%, respectively. However, due to the poor sensitivity, this prediction model does not meet the clinical standards set by Dr. Bariciak.

4.2.2.2 MOD_TWINGEST DT RESULTS

The attribute(s) with nominal values grouped were MATAGE, PPRETERM, PTERM and PARITY. Out of the 20 input attributes, the C5.0 algorithm used 5-13 attributes to generate the DTs. The most contributing factor in predicting preterm birth in twin pregnancies was

SMOKING. The resulting average test sensitivity was 79.32±5.85% and the average test specificity was 91.97±1.20%. The sensitivity and specificity during training was much higher at sensitivity of 91.02% and specificity of 95.12%. The correct classification rate or accuracy was 90.16±1.42%. The PPV and NPV were 66.85±2.51% and 95.66±1.17%, respectively. The large increase in sensitivity in the MOD_TWINGEST compared to the noMOD_TWINGEST could be attributed to the grouping method used in the MOD set. Grouping nominal values prevents spurious splits and thus over fitting, and therefore leads to simpler networks which able to generalize well on new datasets. The MOD_TWINGEST prediction model is clinically acceptable since the average test sensitivity and specificity exceeds 65% and 85%, respectively, as determined by Dr. Bariciak.

Table 4.10: MOD_TWINGEST decision tree results

Train	Sensitivity	0.9444	0.9068	0.8765	0.8696	0.9753
	Specificity	0.9611	0.9421	0.9623	0.9572	0.9284
	PPV	0.8315	0.7604	0.8256	0.8046	0.7349
	NPV	0.9884	0.9804	0.9746	0.9731	0.9946
Test	Sensitivity	0.7578	0.7840	0.7578	0.7654	0.9753
	Specificity	0.9119	0.9372	0.9321	0.9221	0.9284
	PPV	0.6354	0.7175	0.6932	0.6667	0.7349
	NPV	0.9490	0.9552	0.9500	0.9508	0.9946
Train	Sensitivity	0.8820	0.9321	0.8944	0.9198	0.9006
	Specificity	0.9421	0.9573	0.9497	0.9510	0.9610
	PPV	0.7553	0.8162	0.7826	0.7926	0.8239
	NPV	0.9753	0.9858	0.9780	0.9831	0.9795
Test	Sensitivity	0.7963	0.8447	0.7963	0.7081	0.7963
	Specificity	0.9158	0.9170	0.9095	0.9346	0.9171
	PPV	0.6582	0.6733	0.6418	0.6867	0.6615
	NPV	0.9567	0.9668	0.9564	0.9405	0.9567

To evaluate the effects on the prediction model for the case of parous vs. nulliparous women, the MOD_TWINGEST DT was further split into MOD_TWINGEST_parous and

MOD_TWINGEST_nulliparous, respectively. Both the sensitivity of the MOD_TWINGEST_parous set and MOD_TWINGEST_nulliparous were higher compared to the MOD_TWINGEST set with an average test sensitivity of 80.56±1.65% and 83.10±1.79% respectively. The test specificity values remained stable at 89.72±2.75% and 93.11±2.13%, respectively. Thus both prediction models are clinically acceptable.

4.2.2.3 COMPARING BEST PERFORMING nOMOD_TWINGEST VS. MOD_TWINGEST

The best performing DT for noMOD_TWINGEST and MOD_TWINGEST prediction is presented in table 4.7. The best performing DT classifier for the noMOD_TWINGEST set was case 2a and MOD_TWINGEST set was case 3a. The best performing set for MOD_TWINGEST showed higher sensitivity and NPV compared to the noMOD_TWINGEST set with an increase of 0.7267 and 0.1245 respectively, whereas the specificity and PPV values were slightly lower by 0.0893 and 0.1298, respectively. When evaluating the best performing sets overall in terms of sensitivity, specificity, PPV and NPV, the noMOD_TWINGEST set does not meet the clinical standards since the resulting sensitivity is much lower than the acceptable rate of 65%.

Table 4.11: Best performing noMOD_TWINGEST vs MOD_TWINGEST

Train	Sensitivity	0.1667	0.9753
	Specificity	0.9950	0.9284
	PPV	0.8710	0.7349
	NPV	0.8544	0.9946
Test	Sensitivity	0.1988	0.9255
	Specificity	0.9887	0.8994
	PPV	0.7805	0.6507
	NPV	0.8590	0.9835

Further based on the average test sensitivity and average test specificity, the noMOD_TWINGEST model did not meet clinical expectations. The average test sensitivity of the noMOD_TWINGEST set was 15.43±3.01% which is much lower than the clinical standards, while the average test specificity remained high at 99.31±0.29%. Thus the only clinically acceptable model is the MOD_TWINGEST set with an average test sensitivity and specificity of 79.32±5.85% and 91.97±1.20%, respectively.

4.2.2.4 COMPARING PPV AND NPV OF nOMOD_TWINGEST VS. MOD_TWINGEST

The average PPV and NPV for noMOD_TWINGEST and MOD_TWINGEST are presented in table 4.12. Although the PPV and NPV for the noMOD_TWINGEST case were high, further examination of the high PPV indicates that it is due the high test specificity (99.31±0.29%) accompanied by a poor test sensitivity (15.43±3.01%), therefore the PPV in this case is not reliable. Additionally, although the PPV for the MOD_TWINGEST case is low at 67.69±3.17%, this is still higher compared to the previous work of N.Yu which resulted in 43.93% (Yu, 2009). Further, upon analyzing the equation to calculate these predicative values reveals that these values are directly associated to prevalence, and therefore since both the sensitivity and specificity were high in the MOD_TWINGEST case and the prevalence was low, this will result in a low PPV. It is also important to note that a high PPV does not necessarily indicate an accurate prediction. This can be noted in many classifiers with a high PPV but low sensitivity, including the results obtained in the noMOD_TWINGEST case. Therefore, the PPV and NPV values alone should not be used to evaluate a classifier. Although these values can be used to obtain a good direct accuracy of a classifier, it may also be subject to a poor sensitivity value due to low prevalence. However the MOD_TWINGEST case shows consistency in the

NPVs and PPVs between each trial and is accompanied by a high sensitivity and specificity which exceed clinical expectations.

Table 4.12: Comparing PPV and NPV of noMOD_TWINGEST vs MOD_TWINGEST

Train	PPV	0.8351	0.7928
	NPV	0.8533	0.9813
Test	PPV	0.8194	0.6685
	NPV	0.8526	0.9566

4.2.2.5 PREDICTING PRETERM BIRTH IN TWIN GESTATION DISCUSSION

The preprocessed C5.0 based DT model for predicting preterm birth in twin pregnancies non-invasively before 22 weeks gestation was able to achieve higher sensitivity and specificity than any other work in literature to our knowledge. The MOD_TWINGEST set achieved an average test sensitivity and specificity of 79.32±5.85% and 91.97±1.20%, respectively. This was achieved using only 20 attributes. The results exceed the clinical expectation set out by Dr. Bariciak.

Previous similar work for predicting preterm birth in twin pregnancies included invasive methods, and only produced clinically acceptable predictions much later in the gestational period (section 2.3.4). Even though the scope of the previous work conducted by the MIRG group was different, the sensitivity and specificity exceeded those of previous work conducted by the MIRG group. C. Catley focused on predicting high-risk preterm birth in parous women using a 48 variable model and an integrated hybrid system which achieved a sensitivity of 65% and a specificity of 84% with an AUC of 0.80 (Catley, 2007). N.Yu's work focused on both parous and nulliparous women to predict high-risk preterm birth using an integration of decision trees and classification ANN with weight elimination and risk stratification using 19 variables. This model resulted in a test sensitivity of 65.13% and a test specificity of 84.07% for parous cases with an

AUC of 0.8195. However the prediction model for nulliparous cases did not meet the clinical expectations and required further examination (Yu, 2009).

This is the first attempt at creating a prediction model for predicting preterm birth in a high risk population of twin pregnancies including both parous and nulliparous cases within the MIRG team. Many aspects of this model make this prediction model ideal. This classifier required only 20 attributes to be collected by the user. This is comparable to N.Yu's work, although her work only required 19 variables in parous cases and 16 for nulliparous cases. However, N.Yu's work included key variables ('maternal bleeding' and 'previous low birth weight') that are not captured in Ontario's BORN nor Niday database (Yu, 2009). Therefore this is the first prediction model for PTB in twin pregnancies that are specific to Ontario. However it may be further applied as a generalized North American Model.

The PPVs and NPVs provide direct diagnostic meaning to the physicians. Current methods for predicting preterm birth in twin pregnancies which meet clinical standards include cervical length (CL) and fetal fibronectin (fFN). Both these methods are invasive, and in many cases, these methods are only able to provide clinical satisfactory predictions much later in the gestational cycle. The non-invasive prediction model presented in this thesis for predicting preterm birth in twin pregnancies not only exceeds the clinical standards, but it is also resulted in the highest average test sensitivity, specificity, PPV, NPV and AUC of 79.32±5.85%, 91.97±1.20%, 66.85±2.51%, 95.66±1.17%, and 0.9016, respectively in literature. The best indicator to predicting preterm birth was 'smoking', followed by 'maternal substance abuse' and 'previous preterm birth'.

4.3 SINGLETON GESTATION DECISION TREE RESULTS

The BORN database was split into noMOD_SINGLETON DT and MOD_ SINGLETON DT sets. A C5.0 based DT method was applied using repeated 2-fold cross validation method to generate ten test and train datasets (Section 3.6.1.4). The average test sensitivity and specificity did not exceed 65% and 85%, respectively, for both the MOD_ SINGLETON DT and the noMOD_ SINGLETON DT set.

4.1.3 SINGLETON DATASET

A total of 140 575 test cases were evaluated on 125 000 term and 15 575 preterm births. A total 20 attributes were used including MATAGE, PPRETERM, PTERM, PARITY, PREVCS, GENDER, INTBF, SMOKING, MATHP0, MATHP3, MATHP4, MATHP18, MATHP27, REPASS, FIRSTVIS, PRENCLAS, LANGUAGE_up, MATHP_sub, and Diagnosis (for description please see Table 3.4). The prevalence of the dataset was 12.46%.

4.1.3.1 noMOD_SINGLETON DT RESULT

Out of the 20 input attributes, the C5.0 algorithm used between 16-19 attributes to generate the DTs. The most common attributes among all sets were MATH27, INTBF, LANGUAGE_up, MATHP3, MATHP4, MATAGE, PRENCLASS, GENDER, SMOKING, PPRETERM, REPASS, MATHP0 and FIRSTVIS. The most contributing attribute was MATH27 followed by MATH3 and INTBF. The resulting average test sensitivity and specificity were 11.63±1.58% and 74.21±5.36%, respectively. The sensitivity and specificity during training was slightly higher at sensitivity of 14.97% and specificity of 99.71%. The correct classification rate or accuracy was 67.27±4.62%. The PPV and NPR were 5.43±0.74% and 87.08±0.62%, respectively. The noMOD_SINGLETON model is not acceptable, since both the sensitivity and specificity are much lower than the clinical standards.

To examine the effects on the prediction model for the case of parous vs. nulliparous women, the noMOD_SINGLETON DT was further split into noMOD_ SINGLETON_parous and noMOD_SINGLETON_nulliparous, respectively. The sensitivity of the noMOD_SINGLETON_parous set was lower compared to the noMOD_SINGLETON set with an average test sensitivity of 10.69±2.82%, whereas the test sensitivity of the noMOD_SINGLETON_nulliparous increased to 19.76±3.27%, the test specificity value for the noMOD_SINGLETON_parous remained stable at 74.23±4.00%, whereas the noMOD_SINGLETON_nullinuparous decreased to 51.74±6.00%. Due to the poor sensitivity, both prediction models do not meet the clinical standards set by Dr. Bariciak.

4.1.3.2 MOD_SINGLETON DT RESULT

The attributes with the nominal values grouped were MATAGE, PPRETERM, PTERM and PARITY. Out of the 20 input attributes, the C5.0 algorithm used between 18-19 attributes generate the DTs. The attributes that were common among all sets included MATH_sub, INTBF, LANGUAGE_up, PPRETERM, MATHP3, MATHP4, FIRSTVIS, GENDER, MATAGE, MATHP0, PRENCLAS, PREVCS, SMOKING, PTERM, PARITY, MATHP27, REPASS, and MATHP_ment. The most contributing factor was MATH_SUB in predicting preterm birth. The resulting average test sensitivity was higher compared to noMOD_Singleton at 35.57±4.70%, however the average test specificity dropped significantly in comparison to 27.19±3.63%. The sensitivity and specificity during training was higher at sensitivity of 43.84% and specificity of 98.88%. The correct classification rate or accuracy was 28.13±2.95% and the error rate was 71.87±2.95%. The PPV and NPV were 5.76±0.62% and 76.93±2.52%, respectively. The MOD_SINGLETON model is not acceptable, since both the sensitivity and specificity are much lower than the clinical standards.

To examine the effects on the prediction model for the case of parous vs. nulliparous women, the MOD_SINGLETON DT was further split into MOD_ SINGLETON_parous and MOD_SINGLETON_nulliparous, respectively. The sensitivity of the MOD_SINGLETON_parous set was higher compared to the MOD_SINGLETON set with an average test sensitivity of 35.47±12.44%, whereas the test sensitivity of the MOD_SINGLETON_nulliparous was lower at 26.70±7.48%, the test specificity value for the MOD_SINGLETON_parous increased to 41.16±1.77%, whereas the MOD_SINGLETON_nulliparous decreased to 15.84±5.96%. Due to the poor sensitivity, both prediction models do not meet the clinical standards set by Dr. Bariciak.

4.1.3.3 PREDICTING PRETERM BIRTH IN SINGLETON DISCUSSION

The noMOD_Singleton and MOD_Singleton DT prediction models did not produce clinical satisfactory results for predicting preterm birth in singletons. This was expected, in previous work to predict preterm birth in parous and nulliparous women including singleton and twins gestations, the key identifying variables in C.Catley's work for parous women were 'maternal bleeding', followed by 'plurality' and 'previous preterm birth'. N.Yu's work coincided with these findings, for parous cases, all decision trees used 'maternal bleeding', 'plurality' and 'previous preterm birth' (Catley, 2007) , (Yu, 2009). 'Plurality' and 'previous preterm birth' was not applicable in the nulliparous case. Further another variable that ranked highest on the contributing factors was 'previous low birth weight', which again was not applicable in the nulliparous case. In both the Niday and BORN database, the only key identifying variables included were 'previous preterm birth' and 'plurality', both 'maternal bleeding' and 'previous low birth weight' were not included in the BORN nor Niday database. All four variables above

were linked as strong indicators in the prediction of preterm birth, therefore the poorer accuracy and discrimination ability of the singleton model could be attributed to this factor.

