An Integrated Machine Learning Approach to Optimize the Estimation of Preterm Birth

by

Daphne Eu-Jhia Ong, B. Eng

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)

Carleton University
Ottawa, Ontario

July 2015

© 2015, Daphne Eu-Jhia Ong

Abstract

This thesis describes a new methodology used in conjunction with artificial intelligence tools to create multiple models for prediction of preterm birth in obstetrical environments. The data mining approach integrates: Decision Trees (DTs), Artificial Neural Networks (ANNs) - specifically a Feed Forward Back Propagation ANN, and Case Based Reasoning System (CBRS).

This work also introduces a 5by2 cross validation method, assesses two methods of attribute selection, and considers data prevalence (15% and 8.1%) in training and testing networks. Two databases were assessed from two countries: BORN (Canada) and PRAMS (USA).

Best BORN results used selection method 2, had sensitivities of 50.53%, 53.96%, specificities of 91.61%, 95.40%, and area under curves (AUC) of 0.7721, 0.7970 for Parous and Nulliparous cases respectively. Best PRAMS results used selection method 1, had sensitivities of 68.15%, 40.35%, specificities of 64.71%, 94.57%, and area under curves (AUC) of 0.8452, 0.7064 for Parous and Nulliparous cases respectively.

Above all, challenge yourself. You may well surprise yourself at what strengths you have, what you can accomplish. – Cecile Springer

Acknowledgements

First and foremost, I would like to take this opportunity to thank my thesis supervisor, Dr. Monique Frize, for her support, encouragement and guidance. I am very grateful to have had the opportunity to work with you.

I wish to express my sincerest thanks to my co-supervisor Dr. Jeff Gilchrist who always made himself available to answer my questions, provided rich comments, suggestions, and support, which were invaluable for the successful completion of my thesis work.

Another big thank you goes out to my co-supervisor, Dr. Erika Bariciak at the Children's Hospital of Eastern Ontario, for who also always made herself available to answer my questions, for providing useful critiques, for providing necessary information for this thesis work and her support.

Thank you to my family, especially my parents, Kenneth and Judy Ong, for their love and endless support with my academic pursuits; for encouraging me to persevere, think positive thoughts, see the light during its most challenging moments. Also, to my brother, Ryan, thank you for cheering on for me from afar while pursuing your own academic goals. And thanks to my Uncle Wade for always having an open door and for the advice.

I would like to thank the office staff in the Systems and Computer Engineering Department for all of the help and support they provided to me during my time as a graduate student, especially Jennifer Poll, Darlene Hebert, Jessie Cartwright and Jenna

Noseworthy. Another thank you goes out to the Technical Staff members (Danny, Daren, and Khi) in the Department especially when things went wrong with the lab machines!

To my Colleagues, Peers and Friends in IEEE, WIE, WISE, CUBES, EMBS, GSA, and the Canal, MacKenzie and Minto labs, thank you all for your support, encouragement, for sharing your time, your knowledge and wisdom, and for all the great memories and experiences. Thank for being my Carleton Family over the years.

Lastly a special deep thank you for your love, friendship, support, wisdom, and encouragement, in no particular order: HM, SC, SF, SE, AP, KR, MR, JS, JA, ADCC, CH, HL, VJ, EP, AN, RN, GC, MT, NL, RK, KS, AS, MF, DX, NF, LS, QS, AG, LM, JW, MR, SB, LR, AO, PQ, GC, MH and JPS.

Table of Contents

Abstract.....	ii
Acknowledgements	iv
Table of Contents	vi
List of Tables	xii
List of Figures.....	xvi
List of Appendices.....	xvii
List of Acronyms	xviii
1 Chapter: Introduction	1
1.1 Motivation.....	1
1.1.1 Healthcare Perspective.....	1
1.1.2 Technology Perspective	1
1.2 Problem Statement	3
1.3 The Importance of Delivering at Term.....	4
1.3.1 Rate of Preterm Birth.....	5
1.3.2 Health of Preterm Infants.....	6
1.3.3 Cost of Preterm Birth.....	6
1.3.4 Importance of Prediction Models in Perinatal Care	7
1.4 Thesis Objectives	7
1.5 Thesis Outline	8
2 Chapter: Literature Review	10
2.1 Preterm Birth.....	10
2.2 Preterm Birth Classification	11
2.3 Factors Contributing to Preterm Birth.....	12

2.3.1	Demographic and Socioeconomic	13
2.3.2	Behavioral Characteristics	13
2.3.3	Stress	14
2.3.4	Maternal Health	14
2.3.5	Pregnancy History/ Genetic Characteristics	15
2.3.6	Current Fetal Conditions/Pregnancy Characteristics	16
2.3.7	Biological Characteristics	16
2.4	Performance Measures	17
2.4.1	Confusion Matrix (Contingency Table).....	18
2.4.2	Correct Classification Rate (CCR).....	18
2.4.3	Error Rate.....	18
2.4.4	Sensitivity	18
2.4.5	Specificity	19
2.4.6	Likelihood Ratio	19
2.4.7	Prevalence	20
2.4.8	Positive Predictive Value.....	20
2.4.9	Negative Predictive Value (NPV).....	21
2.4.10	Receiver Operating Characteristic (ROC) Curve.....	21
2.4.11	Area Under the Curve (AUC).....	23
2.4.12	Normalization/ Standardization	23
2.4.13	Confidence in the Model.....	24
2.5	Current Knowledge	25
2.6	Critical Analysis for Preterm Birth	25
2.6.1	Risk Factor Scoring.....	25
2.6.2	Screening Test for Preterm Birth	26
2.6.3	Accuracy Prediction of Cervical Length.....	27

2.6.4	Accuracy Prediction fetal fibronectin (fFN)	29
2.6.5	Accuracy/Limitations of standard invasive test methods.....	31
2.7	Current Prediction Models for Pregnancy and Preterm Birth	32
2.7.1	Preterm Birth Prediction	32
2.7.2	Uterine Electromyography ANN Prediction.....	33
2.7.3	Early Prediction of Preterm Birth	33
2.7.4	Automated Risk Assessment Tool for Pregnancy Care	34
2.7.5	Limitations to the Methods	35
2.8	Clinical Decision Support Systems	35
2.8.1	Types.....	35
2.8.1.1	Knowledge Based.....	36
2.8.1.2	Non-Knowledge Based.....	36
2.8.2	Overview of Current CDSS	38
2.8.2.1	Heart Rate Observation System (HeRO).....	38
2.8.2.2	Artemis Analytic System.....	39
2.8.2.3	REAL-Time Risk of Mortality & Instability (REALTOMINS).....	40
2.8.2.4	Physician – Parent Decision Support (PPADS).....	41
2.9	Pattern Classification Methods.....	43
2.9.1	Decision Trees	43
2.9.1.1	Advantages	45
2.9.1.2	Disadvantages.....	45
2.9.2	Artificial Neural Networks	45
2.9.2.1	Advantages	47
2.9.2.2	Disadvantages.....	48
2.9.3	Hybrid ANN	48
2.9.4	Pattern Classification in MIRG.....	48

2.9.4.1	ANN Software and Libraries	48
3	Chapter: Methodology.....	52
3.1	Restating the Problem	52
3.2	Ethics Clearance.....	55
3.3	The Better Outcomes Registry and Network (BORN) Database	55
3.4	The Pregnancy Risk Assessment Monitoring System (PRAMS) Database	56
3.5	Real World Data.....	57
3.6	First Objective	58
3.6.1	Accomplishing Objective 1 – Methodology	60
3.6.1.1	Stage 1: Establish Model and Outcome of Interest	61
3.6.1.2	Stage 2: Data Preprocessing	62
3.6.1.2.1	Step 1: Cleaning Data and Handling Missing Values	63
3.6.1.2.2	Step 2: Feature Selection and Importance.....	64
3.6.1.2.3	Step 3: Preparing Datasets.....	68
3.6.1.3	Stage 3: Decision Trees	71
3.6.1.3.1	Step 1: Preprocessing	71
3.6.1.3.2	Step 2: Software Execution.....	73
3.6.1.3.3	Step 3:Post Assessment – Results and Feature Selection.....	74
3.6.1.4	Stage 4: Artificial Neural Networks:	76
3.6.1.4.1	Step 1: Preprocessing	76
3.6.1.4.2	Step 2: Software Execution and Results	82
3.6.1.5	Stage 5: Hold Out Verification Data:	86
3.6.1.5.1	Step 1: Preprocessing	87
3.6.1.5.2	Step 2: Software Execution and Results	87
3.7	Second Objective	87
3.7.1	Accomplishing Objective 2 – Methodology	88

3.7.1.1	Stage 1: Establish model and outcome of interest	88
3.7.1.2	Stage 2: Data Preprocessing	89
3.7.1.2.1	Step 1: Cleaning Data and Handling Missing Values	89
3.7.1.2.2	Step 2: Feature Selection and Importance	89
3.7.1.2.3	Step 3: Preparing Datasets.....	92
3.7.1.3	Stage 3: Decision Trees	95
3.7.1.3.1	Step 1: Preprocessing	95
3.7.1.3.2	Step 2: Software Execution.....	95
3.7.1.3.3	Step 3:Post Assessment – Results and Feature Selection.....	95
3.7.1.4	Stage 4: Artificial Neural Networks:	96
3.7.1.4.1	Step 1: Preprocessing	96
3.7.1.4.2	Step 2: Software Execution and Results	96
3.7.1.5	Stage 5: Hold Out Verification Data:	97
3.7.1.5.1	Step 1: Preprocessing:	97
3.7.1.5.2	Step 2: Software Execution and Results	97
3.8	Comparative Analysis	97
4	Chapter: Results and Discussion	98
4.1	Prediction of Preterm Birth – BORN Database.....	98
4.1.1	Decision Tree Results	98
4.1.1.1	Method 1 Selection – Compromise:	98
4.1.1.2	Method 2 Selection - Average Usage	102
4.1.2	ANN Results.....	105
4.1.3	Hold Out Verification Results	108
4.2	Prediction of Preterm Birth – PRAMS Database	109
4.2.1	Decision Tree Results	109
4.2.1.1	Method 1 Selection - Compromise.....	110

4.2.1.2	Method 2 Selection – Average Usage	114
4.2.2	ANN Results	116
4.2.3	Hold Out Verification Results	119
4.3	Variable Comparison	121
4.4	Discussion	126
5	Chapter: Concluding Remarks, Contributions to Knowledge and Future Work	
	 135	
5.1	Conclusion.....	135
5.1.1	Accomplished Objectives	135
5.2	Contributions to Knowledge	143
5.3	Future Work	145
5.4	Final Statement.....	147
	References..... 148	
	Appendices..... 166	
	Appendix A – Desktop CPU Testing with ANN Builder	166
	Appendix B – BORN and PRAMS Investigation Into Identical Data within Databases	166
	Appendix A - Desktop CPU Testing with ANN Builder	167
	Appendix B - BORN and PRAMS Investigation Into Identical Data within Databases	169
B.1	Investigation Results – BORN Data.....	170
B.2	Investigation Results – PRAMS Data	171
	Appendix C - Sample of ANN Final Options.....	172

List of Tables

Table 2-1 A 2 by 2 Confusion Matrix.....	18
Table 2-2 AUC Index and its Effectiveness for discrimination.....	23
Table 2-3 The clinical implications of true and false positives vs. true & false negatives (Catley, 2007)	24
Table 2-4 LR results from various studies of CL testing data. Data from (Honest et al., 2009/2012)	28
Table 2-5 LR results from various studies of fFN testing data. Data from (Honest et al., 2009/2012)	30
Table 3-1 BORN Parous List of Variables for DT Input including Outcome Variable. ..	66
Table 3-2 BORN Nulliparous List of Variables for DT Input including Outcome Variable.....	67
Table 3-3 BORN List of Obvious Variables removed from Parous and Nulliparous Datasets.	68
Table 3-4 Summary of Datasets before DTs.....	69
Table 3-5 BORN Variables Inclusion Identification – 1, Include Always based on logic or 2, maybe include depending on % usage.	76
Table 3-6 ANN Builder Formatting Requirements	82
Table 3-7 PRAMS Parous List of Variables for DT Input including Outcome Variable. 90	
Table 3-8 PRAMS Nulliparous List of Variables for DT Input including Outcome Variable.....	91

Table 3-9 PRAMS List of Obvious Variables removed from Parous and Nulliparous Datasets	92
Table 3-10 Summary of datasets before DTs	93
Table 3-11 PRAMS Variables Inclusion Identification – 1, Include Always based on logic or 2, maybe include depending on % usage.....	96
Table 4-1 BORN Train and Test Results Across Ten Generated Trees for Parous Data. .	99
Table 4-2 BORN Train and Test Results Across Ten Generated Trees for Nulliparous Data.	101
Table 4-3 BORN Train and Test Results Across Ten Generated Trees for Parous without Obvious Data.	101
Table 4-4 BORN Train and Test Results Across Ten Generated Trees for Nulliparous without Obvious Data.	101
Table 4-5 BORN Selected DTs and Variable Lists Selected for ANN Analysis for Method 1.....	102
Table 4-6 BORN Variables Selected for ANN Analysis for Method 2. Highlighted variables indicate the ones that are ‘Keep Always’	105
Table 4-7 Verification Results at 15% Prevalence Data – ANN Outcome.	107
Table 4-8 Verification Results at 15% Prevalence Data – Calculated LR values.	107
Table 4-9 Verification Results at 8.1% Prevalence Unseen Data.....	109
Table 4-10 Verification Results at 8.1% Prevalence Unseen Data – Calculated LR values	109
Table 4-11 PRAMS Train and Test Results Across Ten Generated Trees for Parous Data.	111

Table 4-12PRAMS Train and Test Results Across Ten Generated Trees for Nulliparous Data.	112
Table 4-13 PRAMS Train and Test Results Across Ten Generated Trees for Parous without Obvious Data.	112
Table 4-14PRAMS Train and Test Results Across Ten Generated Trees for Nulliparous without Obvious Data.	113
Table 4-15PRAMS Selected DTs and Variable Lists Selected for ANN Analysis for Method 1.	113
Table 4-16PRAMS Variables Selected for ANN Analysis for Method 2. Highlighted variables indicate the ones that are ‘Keep Always’.	115
Table 4-17 Verification Results at 15% Prevalence Data – ANN Outcome.	118
Table 4-18Verification Results at 15% Prevalence Data – Calculated LR values.	118
Table 4-19 Verification Results at 8.1% Prevalence Unseen Data.....	120
Table 4-20 Verification Results at 8.1% Prevalence Unseen Data – Calculated LR values	120
Table 4-21Old PRAMS Parous Variable Used on M1RG ANN – Identifying Variables Similar to New PRAMS Variables Used with ANN Builder.	121
Table 4-22 Old PRAMS Nulliparous Variable Used on M1RG ANN – Identifying Variables Similar to New PRAMS Variables Used with ANN Builder.	122
Table 4-23BORN Parous Variable List from M1 and M2 – Identifying Overlapping Variables in the PRAMS dataset used in M1 & M2.	123
Table 4-24 BORN Nulliparous Variable List from M1 and M2 – Identifying Overlapping Variables in the PRAMS dataset used in M1 & M2.	124