4.2 PERINATAL DECISION SUPPORT SYSTEM (PEDSS) FRAMEWORK RESULTS

4.2.1 PLATFORM SELECTION

The factors leading to successful clinical decision support implementation includes a platform that supports collaboration, audience targeting (role based access), notifications/alerts, workflow integration, remote access, real-time acquisition and update of data, integration to other services (i.e. EMR), and mobility. Further our system must be highly secure to house patient sensitive information, and permit access to only authorized users. It must also be low cost to implement, deploy and maintain (Berner, 2009) (Frize, 2005) (Peleg & Tu, 2011).

Based on the criteria's for successful CDSS implementation, two CMSs in the market resulted as our two top candidates: Drupal (Open Source) and SharePoint (Commercial). Drupal was leveraged in the previous thesis work of S.Weyand to build the PPADS tool (Weyand, 2011). However, for this thesis work, SharePoint was chosen as the platform of choice. Although Drupal is lower cost compared to SharePoint, SharePoint offers integration to enhanced workflow support. Further, many of our health service providers in the Champlain region are already utilizing a SharePoint based platform (Microsoft Pinpoint, 2010), (Champlain LHIN , 2012), (The Health Council of Canada, 2012), (EORLA Winder Communication, 2010).

4.2.1.1 INTRODUCTION TO SHAREPOINT 2010 CONTENT MANAGEMENT SYSTEM

The Microsoft based SharePoint platform provides an appealing and secure user friendly graphical interface, and when correctly implemented it can be both PIA (Privacy Impact Assessment) and TRA (Threat-Risk Assessment) approved, permitting the entry of patient sensitive information (Office of the Commissioner of Lobbying of Canada, 2010), (Federal Trade Commission, 2012), (Hietala, 2010). The SharePoint platform allows for rapid development and deployment of secure applications that enable the exchange of electronic submissions, digital workflow, and secured information storage. The SharePoint space can be tailored to meet the needs of a specific user by altering simple, easily maintained rule sets. The functionality of the SharePoint framework can be extended to meet hospital specific requirements through the use of custom developed components or using commercial off-the-shelf software.

The SharePoint space also supports role based access (RBAC) where in order to access the system, the user must enter a unique user identification and password. Session timeouts after a pre-set period of inactivation, as well as audit trail with an electronic signature and date-time stamp can also be incorporated.

4.2.1.2 SYSTEM COMPONENT SELECTION

Table 4.13 below provides commercially available Microsoft based components to build web based PEDSS. It is important to choose Microsoft based components in order to be compatible with the existing infrastructure. The current See5 application (section 3.6.2.5) executes on a Microsoft Excel document and the CBRS tool uses a Microsoft Access database.

Table 4.13 Components for a web-based PEDSS

Components	Commercially Available	Microsoft Based
	Microsoft Forefront Unified Access Gateway 2010	Yes
	SharePoint 2007/2010	Yes
	Nintex Workflows 2007/2010	Yes
	SQL Database 2008/2008 R2	Yes
	Active Directory 2003/2008	Yes
	Windows Server 2012, Web Front End, Development, Virtual Servers	Yes
	Microsoft InfoPath 2007/2010	Yes

4.2.2 FRAMEWORK OVERVIEW

Advanced features of the content management system were utilized for this thesis work to develop the PEDSS tool including audience targeting, where filters were set up for the varying user groups. All groups of users will access the tool using the same Uniform Resource Locator (URL). Based on the role of the user logged in, the user is presented with filtered views on what is relevant to them. Upon logging in, the user will be required submit a set of credentials: username and password. The login screen for the PEDDS tools is presented below in Figure 4.1.

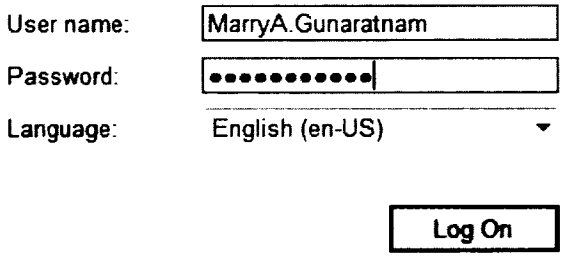

Figure 4.1 Log-in screen to PEDSS

4.2.2.1CLINICIAN'S TOOL

The clinician tool homepage delivers all content relevant to the clinician. A snapshot of the clinical tool homepage is presented in Figure 4.2. The clinician has permissions to add/edit and view content on PEDSS. The clinician's home page consists of an announcements section, a shared calendar of upcoming events, a shared discussion board, a shared documents repository, a patient database, and the perinatal clinical decision support tools: (1) Neonatal Mortality Risk Assessment and (2) Preterm Birth Predictor in Twins Pregnancies.

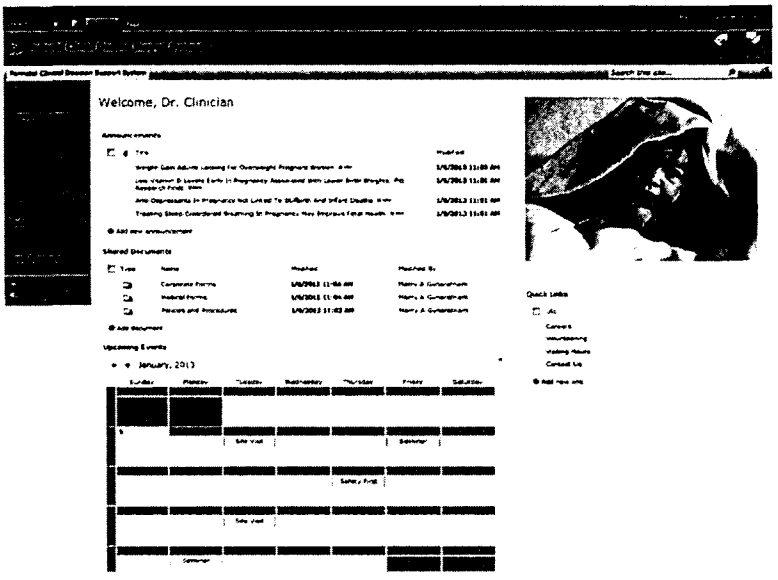

Figure 4.2 PEDSS Homepage

The announcement section contains alerts and reminders relevant to the clinician user community. The clinician may add new announcements.

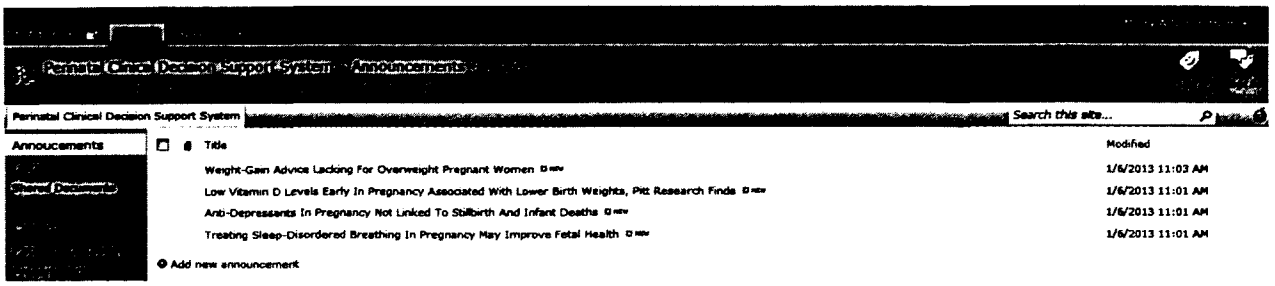

Figure 4.3 PEDSS Announcement Section

The shared calendar is used to post any relevant information seminar or events of interest to the clinicians. The clinicians may add new events.

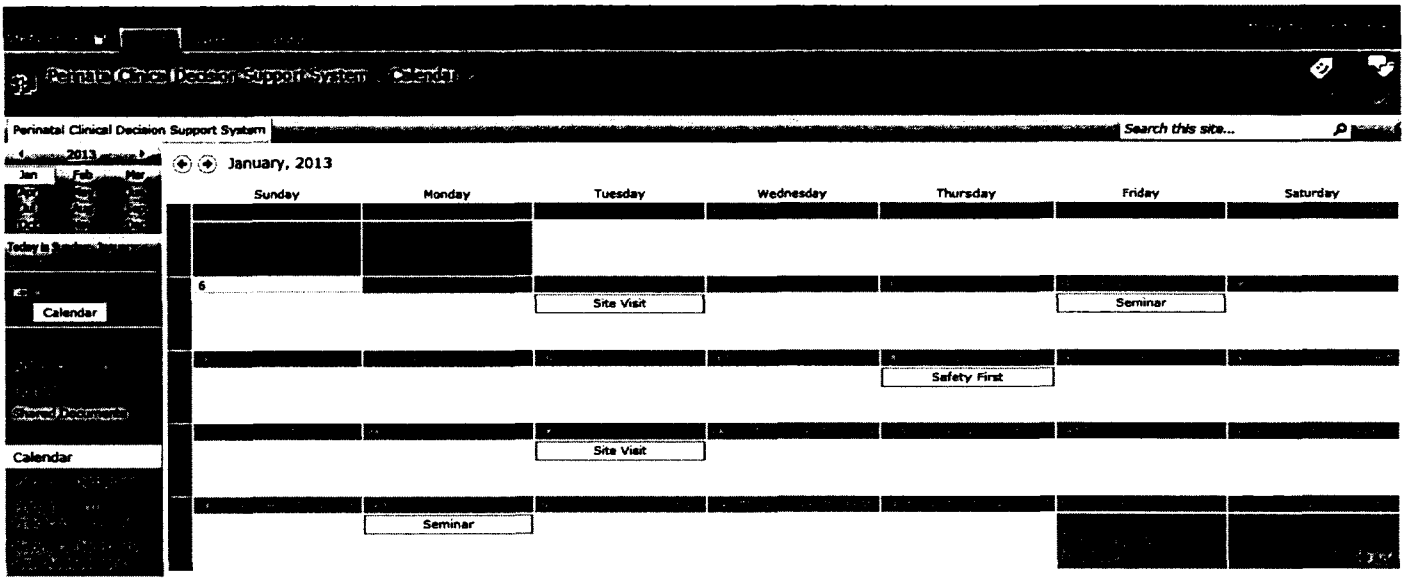

Figure 4.4 PEDSS Shared Calendar Section

The shared discussion board is used to posts any discussion topics which may be relevant to the clinicians. The clinicians may use this to collaborate and gather feedback from other clinicians. The clinicians may add new discussion topics.

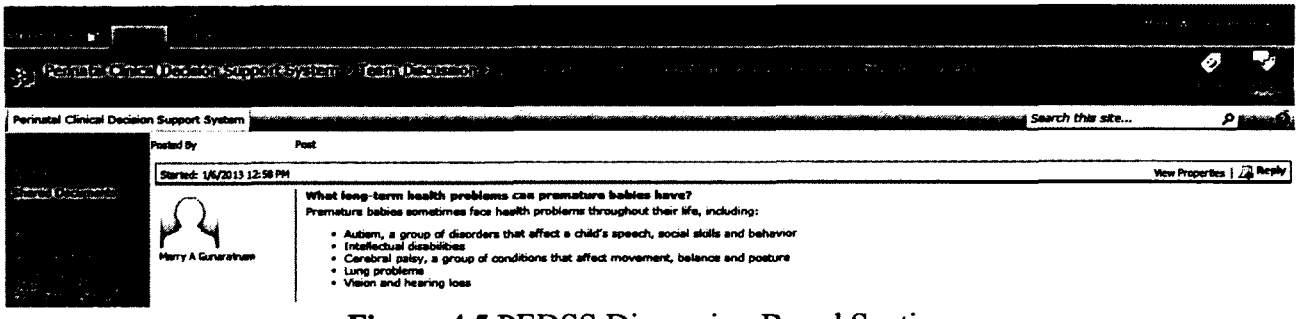

Figure 4.5 PEDSS Discussion Board Section

The shared document repository contains files and folders relevant to the clinicians. Clinician's may add, edit or delete documentation. The metadata of the following is captured: Title, Created By, Created, Modified By, and Modified.

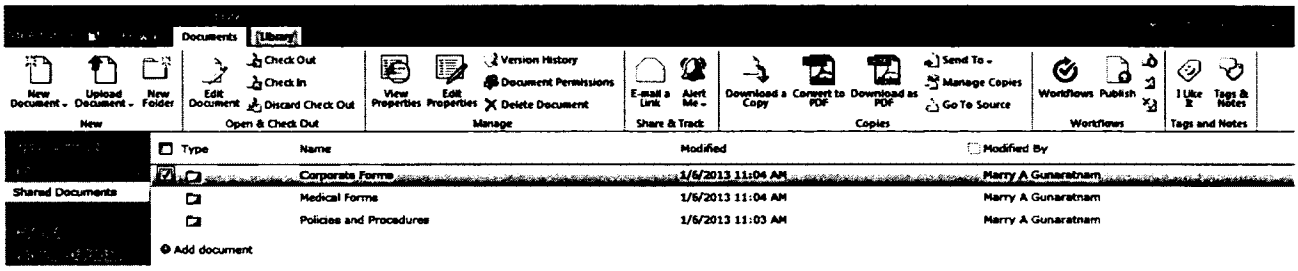

Figure 4.6 PEDSS Shared Document Repository

The patient's database contains the patient files in ascending order as shown in figure 4.7. Each patient file contains the mother's patient ID, name, date of birth, gestational age, gender of newborn, admission date, and risk information of important outcomes including risk of preterm birth in twin pregnancies, and if applicable, risk of neonatal mortality and any relevant patient documentation. The patient file also includes a free-text clinician's note section, where clinicians' can write a notation specific to the patient. The patient file form was created using InfoPath 2010 and was web-enabled and published to the SharePoint environment. This form is shown in figure 4.8.