Table 4-25 PRAMS Parous Variable List from M1 and M2 – Identifying Overlapping Variables in the BORN dataset used in M1 & M2.	125
Table 4-26 PRAMS Nulliparous Variable List from M1 and M2 – Identifying Overlapping Variables in the BORN dataset used in M1 & M2.	125
Table A-1 Desktop Machine CPU Specifications	167
Table A-2 ANN Builder Run Information Using Sample BORN Data.....	167
Table A-3 ANN Builder Run Information Using Sample PRAMS Data	167
Table A-4 ANN Builder Run Information Using Sample Old PRAMS Data	168
Table B-1 Uniqueness in the Original Untouched Data Set	170
Table B-2 Uniqueness within Cases to be Investigated.....	170
Table B-3 Uniqueness before Data Normalization.....	170
Table B-4 Uniqueness after Data Normalization.....	170
Table B-5 Uniqueness in Data Prior to ANN analysis for Method 1	170
Table B-6 Uniqueness of Hold Out Verification Data for Method 1	170
Table B-7 Uniqueness in the Original Untouched Data Set	171
Table B-8 Uniqueness within Cases to be Investigated.....	171
Table B-9 Uniqueness before Data Normalization.....	171
Table B-10 Uniqueness after Data Normalization.....	171
Table B-11 Uniqueness in Data Prior to ANN analysis for Method 1	171
Table B-12 Uniqueness of Hold Out Verification Data for Method 1	171

List of Figures

Figure 2-1 ROC curve and the different points of significance.....	22
Figure 2-2 A diagram of a simple decision tree, based on (Duda et. al., 2001).	44
Figure 2-3 A simple three layer ANN, one a hidden layer.	47
Figure 3-1 Steps to randomize a dataset using repeated 2-fold cross validation technique	72
Figure 3-2 Screenshot of the CBR Interface.....	79
Figure 3-3 5-by-2 Cross Validation Method for ANN	81
Figure 3-4 Shows a screen shot of the details for a sample of the final options used for best networks for BORN Parous Method 2 Option 0, Good Network 1, at 15% prevalence.	83

List of Appendices

Appendix A – Desktop CPU Testing with ANN Builder	167
Appendix B – BORN & PRAMS Investigation into Identical Data within Databases...	169
B.1 Investigation Results – BORN Data Sub-Appendix.....	170
B.2 Investigation Results – BORN Data Sub-Appendix.....	171
Appendix C – Sample of ANN Final Options	172

List of Acronyms

ACC	Accuracy
ANN	Artificial Neural Network
ANN RFW	Artificial Neural Network Research Framework
APGAR	Activity, Pulse, Grimace, Appearance and Respiration
AUC	Area Under Curve
BASH	Bourne-Again Shell
BORN	Better Outcomes Registry Network
BP	Blood Pressure
BPD	Bronchopulmonary dysplasia
CBR	Case-Based Reasoning
CCR	Correct Classification Rate
CDR	Clinical Data Repository
CDSS	Clinical Decision Support System
CHEO	Children's Hospital of Eastern Ontario
CL	Cervical Length
CMS	Content Management System
CSV	Comma Separated Value
DT	Decision Tree
EAV	Entity Attribute Value
ECMS	Enterprise Content Management System
ELBW	Extremely Low Birth Weight
EMR	Electronic Medical Record
FANN	Fast Artificial Neural Network
fFN	Fetal Fibronectin
FN	False Negative
FP	False Positive
GUI	Graphical User Interface
HRC	Heart Rate Characteristics
HeRO	Heart Rate Observation system
ICU	Intensive Care Unit
ID3	Iterative Dichotomiser
IVH	Intraventricular Hemorrhage
JDC	Java Database Connectivity
k-NN	k-Nearest Neighbour
LBW	Low Birth weight
LOS	Length of Stay

LR	Likelihood ratio
MCC	Matthews Correlation Coefficient
MIRG	Medical Information technologies Research Group
MLP	Multilayer Perceptron
MSE	Mean Squared Error
NICU	Neonatal Intensive Care Unit
NPV	Negative Predictive Value
PBNN	Pruning Based Neural Network
PMODS	Physician MOther Decision Support
PPADS	Physician PArent Decision Support
PPV	Positive Predictive Value
PRAMS	Pregnancy Risk Assessment Monitoring System
PTB	Preterm Birth
REALTROMINS	Real-Time Risk of Mortality and Instability
RFW	Research Framework
ROC	Receiver Operating Characteristic
SQL	Structured Query Language
TBNN	Tree Based Neural Network
TN	True Negative
TP	True Positive
VLBW	Very Low Birth Weight
WO	Without obvious

1 Chapter: Introduction

1.1 Motivation

1.1.1 Healthcare Perspective

Canada is internationally known for having good health care. Still there are many ways that existing patient outcomes can be enhanced and expanded to help reduce demands on the system. Key to this is the ability to take preventative measures, and advance these measures with medical technology in the current health care system. Preventative measures have been proven to be key within the last 5-10 years, and during pregnancy these measures have lead to lower mortality and morbidity rates in newborns (Lim et al., 2009).

The ability to predict and detect early complications during a pregnancy allows better management and outcomes for both mother and baby, allowing health care providers to be more efficient in the allocation of their resources and provide timely response to varying medical problems. Of the numerous diagnoses, the ability to predict preterm birth is of particular importance. Preterm birth correlates with infant mortality and increased long-term health problems; henceforth it is considered a top Canadian Neonatal health priority (March of Dimes et al., 2012) (March of Dimes Foundation, 2015).

1.1.2 Technology Perspective

Even with advances in medical technologies and aids for hospitals and clinics, for example Electronic Medical Records (EMRs), the health care domain still remains a paper-intensive and non-automated industry. Despite the resistance to move away from

paper processes, research and development of systems and platforms that can share/exchange information have surfaced and have been implemented in medical environments. For example The Canadian Neonatal Network, The Pregnancy Risk Assessment Monitoring System (PRAMS) Database, and The Better Outcomes Registry and Network (BORN) Database (The Canadian Neonatal Network, 2009) (Centers for Disease Control and Prevention, 2009) (The Mothers Program, 2013). Such systems collect data that can be accessed and shared in many places by clinicians and researchers alike. Thus, would allow for health care providers access to various clinical applications including prediction and assessment tools in order to determine appropriate care protocols and/or medication administration (developed by the researchers) (Kluge, 2007), (Gullo & Lumb, 2005).

Current processes of diagnoses during pregnancy require and rely on the interpretative skills of physicians and health care workers. Newer concepts are being researched and looking at the use of clinical decision support systems (CDSS), making use of computer software designed to aid physicians with decision-making tasks. CDSS utilize information from past cases, so one can make predictions.

For a number of years, the Medical Information technology Research Group (MIRG) has been actively involved in the development of Clinical Decision Support Systems that are being applied to health care problems. MIRGs former students and primary investigators have explored several data mining tools (e.g. from MATLAB, to open source software) to analyze and create prediction models for various medical issues. Over the years MIRG has identified Artificial Neural Networks (ANNs) as a good pattern classification approach, and more recently the benefits of using a hybrid approach,

combining ANNs with the use of Decision Trees (DTs), to create prediction models for a number of medical outcomes. This methodology is being applied in this thesis work to model preterm birth predictions.

1.2 Problem Statement

Several problems are associated with clinical perinatal care. The ability to make accurate and early diagnosis of problems can allow earlier care action and help improve patient outcomes (Angus et al., 2006) (Honest et al, 2012, Behrman et al., 2007). A prediction model's purpose is to successfully predict an outcome of choice; in a clinical setting, a prediction model does so while meeting clinical criteria and expectations. This thesis work includes prediction models for the identification of a single perinatal problem using two different datasets to determine one outcome: preterm labour in pregnancy. Patient data from two databases are assessed: the Pregnancy Risk Assessment Monitoring System (PRAMS) from the USA, and the Better Outcomes Research Network (BORN) from Ontario, Canada. The primary objective of this work was to incorporate the advantages of hybrid decision models (Case Based Reasoning (CBR), Decision Trees (DTs), and Artificial Neural Networks (ANNs)) to create a system that can predict preterm birth at an early stage while maintaining high sensitivity, specificity, positive predictive, negative predictive values and likelihood ratios (positive and negative). A new ANN method was introduced to replace a previous one. Once the objective was met with the use of the BORN database, a comparative analysis was to be performed on the results, with the previously studied PRAMS database (Catley, 2007 and Yu, 2009) to determine the differences and highlight key decision parameters for predicting preterm

labour. This approach was to be reapplied to a more recent version of the PRAMS database. Once completed, a comparative analysis was to be performed again, with previous results and with BORN dataset results.

1.3 The Importance of Delivering at Term

According to statistics Canada, over 380,000 infants were born in Canada 2012/2013, and according to the Centre for Disease Control (CDC), over 3.9 Million in the United States in 2012 (Martin et al., 2013)(Statistics Canada, 2014). Though pregnancy can be a happy time for families, it is also a time to be cautious, as pregnancy can be a period of high risk for both the fetus and the mother. Due to this, many women seek prenatal care to ensure their pregnancy term progresses smoothly. An ideal term is about 280 days (40 weeks) counting from the last menstrual period (Segal & Norwitz, 2004). During this term, the developing fetus is carried inside the mother's uterus, an area that provides shelter for growth and development. It is important for the fetus to remain inside the uterus until completion of its term, as each day, cells, tissues and organs are undergoing development and maturation for survival in an external environment (Segal & Norwitz, 2004)(March of Dimes et al., 2012) (March of Dimes Foundation, 2015). Unfortunately not all pregnancies result in birth at term, and the infant is delivered prematurely. Such occurrence is called preterm birth (PTB), which is defined as "birth of a baby before 37 weeks (259 days)" (Oxford University Press, 2007).

In some countries, where there are less trained professionals, rates of both live births and mortality are higher. In general high live birth and mortality rates have often been found associated with preterm births, thus highlighting the importance of action and care for premature births globally (Gorthi et al., 2009), (Janghel et al., 2009). Being able

to predict preterm birth would allow a better management and outcome, and efficient allocation of resources in a timely manner, effectively helping to reduce mortality rates, potentially worldwide. Further, correctly predicting preterm birth in a non-invasive manner could reduce the need for costly screening tests (Honest et al., 2012)(Behrman et al., 2007) (fFNTTest.com, 2010).

1.3.1 Rate of Preterm Birth

Over the last decade many studies have been dedicated to improving the knowledge about preterm birth (Joseph et al., 1998, Honest et al., 2012)(Behrman et al., 2007). Unfortunately, even with the advancement of prenatal care technology as well as the implementation of educational programs, the rate continues to rise.

In 2010 the preterm birth rate in Canada was 7.8%, with an estimate of 29,800 infants born preterm; 600 of those died from complications (March of Dimes et al., 2012) (March of Dimes Foundation, 2015). For the United States in 2010, the preterm birth rate was 12.03%, with an estimate of 517,400 preterm births; 5,800 were deaths from complications (March of Dimes et al., 2012) (March of Dimes Foundation, 2015).

Though it is lower than it was in 2006-2007, when preterm birth had a drastic rise, increasing by over 19%, the rate is still very high which is a great concern (CIHI, 2009). The rates in the United States are even higher than in Canada, likely due to its type of population. Additionally, high rates are a global concern. Worldwide in 2010, 135 million babies were born, 15 million of them born premature, and of those, 1.1 million died. A global call to action was made with the report *BORN too Soon* (March of Dimes et al., 2012).

1.3.2 Health of Preterm Infants

The earlier a baby is born during pregnancy, the more likely the baby will have health problems. A hospital's neonatal intensive care unit (NICU), the part of the hospital that takes care of sick newborns, becomes the initial environment the newborn is exposed to. The NICU houses technologies and resources vital for care, such as incubators, oxygen support, and pharmaceuticals. Due to advances in medical care, premature babies today are more likely to survive than in the past; however this also depends on the severity of the case; the magnitude of an infant's initial health complications can vary (Behrman et al., 2007) (March of Dimes et al., 2012) (March of Dimes Foundation, 2015). Some of these problems include breathing problems (apnea, or respiratory distress syndrome (RDS)), internal bleeding in the brain (intraventricular hemorrhage, IVH), heart problems/poor blood circulation (Patent ductus arteriosus (PDA)), feeding problems, temperature control issues, liver failure, seizures and infections (Petrini et al., 2008) (Behrman et al., 2007).

The earlier the occurrence of preterm birth, the more severe the situation and complications can be, placing the neonate at a higher risk of mortality or morbidity. Late preterm birth (birth at 34-37 weeks gestation) still presents a higher risk of morbidity than an infant born at term (Kumar & Suresh, 2008) (Tomashek et al., 2007) (Behrman et al., 2007).

1.3.3 Cost of Preterm Birth

In 2007, the US Institute of Medicine estimated that the yearly societal cost of preterm birth was more than \$26.2 billion (Rochman, 2012)(March of Dimes et al., 2012)

(March of Dimes Foundation, 2015). The March of Dimes report found that the expense of delivery is up to 12 times more for a premature birth than for a healthy, full term birth.

In 2005–2006, in Canada, the average hospital cost for typical singleton newborns was approximately nine times higher for preterm newborns than for those born full-term (Lim et al., 2009). The average hospital cost and length of stay increased with decreasing gestational age. Singletons born at extremely preterm gestational ages (<28 weeks) had the highest average cost (\$84,235) and stayed in the hospital an average of 40 times longer than singletons born at full-term (Lim et al., 2009).

1.3.4 Importance of Prediction Models in Perinatal Care

To be able to reduce complications associated with preterm birth, it is important to be able to identify risk factors and initiate appropriate care measures in a timely manner to ensure effective treatment (Allen et al., 2002).

Predicting preterm birth early, if accurate enough, could reduce the need for multiple screening tests (for example measuring cervical length (CL) and testing levels of fetal fibronectin (fFN)). Additionally this would avoid added stressors to expectant mothers, which is commonly encountered with screen testing. Physicians would also be able to 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 may help to decrease the overall preterm birth rates.

1.4 Thesis Objectives

The first objective of this thesis is to develop and evaluate a preterm birth prediction ANN model for the BORN database with machine-learning tools using a DT-ANN hybrid approach. This DT-ANN Hybrid method developed by MIRG, will be

applied using decision trees to help select the features that may be the most important to predict preterm birth, to allow the ANN to work with less features than a complete set, and Artificial Neural Networks to actually classify the cases and evaluate how predictive the models are. To reduce bias, a 5by2 cross validation technique was applied at all stages (Dietterich, 1998). This work also made use of the ANN Builder tool developed by H. Martirosyan, in place of the previous ANN tool, MIRG ANN (Martirosyan, 2014; Rybchynski, 2005). Then a comparative analysis is done of the BORN data results, with the results from PRAMS analysis performed by Nicole Yu, (Yu, 2009).

The second objective of this thesis is to develop and evaluate a preterm birth prediction ANN model with an updated version of the PRAMS database. This was done by applying the same hybrid methodology to, using the new ANN Builder, instead of the MIRG ANN, and compare this prediction model performance to that of N. Yu's work. Additionally, the results of the new PRAMS are compared to the BORN data results. This comparison is to observe the difference between databases that may contain more rich data elements than the other, and how this affects the outcome with the ANN Builder tool.

The immediate results will hopefully help to suggest the type of data that should be collected by PRAMS and BORN programs in order to help predict preterm birth.

1.5 Thesis Outline

This thesis is organized as follows:

Chapter 1: provides motivation, problem statement and the need for prediction models in perinatal care. Thesis objectives are also stated.

Chapter 2: provides a detailed review of current literature, of relevant medical concepts related to preterm birth including environment, decision-making and current decision aids. Additionally a technical background is provided of performance measures and data mining methodologies.

Chapter 3: Describes the methodology used to create the prediction model for preterm birth using decision trees (DT), hybrid ANN (ANN Builder), and Case Based Reasoning (CBR) to replace missing values, applied to the BORN, PRAMS databases.

Chapter 4: presents the results and discussion for predicting preterm birth.

Chapter 5: presents the concluding remarks, contributions to knowledge and future work.

2 Chapter: Literature Review

2.1 Preterm Birth

Despite widespread awareness measures and newer research efforts, over 29, 000 babies are born preterm in Canada and over 517,000 in the US each year (March of Dimes Foundation, 2015). Preterm birth rates in North America are still high, 7.8% in Canada and 12.0% in the United States (March of Dimes Foundation, 2015). Even with advances in medical technology that have helped a majority of preterm babies in this day survive, preterm birth and infant mortality rates remain high.

Interventions during pregnancy are performed to aid and/or reduce risk to both the mother and infant. Such interventions may include caesarean section (surgical operation for delivering an infant through the abdominal wall) or induction (labour is started using artificial means such as drugs that stimulate contraction, or surgical rupture of the membrane) to reduce adverse risk to both parties (Segal & Norwitz, 2004) (McParland et al., 2004). It is reported that about 20% of preterm births are due to situations that include intrauterine growth restriction, preeclampsia, placenta previa, and non-reassuring fetal testing (Segal & Norwitz, 2004), (McParland et al., 2004). Induction or caesarean birth should not be planned before 39 completed weeks unless medically indicated. Other types of preterm birth can be spontaneous (unexplained) preterm labour, due to intra-amniotic infection or preterm pre-labour rupture of the membranes (PPROM) (Segal & Norwitz, 2004).

Since there is a direct correlation between neonatal mortality and preterm birth, finding ways to prevent preterm birth has become a health care priority (March of Dimes

et al., 2012). A global common goal for infant care is to identify key factors that contribute to preterm birth. Many past studies have attempted to predict women who are at risk of preterm birth, however, to date no single mechanism exists that has been proven to be a superior predictor (Honest et al., 2012).

2.2 Preterm Birth Classification

The difficulties with predicting preterm birth are largely due to the difficulties classifying it; preterm birth is a multifactorial problem. Different ways to assess this have been outlined by Yu and Gunaratnam, which show the differences in classification, and thus further the difficulties of assessment (Yu 2009, and Gunaratnam, 2013).

Yu assesses preterm pregnancies for low-risk and high-risk cases. High-risk patients can then be divided into symptomatic and asymptomatic cases. Symptomatic high-risk women have intact membranes but are experiencing symptoms of preterm labour (PTL), i.e. contractions, pelvic heaviness and low back pain, well before complete pregnancy gestation. They are usually admitted to hospital and assessed in emergency obstetric care. Asymptomatic high-risk women, on the other hand, do not experience symptoms of PTL but their lifestyle or pregnancy history suggests an increased risk for PTB (Yu, 2009).

Gunaratnam's work focused on a more general assessment of preterm pregnancies following the World Health Organization (WHO) Guidelines. Outlined in BORN Too Soon, The Global Action Report on Preterm Birth (2012), subcategories of preterm birth are defined based on gestational age. Preterm births are defined 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 with more than 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). The division is as follows: Extremely preterm (<28 weeks), very preterm (28 to <32 weeks), moderate to late preterm (32 to <37 weeks) (March of Dimes et al, 2012).

2.3 Factors Contributing to Preterm Birth

Preterm birth can arise from multiple factors. Certain factors may independently increase the risk of preterm birth, and when combined, these factors can increase the overall risk even more. Preterm birth continues to be a difficult problem because, even though many predictor variables have been indicated to be highly associated with PTB, their use in predicting PTB is not always accurate and to date has had limited success when applied to machine learning approaches (Honest et al., 2012, Yu, 2009). The ability to define the risk factors for preterm birth helps to identify at-risk women, which could allow earlier treatment, preventative and precautionary care. It would also allow one to conduct more studies to identify key factors linked to preterm birth.

Since this is a multifactorial problem, risk factors vary from a combination of maternal demographic, socioeconomic, behavioural, stress, maternal health, pregnancy/genetic history, fetal conditions and biological characteristics (Behrman et al., 2007) (Center for Disease Control and Prevention, 2014). With the large number of predictor variables associated with preterm births, prediction of preterm birth in the low risk, asymptomatic population remains a challenging problem. The following section presents a brief summary of some of these factors.

2.3.1 Demographic and Socioeconomic

The key demographic variables associated with risk of preterm birth that are often observed are maternal age and race. Extremes of maternal age are factors that contribute to low birth weight (LBW) babies and PTB. Women in their teens are more likely to have births complicated by those conditions. Women over the age of 40 are even more likely to have these conditions, due to pre-existing and development of health problems (Behrman et al., 2007).

Race has been seen as a factor that affects preterm birth. Several research models show that African American women are consistently at increased risk of preterm birth compared to native white women. (Goldenberg et al., 2008) (Behrman et al., 2007) (Center for Disease Control and Prevention, 2014). Additionally, studies have identified that low socioeconomic, low educational status, and single maternal status are linked to preterm birth (Goldenberg et al., 2008).

2.3.2 Behavioral Characteristics

Alcohol, tobacco and recreational drugs can have unfavourable effects on the mother and baby, putting pregnant women at higher risk of preterm birth (Behrman et al., 2007). There is a known associated risk with heavier alcohol intake (binge drinking). Additionally, women who smoke expose themselves and their developing infant to a number of toxic chemicals that include nicotine and carbon monoxide, which are both powerful vasoconstrictors (constrict blood vessels, increasing blood pressure), which lead to placental damage and decreased uteroplacental blood flow, which in turn may lead to preterm birth (Goldenberg et al., 2008).

2.3.3 Stress

As mentioned in Chapter 1 (section 1.2.2), the cost of births in the United States is around \$26.2 billion each year. For expectant families, stress can easily escalate from situations such as: not enough income, poverty, and low socioeconomic status. Other issues of psychological and social stress such as divorce, death of a friend or family member, alcohol and drug problems, emotional and physical abuse, partner violence, and racism are strong stress factors that contribute to preterm birth (Latendresse, 2009), (Segal & Norwitz, 2004), (Chaniramani, 2007). Clinical depression also contributes to an elevated risk of preterm birth. Depression is often correlated with increased smoking, drug abuse, and alcohol (Behrman et al., 2007). Additionally, long hours of work and exposure to hard physical labour contribute to a high risk of preterm delivery (Behrman et al., 2007).

2.3.4 Maternal Health

A mother's health during pregnancy can show several indicators as well. A number of maternal illnesses, conditions, and medical treatments are associated with preterm birth (Behrman et al., 2007). Women with a low or a high body mass index (BMI) during pregnancy have a higher risk of having unfavourable conditions during delivery. Women with high BMI are prone to riskier delivery outcomes, contributed by factors such as hypertension (high blood pressure), which can lead to the need for a caesarean section. Women with low BMI are linked to preterm birth, low birth weight, and an increased rate of neonatal mortality (Behrman et al., 2007). As well, women with lower concentrations of iron, folate (vitamin B complex), zinc or even vitamin C, are at a higher risk of preterm birth (Behrman et al., 2007). Insufficient weight-gain is another

preterm birth factor. Underweight women commonly have a lower intake of vitamins and minerals, which may lead to decreased blood flow and increased infections (Behrman et al., 2007). Obese women are likely to have medical conditions like preeclampsia, diabetes, glucose intolerance, hypoglycemia and gestational diabetes mellitus (GDM), all of which are associated with PTB (Behrman et al., 2007).

Additionally, pre-existing hypertension, developed (pregnancy induced) hypertension and lupus erythematosus are conditions that are monitored closely during pregnancy due to associations with PTB (Behrman et al., 2007). Mechanical problems can also arise when there are cervical anomalies (surgery, trauma, or congenital). The cervix is a firm, cylindrical, muscular structure, which forms the canal length between the uterus and the vagina. It is used to retain the fetus and allows its passage and delivery. A cervical cerclage (procedure involves inserting a suture around the cervix in order to prevent preterm delivery (Concise Medical Dictionary, 2007) can be performed to reduce the possibility of preterm cervical dilation and rupture of the membrane.

Thyroid disease and asthma are other maternal medical conditions that have led to preterm birth (Goldenberg et al., 2008).

2.3.5 Pregnancy History/ Genetic Characteristics

Women who have had previous preterm births are at a greater risk of preterm delivery, compared to women who have previously delivered at term. Risk is inversely related to the gestational age of a prior PTB; the lower gestational age of a prior preterm birth, the higher risk of preterm birth for the present pregnancy. Women with a sister(s) who delivered preterm are at 80% risk of also delivering preterm (Svensson et al., 2009).

Underlying genetically linked disorders that lead to preterm birth include diabetes, hypertension, and obesity (Goldenberg et al., 2008) (Behrman et al., 2007).

2.3.6 Current Fetal Conditions/Pregnancy Characteristics

Multiple fetuses inside the uterus limit the amount of room available for fetal development and growth. The increased hormonal and mechanical effects of carrying multiple fetuses could contribute to pregnancy-related complications, which include uterine over distension and contractions leading to preterm birth (Goldenberg et al., 2008). The average gestational age for multiple birth decreases with an increased number of fetuses: for twins, it is 35 weeks, triplets 33 weeks, and quadruplets 29 weeks on average. It is common for birth weights to be decreased as well (Pingili et al., 2009).

Although infertility treatments have allowed thousands of couples who have had difficulty conceiving to fulfill their desire to have children, it is another factor to be aware of. There is evidence of reported association between infertility treatments and preterm birth relating to underlying biological reasons for infertility and subfecundity (long time to becoming pregnant) (Basso and Baird, 2003; Henriksen et al, 1997; Joffe and Li, 1994) (Behrman et al., 2007).

2.3.7 Biological Characteristics

Biological markers can also be used to detect preterm birth. These can be measured in amniotic fluid, urine, cervical mucus, vaginal secretions, serum or plasma, and saliva (Behrman et al., 2007). Fetal Fibronectin (fFN) is a commonly used marker; it is absent from about 24 weeks until close to term, but if detected during a routine screening, between 24-32 weeks, it is an indicator that a woman is at an increased risk of preterm birth. (Goldenberg et al., 2008). Bacterial vaginosis, a disease that occurs due to

the change in the microbial ecosystem in the vagina, can also lead to preterm birth; this can be detected by testing the vaginal pH and is indicated when there is a pH of greater than 4.5. Systemic and intrauterine infections are known to be responsible for a majority of extremely preterm births (Behrman et al., 2007).

There are also many non-genital tract infections associated with preterm birth, which include pyelonephritis and symptomatic bacteriuria, pneumonia and appendicitis (Romero et al., 1989) (Behrman et al., 2007). Additionally, periodontal disease, which leads to intrauterine infections, is associated with preterm births. Viral infections, such as varicella or severe acute respiratory distress syndrome can also lead to preterm birth.

A decidual haemorrhage can cause bleeding in the placenta that can lead to thrombin formation and increased prostaglandin, which can cause cervical change and rupture of the membrane, leading to preterm birth (Allen et al., 2002).

Combinations of biological markers can increase the sensitivity of the prediction by combining risk predictors that address diverse causes of spontaneous and indicated preterm birth (Goldenberg et al., 2001). Predictor values can also be increased by combining selected markers with other factors such as the woman's obstetric history and cervical length (Yost et al., 2004).

2.4 Performance Measures

Performance Measures are key in determining how well a model performs with new cases. Thus once a classification model has been developed, its predictive capabilities must be evaluated by these measures (Catley, 2007).

2.4.1 Confusion Matrix (Contingency Table)

A confusion matrix is a table used to represent or display the number of correct and incorrect predictions made from a model (Tan et. al, 2006, Vidakovic, 2011). Matrices can have an n-by-n display, where n represents the number of classes. The following table represents a confusion matrix for a two-class problem.

Table 2-1 A 2 by 2 Confusion Matrix

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

2.4.2 Correct Classification Rate (CCR)

Also known as the accuracy of a classifier, this is the number of correctly predicted scores by the classifier:

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

2.4.3 Error Rate

The error rate is a measure of incorrect predictions. Prediction model for the training data gives the apparent error, which is often lower than the true error found using validation and test data.

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

2.4.4 Sensitivity

Sensitivity is the True Positive Rate (TPR) and a test's ability to correctly predict a patient with a given condition (Vidakovic, 2011). For example a test with 100%

sensitivity will correctly predict all patients with a specified condition or disease. A test with 80% sensitivity means results are correctly predicting 80% of patients with disease (true positive), while 20% are undetected (or False Negative).

$$Sensitivity = TPR = \frac{TP}{TP + FN}$$

2.4.5 Specificity

Specificity is the True Negative Rate (TNR) and a test's ability to correctly identify patients without a given disease (Vidakovic, 2011). A test with 100% specificity, for example, would correctly predict all patients without the disease. A test with 80% specificity would result in correctly predicting 80% of patients without having the condition (true negative), but 20% of patients would be incorrectly identified as being disease free (false negative).