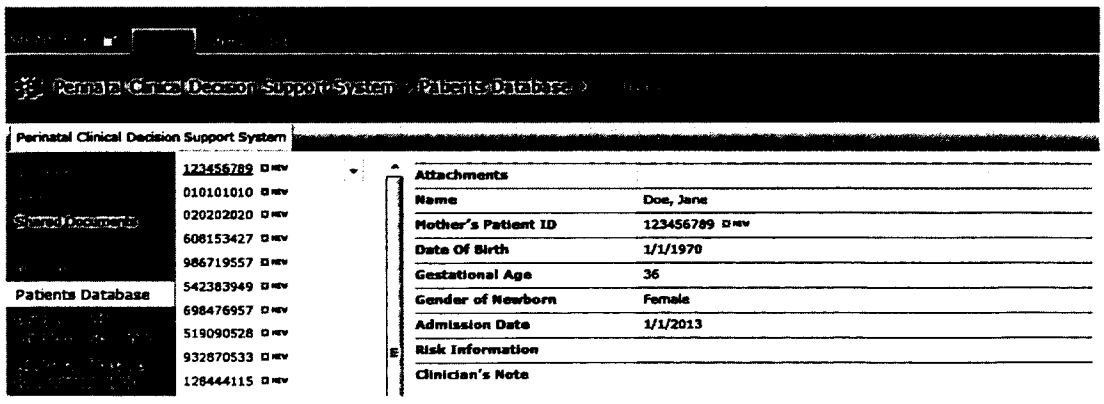

Figure 4.7 PEDSS Patient Database

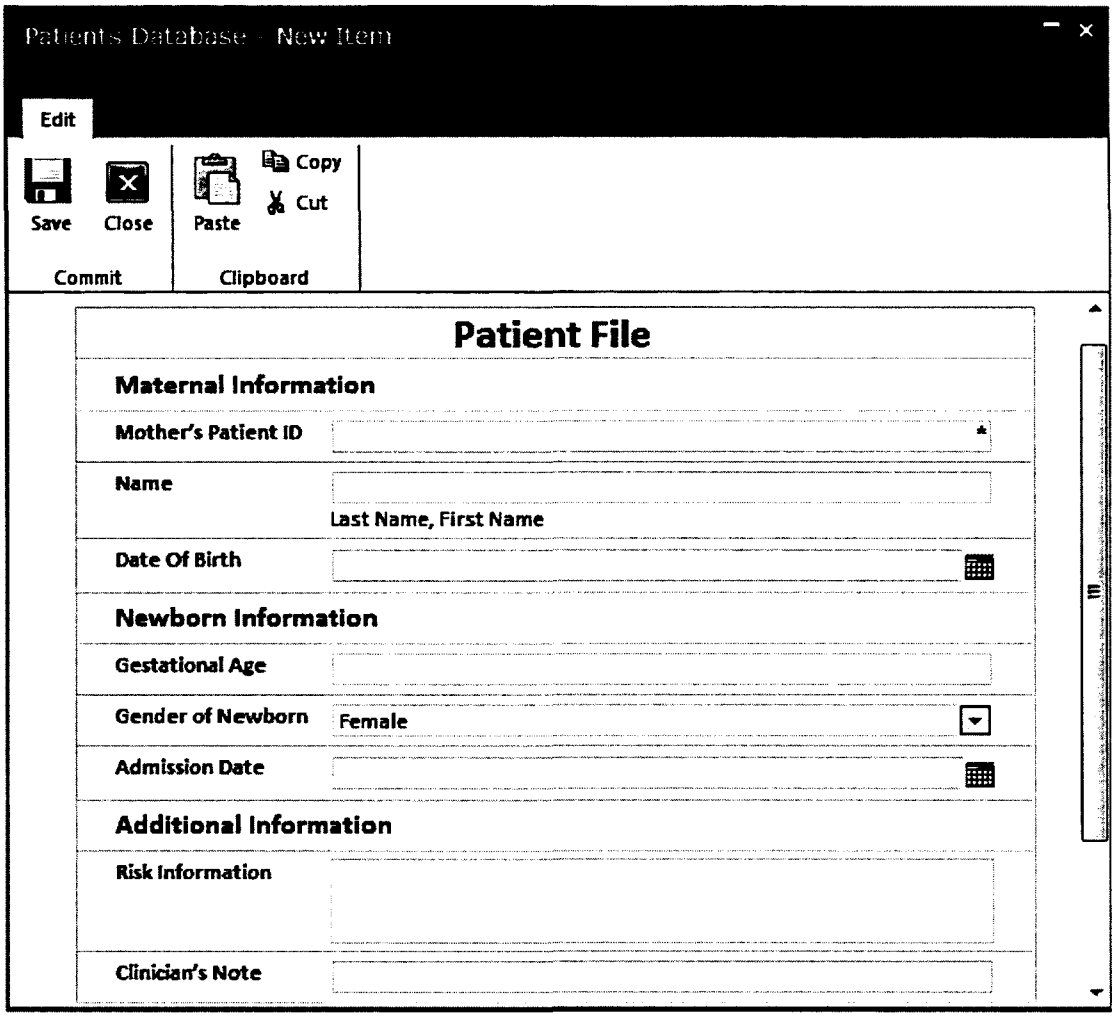

Figure 4.8 Electronic Patient File

There are two perinatal clinical decision support tools available to clinicians, including the Neonatal Mortality Risk Assessment and Preterm Birth Predictor, as follows.

NEONATAL MORTALITY RISK ASSESSMENT

The neonatal mortality risk assessment tool consists of a series of assessment criteria to evaluate the newborn at 10 minutes. All fields are mandatory; however certain fields may be answered as N/A if not applicable. The screener form is shown in figure 4.9.

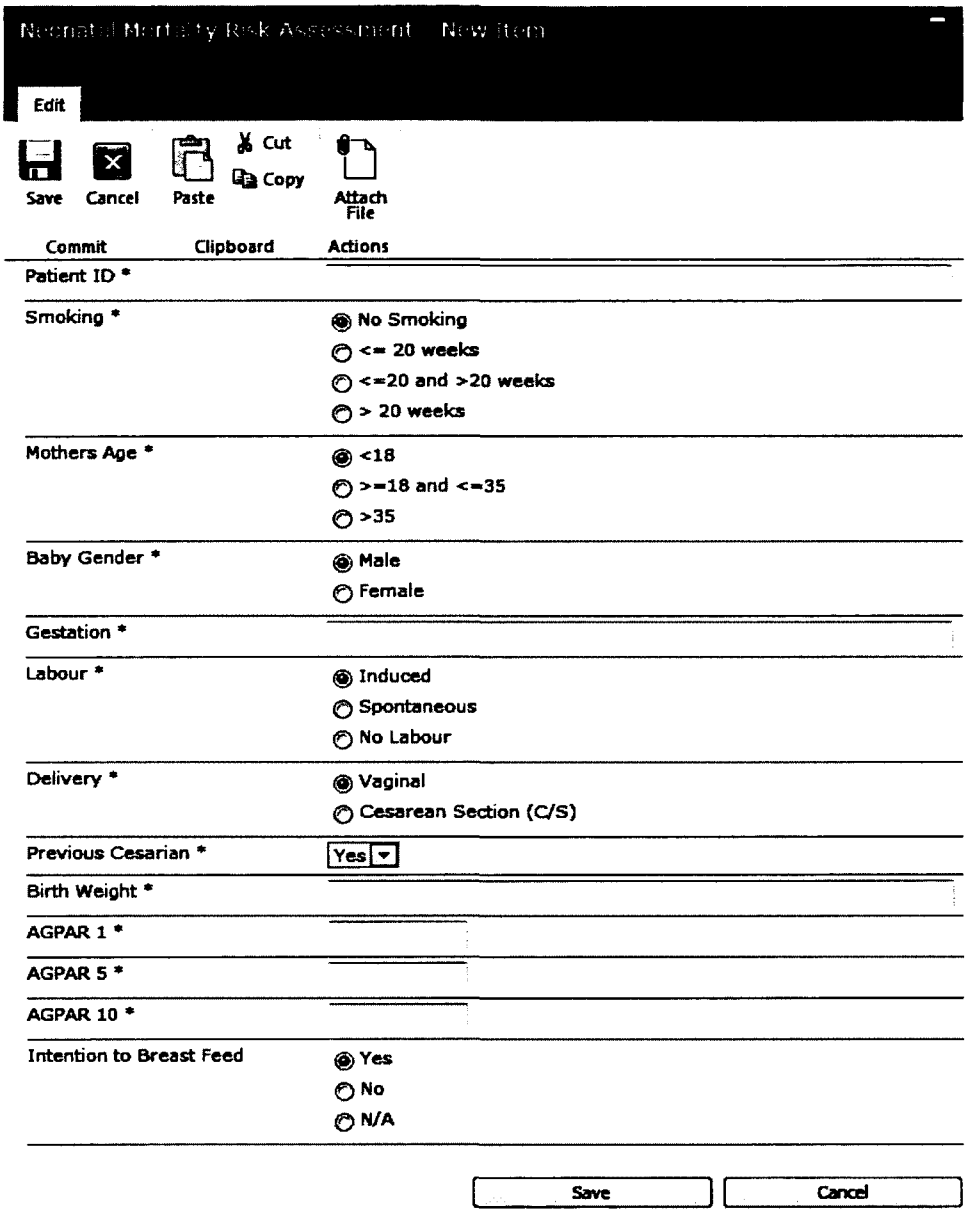

Figure 4.9 Neonatal Mortality Risk Assessment Form – Input Attributes: Smoking, Mother's Age, Baby Gender, Gestation, Labour, Delivery, Previous Caesarian, Birth Weight, AGPAR 1, AGPAR 5, AGPAR 10 and Intention to Breast Feed.

PRETERM BIRTH PREDICTOR IN TWINS PREGNANCIES

The preterm birth predictor tool consists of a series of questions for the mother pregnant with twins. All fields are mandatory; however certain fields may be answered as N/A if not applicable. The screener form is shown in figure 4.10:

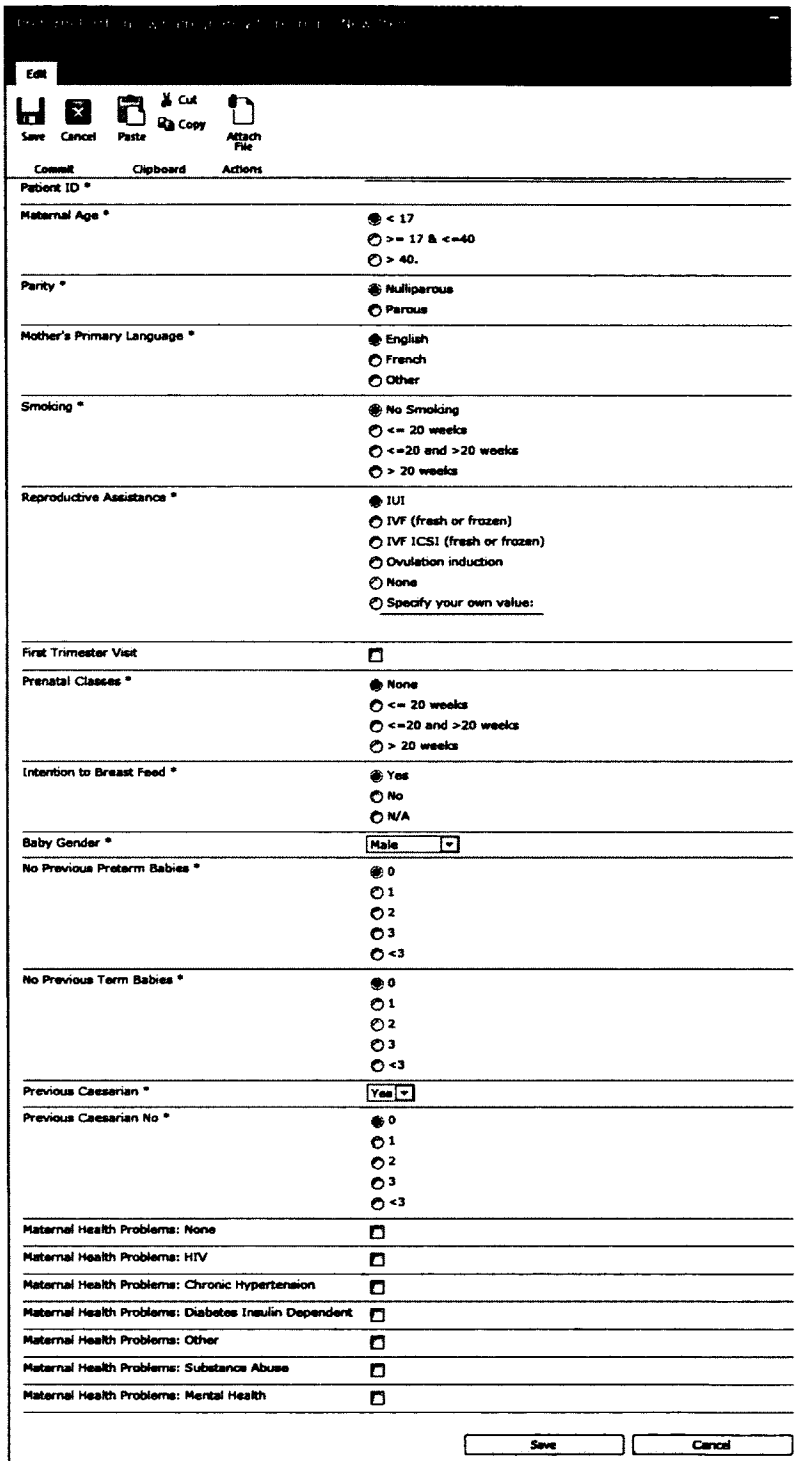

Figure 4.10 Risk of Preterm Birth Prediction in Twins Pregnancy Form – Input Attributes: Mother's Age, Parity, Mother's Primary Language, Smoking, Reproductive Assistance, First Trimester Visit, Prenatal Classes, Intention to Breast Feed, Baby Gender, No Previous Preterm Babies, No Previous Term Babies, Previous Caesarian, Previous Caesarian No., Maternal Health Problems: None, HIV, Chronic Hypertension, Diabetes Insulin Dependent, Other, Substance Abuse and Mental Health.

Once all questions have been answered, the clinician can submit the form. A workflow is then initiated, and an outcome prediction is produced. The workflow is generated from the original C5.0 based decision tree. A sample diagram of a simple C5.0 generated decision tree is given below.

```
AGPAR 5 in {3,4,5,6,7,8,9,10}: Not Applicable (16479.6/22)
AGPAR 5 = 0:
:...AGPAR 1 in {2,3,4,5,6,7,10}: Neonatal Death (0)
:   AGPAR 1 in {0,1}: Neonatal Death (70.2/0.2)
:   AGPAR 1 in {8,9}: Not Applicable (3)
AGPAR 5 = 2:
:...Birth Weight in {<=1000,>1500 and <=2000}: Neonatal Death (11/3)
:   Birth Weight in {>1000 and <=1500,>2000 and <=2500,>2500 and <=3000, >3000
and <=4000,>4000}: Not Applicable (17.1/1)
AGPAR 5 = 1:
:...Gestation in {>30 and <=34,>37}: Not Applicable (3)
    Gestation = >34 and <=37: Neonatal Death (2)
    Gestation = <=30:
    :...Delivery = Vaginal: Neonatal Death (4)
        Delivery = Cesarean Section (C/S): Not Applicable (2)
```

The decision tree can easily be expressed as a set of nested if-statements as follows:

```
if (AGPAR 5 == (3 OR 4 OR 5 OR 6 OR 7 OR 8 OR 9 OR 10))
       return Not Applicable;
if (AGPAR 5 == 0)
{
       if (AGPAR 1== (2 OR 3 OR 4 OR 5 6 OR 7 OR 10))
              return Neonatal Death;
       if (AGPAR 1== (0 OR 1))
              return Neonatal Death;
        if (AGPAR 1== (8 OR 9))
        return Not Applicable
}
if (AGPAR 5 == 2)
{
       if (Birth Weight <=1000 OR >1500 OR <=2000))
              return Neonatal Death;
       if (Birth Weight >1000 OR <=1500 OR >2000 OR <=2500 OR >2500 OR
            <=3000 OR >3000 OR <=4000 OR >4000))
              return Not Applicable;
}
if (AGPAR 5 == 1)
{
       if (Gestation >30 OR <=34 OR >37))
              return Not Applicable;
       if (Gestation >34 OR <=37))
              return Neonatal Death;
       if (Gestation <=30)
          if (Delivery == Vaginal))
              return Neonatal Death;
          if (Delivery == Cesarean Section (C/S)))
              return Not Applicable;
}
```

The nested if statements are used to create the workflow using Nintex. Nintex is a powerful workflow designer add-on to SharePoint. Nintex allows for drag-and-drop functionality, eliminating the need to code. It also integrates to Visual Studio permitting custom actions. Further Nintex can be used to integrate to other web services, including integration to an EMR, as well as allows for automatic notifications and alerts to be integrated. Figure 4.11 shows the associated Nintex workflow for the DT presented above. The parallel action is used to initiate the four main 'If' statements in the DT, next the action set is used to code the 'If' statement, and lastly the outcome as per the DT is presented back to the user using the set field property. The completed workflow is then published to the SharePoint environment.