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

2.4.6 Likelihood Ratio

There are two measures of likelihood ratio: positive and negative (Vidakovic, 2011). This measurement is considered to be more clinically meaningful as a test accuracy measurement (Honest et. al., 2012).

The positive likelihood ratio is the probability of a patient who has the disease testing positive compared to the probability of the patient who does not have the disease testing positive.

$$Likelihood\ Ratio\ Positive\ (LR+) = \frac{Sensitivity}{1 - Specificity}$$

The negative likelihood ratio is the probability of a patient who has the disease testing negative compared to the probability of a patient who does not have the disease testing negative.

$$\text{Likelihood Ratio Negative (LR-)} = \frac{1 - \text{Sensitivity}}{\text{Specificity}}$$

2.4.7 Prevalence

This is used to calculate the size of the population where the disease occurs. This is the number of cases with a disease or condition, over the total number of cases within a specific period (Vidakovic, 2011). The result is expressed as a percentage.

The prior probability that the subject has the disease before the test is calculated as:

$$\text{Prevalence} = \frac{\text{Person with a given health indicator during a specified period}}{\text{Population during the same period}} \times 100$$

2.4.8 Positive Predictive Value

The positive predictive value (PPV) is the probability of having the actual disease (positive outcome), given that the prediction made is positive (Vidakovic, 2011). It is the likelihood that the positive prediction is correct. PPV is proportional to prevalence, and therefore if prevalence is low, this will result in a decrease in PPV. PPV will not be close to the value of 1 (even with a high sensitivity and specificity).

$$PPV = \frac{TP}{TP + FP}$$

$$\frac{\text{Sensitivity} \times \text{Prevalence}}{\text{Sensitivity} \times \text{Prevalence} + (1 - \text{spec}) \times (1 - \text{Prevalence})}$$

2.4.9 Negative Predictive Value (NPV)

The negative predictive value (NPV) is the probability of a non-disease or negative outcome given that the prediction is negative (Vidakovic, 2011). It is the likelihood that the negative prediction is correct. High NVP is an indicator for a greater chance that a negative test has no abnormality. A rare outcome (low prevalence) would indicate a greater assurance that a patient with a negative test is correctly diagnosed. Ennett et al. demonstrated that ANNs work better when the prevalence is at least 15% of cases in a dataset. (Ennett & Frize, 2000, Ennett & Frize, 1998)

$$NPV = \frac{TN}{TN + FN}$$
$$\frac{specificity \times (1 - prevalence)}{specificity \times (1 - prevalence) + (1 - sensitivity) \times prevalence}$$

2.4.10 Receiver Operating Characteristic (ROC) Curve

ROC curves provide a visual display of the trade-off between sensitivity and specificity. The ROC curve is produced by plotting the false positive rate (FPR), which is the same as “1-Specificity”, against the true positive rate (TPR).

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 (Vidakovic, 2011). Figure 2-1 shows important points of the ROC as noted by Fawcett (2006), and as summarized by Yu (2009).

Figure 2-1 ROC curve and the different points of significance.

- At coordinate (0,0) there is no positive classification give and therefore no false positive classification exists. At point (1,1), the classifier assigns a positive class label to all points hence it is where the false positive rate is at it's max.
- Point a, is an ideal point, it is where all the points are classified correctly.
- Point b, has a higher TPR than a, but the FPR has also increased.
- Point c, has a low FPR, but its TPR is low (almost 60%).
- D, the diagonal line, indicates random guessing. Points on this line indicate that TPR and FPR are equal thus classifier performance is random and does not contain useful information.
- Point e, falling below the d line is worse than random guessing. May indicate that useful information is contained in the classifier however the application is incorrect.

Points under different threshold values of the model provide the ROC curve. The curve itself assesses the performance of the model at individual points, but the area under the curve (AUC) assesses the whole model. AUC is a scalar value of the ROC curve and is used as a performance measure between models.

2.4.11 Area Under the Curve (AUC)

The area under the curve is a measure of the classification model's effectiveness in correctly predicting medical outcomes. It is used to compare the performance between classifiers and will always have a value between 0 and 1.

The AUC value can be interpreted as the probability for a classifier to rank a randomly chosen positive case (i.e. death or presence of a disease) higher than a randomly chosen negative case (Fawcett, 2006). AUCs level of effectiveness of outcomes can be seen in Table 2.2 (Ambrosius, 2007).

Table 2-2 AUC Index and its Effectiveness for discrimination

Min	Max	Effectiveness
	≤ 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

2.4.12 Normalization/ Standardization

ANNs require input variables to be rescaled to eliminate the effects of range differences; this is called normalization or standardization. Variables found in medical databases can significantly range in magnitude and meaning, an ANN may calculate larger values are more significant than smaller ones if normalization is not used (Kantardzic, 2002), (Duda et. al., 2001), (Klosgen & Zytchow, 2002). The MIRG ANN Guide (Pan, 2002) suggests a modified Z-score transformation equation to scale values between -1 and 1:

$$\hat{x}_i^n = \frac{x_i^n - \mu_i}{3\sigma_i}$$

2.4.13 Confidence in the Model

Performance measures must be maximized or minimized for varying clinical outcomes; however determining which performance measures to maximize for a condition is a challenge. Identifying a newborn at risk of preterm birth in a 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). Recognition of these conditions may lead to a better allocation of resources and earlier monitoring to ensure a timely response to medical problems.

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

Table 2-3 The clinical implications of true and false positives vs. true & false negatives (Catley, 2007)

	Benefits	Drawbacks
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.

2.5 Current Knowledge

The complexity of the preterm birth problem is evidently clear through its non-linear, multifactorial, and heterogeneous nature. Several factors have been found to increase preterm birth risk, and different ranges of variables contribute. It is thought that, about 25% of all preterm births are associated with health problems that can be identified in the mother (Allen et al. 2002). Recognizing this, we can target these known health problems when attempting to identify risk factors for preterm birth.

Once the risk for preterm birth is established, the measures that can be taken to reduce the occurrence of preterm birth include: receiving early prenatal care, eliminating and avoiding modifiable risks associated with preterm birth (tobacco, drugs, alcohol), eating well (healthy diet), seeking help as soon as one starts noticing signs and symptoms of preterm birth, and a good communication with physicians (Ontario's maternal newborn and early child development resource centre, 2012).

Even though it is still difficult to predict the risk for preterm birth, some of the known symptoms to watch for are: persistent cramps, vaginal bleeding, lower back pain, feeling the baby pushing down, contractions, and increase in the amount of vaginal discharge (Ontario's maternal newborn and early child development resource centre, 2012).

2.6 Critical Analysis for Preterm Birth

2.6.1 Risk Factor Scoring

As detailed in Section 2.3, there are numerous risk factors concerned with preterm birth. It is also possible to have preterm labour with or without many of identified risk factors, however the more risk factors one has, the greater the likelihood of a preterm

pregnancy (Kassel, 2015). Over the years, risk-scoring systems have been developed to predict PTB and to identify factors that are associated with PTB. Higher scores indicate higher risks during pregnancy. A review of risk scoring systems developed over the years was performed in 2004 and showed a need for a better prediction system (Mercer et al, 1996), (Honest et al, 2004). Several of these systems were able to maintain a high specificity (80% and up), but had a low sensitivity value (30%-50% ranges). It has been about 40 years since the first reported scoring system was published (Kaminski in 1973); however, the poor prediction accuracy remains a challenge. Scoring systems identify key risk factors for PTB; however, this problem requires more complex pattern classification systems that a scoring system cannot offer.

2.6.2 Screening Test for Preterm Birth

Screening is a means of early diagnosis and treatment. These tests are often recommended to women who do not have current symptoms, but may be at high risk for preterm birth conditions (e.g. high risk factor score) (Kassel, 2015). There are several types of screening tests and examinations that can be performed to predict preterm birth and they range from several classifications such as: History, Physical Examination, Biochemistry, Microbiology, Physiological, and Ultrasound (Honest et al, 2012). The two most commonly reported screening tests for predicting birth for women are cervical length (CL) measurements and fetal fibronectin (fFN), (described in Sections 2.5.3 and 2.5.4, respectively).

Heterogeneity

It is difficult to do a direct comparison of the results of accuracy in predictions of tests such as fFN and CL because of the heterogeneous nature of studies. These studies

have focused on various population scopes and ranges of the obstetrical problem; for example studies have looked at: asymptomatic patients in low and high-risk populations of singleton births, multiple gestation and symptomatic women with either preterm labor or preterm premature rupture of the membranes (Honest et al, 2012).

Additionally, the cut-off gestational ages for preterm birth differ among studies. Studies have used cut-off gestational ages of <34, <35, <37, or even as low as < 32 weeks (Honest et al, 2012 and 2009). CL screenings also have different assigned values, which associate to different risk indicators of preterm birth (Honest et al, 2012 and 2009). These factors make it difficult to show a direct comparison between prediction accuracies of various studies.

2.6.3 Accuracy Prediction of Cervical Length

The cervix is where the uterus (or womb), opens into the vagina. When a woman is not pregnant the average CL is 40mm to 50mm (4cm to 5cm). Cervix length is expected to shorten as a pregnancy progresses, but the length of 30mm to 35mm is not expected until 32-36 weeks (Preventing Preterm Birth, 2015). At 20 weeks gestation CL should average around 40mm and at 34 weeks gestation 34mm; less than 15mm at 22-24 weeks gestation is abnormal. The risk of spontaneous preterm birth occurring increases as the length of the cervix decreases (Women's Imaging Services, 2012). One of the most accurate ways to measure this length is by transvaginal ultrasound (Jordan, 2013). If a woman is known to be at high risk tests should be taken at 14-24 weeks gestation, otherwise a general routine test should be taken at 18-22 weeks gestation (Women's Imaging Services, 2012). On the other hand other sources indicate that testing at 20-24

weeks gestation is a good predictor; however recommendation regarding length cut-off occur (Hassan et al, 2000), (Preventing Preterm Birth, 2015).

Even with these known reported values, research has continued to make an attempt at further defining and calculating the accuracy of this predicting test. For example: Hassen (2000) concluded prediction accuracy for PTB ≤ 32 weeks with a CL of ≤ 15 mm had a PPV of 47.6%, a NPV of 96.7%, sensitivity of 8.2%, and specificity of 99.7% (n=6877). The sensitivity is incredibly low, and only increased to 14.7% for CL measurements ≤ 25 mm (Hassan et al, 2000). Another study reports accuracy prediction (cut-off times of ≤ 32 weeks, ≤ 35 weeks, ≤ 37 weeks) with a sensitivity between 13–44%, specificity between 90–99%, PPV between 15–47% and NPV between 80–98% (n=760) (Hibbard et al., 2000). The sensitivity range does not exceed 50%; so many preterm births were not successfully identified. In a study done in 2005, they achieved a sensitivity of 46%, specificity of 94%, PPV of 67% and NPV of 87% (n=384 and prevalence = 21.4%) (Gomez et al., 2005). This shows the risk for preterm birth is inversely proportional to the CL, however results vary widely.

More recently, the likelihood ratio (positive and negative), calculated from sensitivities and specificities, are used to identify clinical usefulness. A systematic review by Honest et al, 2012 assessed numerous research studies performed on various testing methods used to predict preterm birth, evaluating usefulness of tests. This included the CL measurement test. The following table, Table 2.4, shows sample of the data collected. Usefulness identified with LR + >5 and LR- <0.2 .

Table 2-4 LR results from various studies of CL testing data. Data from (Honest et al., 2009/2012)

Tests (thresholds)	No. of studies	LR- (95% CI)	LR+ (95% CI)
Prediction of spontaneous preterm birth in asymptomatic women before 34 weeks' gestation			
Cervical length (15 mm @14–20 weeks)	3	0.89 (0.82, 0.97)	142.86 (3.58, 5709.07)
Cervical length (20 mm @14–20 weeks)	3	0.90 (0.83, 0.98)	35.36 (4.32, 289.68)
Cervical length (25 mm @14–20 weeks)	3	0.80 (0.71, 0.90)	13.38 (6.90, 25.96)
Cervical length (20 mm @20–24 weeks)	2	0.79 (0.72, 0.87)	7.64 (5.21, 11.20)
Cervical length (25 mm @ 20–24 weeks)	2	0.68 (0.60, 0.78)	4.68 (3.64, 6.03)
Cervical length (22 mm @20–24 weeks)	5	0.74 (0.51, 1.08)	4.51 (1.16, 17.64)
Cervical length (30 mm @14–20 weeks)	3	0.81 (0.68, 0.97)	2.48 (1.19, 5.19)
Cervical length (30 mm @20–24 weeks)	5	0.60 (0.50, 0.73)	2.28 (1.91, 2.71)
Prediction of spontaneous preterm birth in asymptomatic women before 37 weeks' gestation.			
Cervical length (32.5 mm @20–24 weeks)	1	0.33 (0.17, 0.66)	3.99 (2.84, 5.62)
Prediction of spontaneous preterm birth in symptomatic women within 7 days of testing.			
Cervical length (15 mm)	6	0.026 (0.004, 0.182)	8.61 (6.65, 11.14)
Prediction of spontaneous preterm birth in symptomatic women before 34 weeks' gestation.			
Cervical length (30 mm)		0.300 (0.083, 1.072)	1.88 (1.36, 2.59)
Prediction of spontaneous preterm birth in symptomatic women before 37 weeks' gestation.			
Cervical length (18 mm)	1	0.350 (0.170, 0.700)	3.36 (1.73, 6.54)
Cervical length (30 mm)	3	0.290 (0.150, 0.580)	2.29 (1.68, 3.12)

With several studies assessed, and with differing criteria and thresholds, CL measurement doesn't measure up to standards set (Honest et al., 2012). Though this method is a clinical standard, the literature review suggests that stronger methods of prediction are needed for preterm births.

2.6.4 Accuracy Prediction fetal fibronectin (fFN)

Fetal Fibronectin is a special 'glue-like' protein that holds the baby in place in the womb by bonding to the uterus. It is found between maternal – fetal interface and in low levels of vaginal secretions (fFNTTest.com, 2010), (Skandhan & Dawes, 2015). fFN is detectible in vaginal secretions in the beginning of the pregnancy when the bond is first forming and again at the end of the pregnancy, when it begins to naturally break down (around 35 weeks gestation), indicating when the body is preparing for delivery (fFNTTest.com, 2010). Thus makes it a good indicator of preterm birth, if detected early

enough. To perform this test, a swab of vaginal secretion is taken and tested for level of fibronectin. It is normally present up to 24 weeks of pregnancy, and then again three weeks before delivery (after 34 weeks gestation), (fFNTTest.com, 2010). This test is identified as one of the most powerful predictors <32 weeks gestation (Goldenberg et al., 1998). Levels of $\geq 50\text{mg/mL}$ at or after 24 weeks is abnormal and marked, suggesting there is an increased risk of spontaneous preterm birth. This test is useful after 24 weeks, but the test is not useful before 24 weeks or after 34 weeks (Berghella et al, 2008) (Goldenberg et al., 2005).