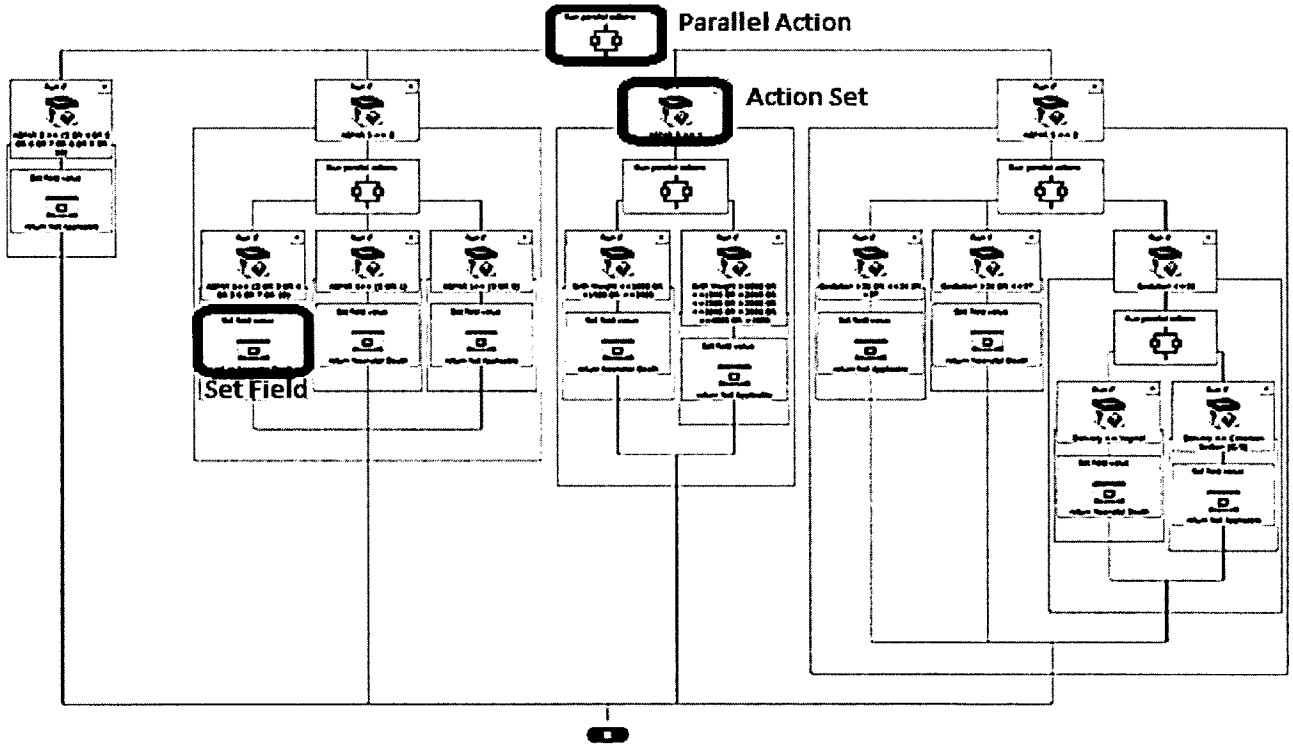

Figure 4.11 Workflow Diagram Using Nintex Workflows 2010

In the case of neonatal mortality, a decision tree was generated at each trial, and a total 10 trials were executed in the MOD_NEONATALMORT set, thus resulting in 10 decision trees. A

workflow is created for every DT, and thus 10 DTs equates to 10 workflows. An additional workflow using a vote count algorithm was implemented to predict the outcome prediction score. In total of (#of DTs) + 1 workflow are published, which in this case equates to 10+1=11 workflows. This workflow is automated, and produces results in minutes.

The vote count algorithm includes a DT_counter and outputs an outcome prediction score. The DT_counter increases by a factor of 10 for every positive outcome per DT. The outcome prediction score = 100 - DT_counter, thus if 2 out of 10 DTs indicated a positive outcome, then the DT_counter equals 20, and the outcome prediction score equals 100-20=80. This would indicate that the chance of a negative outcome based on the 10 DTs is 80%. The outcome prediction score will assist clinicians and the parents in the decision making process if a patient is at high risk for any of the conditions screened. A screenshot of the outcome predictions is shown in figure 4.12.

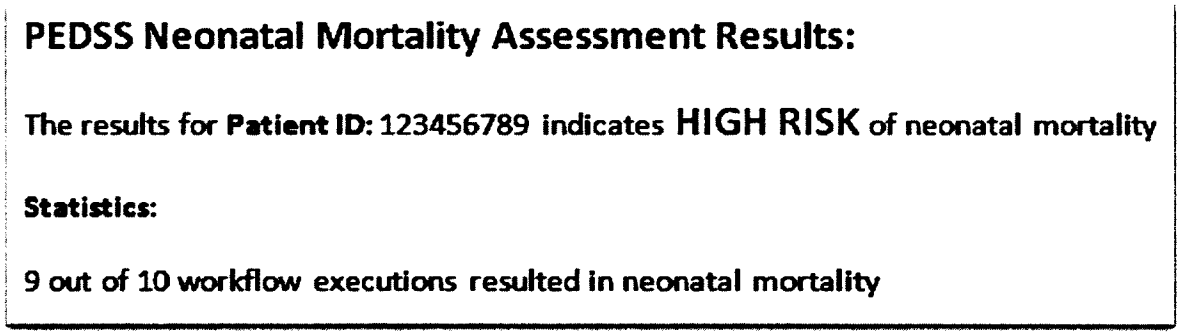

Figure 4.12 Prediction Outcome Notification Message

In the case where a neonate has been assessed as being at high risk of neonatal mortality, one of the most difficult decisions is deciding on a change in direction of care. In this case, S.Weyand's decision tool may be used (Weyand, 2011). S.Weyand's work included a decision support instrument in order to aid the decision making regarding change in the direction of care.

It is important to note that the decision support tool is used to inform parents and physicians, assist in the decision making process by organizing decisional information, determine the information required for the decision making and bring to attention any additional information which may be required for the decision to be made. Thus it does not make the decision or produce a finalized conclusion (Weyand, 2011). S.Weyand's decision support tool consisted of 6 steps which address various aspects of the decision that would need to be made and provides information about various options including the pros and cons of each option.

CHAPTER 5 CONCLUDING REMARKS, CONTRIBUTING KNOWLEDGE AND FUTURE WORK

5.1 CONCLUSION

An improved classification method has been derived to improve the prediction of two distinct medical problems non-invasively: neonatal mortality with information available at 10 minutes after birth and preterm birth in twin pregnancies before 22 weeks. As for the next step, in order to bring the derived classification model to clinical use in an obstetrical environment, a framework for a secure web-based Perinatal Decision Support System (PEDSS) to assist the clinicians and parents in predicting the likelihood of neonatal mortality and preterm labour in twin pregnancies was presented. In the process of resolving the current issues, this thesis work has raised questions that require further assessment. This is discussed in the conclusion, contributing knowledge and future work sections of this chapter.

5.1.1 ACCOMPLISHING PRELIMINARY OBJECTIVES

The first objective of this thesis was to assess varying data mining methods with an aim to improve the classification of neonatal mortality and preterm birth in twin pregnancies. Various models were analyzed including decision trees (DTs) and hybrid ANN to see which produced better outcomes. In order to accomplish the first objective, an improved data processing step was introduced in the DT model (section 3.6.1.2). Concluding from the first objective, it was found that the data processing method introduced in this thesis work may be applied to many other multi-factorial medical problems to improve its classification. The new approach has provided several improvements to better predict medical problems. This thesis work has established the following:

1. Pre-processed datasets run against C5.0 algorithm produced decision trees superior to the DT-ANN hybrid method. This was shown in the increased accuracy of the C5.0 based method. This method should be examined as a potential principal method for future research work.
2. Created a neonatal mortality prediction system for newborn to be assessed with data available from the first 10 minutes after birth non-invasively with acceptable discrimination, exceeding the results of current standard predictions. This screener tool may be used as alternative to costly and invasive methods.
3. Created a preterm birth prediction system for a high risk population (for women pregnant with twins) non-invasively before 22 weeks gestation exceeding the results of current standard predictions. The high risk population identified is likely subject to costly and invasive screening methods (CL and fFN). This screener tool may be used as an alternative to such methods.

4. The previous neonatal prediction method only focused on newborns after admission to NICU. This is the first attempt at predicting neonatal mortality in a heterogeneous population with data available at 10 minutes after birth. To eliminate cases where neonates did not survive prior to 10 minutes, the cases with AGPAR 10 score less than 1 (death) were removed.
5. The prediction of preterm birth in women pregnant with twins included both parous and nulliparous mothers with data available before 22 weeks gestation. The resulting performance outcome was better than all previous similar studies in the literature. Previous work conducted by the Medical Information Technology and Research Group (MIRG) focused on parous and nulliparous mothers, where results for nulliparous women indicated poorer performance results compared to parous cases.
6. Preliminary results of women pregnant with singleton were poor compared women pregnant with twins. This was expected due to limited variables in both the Niday and BORN databases. Both 'maternal bleeding' and 'previous low birth weight' were not included in the BORN or the Niday databases; the only key identifying variables included 'previous preterm birth' and 'plurality. All four variables above were linked as strong indicators in the prediction of preterm births; therefore, the poorer accuracy and discrimination ability of the singleton model could be attributed to this factor.
7. Several improvements were made compared to past models: For the neonatal mortality case, the prediction of neonatal mortality non-invasively was reduced to data available at 10 minutes after birth using only 13 attributes, whereas the previous models required up to 12 hours from birth using 3 variables derived from invasive methods. The ability to predict preterm birth in twin pregnancies was extended to include both parous and

nulliparous cases and was reduced to 22 weeks using only 20 attributes. Whereas previous prediction models predicted preterm birth in parous women pregnant with twins using 19 attributes at 23 weeks gestation with lower sensitivity and specificity.

8. The need for improved data pre-processing steps was emphasized (section 3.6.1.2): nominal inputs with many variables were grouped as appropriate according to literature to prevent spurious splits of decision trees. Attribute(s) with nominal values grouped improved the performance outcomes in both the neonatal mortality and preterm birth cases.
9. The positive predictive value (PPV) is the likelihood that that the positive prediction is correct. Previous models by N.Yu resulted in low PPV for neonatal mortality (PPV=0.34) and preterm birth (PPV=0.43). Thus, the previous models would have required a more reliable follow up assessment which may in the form of invasive test methods such as CL or fFN. These are the first prediction models which achieved a high PPV, while meeting the clinical standards for sensitivity and specificity. The average test PPV for the MOD_Neonatal DT, MOD_Neonatal Hybrid ANN and MOD_TWINGEST are 72.35$\pm$13.91%, 71.14$\pm$1.64% and 80.0%$\pm$14.66%, respectively.

The second objective of the thesis was to derive a conceptual framework a secure web-based Perinatal Decision Support System (PEDSS). A method was developed to convert the resulting C5.0 decision tree to commercial code. This thesis work is the first to present a framekwork for a secure commercial off-the-shelf solution that could be used to predict neonatal mortality 10 minutes after birth and preterm birth in twin pregnancies before 22 weeks gestation. This framework may be extended for other medical problem(s) in a perinatal environment. This thesis work has established the following:

1. A conceptual framework for a secure web-based Perinatal Decision Support System (PEDSS) was presented which provides audience targeted information and risk prediction of neonatal mortality and preterm birth in twin pregnancies.

2. A four step process for developing the framework of a Perinatal Decision Support System (PEDSS) was presented. The first step is the identifying the platform, the second step is identifying system composition including hardware/software and communication architecture, the third step is system integration, and fourth step is knowledge maintenance.

3. A high level overview of commercially available components to build a complete secure web-based perinatal decision support system was developed. When properly implemented, this system shall be both PIA (Privacy Impact Assessment) and TRA (Threat Risk Assessment) approved permitting the entry of patient sensitive information.

4. A web based framework of a screener tool for non-invasively identifying the risk of neonatal mortality with data available 10 minutes after birth and preterm birth in twin pregnancies prior to 22 weeks gestation was presented.

5.2 CONTRIBUTIONS TO KNOWLEDGE

1. Developed an improved risk assessment model for the prediction of neonatal mortality in newborns with data available at 10 minutes after birth using DTs. The model's average test sensitivity and specificity was $62.24 \pm 3.28\%$ and $99.95 \pm 0.03\%$, respectively.

2. Developed an improved risk assessment model for the prediction of preterm birth in twin pregnancies before 22 weeks gestation using DTs. The model's average test sensitivity and specificity was $79.32 \pm 5.85\%$ and $91.97 \pm 1.20\%$, respectively.

3. Demonstrated superior performance of pre-processed dataset using C5.0 based DT to predict neonatal mortality compared to DT-ANN Hybrid and ANN weight elimination methods.

4. Showed that dataset for C5.0 required less processing and adjusting of parameters, and performs well with large datasets in a shorter timeframe compared to DT-ANN Hybrid and ANN weight elimination methods. There is no need to normalize the data, among other time consuming pre-processing steps to create a classifier.

5. A conceptual framework for a secure web-based Perinatal Decision Support System (PEDSS) which included a screener tool for assessing risk for neonatal mortality and preterm birth in twin pregnancies was presented. This generalized framework may be further extended for assessing the likelihood of many other medical outcomes.

5.3 FUTURE WORK

1. Further work is needed to improve the prediction of preterm birth in women pregnant with singleton. Additional variables are required to improve the results, specifically, 'maternal bleeding' and 'previous low birth weight'

2. Apply the preprocessing step to the datasets of other perinatal databases on a national, international or global level to assess the DT-ANN algorithm on different perinatal environments.

3. Throughout the duration of this thesis work, applications for ethical clearance were made and revised for use the BORN database. Although we had been granted access to this database in November 2012, many indicative variables included contained corrupted values in the dataset.

Further examination of the variables may reveal better improvements for the prediction of preterm births among women pregnant with a singleton.

4. Implement the proposed Perinatal Decision Support System (PEDSS) in a clinical setting, and perform a usability study to determine the usefulness and ease of use. Determine the physician views and any area of concern such as: security, privacy, ethical dilemmas etc.

5. In conjunction with physicians, implement a clinical trial of the Perinatal Decision Support System (PEDSS) and compare the screener tool findings with clinical judgment and actual maternal outcomes.

6. Use DT method with nominal values grouped in the BORN database to evaluate the risk of preterm birth in nulliparous women with a singleton birth to determine whether the performance can be improved compared to past methods.

5.4 FINAL STATEMENT

Preliminary results of the C5.0 based DT model proved to be successful for non-invasively predicting in a group of new, undifferentiated patients, including the case of: neonatal mortality using data available at 10 minutes after birth, and preterm birth in twin pregnancies using data available at 22 weeks gestation. The knowledge gained in this thesis work has proven to be advantageous to not only the data mining and artificial intelligence sector, but as well as the medical field, especially in perinatal care. Hopefully, one day the proposed perinatal decision support system is seen in clinical use and its potential benefits are experienced by those in need.

References

Alere's Women's & Children's Health, (2008). What is Preterm Labour?. Last accessed: December17, 2011 from http://www.sidelines.org/62188_Preterm_english.pdf

Allen N., et al. (2002). Preterm Birth: Making a Difference. Best Start 2002. Last accessed: December 17, 2011 from http://www.beststart.org/resources/rep_health/pdf/Preterm_mstr.pdf

American College of Surgeons. (2006). Ceasarean Childbirth. Last accessed: December 17, 2011 from http://www.facs.org/public_info/operation/cesarean.pdf

Anyanwu, M.N., and Shiva, A.G., (2009). Comparative Analysis of Serial Decision Tree Classiication Algorithms. *International Journal of Computer Science and Security*. 3, 230-237

Apte, C., et al. (2002). Data Intensive Analytics for Predictive Modelling. Last accessed: December17, 2011 fromhttp://www.research.ibm.com/dar/papers/pdf/dar_overview_ibmjrd.pdf

ARTEMIS. (2008). Health Informatics Research. *UOIT Health Informatics Research*. Last Accessed: February 10, 2013 from http://hir.uoit.ca/cms/?q=node/24

Berner, E.S. (2007). *Clinical Decision Support System.* New York: Springer Science + Business Media LCC,

Berner, E.S. (2009). *Clinical Decision Support Systems: State of Art.* Rockville: AHRQ Publication,

Blackmore, C.A., et al. (2011). Preterm Birth. *CDC's Public Health and Surveillance for Women, Infants and Children.* Last accessed: December17, 2011 from http://www.cdc.gov/reproductivehealth/ProductsPubs/DatatoAction/pdf/birout2.pdf

Blickstein, I., and Keith, L. Multiple Pregnancy: Epidemiology, Gestation, and Perinatal Outcome, 471-473. United Kingdom: Informa UK Ltd, 2005.

BORN Ontario. (2010). BORN Bulletin. 1(1). Last Accessed: January 21, 2013 from http://www.bornontario.ca/_documents/Publications/BORN%20Ontario%20Bulletin%20-%20January%202010.pdf

BORN Ontario. (2013). Data Dictionary. Last accessed: May, 4, 2013 from http://www.bornontario.ca/data/data-dictionary

Canadian Institute for Health Information. (2006). *Giving Birth in Canada: The Costs.* Last accessed: December17, 2011 fromhttp://publications.gc.ca/collections/Collection/H118-38-2006E.pdf

Catley, C.A. (2007). An Integrated Hybrid Data Mining System for Preterm Birth Risk Assessment Based on a Semantic Web Services for Healthcare Framework. Ph.D. diss. Carleton University.

Catley, C., Frize, M., Walker, R.C., and Petriu, D.C. (2005). Predicting Preterm Birth using Artificial Neural Networks. *CMBS.* 103-108

Catley, C., Petriu, D.C., and Frize, M. (January 2004). Software Performance Engineering of a Web service-based Clinical Decision Support infrastructure. *WOSP.* 29, 130-138

CBC News. (2009). *Canada ranked 25th in the world infant mortality.* Last accessed: December17, 2011 fromhttp://www.cbc.ca/news/health/story/2009/11/04/infant-mortality-canada.html

Centre for Disease Control and Prevention. (2009). How Prevalence is Calculated. Last Accessed: April 12, 2012 from http://www.cdc.gov/pednss/how_to/read_a_data_table/calculating_prevalence.htm

Chakravarti, B., and Dr. Battacharyya, S. (2006). Mobile Based Clinical Decision Support System, Provider & Public Health Group,1-7. Last Accessed: April 12, 2012 from http://www.sbbhattacharyya.info/Mobile%20Based%20Clinical%20Decision%20Support%20System.pdf

Champlain LHIN. (2012). Champlain Diabetes Services. Last accessed: May, 4, 2013 from http://www.champlaindrcc.ca/health_care_providers.html

Chen, D., and Burrel, P. (2011). Case-Based Reasoning System and Artificial Neural Networks: A Review. *Neural Comput& Application.* 10, 264–276

Child Health Research Project Special Report. (1999). *Reducing Perinatal and Neonatal Mortality.* 3, 2-8.

Chronaki, C., et al. (1997). A Web-based Environment for Medical Collaboration in a Regional Healthcare Network.*Proceedings of Car 97,* 656-661.

Conde-Agudelo, A. and Romero, R. (2010). Cerivicovaginal Fetal Fibronecting for the Prediction of Spontaneous Preterm Birth in Twins Pregnancies: A Systematic Review and Meta-Analysis. *Maternal Fetal Neonatal Analysis.* 23(12):1365-1376.

Conde-Agudelo, A., Romero, R., and Yeo, L. (2010). Transvaginal sonographic cervical length for the prediction of spontaneous preterm birth in twin pregnancies: a systematic review and metaanalysis. American Journal of Obstetrics and Gynecology. 203(2):1281-12

Conference Board of Canada. (2001). *Infant Mortality.* Last accessed: December17, 2011 from http://www.conferenceboard.ca/hcp/Details/Health/infant-mortality-rate.aspx

Cotea, C., and Jiwani, S. (2003). A Case-Based Reasoning (CBR) System For the Neonatal Intensive Care Unit (NICU). Ottawa: University of Ottawa.

Devedzic, V. (2000). Knowledge Discovery and Data Mining in Databases.*School of Business Administration, University of Belgrade, Yugoslavia.*Last accessed: December17, 2011 from ftp://cs.pitt.edu/chang/handbook/34.pdf

Domingo-Ferrer, J., and Solanas, A. (2008). A measure of variance for hierarchical nominal attributes. *Information Sciences.* 178:4644-4655

Dunn, S. et al., (2011). 2008 Niday Perinatal Database quality audit: report of a quality assurance project. *Chronic Diseases and Injuries in Canada.*32(1):32-42

Durai, S.A, and Saro, E.A. (2006). Image Compression with Back-Propagation Neural Network using Cumulative Distribution Function. World Academy of Science, Engineering and Technology. Last Accessed: April 12, 2012 from http://www.waset.org/journals/waset/v17/v17-12.pdf.

EMC. (2006). Knowledge Management and Collaboration Platform for HealthCare Enviroments. Last accessed: December17, 2011 from http://www.emc.com/collateral/software/solution-overview/h2211-healthcare-doc-knowledge-mgt-collab-platform-ss.pdf

Eom, J.H., Kim, S.C., and Zhang, B.T. (2008). AptaPEDSS-E: A classifier ensemble-based perinatal decision support system for cardiovascular disease level prediction, *Expert System with Application,* 34, 2465-2479.

EORLA – Eastern Ontario Regional Laboratory Association. (2010). EORLA Winder Communication. Last accessed: May, 4, 2013 from http://www.eorla.ca/articles/10/02/01/eorla-winter-communication-february-2010

Estrabrooks, A. (2000). A Combination of Scheme For Inductive Learning From Imbalanced Data Sets. MSc. diss. Dalhouise University.

Fayyad, U. (1996). From Data Mining to Knowledge Discovery in Databases.*American Association for Artificial Intelligence.*Last accessed: December17, 2011 fromhttp://www.aaai.org/aitopics/assets/PDF/AIMag17-03-2-article.pdf

Federal Trade Commission. (2012). Federal Trade Commission Privacy Impact Assessment for the: SharePoint 2010 Pilot. Last accessed: May, 4, 2013 from http://www.ftc.gov/os/2012/04/1204sharepointpia.pdf

Fernandez-Carrocera, L.A., Guevara-Fuentes, C.A. and Salinas-Ramirez, V. (2011). Risk

Factors Associated with Mortality in Infants Weighing Less Than 1500g using CRIB II Scale. Boletin Medico Del Hospital Infantil De Mexico. 68(5):330-336.

Fairchild K.D., and Aschner J.L. (2012). HeRO Monitoring to Reduce Mortality in NICU Patients. Research and Reports in Neonatology. 2:65-75.

Frawley, W.J., et al. (1992). Knowledge Discovery in Databases: An Overview. *AI Magazine.* Last accessed: December17, 2011 fromhttp://www.aaai.org/ojs/index.php/aimagazine/article/view/1011/929

Frize, M., Weyand, S., Bariciak, E., Dunn,S.and S.Tozer. Combined PhysicianParent Decision Supporttool for the Neonatal Intensive Care Unit. Proc.MeMeA 2011, Bari, Italy.

Frize, M., Yu, and Weyand, S. (2011). Effectiveness of a hybrid pattern classifier for medical applications. *Int.J.HybridIntell.Syst.* 8, 71-79

Frize, M., Yang, L., Walker, R.C., and O'Conner, A.M. (June 2005). Conceptual Framework of Knowledge Management for Ethical Decision-Making Support in Neonatal Intensive Care, *IEEE Transanctions on Information Technology in Biomedicine*, 9, 205-215

Fujitsu Comuputer Products of America Inc. (2010). HealthCare Document Imaging Trend Report. Last accessed: December 17, 2011 from http://www.fujitsu.com/no/Images/healthcare-document-imaging-trend_wp.pdf

Gera, T. and Ramji, S. (2001). Early Predictors of Mortality in Very Low Birth Weight Neonates. *Indian Pediatrics*. 38:596

Gilchrist, J. (2012). Clinical Decision Support System Using Real-Time Data Analysis for a Neonatal Intensive Care Unit. PhD. diss. Carleton University.

Goldenberg, et al. (2008). Epimeiology and causes of preterm birth. *The Lancet,* 371,75-84

Griffin, M.P., and Mooreman, J.R. (2005). Heart Rate Characteristics and Laboratory Tests in Neonatal Sepsis. Journal of the American Academy of Pediatrics. 115:937-41.

Hannon Hill Corporation. (2010). Content Management For Healthcare Organization. Last accessed: December 17, 2011 from http://www.hannonhill.com/downloads/pdf/white-papers/Healthcare_and_CMS.pdf

Heidi L., et al. (2011). "Heart Rate Observation System effectively reduces NICU infant mortality", The Journal of Pediatrics. Last Accessed: Feburary 10, 2013 from http://www.news-medical.net/news/20110826/Heart-Rate-Observation-System-effectively-reduces-NICU-infant-mortality.aspx

HERO. Medical Predictive Science Corporation. (2011). Last Accessed: Feburary 10, 2013 from http://www.heroscore.com/hero.htm

Hietala, J. (2010). Security Concern and Approaches for SharePoint. *Planet Technologies.* Last accessed: May, 4, 2013 from http://www.cospug.com/Sessions/SharePoint%20Security%20Concerns%20and%20Approaches.pdf

Huser, V. et al. (2011). Implementation of workflow engine technology to delivery basic clinical decision support functionality. *BMC Medical Research Methodology, 11, 1-19*

Hueston, W.J. (1998). Preterm Contractions in Community Settings: II. Predicting Preterm Birth in Women Pregnant with Preterm Contractions. *Obstetrics and Gynecology.* 92(1):43-36.

Hunter, J., et al. (2003). NEONATE: Decision Support in the Neonatal Intensive Care Unit – A Preliminary Report. *Artificial Intelligence in Medicine.* 2780:41-45.

IBM InfoSphere. (2010). Leveraging key data to provide proactive patient care. Last accessed: May, 4, 2013 from http://www-01.ibm.com/software/success/cssdb.nsf/CS/SSAO-8BQ2D3?OpenDocument&Site=corp&cty=en_us

InfoSphere Streams for Smarter Healthcare. (2013). *IBM InfoSphere.* Last accessed: May, 4, 2013 from http://www-01.ibm.com/software/data/infosphere/stream-computing/smarter-healthcare.html

Intapong, P., et al. (2010). Modular Web-Based Collaboration Platform.*International Journal of Advanced Science and Technology*, 22, 27-46.

Institute Advanced Medical Education. (2013). Echogenic Fetal Kidneys: Differential Diagnosis and Postnatal Outcome. Last Accessed: January 21, 2013 from https://iame.com/online/cervical_length/content.php

Jayalakshmi, T., and Dr. Santhakumaran, A. (February 2011). Statistical Normalization and Back Propagation for Classification. *International Journal of Computer Theory and Engineering.* 3, 1793-8201

Jha, G.K. Aritificial Neural Networks. (2011). Date Last Accessed: April 12th, 2012 http://www.iasri.res.in/ebook/EB_SMAR/e-book_pdf%20files/Manual%20IV/3-ANN.pdf

Kid's Health. (2011). About Preemies. Last accessed: December17, 2011 from http://kidshealth.org/parent/growth/growing/preemies.html

Lalkhem, A.G., and McCluskey, A. (2008). Clinical tests: sensitivity and specificity. *Continuing Education in Anaesthesa.Critical Care &Pain,* 8, 221-223

Lauer, J.A., et al. (2003). Determinants of Ceasarean section rates in developed countries: supply, demand and opportunities for control. World Health Report. Last accessed: December17, 2011 fromhttp://www.who.int/healthsystems/topics/financing/healthreport/29DeterminantsC-section.pdf

Lee, K.A., and Gay,C.L. (2004). Sleep in late pregnancy predicts length of labour and type of delivery. *American Journal of Obstetrics and Gynecology.* 191(6):2041-6.

Lim, G. et al. (2009). Hospital Costs for Preterm and Small-or-Gestational Age Babies in Canada. *Healthcare Quarterly.* 12, 21-24

Lim, A.C et al. (2011). Cervical Length Measurement for the Prediction of Preterm Birth in Twins Pregnancies: A Systematic Review and Bivariate Meta-analysis. *Ultrasound in Obstetrics and Gynecology.* 38(1):10-17.