There is a wide range of accuracy for results, for the prediction of preterm birth with fFN for various studies over the years. However, it does align with a statement made by Leitch and Kaider in 2003: “due to the heterogeneous results in different studies, caution should be taken when they are applied to a specific population” (Leitch & Kaider, 2003). The best prediction results for fFN in their study had a sensitivity of 52% [95 CI: 42 –63] and a specificity of 85% (95 CI: 82–87) (Leitch & Kaider, 2003). Sensitivity and specificity results from a more recent study, (Study 2), fall within confidence intervals: 45.2 [95 C I:34.3-55.3] and 78.8 [95CI:71.8-85.2] respectively (Incerti et al., 2007). These results indicate that not much improvement has been made to the prognostic capability of fFN.

Like CL, fFN was a test included in the systematic review by Honest et al., 2012. The following table shows sample data collected on various fFN studies for predicting preterm birth.

Table 2-5 LR results from various studies of fFN testing data. Data from (Honest et al., 2009/2012)

Tests (thresholds)	No. of studies	LR- (95% CI)	LR+ (95% CI)
Prediction of spontaneous preterm birth in asymptomatic women before 34 weeks' gestation			
Cervico-vaginal fetal fibronectin	7	0.69 (0.56, 0.85)	10.18 (6.56, 15.80)
Prediction of spontaneous preterm birth in asymptomatic women before 37 weeks' gestation.			
Cervico-vaginal fetal fibronectin	15	0.94 (0.83, 1.07)	26.38 (1.73, 402.99)
Prediction of spontaneous preterm birth in symptomatic women within 7 days of testing.			
Cervico-vaginal fetal fibronectin	18	0.237 (0.067, 0.832)	3.52 (2.36, 5.23)
Prediction of spontaneous preterm birth in symptomatic women before 34 weeks' gestation.			
Cervico-vaginal fetal fibronectin	8	0.332 (0.190, 0.582)	3.98 (2.73, 5.80)
Prediction of spontaneous preterm birth in symptomatic women before 37 weeks' gestation.			
Cervico-vaginal fetal fibronectin	31	0.129 (0.051, 0.324)	7.97 (4.88, 13.03)

With several studies assessed, and with differing criteria and thresholds, fFN test shows strong predictive power, but not strong enough as a stand-alone prediction tool. A positive fFN test does not necessarily indicate preterm birth, as most of these studies report a low sensitivity <50%, indicating that the majority of PTB cases are not identified. Thus the need for a better prognostic system is clear.

2.6.5 Accuracy/Limitations of standard invasive test methods

Current standard prediction methods offered to high-risk patients are costly and effective only after 23 weeks gestation. The challenge has been to compare past studies' performance outcomes. As mentioned in the previous section, heterogeneity is amongst one of the many reasons that it is difficult to make accurate predictions.

A clear need exists for a non-invasive prediction tool that can predict preterm birth at <23 weeks gestation in pregnant women, while meeting clinical standards. As predicating preterm birth in women is controlled by several factors, this thesis will explore the outcome of study data from a new population of pregnant women and look at

a new set of factors. This will be done by applying a pre-existing methodology (using two data mining techniques including decision trees and hybrid ANNs) to predict this complex outcome, and comparing it to the previously analyzed population and its factors.

Honest et al., 2012 claims that there is no good test for predicting preterm birth, suggesting using one or more tests combined may provide more accurate predictions. A study by DeFranco et al., 2013 suggests that CL and fFN combined make a good prediction. This thought also researched by Goldenberg et al., 2001 and Yost et al., 2004.

2.7 Current Prediction Models for Pregnancy and Preterm Birth

2.7.1 Preterm Birth Prediction

Highlighted in Yu's thesis is a study (1992- 1993), which provides a model of preterm birth prediction (≤ 37 weeks) using "Learning from Example with Rough Sets" (LERS). A holdout method was applied to two large prenatal databases where half the data was used to train the model and the other half of the data was used for testing. These datasets produced rules using biophysical and symptom variables. The first dataset resulted in 350 rules and the second 170 rules. To create these models a large number of variables were used: 51 for dataset 1 and 77 for dataset 2. The performance of this study was measured by accuracy, False Positive, False Negative and unclassified percentages. Dataset 1 had an accuracy of 88.83%, FP of 6.34%, FN of 7.09% and 0.44% were unclassified. Dataset 2 had an accuracy of 59.28%, FP of 12.15%, FN of 25.29% and 3.28% were unclassified. These results may be considered biased, as the population was limited; a majority of the 18,890 subjects were Caucasian, married and well-educated women in their late 20s. Additionally, data was collected from a home uterine monitoring

vendor, where patients had either already experienced PTB or were at high risk for it. The sensitivity and specificity could not be calculated since its true categories (full term or preterm) were not provided (Joseph et al., 1998).

2.7.2 Uterine Electromyography ANN Prediction

This is a methodology that combines the use of pattern classification and the use of biomedical instrumentation. Uterine contractions produce electrical activity that can be monitored using electromyography – this is called uterine electromyography (uEMG). For a set of uEMG parameters, uEMG values were calculated and then used as inputs for the ANN. Patient cases were divided into women who delivered to term [term labouring (TL): $n = 75$; term non-labouring (TN): $n = 59$] and women who delivered preterm [preterm labouring (PTL): $n = 13$; preterm non-labouring (PTN): $n = 38$]; labouring and non-labouring refer to clinical diagnosis of labour by a physician (Maner & Garfield, 2007). This resulted in classification rates (the number of correctly predicted scores by the classifier) of: TL (79.0%), PTL (92.0%), TN (86.0%) and PTN (71%) (Maner & Garfield, 2007). This study shows promising results, however the sample size is small ($n = 192$).

2.7.3 Early Prediction of Preterm Birth

In 2007, Tan *et al.* used multivariate logistic regression to create a model for the prediction of preterm birth (≤ 32 weeks) (Tan et al., 2007). Considered a restrictive outcome prediction method, it was used because neonatal mortality and morbidity is higher in gestation weeks 20-31 than for later preterm birth. This group used United States 1996 birth registration data for singletons and the 1995–1997 linked birth/infant death dataset for multiple births, applying the multivariate logistic regression model to

predict preterm birth for singleton, twin and triplet pregnancies. The resulting sensitivity, specificity, PPV, and NPV for singletons were 24.58%, 93.54%, 5.91%, and 98.69%, respectively; for twins 64.66%, 57.04%, 16.29% and 92.59%, respectively; and for triplets were 63.57%, 53.58%, 42.96%, and 72.78%, respectively (Tan et al., 2007). Specificity results are high, however sensitivity values are quite low; even in the cases of multiple births; they are too low to be clinically useful (sensitivity > 65%, specificity >85%).

2.7.4 Automated Risk Assessment Tool for Pregnancy Care

In a 2009 study performed by Gorthi *et al.* (2009), their research group developed an ID3 based Clinical Decision Support System (CDSS) to provide health care workers with a pregnancy criticality risk scale in order to help with the decision-making process. The tool classifies the pregnancy as normal, moderate risk, or high risk. Due to the lack of availability of real patient cases under each risk category, cases were generated using a rule-based engine, for a total of 240 generated cases. Decision Trees, specifically Classification & Regression Trees (CART), were the selected algorithm and were also used to identify critical parameters in the decision making process. A CART is a combination of classification trees (where the predicted outcome is a class to which the data belongs e.g. colour: 1, 2, or 3) and regression trees (where the predicted outcome can be considered a real number – e.g. a patients length of stay in a hospital). Based on a literature survey and clinician input, a total of 31 features were selected for analysis. Using DTs for feature analysis it was later found that only 22 features should be involved in classification. Cases were divided into sets of 200 for training and for 40 testing, and a five-fold cross-validation method was used for classification efficiency. The classifier

reported an accuracy of 82.5%. Sensitivity and specificity were not reported. (Gorthi *et al.*, 2009).

2.7.5 Limitations to the Methods

Current clinical standard prediction methods like fFN and CL are costly, including cost of follow up tests, and they are more suitable for patients who are considered symptomatic or high-risk asymptomatic, and are mainly effective in the middle of the pregnancy. These standard methods are somewhat invasive, which can become additional stressors to an expectant mother during an already emotional period. Moreover, these methods vary over a large range of sensitivity values (under 50% - meaning less than 50% of the time preterm birth is correctly predicted). The machine learning models appear to be promising, and for the most part are non-invasive, however, reporting results must include sensitivity and specificity and meet clinical standards or expectations. Clearly there is still insufficient prediction capability. With the low sensitivity of current methods, it is evident that a complex relationship exists between the predictor variables being used and the outcome of interest. This thesis explores the application of non-linear modeling techniques, such as ANN and decision trees, to estimate the outcome of PTB.

2.8 Clinical Decision Support Systems

2.8.1 Types

CDSSs can be classified as one of two types, either a knowledge-based system or a non-knowledge based system. To date, most systems are knowledge-based systems; they contain rules, guidelines or compiled knowledge. The ultimate goal of decision

support systems, not only limited to health care, is to support and enhance human thinking (Berner, 2007).

2.8.1.1 Knowledge Based

A knowledge based decision support system consists of three parts: the knowledge-base, an inference engine (or reasoning engine) and a mechanism to communicate. The knowledge base component consists of rules more commonly in an “IF-THEN” form. So “IF” a condition occurred or was met, “THEN” it prompts another occurrence (Berner, 2007). Examples would be a rule designed to prevent duplicate test ordering, probabilistic association between drug and food interactions, signs or symptoms and diagnosis. The inference engine contains the formulas that combine the rules with patient data in the knowledge-base. The communication mechanism is what transports patient data into the system and the results out of the system shown to the user, who makes a final decision.

A CDSS that is integrated with an EMR has the ability to auto-populate patient data into appropriate fields from the information available in the EMR. Otherwise, in a stand-alone system, the patient data needs to be entered into the system by a user. The system output can be in the form of notifications, alerts or emails (Berner, 2007).

2.8.1.2 Non-Knowledge Based

A non-knowledge-based decision support system uses artificial intelligence, in the form of machine learning, which allows a computer to learn from the past. It provides the capability to learn from past knowledge or to analyze and detect patterns in clinical

data. Common types of non-knowledge-based systems include artificial neural networks (Berner, 2007).

Numerous approaches are being researched and are looking at the use of clinical decision support systems (CDSS), using computer software, designed to aid physicians with decision-making tasks. A CDSS utilizes information from past cases, so one can predict the type of care needed throughout a new pregnancy case. A CDSS was designed to be a stand-alone system or part of non-commercial computer- based patient record system. It is only recently that people have started to include CDSS into computer-based patient record systems (Berner, 2007). Examples of such systems are discussed in Section 2.7.2.

Studies have shown that using a CDSS has led to enhancements in clinical performance patient outcomes (Child Health Research Project Special Report, 1999). For example, Electronic Medical Records (EMRs) are designed to improve both accessibility and legibility of information (Berner, 2009). EMRs are considered to be foundation for providing quality healthcare. A CDSS is essential to archive the full potential of an EMR (Berner, 2009). CDSSs are designed to intelligently filter information in order to aid the decision-making process in clinical settings and are used to gather large sets of data and manage medication actions effectively. Implementing and utilizing a CDSS will help reduce the occurrence of practice errors and provide higher quality of care (Peleg & Tu, 2011). See Section 2.7.2 for examples.

A few different systems have been developed and they use many types of machine learning algorithms. Some systems have already been applied and tested, and further work is being done to improve the quality, speed and accuracy of prediction

results, to be able to fully assist physicians in times of need, to help predict and detect early complications for pregnancy. HeRO, Artemis, and REALTOMINS are a few examples to name a few, these and others will be discussed in the next section.

Preterm birth prediction research is still in progress, however researchers (like the MIRG team) are looking at ways to use a CDSS and applying machine-learning algorithms to help make this prediction.

2.8.2 Overview of Current CDSS

2.8.2.1 Heart Rate Observation System (HeRO)

Researchers at the University of Virginia, led by J. Randall Moorman, have developed a system, (HeRO), that focuses on the early detection of neonatal distress in very low birth weight (VLBW) infants, by detecting abnormal heart rate characteristics (HRC), (Moorman *et al.*, 2011). Abnormal heart rate characteristics could signify illness. The HeRO system is non-invasive and generates a numeric score that quantifies the prevalence of abnormal patterns in each patient's heart rate. It makes use of mathematical and statistical tools, and analyses heart rate characteristics with: time-domain parameters (mean and variance of the heart rate signal), frequency-domain parameters (band specific variance), phase domain (phase of waveforms), signal quality index, entropy estimation and deceleration detection using wavelet transform based algorithms. These are then combined through multivariate statistical methods like logistic regression, k-nearest neighbor, and other techniques (Moorman *et al.*, 2011). The HeRO system continuously acquires records and measurements to analyze heart rate patterns and provides a real-time display of the HeRO score.

Moorman *et al.* (2011) conducted a controlled clinical trial of 3003 VLBW infants in 9 neonatal intensive care units, between April 2004 and May 2010. Infants in the NICUs were split into two groups: one with heart rate characteristics (HRC) monitoring display and the other one without HRC displayed. Results showed that there was a reduced mortality rate for groups with HRC monitoring displayed “from 10.2% to 8.1%, and there was a trend towards an increase for days alive and ventilator free: 2.3 days more compared to the control group (no HRC display)” (Moorman *et al.*, 2011b). This trial shows that HeRO, helps to improve neonatal outcomes.

2.8.2.2 Artemis Analytic System

The Artemis Analytic System was designed and piloted in collaboration with researchers from the University of Ontario Institute of Technology (UOIT), IBM T.J. Watson Research Center, and The Hospital for Sick Children in Toronto, Canada. The Artemis system was first implemented at the Hospital of Sick Children in August 2006 and is still in use (IBM InfoSphere, 2013).