Lim, G. et al. (2009). Hospital Costs for Preterm and Small-or-Gestational Age Babies in Canada. *Healthcare Quarterly.* 12, 21-24

Lisbboa, P.J.G. (2011). A review of evident of health benefit from artificial neural networks in medical intervention. *Neural Networks.* 2,11-39

Lurie,I. (2002) A Web Content Management Blueprint. Portent. Last accessed: December17, 2011 from http://www.portent.com/library/cmsexplained.pdf

MacDorman, M.F., et al. (2011), Infant Mortality. *CDC's Public Health and Surveillance for Women, Infants and Children.* Last accessed: December17, 2011 from http://www.cdc.gov/reproductivehealth/ProductsPubs/DatatoAction/pdf/birout6.pdf

Mackinnon, M.J. and Glick, N. (1999). Data Mining and Knowledge Discovery in Databases – An Overview. *Australian Statistical Publishing Associatio. 31, 225-275*

Maimon, O., and Rokach, L. (2005). Introduction to Knowledge Discovery in Databases. *DATA MINING AND KNOWLEDGE DISCOVERY HANDBOOK.* Last accessed: December17, 2011 from http://www.springerlink.com/content/g1384q4ww000875t/

Maiya, P., Nagashree, S., and Shaik, M.S. (2001). Role of Score for Neonatal Acute Physiology (SNAP) in Predicting Neonatal Mortality. *Indian J Pediatrics.* 68(8):829-834

Mella, M. and Berghella, V.(2009). Prediction of Preterm Birth: Cervical Sonography. Seminars in Perinatology. 317-323

Microsoft Pinpoint. (2010). Champlain LHIN & Winchester District Memorial Hospital Physician Collaboration Portal. Last accessed: May, 4, 2013 from http://pinpoint.microsoft.com/en-CA/services/navantis-4294997373-4295480255

Milan, A. (2011). Mortality Overview, 2007. Last accessed: January 21, 2013 from http://www.statcan.gc.ca/pub/91-209-x/2011001/article/11524-eng.htm

Miller, K. (2010). Clinical Decision Support. Biomedical Computation Review. Last accessed: December17, 2011 from http://biomedicalcomputationreview.org/6/1/7.pdf

Mohkam, M. et al. (2010). A Comparison of CRIB, CRIB II, SNAP, SNAP II, and SNAP-PE Scores for the Prediction of Mortality in Critically Ill Neonates. *Medical Journal of Islamic Republic of Iran.* 24(4)193-199.

Musen, M.A., Shahar, Y., and Shorliffe, E.H. (2006). Clinical Decision Support Systems. *Biomedical Informatics: Computer Applications in Healthcare and Biomedicine.* New York: Springer. 600-631

Myles, A.J. (2004). An Introduction to Decision Tree Modeling. *Journal of Chemometrics,* 18, 275-285.

N.Allen., et al. (2002).Preterm Birth: Making a Difference. Last accessed: December17, 2011 from http://www.beststart.org/resources/rep_health/pdf/Preterm_mstr.pdf

National Union of Public and General Employees. (2011). OECD: Canada's Infant Mortality Rate Worsens. Last accessed: December17, 2011 fromhttp://www.nupge.ca/content/3272/oecd-canadas-infant-mortality-rate-worsens

National Union of Public and General Employees. (2011). *OECD: Canada's Infant Mortality Rate Worsens.* Last accessed: December17, 2011 fromhttp://www.nupge.ca/content/3272/oecd-canadas-infant-mortality-rate-worsens

NEONATE Effective Decision Support in the Neonatal Intensive Care Unit. ESRC/EPSRC PACCIT Programme. (2003). Last Accessed: Feburary 10, 2013 from http://inf.abdn.ac.uk/research/neonate/

Niday Perinatal Database Project. (2008). Annual Perinatal Statistical Report 2007-08. Last Accessed: April 12, 2012 from https://www.nidaydatabase.com/info /pdf/PPPESO%20Annual%20Perinatal%20Statistical%20Report%202007-08.pdf,

Norton. A, (2010). Infertility treatments may raise preterm birth risk. *Reuters.* Last accessed: May, 4, 2013 from http://www.reuters.com/article/2010/03/18/us-infertility-treatments-idUSTRE62H3S020100318

O'Conner, P.J. et al. (2011). Impact of Electronic Health Record Clinical Decision Support on Diabetes Care: A Randomized Trial". *Annals of Family Medicine.* 9, 12-21

Office of the Commissioner of Lobbying of Canada. (2010). Privacy Act Annul Report 2010-2011. Last accessed: May, 4, 2013 from https://ocl-cal.gc.ca/eic/site/012.nsf/eng/00478.html

Oliveira, T.A., et al. (1998). Fetal Fibronectin as a Predictor of Preterm Delivery in Twin Gestations. *International Journal of Gynaecology and Obstetrics.* 62(2):135-139

Oliveira, T.A., et al. (1999). Detection of Fetal Fibronectin in Twin Pregnancies in Relation to Gestational Age. *Sao Paulo Med.* 117(3):121-124

Ontario's maternal newborn and early child development resource centre. (2004). *Preterm Labour Signs & Symptom..*Last accessed: December17, 2011 from http://www.beststart.org/resources/rep_health/pdf/preterm_broch_eng_04.pdf

Oracle Data Mining Concepts. (2005). Classification. Last Accessed: April 12, 2012 from http://docs.oracle.com/cd/E11882_01/datamine.112/e16808/classify.htm

Patient Co.Uk. (2010). *Stillbirth and Neonatal Death.*Last accessed: December17, 2011 from http://www.patient.co.uk/doctor/Stillbirth-and-Neonatal-Death.htm

Peleg, M., and Tu, S. (2011). Decision Support, Knowledge Representation and Management in Medicine. Stanford Centre for Biomedical Informatics Research. Last accessed: December17, 2011 from http://bmir.stanford.edu/file_asset/index.php/1009/SMI-2006-1088.pdf

Peng, W., et al. (2006). An Implementation of ID3 – Decision Tree Learning Algorithm. *University of New South Wales*. Last accessed: December17, 2011 from http://web.arch.usyd.edu.au/~wpeng/DecisionTree2.pdf

Pepler, P.T., Uys, D.W. and Nel, D.G. (2012). Predicting Mortality and Length-of-Stay for Neonatal Admissions to Private Hospital Neonatal Intensive Care Units: a Southern African Retrospective Study. *African Health Sciences*. 12(2):166-173.

Public Health Agency of Canada. (1996). *Measuring Up – A Health Surveillance Update on Canadian Children and Youth*. Last accessed: December17, 2011 fromhttp://www.phac-aspc.gc.ca/publicat/meas-haut/mu_c-eng.php

Rasmussen, L.V., Oberg, R., and Starren, J.B., (2011). Implementaton of workflow engine technology to deliver basic clinical decision support functionality. BMC Medical Research Methodology. Last Accessed: April 12, 2012 from http://www.biomedcentral.com/content/pdf/1471-2288-11-43.pdf

REALTROMIS, Inc. – Product Pipeline Analysis. GlobalData. (2013). Last Accessed: Feburary 10, 2013 from http://www.marketresearch.com/GlobalData-v3648/REALTROMINS-Product-Pipeline-7295746/

Rowely, D.L.,et al. Neonatal and Postneonatal Mortality. (2011). *CDC's Public Health and Surveillance for Women, Infants and Children.* Last accessed: December 17, 2011 from http://www.cdc.gov/reproductivehealth/ProductsPubs/DatatoAction/pdf/birout7.pdf

Rulequest Research. (2012). See5: An Informal Tutorial. Date Last Accessed: April 12th, 2012 http://www.rulequest.com/see5-win.html#.names

Rulequest Research. (2010). Is See5/C5.0 better than C4.5? Last Accessed: February 10, 2013 from http://rulequest.com/see5-comparison.html

Ruiz, R.J., Fullerton J., and Brown C.E.L., (2004). The Utility of fFN for the Prediction of Preterm Birth in Twin Gestations. *Journal of Obstetrics, Gynecology and Neonatal Nursing.* 33(4):446-454.

Rumeau-Rouquette, C., et al. (1974). Prediction of perinatal mortality in early pregnancy. *J.Perinat Med.* 2, 196-207.

Rybchynski, Dawn, Design of an Artificial Neural Network Research Framework to Enhance the Development of Clinical Prediction Models. MASc Thesis, University of Ottawa, 2005.

Schiff, M., and Rogers, C. (1999). Predicting Caesarean Delivery for American Indian Women in New Mexico. *BIRTH.* 26, 226-231

Sectiono, R., and Leow.,W.K. On Mapping Decision Trees and Neural Networks.School of Computing, National University of Singapore. Last accessed: April 29, 2012 from http://www.sciencedirect.com/science/article/pii/S095070519900009X

Silins, J., et al. (1985). Risk factors for perinatal mortality in Canada. *Canadian Medical Association,*133, 1214-1219.

Singer, E. et al. (2007). Accuracy of Fetal Fibronectin to Predict Preterm Birth in Twin Gestations with Symptoms of Labor. Department of Obstetrics and Gynecology and Division of Maternal Fetal Medicine. 109(5):1083-1087.

Sitecore. (2009). Choosing the Best Web Content Management Solution. Last accessed: December17, 2011 fromhttp://www.awareweb.com/~/media/files/aware/com/docs/ChoosingtheBestWebCMS.pdf

Srihari, S. (2007). Introduction to Pattern Recognition. *Department of Computer Science and Engineering, University of Buffalo.* Last Accessed: January 21, 2013 from http://www.cedar.buffalo.edu/~srihari/CSE555/

Souka, A.P., et al. (1999). Cervical Length at 23 weeks in Twins in Predicting Spontaneous Preterm Delivery. *Obstet Gynecol.* 94(3):450-454

Sprawls, P. (2000). Image Characteristics and Quality. *Sprawls Educational Foundation.* Last accessed: May, 4, 2013 from http://www.sprawls.org/ppmi2/IMGCHAR/

Svensson, A.C., et al. (2009). Maternal Effects for Preterm Births: A Genetic Epidemiologic Study of 630,000 Families. American Journal of Epidemiology. 170(11):1365-72.

Tan P. C. et al. (2008). Predictors for Caesarean delivery and neonatal admission after trial of labour in women with one previous lower segment caesarean scar. *Singapore Med.* Last accessed: August 20th, 2012from http://smj.sma.org.sg/4903/4903a1.pdf

Tanner, L. et al. (2008). Decision Tree Algorithm Predict the Diagnosis and Outcome of Dengue Fever in the Early Phase of Illness. *Neglected Tropical Diseases. 2, 1-8*

The Conference Board of Canada. (2001). Infant Mortality. Last accessed: December17, 2011 fromhttp://www.conferenceboard.ca/hcp/Details/Health/infant-mortality-rate.aspx

The Daily Motion. (2008). *Deaths.* Last accessed: April29, 2012 from.http://www.statcan.gc.ca/daily-quotidien/080114/dq080114b-eng.htm.

The Health Council of Canada. (2012). Building Access to Specialist Care through e-Consultation: The Champlain BASE Project. Last accessed: May, 4, 2013 from http://innovation.healthcouncilcanada.ca/innovation-practice/building-access-specialist-care-through-e-consultation-champlain-base-project

The Mothers Program. (2013). BORN – Better Outcomes Registry & Network. Last Accessed: January 21, 2013 from http://www.themothersprogram.ca/resources-and-information/born

The Star. (2012). Ontario midwives want birth centres for low-risk pregnancies. Last accessed: August 19th, 2012 from http://www.thestar.com/living/article/1135862--ontario-midwives-call-for-birth-centres-for-low-risk-pregnancy-and-infant-delivery

UNICEF. (2010). Statistics. Last Accessed: January 21, 2013 from http://www.unicef.org/infobycountry/canada_statistics.html

Weirich, C. et al. (2005). Neonatal Mortality in Intensive Care Units of Central Brazil. *Rev Saude Publica.* 39(5):775-781

Wennerholm, UB. et al. (1997). Fetal fibronectin, endotoxin, bacterial vaginosis and cervical length as predictors of preterm birth and neonatal morbidity in twin pregnancies. British Journal of Obstetrics and Gynecology. 104(12):1398-1404.

Weyand, A. (2011). Development and Usability Testing of a Neonatal Intensive Care Unit Physician-Parent Decision Support Tool (PPADS). M.A.Sc. diss. University of Ottawa.

Weyand, S., Frize,M., Bariciak, E., and S. Dunn. Development and Evaluation Combined Physician-Parent Decision-Support Tool for the Neonatal Intensive Care Unit. Proc. EMBC, Boston, 2011.

World Health Organization Press. (2006). Neonatal and Perinatal Mortality. Last accessed: December 17, 2011 from http://whqlibdoc.who.int/publications/2006/9241563206_eng.pdf

Yu, N. (2009). An Integrated Decision Tree-Artificial Neural Network Hybrid to Estimate Clinical Outcomes: ICU Mortality and Pre-Term Birth. M.A.Sc. diss. Carleton University.

Zardo, M.S. and Procianoy, R.S. (2003). Comparison Between Different Mortality Risk Scores in a Neonatal Intensive Care Unit. Rev Saude Publica: 37(5):591-596.

Zupan, J. (1994). Introduction to Artificial Neural Network (ANN) Methods: What They Are and How to Use Them. *ActaChimicaSlovenica,* 327-352.

APPENDIX

A 1.0 INFANT MORTALITY

The World Health Organization (WHO) predicts that the infant mortality rate around the world is ~9 million, where nearly all deaths occur in developing countries. The infant mortality rate in Canada is shockingly high in relation to its level of socio-economic development. Although the mortality rate has declined over the past few decades, other countries of similar socio-economic status have done better (Conference Board of Canada, 2001). Preterm newborns of low birth weight (LBW) weighing less than 1500g are at increased risk of infant mortality and result in nearly 70% of neonatal death in developing countries (Child Health Research Project Special Report, 1999).

A 1.1 CLASSIFICATION

Infant mortality (<1 year) is further classified into perinatal (>22 weeks of gestation) which includes still births (fetus death, born dead), neonatal (<28 days) and postnatal (28-364 days) periods. Neonatal period can be further subdivided into early neonatal (<7 days) and late neonatal (7-<28 days) periods (Rowley et al., 2011). Perinatal mortality commonly occurs as a result of congenital malformations, pregnancy-related complications (placenta previ or abruption placentae), delivery related complications (intrapartum asphyxia), birth trauma and infectious disease (Child Health Research Project Special Report, 1999). Neonatal mortality typically occurs as a result of surrounding events during the prenatal period and delivery. Neonatal mortality and perinatal mortality account for approximately 70% of all infant deaths. Postnatal mortality accounts for the remaining 30% of all infant deaths and occurs as a result of events that arise after delivery such as environmental factors (Rowely et al., 2011).

Infant mortality may occur due to a combination of risk factors including the maternal health, accessibility to medical care, and socioeconomic conditions. Thus, the infant mortality rate is an important indicator which represents the overall health of the nation (MacDorman, 2011).

A 1.2 FACTORS CONTRIBUTING TO INFANT MORTALITY

Infant mortality is strongly correlated to structural factors including economic development, general living conditions, social well-being and quality of the environment (Conference Board of Canada, 2001). Low birth weight has also been linked to increased mortality and morbidity rates among infants (Public Health Agency of Canada, 1996). The infant mortality rate among the aboriginal community in Canada is twice as high compared to non-aboriginal populations (National Union of Public and General Employees, 2011). Researchers have excluded genetics as being a factor in increased infant mortality rates among natives, and pointed to common socioeconomic factors among these groups as risk factors instead, which include low household income, poor water quality, substandard housing and lack of healthcare (CBC News, 2009).

Maternal factors which contribute to the risk of infant mortality include age, education, marital status, family income, access to medical care, and substance abuse (cigarettes, alcohol and drugs) during pregnancy. Other variables related to infant mortality include birth order, previous history of infant or fetal loss, adequacy of prenatal care, total gestational period, birth weight, Apgar score, and plurality (MacDorman, 2011). Please refer to appendix F 2.0 Table E.1 for a complete list of risk factor variables associated with infant mortality available in the BORN and Niday database.

A 1.2.1 MATERNAL HEALTH & BEHAVIOURAL CHARACTERISTICS

Maternal health factors that may increase the risk of infant mortality include the following: mother's BMI<30, chronic diseases (i.e. diabetes, renal failure, hypertension, haemoglobinopathy, rhesus disease, thrombophialsis, antiphospholipid syndrome), infection (erythmiainfectriosum, varicella, measles) and substance abuse (i.e tobacco, alcohol and drugs) (Patient Co.Uk, 2010).

A 1.2.2 DEMOGRAPHIC CHARACTERISTICS

Racial and ethnic differences also play a role in infant mortality. However, race and ethnicity is not an etiological risk factor. Certain racial and ethnic groups are exposed to various risk factors socially, culturally, environmentally and economically which may put their newborns at greater risk of infant mortality. The infant mortality rate is highest among African, Native American, and Puerto Rican mothers, and is lowest among Asian mothers (MacDorman, 201).

A 2.0 Risk Factors Associated with Infant, Neonatal, Perinatal, and Postnatal Mortality

Table A.1 presents a categorized list of potential risk factors associated with infant, neonatal, perinatal and postnatal mortality derived from medical literature found in the BORN, and Niday (Rowely et al., 2011):

Table A.1 - Potential risk factors associated w/ infant, neonatal, perinatal & postnatal mortality

Risk Factor	Available before 23 week Gestation	Factor Present in	
		BORN	Niday
BMI <30			
Low Birth Weight		X	X

Diabetes			
Renal Failure			
Hypertension			
Haemoglobinopathy			
Rhesus disease			
Antiphospholipid Syndrome),			
Varicella			
Measles			
Alcohol		X	X
Tobacco			
Drugs		X	
Living Conditions			
Quality of the Environment			
Social Well Being			
Low Household Income			
Poor Water Quality			
Substandard Housing			
Lack of Healthcare			
Ethnicity - Black			
Ethnicity - Native American			
Ethnicity - Puerto Rican			
Birth Weight <1000g		X	X
Birth Weight >3000-4500g			
Complications on Delivery		X	X
Congenital Anomalies			
Birth Defects			
Unskilled Labour Class			
Rural Area			
<20 Years of Age			
Multiparous (parity of 2)		X	X
Complications of the Placenta		X	X
Short Gestation			
Low Birth Weight		X	X
Poor Hygiene			
Complications Cord or		X	X
Membrane			
Respiratory Distress Syndrome			
Sudden Infant Death Syndrome			
Infections			

A 2.1 Infant Mortality Ratios and Rates

Fetal death ratio

$$\frac{\text{Fetal deaths}}{\text{Live births}} \times 1000$$

Fetal death rate

$$\frac{\text{Fetal deaths}}{\text{Total births}} \times 1000$$

Fetal death rate, weight-specific

$$\frac{\text{Fetal deaths weighing 1000g and over}}{\text{Total births weighing 1000 g and over}} \times 1000$$

Early neonatal mortality rate

$$\frac{\text{Early neonatal deaths}}{\text{Live births}} \times 1000$$

Early neonatal mortality rate, weight-specific

$$\frac{\text{Early neonatal deaths of infants weighing 1000 g and over at birth}}{\text{Live births weighing 1000 g and over}} \times 1000$$

Perinatal mortality ratio

$$\frac{\text{Fetal deaths and early neonatal deaths}}{\text{Live births}} \times 1000$$

Perinatal mortality rate

$$\frac{\text{Fetal deaths and early neonatal deaths}}{\text{Total births}} \times 1000$$

Perinatal mortality rate, weight-specific

$$\frac{\text{Fetal deaths weighing 1000 g and over, plus early neonatal deaths of infants weighing 1000 g and over at birth}}{\text{Total births weighing 1000 g and over}} \times 1000$$

Neonatal mortality rate

$$\frac{\text{Neonatal deaths}}{\text{Live births}} \times 1000$$

Neonatal mortality rate, weight-specific

$$\frac{\text{Neonatal deaths of infants weighing 1000 g and over at birth}}{\text{Live births weighing 1000 g and over}} \times 1000$$

Infant mortality rate

$$\frac{\text{Deaths under one year of age}}{\text{Live births}} \times 1000$$

Infant mortality rate, weight-specific

$$\frac{\text{Infant deaths among live births weighing 1000 g and over at birth}}{\text{Live births weighing 1000 g and over}} \times 1000$$

B 1.0 PRETERM BIRTH

B 1.1 EPIDEMIOLOGY OF PRETERM BIRTHS

Obstetric signs that have led to preterm birth include (1) delivery for maternal or fetal indications where the labour is induced or delivered through caesarean section, (2) spontaneous preterm birth with intact membrane, and (3) preterm premature rupture of the membrane (PPROM) irrespective of vaginal or caesarean section delivery. It is estimated that 30-35% of preterm births are indicated, 40-45% are spontaneous preterm labour, and 25-30% tail PPROM. The contribution of factors associated with preterm births varies by ethnic groups. It has been observed that spontaneous preterm birth commonly occurs in white women, whereas PPROM predominates in black women (Goldenberg et al., 2008).

B 1.1.1 HEALTH OF PRETERM INFANTS

Babies born prematurely are prone to several health problems due to incomplete development, and are at higher risk of morbidity and mortality compared to full-term babies of normal weight. Often, the earlier the baby is born is proportional to an increased risk in the severity of long and short term complication. Long term complication may include blindness, trouble walking and learning challenges (Alere's Women's & Children's Health, 2008). Short term complications vary and may include respiratory distress, internal bleeding, poor circulation, liver failure, seizures, temperature instability, gastrointestinal complications etc... (Yu, 2009). Preterm babies are often transferred to the Neonatal Intensive Care Units (NICU) shortly after birth (Ontario's maternal newborn and early child development resource centre, 2004). The average hospital cost of a baby admitted to NICU in 2002-2003 was $9700 (Canadian Institute

for Health Information, 2006). NICU provide preterm infants with proper temperature and nutrition needed for proper growth and development (Kid's Health, 2011).

B 1.1.2 IMPACT ON SOCIETY

Preterm babies often tend to use more hospital resources and require special monitoring and care during the first critical days/weeks in their lives. Often the use of specialized equipment including respirators, monitors, intravenous pumps and kidney dialysis machines are required for the survival of preterm infants. The length of stay of preterm infants is typically higher than term infants. All of these factors contribute greatly to increased healthcare costs. Further, the differential health consequences of each preterm baby typically extend far beyond infancy and childhood (Lim et al., 2009).

Often, hospital costs decrease as the birth weight and gestational age increase, where the smallest infants commonly require the longest length of stay and thus have the highest hospital costs. Newborns weighing 2,500g costs approximately $1000 in hospital resources and on average required only 2 hospital days, whereas newborns weighing less than 750g cost more than $117 000 in hospital resources and on average required more than 104 hospital days. The average hospital cost for twins born preterm is considerably higher than singleton birth (Goldenberg et al., 2008). It is estimated that in Canada, each preterm baby will use $676 800 in healthcare services throughout its lifetime (Allen et al. 2002).

Preterm babies who survive with a disability will require extensive community resources and support to achieve optimal quality of life including educational support, social services, respite care for the family and supportive housing and transportation (Allen et al. 2002). A preterm infant surviving a disability may also require special assistive devices such as a

wheelchair, where a portion of the costs may be reimbursed through various financial assistance programs, but in many cases, the family bears much of these costs (Allen et al. 2002).

In addition to the various health problems many preterm infants are prone to, the family of a premature baby undergo emotional distress due to the uncertainty of their baby's future and financial costs. Mothers often experience emotional distress and depression due to the birth and hospitalization of the baby. Parents also face challenges due to prolonged hospitalization of the baby which separates the baby from the parents during the critical newborn period (Allen et al. 2002). Further, the financial burden may impact and limit the family's social life and lead to difficulty in maintaining proper employment and income. The strength of the family is also challenged when taking care of a baby born preterm.

B 2.0 Risk Factors Associated with Preterm Births

Table B.1 lists a categorized list of potential risk factors associated with preterm birth derived from medical literature found in the BORN, Niday and PRAMS database:

Table B.1 - Potential risk factors associated with preterm birth

Risk Factor	Available before 23 week gestation	Factor Present in		
		[redacted]		[redacted]
Age <18 or >35				
Ethnicity [Afro-American]				X
Single				X
Education [Not Complete HS]				X
Low Income				X
Living in Poverty				X
Work Condition (standing >3hr)				
Stress		X		X
Previous PTB		X	X	X
History of Infertility Problems				X
Sister/Grandmother with PTB				

BMI High			
BMI Low <20			
Low Iron			
Low Folate			
Low Zinc			
Insufficient Weight Gain			
Diabetes	X	X	X
Chronic Hypertension			
Cardiac Disease	X		
Chronic Pulmonary Disease			
Asymptomatic Bacteriuria			
Genetic Tract Abnormalities			
History of STD.			
Abnormally Shaped Uterus			
Excessive Uterine Contraction			
>2 Third Trimester Abortions			
>2 Miscarriages			
Vaginal Spotting (light)			
Fetal Fibronectin			
Salivary Estriol			
Cervical Dilation <25 mmm			
Cervical Anomalies			
Bacterial Vaginosis (pH >4.5)			
Presence of Pyelonephritis			
Presence of Pneumonia			
Presence of Appendicitis			
Asthma	X		
Thyroid diseases			
Severe ARDS			
Severe Urinary Tract Infections			
Antepartum Hemorrhage			
Multiparous	X	X	X
Nulliparous			
Late or No Prenatal Care	X	X	
Vaginal Bleeding (1st Trimester)			
Vaginal Bleeding (2nd Trimester)			
Low Pregnancy Weight			
<62" Height			
Smoking	X	X	X
Exposure to Tobacco Smoke			
Alcohol Cons. Before 3rd Trimester			X
Cocaine			
Drug Use (Heroin, Marijuana)	X		

C 1.0 Decision Tree Results

C1.1 nOMOD_NEONATALMORT DT RESULTS

C5.0 Decision Tree Results

Cases with Mortality:	24	***Confusion Matrix***
Cases without Mortality:	232	***TP*** ***FP***
Total cases:	256	***FN*** ***TN***
Minority dataset:	**9%**	

Repeated 2-fold Cross Validation Results

	Train	Test	Std Dev
Accuracy:	1.00	**1.00**	0.00
Sensitivity:	0.62	**0.57**	0.07
Specificity:	1.00	**1.00**	0.00
PPV/Precision:	0.92	**0.85**	0.08
NPV:	1.00	**1.00**	0.00
Recall:	0.62	**0.57**	0.07
F1 score:	0.73	**0.68**	0.05
MCC:	0.75	**0.70**	0.05
SS1:	0.62	**0.57**	0.07

C 1.2 MOD_NEONATALMORT DT RESULTS

C5.0 Decision Tree Results

Cases with Mortality:	24	***Confusion Matrix***
Cases without Mortality:	232	***TP*** ***FP***
Total cases:	256	***FN*** ***TN***
Minority dataset:	**9%**	

Repeated 2-fold Cross Validation Results

	Train	Test	Std Dev
Accuracy:	0.9979	**0.9974**	0.00
Sensitivity:	0.7323	**0.7013**	0.04
Specificity:	0.9997	**0.9993**	0.00
PPV/Precision:	0.9380	**0. 8817**	0.08
NPV:	0.9983	**0.9981**	0.00
Recall:	0.7323	**0.7013**	0.04
F1 score:	0.8204	**0.7781**	0.03
MCC:	0.8262	**0.7830**	0.03
SS1:	0.7320	**0.7006**	0.04

C 1.3 noMOD_TWINGEST DT RESULTS

C5.0 Decision Tree Results

Cases with Mortality:	24	***Confusion Matrix***
Cases without Mortality:	232	***TP*** ***FP***
Total cases:	256	***FN*** ***TN***
Minority dataset:	**9%**	

Repeated 2-fold Cross Validation Results

	Train	Test	Std Dev
Accuracy:	0.8528	**0.8515**	0.0053
Sensitivity:	0.1585	**0.1542**	0.0301
Specificity:	0.9937	**0.9931**	0.0029
PPV/Precision:	0.8351	**0.8194**	0.0650
NPV:	0.8533	**0.8526**	0.0044
Recall:	0.1585	**0.1542**	0.0301
F1 score:	0.2661	**0.2585**	0.0435
MCC:	0.2997	**0.2906**	0.0394
SS1:	0.1522	**0.1473**	0.0298

C 1.4 MOD_TWINGEST DT RESULTS

MOD_MULTGEST RESULTS

Cases with Mortality:	24	***Confusion Matrix***
Cases without Mortality:	232	***TP*** ***FP***
Total cases:	256	***FN*** ***TN***
Minority dataset:	**9%**	

Repeated 2-fold Cross Validation Results

	Train	**Test**	Std Dev
Accuracy:	0.9443	**0.8983**	0.0074
Sensitivity:	0.9102	**0.7932**	0.0585
Specificity:	0.9512	**0.9197**	0.0120
PPV/Precision:	0.7928	**0.6685**	0.0251
NPV:	0.9813	**0.9566**	0.0117
Recall:	0.9102	**0.7932**	0.0585
F1 score:	0.8467	**0.7242**	0.0236
MCC:	0.8291	**0.6807**	0.0364
SS1:	0.8614	**0.7129**	0.0506

C 1.5 noMOD_SINGLETON DT RESULTS

C5.0 Decision Tree Results

Cases with Mortality:	24	***Confusion Matrix***
Cases without Mortality:	232	***TP*** ***FP***
Total cases:	256	***FN*** ***TN***
Minority dataset:	**9%**	