This system acquires and stores clinical data in the NICU for the purpose of real-time analysis. “A series of clinical rules are executed online and in real-time to detect medical conditions including IVH, nosocomial infection, periventricular leukomalacia and pneumothorax” (IBM InfoSphere, 2013). It’s framework uses Capsule Tech Inc. DataCaptor for collecting physiological patient data such as the heart rate, respiratory rate and blood oxygen saturation. ARTEMIS utilizes IBM’s InfoSphere Stream Computing system for the purpose of data processing along with the Stream Processing Language (SPL) programming language for programming and controlling. “Knowledge extraction

is performed using multidimensional, temporal and data stream data mining frameworks developed by the Artemis team” (Blount *et al.*, 2010), (Thommandram *et al.*, 2014).

Reported in 2014, Thommandram *et al.* used the Artemis system to classify neonatal spells. These are cardiorespiratory events that occur in neonates with various occurrences of suspended breathing; a fall in heart rate and decrease in blood oxygen saturation (Thommandram *et al.*, 2014). The system detected and identified respiratory pauses by measuring the intervals between breaths. Additionally, threshold breach detectors and algorithms detecting change events in individual signals were used to identify changes in heart rate and blood oxygen. To then classify spells, events detected in the individual signals were synchronized together and a rule-based classifier was applied. Performance of event detection algorithms and the rule-based classifier was evaluated using a 24-hour sample of data from a neonatal patient who was reported to be experiencing spells. The researchers then compared the results of the system to the manual detection of events by a neonatologist. Accuracies are reported as 98.9%, 97.8% and 98.3% for detecting pauses in breathing, changes in heart rate and blood oxygen saturation, respectively. Neonatal spells were classified with 98.9% accuracy (Thommandram *et al.*, 2014).

2.8.2.3 REAL-Time Risk of Mortality & Instability (REALTOMINS)

Developed at the University of North Carolina in Chapel Hill, the REALTROMINS (REAL-Time Risk of Mortality and Instability) system uses real-time predictive analytics to identify high-risk patients in NICUs and Pediatric Intensive Care Units (PICUs) and presents information to physicians through a user interface (REALTROMINS Inc, 2014).

This system collects attributes such as physiologic signals (e.g. ECG), physiologic based measures of organ function (e.g. serum glucose), demographic (e.g. age) and diagnosis related (e.g. post-operative status) predictors of mortality. It uses the Health Level 7 (HL7) messaging standards, to collect these parameters (HL7, 2009).

Variations in physiological signals are calculated and analyzed consistently to determine if a particular variable is within a critical interval; if so, it is recorded as an occurrence of an event for a patient. Over time, these events are collected and presented to clinicians in graphic form to illustrate the risk of mortality. Models that estimate the risk of mortality can be created from the data collected using neural networks in SAS Enterprise Miner software; the information is presented to physicians real-time with an update of every two minutes (Kocis and Kocis, 2012).

To evaluate the performance of the REALTROMINS system, the data of 28 patients of the Pediatric Intensive Care Unit, half of who died, were used to create mortality models with ANNs. The first four hours of the admission data for all patients were partitioned, creating 1680 “survive” packets and 1680 “died” packets: 3360 records in total; 80% of packets were used for the training and a random 20% hold out sample of packets were used for validating the network. The reported predictive capabilities of the model were 98.44% sensitivity and 98.88% specificity on the training set, with identical results for the validation set (Kocis and Kocis, 2012).

2.8.2.4 Physician – Parent Decision Support (PPADS)

The CDSSs presented so far provide outcome estimates that are available for clinicians to view; none of them provide decision-support or other information for facilitating parent decision-making.

To improve shared decision-making in the NICU, MIRG at Carleton University developed the PPADS tool. It is a combined decision aid tool to be accessible both for health care professionals and for parents to help them to make more informed care decisions, especially in difficult cases. The PPADS tool currently provides predictions for mortality and aims to enhance the evidence upon which critical decisions are made in the NICU (Frize et al., 2011, and 2013, Weyand, 2011).

The PPADS tool is developed using the Drupal content management system. It consists of two distinct user interfaces: a clinician and a parent interface for patients in the NICU. The parent interface provides information about the infant with risk estimations, and provides a decision support module to help parents clarify and communicate their personal values associated with the different aspects of options available to them. The clinician interface contains the list of all patients with corresponding files of admission and various medical details including outcome predictions (Weyand, 2011).

The first phase of the usability study conducted by our group, involved parents who had a neonate who graduated (survived) from the NICU. Eight of the parents who participated in the study found the tool easy to use and they completed the required tasks without any miss-clicks or system navigation errors. Furthermore, the participants indicated they would use the tool if given the opportunity and provided rich comments on how to improve the tool; adding more information was the most common suggestion (Weyand 2011, Frize *et al.*, 2011), (Frize *et al.*, 2013).

Further, a usability case study was performed by Martirosyan, which involved parents who had a neonate who did not graduate (did not survive) from the NICU. The

two parents who participated provided feedback with suggestions for improvement, which were encouraging, regarding the usefulness of the PPADs tool (Martirosyan, 2014). MIRG is developing a similar support system for perinatal care scenarios, Physicians – Mother Decision Support (PMODS). (Martirosyan et. al., 2015)

2.9 Pattern Classification Methods

Pattern classification is a branch of machine learning, which focuses on recognition of patterns and regularities in data. It consists of the theory and the algorithms of assigning abstract objects into distinct categories, with categories known in advance. These methods have been applied for information retrieval, data mining, document image analysis and recognition, computational linguistics, forensics, biometrics and bioinformatics (Srihari, 2007). This thesis uses two pattern classification methods: Decision Trees (DTs) and Artificial Neural Networks (ANNs), in a hybrid approach. DTs are used as a feature selection algorithm, which attempts to prune out redundant or irrelevant features, and ANNs provide a pattern classification approach that extracts crude domain knowledge from a dataset created using features based on the results of a DT.

2.9.1 Decision Trees

A decision tree, visually comparable to an actual tree, is composed of the root node, subsequent nodes, and terminal leaves. The tree starts with a root node at the top, which is then split into branch nodes, which lead to successive splitting criteria (node representing a choice of alternatives, often within threshold values), and then terminates with a leaf node, which has no further splits (options), which represents a final outcome or decision. See Figure 2-2 for an example of a Decision Tree.

Figure 2-2 A diagram of a simple decision tree, based on (Duda et. al., 2001).

Decision trees are a favoured data mining method because of how easily its structure can be interpreted. They can be interpreted by anyone, such as business end-users and analysts to understand the models, versus neural networks, which are more difficult to interpret and understand.

There are a number of algorithms that can be used to create a DT; the ones that are more frequently used are Classification And Regression Trees (CART), ID3, C4.5 and C5 (Duda et. al., 2001, Rulequest Research, 2010, 2012). Algorithms of DTs follow the same core steps, but the method of how to differentiate features is unique to each method. This thesis work uses C5.0 (or See5), which is an improvement of C4.5, which was a successor of ID3 (Duda et. al., 2001).

2.9.1.1 Advantages

Not only are DTs easier to interpret and understand; they can handle metric data (numerical, real values) and non-metric (categorical, descriptions), both a frequent occurrence in medical data. They are also able to handle missing values and nominal type data. This would be easier for physicians to comprehend compared to other data mining methods. DTs are useful for rapid analysis and little data preparation is required (data not need to be normalized or missing data replaced). DTs can handle large amounts of data in short time compared to other classifications. A logical approach does not depend on the statistical distribution of the attribute or its independence (Kantardzic, 2002); it can often be developed using common statistical techniques (Peng, 2006). Additionally See5 is a quick tool, which can classify over 90,000 cases from 1 to 10 seconds (Yu, 2009).

2.9.1.2 Disadvantages

Even with the many advantages there are some drawbacks to using DTs. DTs can over fit the data, and create complex trees that do not generalize well. They are not robust; they are sensitive to the presence of outliers; the accuracy can be affected by noise, training cases, and size of training data (Mitra & Acharya, 2003). A small change in the size of the dataset could also result in generating a different tree. Therefore the prediction of results may not be stable when validating multiple sets. DTs are unable to handle complicated relationships between features (Mitra & Acharya, 2003). In some of MIRG's experiments on predicting mortality, DTs were found to perform much less well than ANNs. (Frize et al, 2013)

2.9.2 Artificial Neural Networks

Artificial Neural Networks (ANNs) are powerful non-linear mapping structures,

used to model unknown relationships. It makes use of a system structure similar to that of biological neurons (Kantardzic, 2002), mimicking or functioning similar to the human brain; ANNs can solve problems that involve data that is complex, non-linear, imprecise and/or noisy (Jha, 2011). The human brain is a collection of over 10 billion interconnected neurons with the capability 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). Processing components of ANNs consist of inputs, weights, and output values; they provide a mathematical model of many neurons connected in a network (Mitra & Acharya, 2003). Since it is modeled after the biological nervous system, processing elements of neural networks referred to as nodes, and connection weights represent the synapses, non-linear characteristics exhibited by neurons given by a transfer function. The weighted sum of input signal by a transfer function is used to compute impulse of neuron. This is improved by weights, being adjusted according to the learning algorithm selected.

The ANN is trained to detect a pattern between the input data and the related output value from a dataset. After training the dataset, the ANN can be used to predict the result of a new input data (Jha, 2011). There are several types ANN models, as for example: feedforward, recurrent, probabilistic networks. The ANN used for this work uses the Fast Artificial Neural Network library implements multilayer feed-forward ANN (FANN, 2014) (Martiosyan, 2014). An example of a multilayer ANN is shown in Figure 2-3.

Figure 2-3 A simple three layer ANN, one a hidden layer.

2.9.2.1 Advantages

Unlike decision trees, ANNs can generalize data with noise. It does not require prior knowledge, and thus requires less statistical training in order to develop (Berner, 2007). ANNs have the ability to detect complex relationships between variables that are independent and dependent and are able to detect potential interactions with a predictor variable. This functionality eliminates the need to have the “IF-THEN” type rule, which would have to be collected by clinicians. The function of each neuron in each layer may be calculated in parallel, making these “parallel networks”, which allows the processing speed to be increased greatly by using computers with multiple processors. Based on type of data to be analyzed or type of output required, ANNs can apply different training algorithms to develop models (Kantardzic, 2002) (Duda et. al., 2001). An ANN can also improve the results each time it is trained, due to its dynamic structure (Berner, 2007).

2.9.2.2 Disadvantages

The input data has to be first analyzed and processed before being fed into the ANN, including normalizing the data. Maximum and minimum values are required to be known prior to normalization. This poses a challenge for a real time environment, since that data is not known in advance and thus can only be calculated when available (Mitra & Acharya, 2003). ANNs do not handle missing data well; so missing data needs to be replaced with imputed values. Though it is possible to determine most contributing input, resulting formulas and weights are often not interpretable.

2.9.3 Hybrid ANN

The idea behind hybrid methods is to combine advantages of different pattern classification approaches while limiting the shortcomings. An ANN is a “black box” classifier, with knowledge found within weights, biases and connection links (Li & Wang, 2004). An ANN benefits with a pattern classification approach that extracts crude domain knowledge from the dataset such as a DT. Combining an ANN with a knowledge-based network allows the generalization capability of ANN to be improved (Mitra & Acharya, 2003).

2.9.4 Pattern Classification in MIRG

2.9.4.1 ANN Software and Libraries

MIRG ANN

Our research group has been using ANN's for predicting various medical outcomes. These include: mortality, length of stay, and other medical complications related to perinatal care. The ANN Research Framework (MIRG ANN RFW), was

developed using MATLAB's Neural Network Toolbox as a basis for implementing multi-layer feed forward ANN's, by previous students (Ennett et al., 2004).

The MIRG ANN RFW has the ability to modify the number of hidden nodes from zero to $2n+1$ as an automated feature, where n is the number of input nodes, in the single-layer or multi-layer networks. This was followed by the implementation of an automated parameter update, for several of the ANN driving parameters (Rybchynski, 2005). Parameter values were modified individually: A single parameter gets adjusted until the best parameter value (best result for stopping criteria) is met, while other parameters remain constant. The best parameter value for each is kept for the next step (Rybchynski, 2005).

A logarithmic sensitivity index was used to stop the training of the network; interrupting the training process when no improvements in the index were made in 500 epochs sub sequentially (Ennett et al., 2002).

$$index = -Sensitivity^n \cdot \log(1 - Sensitivity \cdot Specificity)$$

Once training is complete, the results files are processed to determine the best ANN Model. Afterwards, weights of a selected model are manually extracted and given to the prediction tool implemented in MATLAB. Data needs to be passed to the tool manually for each patient in order to make risk estimations for each (Rybchynski, 2005).

Despite having various features and functionalities added by MIRG students to the basic MATLAB tool, the manual steps involved in the risk estimation by the MIRG ANN RFW posed difficulties for the integration of ANN models into the systems and models that require real-time analysis (Martirosyan, 2014). As well, the RFW and the prediction tool developed in MATLAB, requires a commercial license to use.

FANN Library (ANN Builder)

Martirosyan's thesis work explored several open sources libraries to create ANN models, and make predictions in a real time environment. Martirosyan found the Fast Artificial Neural Network (FANN) library to be suitable to MIRG's work for the following reasons (Martirosyan, 2014):

- Firstly, the library implements feed-forward networks, which our research group has identified to be well performing machine learning methods for our medical datasets;
- Secondly, the fast performance is one of the main features of the library, which is important in processing and analyzing real-time data;
- Lastly, the library is implemented in the C programming language which makes the FANN library based applications compatible with many software environments and portable to many different computer architectures or platforms (Nissen, 2007), (FANN, 2014).

The FANN library, used to develop the ANN Builder, implements a feed-forward ANN and allows for the training and testing of networks with various architectural options, to select activation, training and the parameters that drive the learning of the ANN (FANN, 2014 and Martirosyan, 2015). The ANN Builder automates the process of selecting options, the adjustment of parameter values for creating ANN models, and testing of all combinations of training algorithms, activation functions, and network parameters. Built into the automation is a selection of modes, which uses a fraction of all possible combinations. Changing modes increases the number of combination parameters performed and evaluated. The more combinations used, the longer the computation

process takes. There are approximately 96,000 combinations, and three mode settings (fast, medium, and slow). The ANN Builder tool was designed to be generic and not specific to the scenarios modeled here, that is, it is designed to be domain or data independent. For example, NICU mortality has been modeled previously with the ANN Builder. More details of functions and workings of the ANN Builder can be found in Martirosyan's thesis (Martirosyan, 2014).

With the MIRC ANN Builder, mortality models with the average of 81% sensitivity and 98% specificity results were developed using 48-hour summation neonatal mortality data. The results of the models, developed and selected according to their sensitivity and specificity performance, exceeded the clinically usefulness criteria (Sensitivity ≥ 60 , Specificity ≥ 90) and the results of the previous ANN application, MIRC ANN RFW, (Sensitivity = 68, Specificity = 97) (Gilchrist, 2012, Frize et al., 2013).