Repeated 2-fold Cross Validation Results

	Train	Test	Std Dev
Accuracy:	0.90	**0.67**	0.05
Sensitivity:	0.15	**0.12**	0.02
Specificity:	1.00	**0.74**	0.05
PPV/Precision:	0.86	**0.05**	0.01
NPV:	0.90	**0.87**	0.01
Recall:	0.15	**0.12**	0.02
F1 score:	0.25	**0.07**	0.01
MCC:	0.32	**-0.11**	0.03
SS1:	0.15	**-0.14**	0.04

C 1.6 MOD_SINGLETON DT RESULTS

C5.0 Decision Tree Results

Cases with Mortality:	24	***Confusion Matrix***
Cases without Mortality:	232	***TP*** ***FP***
Total cases:	256	***FN*** ***TN***
Minority dataset:	**9%**	

Repeated 2-fold Cross Validation Results

	Train	Test	Std Dev
Accuracy:	0.93	**0.28**	0.03
Sensitivity:	0.44	**0.36**	0.05
Specificity:	0.99	**0.27**	0.04
PPV/Precision:	0.85	**0.06**	0.01
NPV:	0.94	**0.77**	0.03
Recall:	0.44	**0.36**	0.05
F1 score:	0.54	**0.10**	0.01
MCC:	0.55	**-0.43**	0.08
SS1:	0.43	**-0.37**	0.04

D 1.0 Hybrid ANN Results

D 1.1 nOMOD_NEONATALMORT HYBRID ANN RESULTS (TRAIN)

Repeated 2-fold Cross Validation Results

	Verification	Std Dev	Confidence Interval (95%)
Accuracy:	0.6883	0.1722	0.11
Sensitivity/Recall:	0.3161	0.2048	0.13
Specificity:	0.7364	0.2202	0.14
PPV/Precision:	0.1495	0.0261	0.02
NPV:	0.8930	0.0059	0.00
F1 score:	0.1816	0.0161	0.01
MCC:	0.0467	0.0283	0.02
SS1:	0.0525	0.0304	0.02

Fold 1A		**Fold 1B**		**Fold 2A**		**Fold 2B**		**Fold 3A**	
11124	1430	3448	462	13020	1586	13047	1497	12738	1559
3342	485	11123	1346	1484	341	1486	349	1754	330
Accuracy:	0.709	*Accuracy:*	0.293	*Accuracy:*	0.813	*Accuracy:*	0.818	*Accuracy:*	0.798
Sens:	0.2533	*Sens:*	0.7445	*Sens:*	0.1770	*Sens:*	0.1891	*Sens:*	0.1747
Spec:	0.7690	*Spec:*	0.2366	*Spec:*	0.8977	*Spec:*	0.8977	*Spec:*	0.8790
PPV:	0.1267	*PPV:*	0.1079	*PPV:*	0.1868	*PPV:*	0.1902	*PPV:*	0.1583
NPV:	0.8861	*NPV:*	0.8818	*NPV:*	0.8914	*NPV:*	0.8971	*NPV:*	0.8910
F1 score:	0.169	*F1 score:*	0.189	*F1 score:*	0.182	*F1 score:*	0.190	*F1 score:*	0.166
MCC:	0.017	*MCC:*	-0.014	*MCC:*	0.076	*MCC:*	0.087	*MCC:*	0.051
SS1:	0.022	*SS1:*	-0.019	*SS1:*	0.075	*SS1:*	0.087	*SS1:*	0.054
Fold 3B		*Fold 4A*		*Fold 4B*		*Fold 5A*		*Fold 5B*	
12923	1504	*9758*	1134	*5972*	629	*12688*	1521	*12182*	1460
1622	330	*4810*	679	*8497*	1281	*1857*	315	*2310*	427
Accuracy:	0.809	*Accuracy:*	0.637	*Accuracy:*	0.443	*Accuracy:*	0.794	*Accuracy:*	0.770
Sens:	0.1799	*Sens:*	0.3745	*Sens:*	0.6707	*Sens:*	0.1716	*Sens:*	0.2263
Spec:	0.8885	*Spec:*	0.6698	*Spec:*	0.4127	*Spec:*	0.8723	*Spec:*	0.8406
PPV:	0.1691	*PPV:*	0.1237	*PPV:*	0.1310	*PPV:*	0.1450	*PPV:*	0.1560
NPV:	0.8958	*NPV:*	0.8959	*NPV:*	0.9047	*NPV:*	0.8930	*NPV:*	0.8930
F1 score:	0.174	*F1 score:*	0.186	*F1 score:*	0.219	*F1 score:*	0.157	*F1 score:*	0.185
MCC:	0.067	*MCC:*	0.029	*MCC:*	0.055	*MCC:*	0.041	*MCC:*	0.057
SS1:	0.068	*SS1:*	0.044	*SS1:*	0.083	*SS1:*	0.044	*SS1:*	0.067

D 1.2 nOMOD_NEONATALMORT HYBRID ANN RESULTS (TEST)

Repeated 2-fold Cross Validation Results

	Verification	Std Dev	Confidence Interval (95%)
Accuracy:	0.7931	0.0830	0.05
Sensitivity/Recall:	0.2185	0.1297	0.08
Specificity:	0.8655	0.1083	0.07
PPV/Precision:	0.1814	0.0572	0.04
NPV:	0.8984	0.0051	0.00
F1 score:	0.1862	0.0550	0.03
MCC:	0.0811	0.0531	0.03
SS1:	0.0840	0.0533	0.03

Fold 1A		**Fold 1B**		**Fold 2A**		**Fold 2B**		**Fold 3A**	
6587	717	7876	771	6543	757	6530	728	6526	765
700	18	6695	1037	711	179	717	216	724	175
Accuracy:	0.823	*Accuracy:*	0.544	*Accuracy:*	0.821	*Accuracy:*	0.824	*Accuracy:*	0.818
Sens:	0.0245	*Sens:*	0.57356	*Sens:*	0.1912	*Sens:*	0.2288	*Sens:*	0.1862
Spec:	0.9039	*Spec:*	0.54053	*Spec:*	0.9020	*Spec:*	0.9011	*Spec:*	0.9001
PPV:	0.0251	*PPV:*	0.13412	*PPV:*	0.2011	*PPV:*	0.2315	*PPV:*	0.1947
NPV:	0.9018	*NPV:*	0.91084	*NPV:*	0.8963	*NPV:*	0.8997	*NPV:*	0.8951
F1 score:	0.025	*F1 score:*	0.217	*F1 score:*	0.196	*F1 score:*	0.230	*F1 score:*	0.190
MCC:	-0.072	*MCC:*	0.072	*MCC:*	0.095	*MCC:*	0.131	*MCC:*	0.088
SS1:	-0.072	*SS1:*	0.114	*SS1:*	0.093	*SS1:*	0.130	*SS1:*	0.086
Fold 3B		*Fold 4A*		*Fold 4B*		*Fold 5A*		*Fold 5B*	
6530	755	*6493*	793	*6564*	724	*6538*	763	*6543*	754
719	187	*721*	183	*716*	187	*705*	183	*710*	184
Accuracy:	0.820	*Accuracy:*	0.815	*Accuracy:*	0.824	*Accuracy:*	0.821	*Accuracy:*	0.821
Sens:	0.1985	*Sens:*	0.1875	*Sens:*	0.2053	*Sens:*	0.1934	*Sens:*	0.1962
Spec:	0.9008	*Spec:*	0.9001	*Spec:*	0.9016	*Spec:*	0.9027	*Spec:*	0.9021
PPV:	0.2064	*PPV:*	0.2024	*PPV:*	0.2071	*PPV:*	0.2061	*PPV:*	0.2058
NPV:	0.8964	*NPV:*	0.8912	*NPV:*	0.9007	*NPV:*	0.8955	*NPV:*	0.8967
F1 score:	0.202	*F1 score:*	0.195	*F1 score:*	0.206	*F1 score:*	0.200	*F1 score:*	0.201
MCC:	0.101	*MCC:*	0.091	*MCC:*	0.107	*MCC:*	0.099	*MCC:*	0.100
SS1:	0.099	*SS1:*	0.088	*SS1:*	0.107	*SS1:*	0.096	*SS1:*	0.098

D 1.3 MOD_NEONATALMORT HYBRID RESULTS (TRAIN)

Repeated 2-fold Cross Validation Results

	Verification	Std Dev	Confidence Interval (95%)
Accuracy:	0.76	0.05	0.03
Sensitivity/Recall:	0.6544	0.0593	0.0367
Specificity:	0.8007	0.0842	0.0522
PPV/Precision:	0.5672	0.0905	0.0561
NPV:	0.8646	0.0081	0.0050
F1 score:	0.60	0.02	0.02
MCC:	0.44	0.05	0.03
SS1:	0.46	0.03	0.02

Fold 1A		**Fold 1B**		**Fold 2A**		**Fold 2B**		**Fold 3A**	
10113	1630	8524	1244	10399	1778	10212	1610	10942	1878
1872	2734	3434	3146	1528	2644	1804	2722	1034	2495
Accuracy:	0.786	*Accuracy:*	0.714	*Accuracy:*	0.798	*Accuracy:*	0.791	*Accuracy:*	0.822
Sens:	0.6265	*Sens:*	0.7166	*Sens:*	0.5979	*Sens:*	0.6283	*Sens:*	0.5705
Spec:	0.8438	*Spec:*	0.7128	*Spec:*	0.8719	*Spec:*	0.8499	*Spec:*	0.9137
PPV:	0.5936	*PPV:*	0.4781	*PPV:*	0.6337	*PPV:*	0.6014	*PPV:*	0.7070
NPV:	0.8612	*NPV:*	0.8726	*NPV:*	0.8540	*NPV:*	0.8638	*NPV:*	0.8535
F1 score:	0.610	*F1 score:*	0.574	*F1 score:*	0.615	*F1 score:*	0.615	*F1 score:*	0.631
MCC:	0.462	*MCC:*	0.388	*MCC:*	0.479	*MCC:*	0.472	*MCC:*	0.521
SS1:	0.470	*SS1:*	0.429	*SS1:*	0.470	*SS1:*	0.478	*SS1:*	0.484
Fold 3B		*Fold 4A*		*Fold 4B*		*Fold 5A*		*Fold 5B*	
9021	1363	*9164*	1435	*8473*	1198	*8059*	1148	*10956*	1842
2946	3018	*2791*	2959	*3515*	3162	*3896*	3246	*1032*	2518
Accuracy:	0.736	*Accuracy:*	0.742	*Accuracy:*	0.712	*Accuracy:*	0.691	*Accuracy:*	0.824
Sens:	0.6889	*Sens:*	0.6734	*Sens:*	0.7252	*Sens:*	0.7387	*Sens:*	0.5775
Spec:	0.7538	*Spec:*	0.7665	*Spec:*	0.7068	*Spec:*	0.6741	*Spec:*	0.9139
PPV:	0.5060	*PPV:*	0.5146	*PPV:*	0.4736	*PPV:*	0.4545	*PPV:*	0.7093
NPV:	0.8687	*NPV:*	0.8646	*NPV:*	0.8761	*NPV:*	0.8753	*NPV:*	0.8561
F1 score:	0.583	*F1 score:*	0.583	*F1 score:*	0.573	*F1 score:*	0.563	*F1 score:*	0.637
MCC:	0.407	*MCC:*	0.408	*MCC:*	0.389	*MCC:*	0.369	*MCC:*	0.527
SS1:	0.443	*SS1:*	0.440	*SS1:*	0.432	*SS1:*	0.413	*SS1:*	0.491

D 1.4 MOD_NEONATALMORT HYBRID RESULTS (TEST)

Repeated 2-fold Cross Validation Results

	Verification	Std Dev	Confidence Interval (95%)
Accuracy:	0.8333	0.0049	0.00
Sensitivity/Recall:	0.60741	0.00588	0.00
Specificity:	0.91300	0.00651	0.00
PPV/Precision:	0.71135	0.01638	0.01
NPV:	0.86834	0.00329	0.00
F1 score:	0.66	0.01	0.00
MCC:	0.55	0.01	0.01
SS1:	0.52	0.01	0.01

Fold 1A		**Fold 1B**		**Fold 2A**		**Fold 2B**		**Fold 3A**	
5530	799	5516	873	5488	814	5502	853	5584	818
545	1300	498	1288	591	1281	509	1311	476	1296
Accuracy:	0.836	*Accuracy:*	0.832	*Accuracy:*	0.828	*Accuracy:*	0.833	*Accuracy:*	0.842
Sens:	0.6193	*Sens:*	0.5960	*Sens:*	0.6115	*Sens:*	0.6058	*Sens:*	0.6131
Spec:	0.9103	*Spec:*	0.9172	*Spec:*	0.9028	*Spec:*	0.9153	*Spec:*	0.9215
PPV:	0.7046	*PPV:*	0.7212	*PPV:*	0.6843	*PPV:*	0.7203	*PPV:*	0.7314
NPV:	0.8738	*NPV:*	0.8634	*NPV:*	0.8708	*NPV:*	0.8658	*NPV:*	0.8722
F1 score:	0.659	*F1 score:*	0.653	*F1 score:*	0.646	*F1 score:*	0.658	*F1 score:*	0.667
MCC:	0.553	*MCC:*	0.548	*MCC:*	0.534	*MCC:*	0.553	*MCC:*	0.568
SS1:	0.530	*SS1:*	0.513	*SS1:*	0.514	*SS1:*	0.521	*SS1:*	0.535
Fold 3B		*Fold 4A*		*Fold 4B*		*Fold 5A*		*Fold 5B*	
5427	852	*5527*	852	*5547*	826	*5537*	833	*5521*	846
600	1296	*480*	1315	*533*	1269	*518*	1286	*509*	1299
Accuracy:	0.822	*Accuracy:*	0.837	*Accuracy:*	0.834	*Accuracy:*	0.835	*Accuracy:*	0.834
Sens:	0.6034	*Sens:*	0.6068	*Sens:*	0.6057	*Sens:*	0.6069	*Sens:*	0.6056
Spec:	0.9004	*Spec:*	0.9201	*Spec:*	0.9123	*Spec:*	0.9145	*Spec:*	0.9156
PPV:	0.6835	*PPV:*	0.7326	*PPV:*	0.7042	*PPV:*	0.7129	*PPV:*	0.7185
NPV:	0.8643	*NPV:*	0.8664	*NPV:*	0.8704	*NPV:*	0.8692	*NPV:*	0.8671
F1 score:	0.641	*F1 score:*	0.664	*F1 score:*	0.651	*F1 score:*	0.656	*F1 score:*	0.657
MCC:	0.525	*MCC:*	0.562	*MCC:*	0.546	*MCC:*	0.551	*MCC:*	0.552
SS1:	0.504	*SS1:*	0.527	*SS1:*	0.518	*SS1:*	0.521	*SS1:*	0.521