3 Chapter: Methodology

This chapter restates the problem studied in this thesis work and provides a description of the perinatal databases used. The methodology used for the development of a prediction model for preterm birth in pregnancies of two databases is discussed. Analysis strategies to compare the performance obtained with various dataset are also discussed. This thesis introduces a new method of feature selection, not previously applied in the works of Catley (Catley, 2007), Yu, (Yu, 2009) and Gunaratnam (Gunaratnam, 2013).

3.1 Restating the Problem

Primarily there are two key factors that have led to an interest in the prediction of preterm birth: (i) the current poor ability to correctly classify positive outcomes, and (ii) the complexity that exists between variables used for predictions. Insufficient availability of predictor variables and the rare occurrence of medical conditions play a significant role in the success rate in identifying outcomes (Yu, 2009). The relationship between variables has not been successfully studied by current methods.

The focus of this work was to attempt to optimize the prediction of preterm birth by combining the advantages of DTs and ANNs in a hybrid approach, and apply these to new databases as well as evaluate the performance of the new models created. This approach had been designed to minimize the number of input variables of a large database (reducing complexity of the problem) and optimize the sensitivity while maintaining sufficient specificity. The databases assessed are the Better Outcomes Registry and Network (BORN), and Pregnancy Risk Assessment Monitoring System

(PRAMS), which are defined in following sections (Sections 3.3 and 3.4).

The first objective was to evaluate a preterm birth prediction ANN model by applying an integrated DT- ANN hybrid system with a new unseen database, eliminating variables that did not affect the prediction, and comparing these results to the results performed on previously accessed databases. This DT – ANN hybrid system was applied to a new large database (BORN – see 3.3) to estimate preterm birth. The main issues are: poor sensitivity of the common fFN and CL prediction methods, biased and poor estimations made by pattern classification approaches attempted to date, and a significantly large number of predictor variables (Yu, 2009) (Honest et al., 2012) (DeFranco et al., 2013).

The second objective was to evaluate another ANN model by using the DT- ANN hybrid system, with a new PRAMS database (2009 – 2011) (see 3.4) and compare this to previous PRAMS results (2002 – 2004) and to the new BORN database results (2010 – 2012).

MIRG previous studies focused on artificial intelligence based techniques using artificial neural networks (Catley, 2007), and later a hybrid approach using DTs and ANNs (Yu, 2009). Previous research on Preterm Birth focused on non-knowledge based techniques. Further studies applied a knowledge-based approach to the DT-ANN technique, incorporating a better validation strategy (a cross validation method to reduce bias, explained later in the chapter) (Yu, 2009).

The model derived earlier (Catley, 2007), utilized a method where network parameters were extracted from a classification based ANN; this was followed by applying a risk – stratification ANN to estimate preterm birth. That method, however,

only used Parous cases and USA specific variables as seen in the PRAMS dataset. The Parous classification refers to infants born to mothers with previous children versus Nulliparous classification where infants born to mothers with no previous children. This is significant because cases involving a previous pregnancy that resulted in a preterm birth are known to be high risk for a preterm birth outcome in subsequent pregnancies. Using the older PRAMS database, the system processed test data through 3 passes: one to maximize sensitivity, one to maximize specificity, and a third to classify ambiguous cases using a simple DT voting algorithm. The results had a sensitivity of 64%, a specificity of 84% and AUC of 0.7795, but required 48 variables to maintain such a performance (Catley, 2007) Though these results were acceptable by Catley's clinical advisor, Dr. Robin Walker, the drawback was the large number of variables, and it was not applicable to nulliparous cases.

Yu (2009) used a hybrid method integrating DTs, classification ANN with weight elimination and risk stratification. The hybrid classifier used decision trees to eliminate variables with little impact on predicting the outcome in question. The remaining variables were then processed through an ANN with weight elimination (ANN-we). This approach achieved a sensitivity of 65.13%, specificity of 84.07% and AUC of 0.8195. This method used 19 variables and included parous cases. Nulliparous cases were analyzed in a separate model; however its performance was a little less accurate than for parous cases, which is to be expected.

In this new work, improvements were made to the methodology used in Objectives 1 and 2. A few enhancements were designed: (a) change the scope to apply to a more general population of pregnancies, exploring parous and nulliparous scenarios and

considering non clinically obvious options (excluding pregnancy complications already known to lead to preterm birth); (b) introduce a new feature selection approach, to minimize the number of variables and aid with time constraints; (c) introduce a new validation strategy (5by2 cross validation); and (d) introduce a new ANN design for the hybrid approach. Such modifications should maintain past performance results or preferably exceed them. A General North American Model is needed and we explored the differences between two databases, one using very extensive USA variables and the other using more simplistic Ontario variables.

3.2 Ethics Clearance

This thesis work was approved by the Carleton University Research Ethics Committee and by the CHEO Ethics Review Board. The databases used in this thesis work were: (1) Better Outcomes Registry & Network (2010 - 2012) database and the (2) Pregnancy Risk Assessment Monitoring System (2002-2004) and (2009-2011). A contract was signed protecting the intellectual property of the data before accessing both databases for this thesis.

3.3 The Better Outcomes Registry and Network (BORN) Database

The BORN database contained over 669,000 cases with over 220 variables relating to census demographics, maternal and infant information and medical conditions, between the years of 2010 – early 2012.

Created on January 25th 2010 under the direction of the Ministry of Health and Long Term Care, the database holds maternal-child health data 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). BORN's vision is to obtain knowledge required to sustain lifelong health, and its mission is to facilitate and improve the care provided to maternal-fetal and child populations through linking the information and providers to recognize gaps in care. The database contains accurate, trusted and timely information in order to contribute to the healthcare system (The Mothers Program, 2013).

Data obtained by BORN is entered retrospectively and mainly focuses on maternal and newborn outcomes, midwifery information, congenital anomalies surveillance, newborn screening and prenatal screening. Health care providers and policy makers may use the data 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 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, there is a guideline, which describes the variable and value types. The value(s) are then converted into a numeric code. The BORN database contains over 290 variables, which are entered into the registry by healthcare providers after the completion of the pregnancy (BORN Ontario, 2013).

3.4 The Pregnancy Risk Assessment Monitoring System (PRAMS) Database

The PRAMS database contains over 113,000 cases with over 300 variables relating to demographics, maternal and infant behaviour and medical conditions between the years of 2002-2004; a newer database was obtained which contains data collected between 2009 and 2011.

The PRAMS database is a national database from 32 participating states of America and is collected by the Center for Disease Control and Prevention (CDC) in Atlanta, USA. The PRAMS database was established by the CDC in 1987, using state-specific, population-based sampling; data is collected on maternal and infant behaviour and experiences of prenatal, birth and post-partum (Centers for Disease Control and Prevention, 2009). It is designed to supplement data from vital records, and is used to monitor changes in maternal and child health indicators. The questionnaire is completed after birth by parents and contains questions pertaining to pre-pregnancy, pregnancy and post-partum behavior. The PRAMS database combines birth certificate variables (taken at the time of birth) as well as variables collected from the PRAMS questionnaire (collected a few months after delivery). PRAMS data collection methodology begins with a mail-in questionnaire sent 2- 4 months after delivery. Non-responders are followed up by telephone interview (Centers for Disease Control and Prevention, 2009).

3.5 Real World Data

Real world data is often difficult to obtain for research, especially medical data, due to privacy concerns. Typically data is generated to simulate patient data and is used for research and assessment. In this research, having access to real anonymized patient data from the BORN and PRAMS databases is unique in that the techniques can be applied to a real problem. However, with real world data, one has to deal with the accompanying issues not encountered with simulated data, that is, the data may not be clean, less voluminous, sparse, and may not be unique. This is a common issue when dealing with “big data”. But in our case, having real data of high quality enabled to work on a real life issue such as predicting preterm birth.

3.6 First Objective

The first objective was to perform a DT-ANN hybrid analysis on the BORN database to model the prediction of preterm birth, with an ANN. Details of this objective and sub objectives are as follows:

1a) Extend the Model to four scenarios of cases:

Preliminary results for Parous, Nulliparous, Parous without Obvious Variables and Nulliparous without Obvious variables outcome estimations are to be established. Catley's work focused on Parous cases which limits the generalized application of the model. The model should be more applicable to a generalized population of pregnant women, including those women who have no information about outcomes from prior pregnancies. In order to see if there is something beyond current clinician knowledge that the hybrid method may pick up, the idea of modeling cases other than simple Parous and Nulliparous was explored. To do so, we've chosen to observe the outcome of removing "clinically obvious" variables from parous and nulliparous identified cases, since these variables represent information about the pregnancy that are so significantly associated with preterm birth that the healthcare team would most likely classify the pregnancy as high risk based on knowledge of that information alone. These clinically obvious variables were identified and selected in consultation with our clinical partner.

1b) Find minimal number of variables required while keeping high sensitivity and maintain specificity:

Preterm Birth is a multifactorial problem and many predictor variables have been associated with this outcome. In Yu's thesis, she needed 19 variables to meet clinical expectations, a number which was significantly reduced from Catley's 48 variables from

the PRAMS database. To optimize a physician's time and reduce the complexity of the problem, this work attempts to apply previously successful methods and find the minimal number of input variables required for BORN datasets. This work is performed by using DTs to select significant variables prior to performing an analysis with ANNs. The method used by Yu for feature selection was used; additionally a new method of feature selection was introduced applying the use of percentage usage of a variable to create a DT.

1c) Introducing a 5by2 cross validation strategy:

A repeated 5by2-fold cross validation method was applied, throughout the process of the DT-ANN methodology, to reduce the bias of the results (Dietterich, 1998). The advantage is that all of the data values can be used for both training and testing, making sure that key variables are not left out.

1d) Introduce and test the ANN Builder in place of the MIRG ANN:

A newly developed automated ANN design replaced the previously used MIRG ANN. The MIRG ANN is a MATLAB based tool, which was unsuitable for real-time estimations and requires a commercial license for use, whereas the new ANN Builder is an open-source tool, developed with the FANN Library (FANN, 2014). Martirosyan's thesis work on developing the ANN Builder, and using a small NICU sample set of data, showed that it outperformed the MIRG ANN. This thesis work, in addition to applying the hybrid DT-ANN methodology, introduced and tested the new ANN with large datasets. The hope was to see if the BORN variables achieve a close or equivalent performance with past PRAMS results, (sensitivity of 64%, a specificity of 84% and AUC of 0.7795). By improving these aspects of the prediction tool and applying it to

other databases, comparisons can be made. This aims to reduce the complexity of the problem and helps towards advancing the generalization of a model to a general obstetrical population. This tool was designed to be of generic use, and has been used successfully to model the prediction of NICU mortality data (Martirosyan. 2014).

MIRG's research collaboration extends beyond Canadian Borders – with access to provincial databases (BORN) and USA databases (PRAMS). Results were assessed, looking at how selected variables compare and perform across the two databases, and between the two countries. This is expected to help identify key variables that can be used for a General Model, and help improve the information that will guide database developers on what variables are important for a database to collect.

3.6.1 Accomplishing Objective 1 – Methodology

Stage 1: Establish Model and Outcome of Interest

Stage 2: Data Preprocessing

Step 1: Cleaning Data and Handling Missing Values

Step 2: Feature Selection and Importance

Step 3: Preparing Datasets

- Dividing database into Parous, Nulliparous, Parous without Obvious and Nulliparous without Obvious sets, and preparing sets for data analysis at different stages:
 - i. Hold Out Testing for later Verification,
 - ii. ANN Analysis Data and;
 - iii. Data for Decision Tree Analysis
- Setting Prevalence

Stage 3: Decision Trees

Step 1: Preprocessing

- 5by2 cross validation to create train and test datasets.

Step 2: Software Execution

- Using See5 to generate Decision Trees

Step 3: Post Assessment – Results and Feature Selection

- i. Method 1 Selection – Compromise
- ii. Method 2 Selection – Average Usage

Stage 4: Artificial Neural Networks:

Step 1: Preprocessing

- i. Identify Significant Variables
- ii. Normalize/Standardize
- iii. Handle Missing Values
 - o Database Formatting
 - o Using the MIRG Case Based Reasoning System (CBRS)
- iv. ANN Builder Data Preparation
 - o Data Formatting
 - o 5by2 Cross-validation and Prep files

Step 2: Software Execution and Results

- ANN Model Selection Tool
- Performance Measures Calculation Tool
- ANN Cluster System Use
- Preliminary Results Collection

Stage 5: Hold Out Verification Data:

Step 1: Preprocessing:

- i. Identify Significant Variables
- ii. Normalize/Standardize
- iii. Handle Missing Values
 - o Database Formatting
 - o Using the MIRG Case Based Reasoning System (CBRS)
- iv. ANN Builder Data Preparation
 - o Data Formatting
 - o 5by2 Cross-validation and Prep files

Step 2: Software Execution and Results

- ANN Builder – Verification Mode
- Results Scripts.

3.6.1.1 Stage 1: Establish Model and Outcome of Interest

Preterm Birth can be classified if the gestational age at birth was less than 37 complete weeks. Our goal is to find factors that will help with identifying risk of preterm birth before the pregnancy reaches 23 weeks of gestation. With our clinical advisor Dr. Erika Bariciak, the list of variables was reduced, to select target variables that would aid with this identification, and be known before 23 weeks into the pregnancy. Catley (2007) indicated the need for prediction before 23 weeks of gestation. The reason for the choice

of 23 weeks is that infants can be viable as of week 23 and thus it is important to assess if a premature birth will happen before this period of gestation (Catley, 2007). Consequently, selected input variables should be available early in the pregnancy (< 23 weeks). The earlier outcome prediction allows for more time for the physician to implement a treatment plan.

As this study was to include a generalized population, variables selected were identified as parous (infants born to mothers with previous children) or nulliparous (infants born to mothers with no previous children). This separation was made due to the fact that certain variables found in Parous are irrelevant for Nulliparous cases and some of the information learned from a previous pregnancy can be highly predictive of subsequent similar outcomes. Additionally, our research group wanted to observe the outcome of removing “clinically obvious” variables from parous and nulliparous identified cases, to help determine how much extra clinical information this type of tool might add to the healthcare team managing pregnancies. Therefore four scenarios were to be modeled: Parous, Nulliparous, Parous without Obvious variables (ParousWO), and Nulliparous without Obvious variables (NulliparousWO). These were to be divided into four datasets to be analyzed.

3.6.1.2 Stage 2: Data Preprocessing

Data pre-processing is the first step in knowledge discovery. This step is performed before the creation of a pattern classification model. The preprocessing step is used to clean the raw data in the database. This is due to the potential existence of error types within a database, which could be: inconsistent data, missing values, non-relevant

details etc. Therefore, measures are taken to correct them. Measures to correct these errors include: eliminating irrelevant cases/attributes and/or imputing missing data.

3.6.1.2.1 Step 1: Cleaning Data and Handling Missing Values

The C5.0 algorithm is capable of handling missing values. A value that is missing or unknown was replaced with ‘?’ and a value that is not applicable to that particular case was denoted as ‘N/A’. The C5.0 algorithm takes the weighted average of the predictions across all branches upon encountering missing data (Rulequest Research, 2012).

However, in the event of a large number of values missing per case or per attribute, the following techniques were applied to this thesis work in order to handle missing information as Yu, 2009, applied in her thesis.

Eliminate Cases: Individual cases with more than 50% of variables missing were deleted. The outcome of interest for preterm birth is term or preterm; if the outcome was missing, the case was deleted.

Eliminate Attributes: If more than 50% of data was missing for a particular attribute, the attribute itself was deleted.

Imputation of Data: If data was missing because the information did not apply to that particular variable, then it may be corrected. For example, if the question was “Did you previously have caesarean?” and the answer was ‘No’, and another question was “How many caesareans have you previously had?” and the answer was blank, this may be corrected to ‘0’.

For the BORN database, eliminating cases and attributes was applied. A MATLAB script was developed to read the dataset as a csv file and eliminate cases with > 50% missing variables, and outputting the results in a new csv file.

3.6.1.2.2 Step 2: Feature Selection and Importance

One of the main tasks of analyzing a dataset for this thesis was to determine variables that are likely to be predictive. Variables that were not available before 23 weeks gestation were not considered. As mentioned previously, four scenarios were to be modeled: Parous, Nulliparous, Parous without Obvious variables (ParousWO), and Nulliparous without Obvious variables (NulliparousWO).

The Parous dataset originally contained 55 variables from the BORN database for the prediction of preterm birth. The list of variables can be seen below in Table 3.1.

The Nulliparous dataset originally contained 51 variables from the BORN database for the prediction of preterm birth. The list of variables can be seen below in Table 3.2.

The ParousWO dataset originally contained 47 variables from the BORN database for the prediction of preterm birth. The list of variables can be derived from the list of Parous variables Table 3.1, but eliminating the obvious variables, listed in Table 3.3.

The NulliparousWO dataset originally contained 44 variables from the BORN database for the prediction of preterm birth. The list of variables can be derived from the list of Nulliparous variables Table 3.2, but eliminating the obvious variables, listed in Table 3.3.

There can be a number of advantages and disadvantages of having too many or too few input variables. Having a larger number of variables may result in an improved

classifier, but it may also increase complexity, and lead to added computational costs and noise. Contrarily, too few variables may not provide enough information for proper classification. Less input variables help reduce complexity of a database and helps improve the generalization of a classifier, but it is just as important to keep the performance of the classifier high; ideally feature sets should only include variables that are important (Kantardzic, 2002).

Note on Census data exclusion:

We determined we could not use some of the BORN Census data variables as they appeared in the dataset. The data, in order to be useful, is linked to maternal postal code, but we were not provided with that information because of how identifiable it would be. We had decided to proceed forward without the census data. The types of Census data omitted included marital status, family size, education level, employment status, ethnic origin, and income. These have been identified as important variables associating with preterm birth indicators as identified by our clinician partner. (E. Bariciak, personal communications, 2015).

Table 3-1 BORN Parous List of Variables for DT Input including Outcome Variable.

BORN Parous Pre DTs		
	Variable	Description
1	MATWGTKG	Mother's weight – kg
2	MULTGEST	Multiple gestation
3	MATAGE	Mother's age (years)
4	PPRETERM	Number previous preterm babies
5	PTERM	Number previous term babies
6	PARITY	Parity
7	PREVCS	Previous C/S
8	GENDER	Baby's sex
9	INTBF	Intention to breastfeed
10	SMOKING	Smoking
11	MATHP1	Maternal health problem - Alcohol dependence syndrome/Alcoholism
12	MATHP2	Maternal health problem - Thyroid disease
13	MATHP3	Maternal health problem - Chronic hypertension
14	MATHP4	Maternal health problem - Diabetes insulin
15	MATHP5	Maternal health problem - Diabetes non-insulin
16	MATHP16	Maternal health problem - Heart disease
17	MATHP17	Maternal health problem - Hepatitis B
18	MATHP18	Maternal health problem – HIV
19	MATHP19	Maternal health problem – Lupus
20	MATHP26	Maternal health problem - Thyroid disease
21	MATHP27	Maternal health problem – Other
22	OBCOMP1	Obstetrical complications – Eclampsia
23	OBCOMP2	Obstetrical complications - Gestational diabetes
24	OBCOMP3	Obstetrical complications - Hypertension (gestational or transient)
25	OBCOMP4	Obstetrical complications – IUGR/SGA
26	OBCOMP5	Obstetrical complications – LGA
27	OBCOMP6	Obstetrical complications – Peridontal infection
28	OBCOMP7	Obstetrical complications - Placenta previa
29	OBCOMP8	Obstetrical complications - Placental abruption
30	OBCOMP9	Obstetrical complications – Pre-eclampsia
31	OBCOMP11	Obstetrical complications - Preterm labour
32	OBCOMP12	Obstetrical complications - Preterm premature
33	OBCOMP13	Obstetrical complications – UTI
34	OBCOMP14	Obstetrical complications - Other cervical/vaginal
35	OBCOMP15	Obstetrical complications – Other
36	REPASS	Reproductive assistance
37	FIRSTVIS	First trimester visit
38	CONGAN52	Anomalies unclassified elsewhere - conjoined
39	CONGAN54	Congenital anomalies – Anomalies unclassified elsewhere – Multisystem
40	CONGAN55	Congenital anomalies - Anomalies unclassified elsewhere - Other syndromes
41	PRENCLAS	Prenatal classes
42	PCSNUM	No. of previous C/S's
43	BMI	Calculated: Mother's BMI
44	LANGUAGE_up	Mother's primary language
45	MATHP_sub	Maternal health problem - Substance use
46	MATHP_ment	Maternal health problem - Mental health
47	CONGAN_CNS	Congenital anomalies - Central Nervous System
48	CONGAN_EYE	Congenital anomalies – Eye/Ear
49	CONGAN_OROFACIAL	Congenital anomalies – Orofacial

50	CONGAN_CARDIAC	Congenital anomalies – Cardiac
51	CONGAN_RES	Congenital anomalies - Respiratory system
52	CONGAN_GAS	Congenital anomalies – Gastrointestinal
53	CONGAN_GEN	Congenital anomalies – Genitourinary
54	CONGAN_MUS	Congenital anomalies – Musculoskeletal
55	CONGAN_CHR	Congenital anomalies – Chromosomal
56	TERM	Outcome

Table 3-2 BORN Nulliparous List of Variables for DT Input including Outcome Variable.

BORN Nulliparous Pre DTs		
	Variable	Description
1	MATWGTKG	Mother's weight – kg
2	MULTGEST	Multiple gestation
3	MATAGE	Mother's age (years)
4	PARITY	Parity
5	GENDER	Baby's sex
6	INTBF	Intention to breastfeed
7	SMOKING	Smoking
8	MATHP1	Maternal health problem - Alcohol dependence syndrome/Alcoholism
9	MATHP2	Maternal health problem - Thyroid disease
10	MATHP3	Maternal health problem - Chronic hypertension
11	MATHP4	Maternal health problem - Diabetes insulin
12	MATHP5	Maternal health problem - Diabetes non-insulin
13	MATHP16	Maternal health problem - Heart disease
14	MATHP17	Maternal health problem - Hepatitis B
15	MATHP18	Maternal health problem – HIV
16	MATHP19	Maternal health problem – Lupus
17	MATHP26	Maternal health problem - Thyroid disease
18	MATHP27	Maternal health problem – Other
19	OBCOMP1	Obstetrical complications – Eclampsia
20	OBCOMP2	Obstetrical complications - Gestational diabetes
21	OBCOMP3	Obstetrical complications - Hypertension (gestational or transient)
22	OBCOMP4	Obstetrical complications – IUGR/SGA
23	OBCOMP5	Obstetrical complications – LGA
24	OBCOMP6	Obstetrical complications – Periodontal infection
25	OBCOMP7	Obstetrical complications - Placenta previa
26	OBCOMP8	Obstetrical complications - Placental abruption
27	OBCOMP9	Obstetrical complications – Pre-eclampsia
28	OBCOMP11	Obstetrical complications - Preterm labour
29	OBCOMP12	Obstetrical complications - Preterm premature
30	OBCOMP13	Obstetrical complications – UTI
31	OBCOMP14	Obstetrical complications - Other cervical/vaginal
32	OBCOMP15	Obstetrical complications – Other
33	REPASS	Reproductive assistance
34	FIRSTVIS	First trimester visit
35	CONGAN52	Anomalies unclassified elsewhere - conjoined
36	CONGAN54	Congenital anomalies – Anomalies unclassified elsewhere – Multisystem
37	CONGAN55	Congenital anomalies - Anomalies unclassified elsewhere - Other syndromes
38	PRENCLAS	Prenatal classes
39	BMI	Calculated: Mother's BMI
40	LANGUAGE_up	Mother's primary language

41	MATHP_sub	Maternal health problem - Substance use
42	MATHP_ment	Maternal health problem - Mental health
43	CONGAN_CNS	Congenital anomalies - Central Nervous System
44	CONGAN_EYE	Congenital anomalies – Eye/Ear
45	CONGAN_OROFACIAL	Congenital anomalies – Orofacial
46	CONGAN_CARDIAC	Congenital anomalies – Cardiac
47	CONGAN_RES	Congenital anomalies - Respiratory system
48	CONGAN_GAS	Congenital anomalies – Gastrointestinal
49	CONGAN_GEN	Congenital anomalies – Genitourinary
50	CONGAN_MUS	Congenital anomalies – Musculoskeletal
51	CONGAN_CHR	Congenital anomalies – Chromosomal
52	TERM	Outcome

Table 3-3 BORN List of Obvious Variables removed from Parous and Nulliparous Datasets.

BORN Obvious Variables		
	Variable	Description
1	MULTGEST	Multiple gestation
2	PPRETERM	Number previous preterm babies
3	OBCOMP1	Obstetrical complications – Eclampsia
4	OBCOMP3	Obstetrical complications - Hypertension (gestational or transient)
5	OBCOMP4	Obstetrical complications – IUGR/SGA
6	OBCOMP9	Obstetrical complications – Pre-eclampsia
7	OBCOMP11	Obstetrical complications - Preterm labour
8	OBCOMP12	Obstetrical complications - Preterm premature

3.6.1.2.3 Step 3: Preparing Datasets

Once the lists of potentially important variables were determined and the BORN database was initially preprocessed, the dataset was partitioned into four sets containing the groupings identified previously: Parous, Nulliparous, ParousWO and NulliparousWO. For additional measures, cases with >50% missing variables were eliminated. It was found that two variables (MATWGTKG and BM1), had >50% missing values, so these were eliminated from further assessment.

Parous dataset therefore had 53 variables and 347,178 total numbers of cases. Nulliparous dataset therefore had 49 variables and 275,540 total numbers of cases. ParousWO dataset therefore had 45 variables and 348,975 total numbers of cases. NulliparousWO dataset therefore had 42 variables and 275,868 cases.

Table 3-4 Summary of Datasets before DTs

PreDT BORN Post Script Processing				
Dataset	No. Features	Positive Cases	Negative Cases	Total Cases
Parous	53	24,709	322,469	347,178
Nulliparous	49	23,497	252,043	275,540
ParousWO	45	24,689	324,286	348,975
NulliparousWO	42	23,513	252,355	275,868

i) Hold Out Testing Data for Later Verification:

Once these datasets were established, a sample size of the most recently entered cases was removed from the dataset to be used later as unseen data for verification testing with ANN results (hold out testing). For BORN datasets, a group of 25,000 cases were selected and removed completely from the pool of cases for the final verification; half (12,500) of these selected cases needed to be positive (preterm) and the other half (12,500) of these cases needed to be negative (term). These cases were stored in a separate file. From the 25,000 cases set aside, cases would be randomly pulled out to create a verification set of 12,500 cases total, with a prevalence set to 8.1%. A selection size of 25,000 cases was set to ensure that there would be enough information to make a verification dataset with the size of 12,500 cases total (a quarter of the ANN Analysis Sample - 50,000 divided by 4 = 12,500 cases). There also needed to be enough cases to set prevalence of the verification set to match the rate of preterm birth in Ontario (8.1%) (CIHI, 2012).

ii) ANN Analysis Data:

From the remaining dataset, a sample of cases was selected for use with the ANN (data to train, test and verify), to select best networks. For BORN, a sample set of 50,000 cases was copied from the dataset and set aside for this initial ANN analysis. However,

this data needs to have a prevalence of 15% for the ANN; for optimal training and network selection, a higher than reported prevalence can be used to get best results (Ennett & Frize, 1998 and 2000). Therefore, 15% of 50,000 was 7,500 cases required for preterm and the remaining 42,500 cases for term (85%).

Using the filtering functions in Microsoft Excel, the 50,000 cases of sample data for ANN were filtered and selected based on cases with the least amount of missing values, and were cases that had been the most recently entered into the database.

The 50,000 case size sample was selected due to restrictions of a tool (CBRS) described later on (section 3.5.1.4.1. iii)) that plays a crucial part in ANN Preprocessing and Preparation.

iii) Data for Decision Tree Analysis

All of the data, (including the 50,000 from step ii) except for the hold out data (the 25,000 case removed for verification sets – step i) was used for Decision Tree analysis. Before proceeding forward, another check for >50% missing variables and values was performed, to ensure there was enough data for proper analysis. Because a 15% prevalence was used for ANN analysis, the data for Decision Tree analysis was set to 15% as well. However with such a large difference in number of preterm cases to the number of term cases there was a need to adjust the number of cases so that the 15% prevalence could be met. The number of cases selected for each scenario (parous, nulliparous and without obvious) was optimized for DTs to analyze the most possible number of cases, with the 15% prevalence.

Ex. Simple math was applied to determine numeric values required (parous numbers used):

$$15\% = 12,000 \text{ preterm cases} / X \text{ total cases; want to determine } X.$$

$$X = 12,000 / 15\%$$

$$X = 80,000$$

Thus 80,000 total cases – 12,000 preterm cases = 68,000 term cases (the number of negative cases required to meet prevalence of positive cases).

Setting Prevalence:

Because training with high prevalence generates good resulting networks (Ennett & Frize 1998 and 2000), 15% was selected to train the ANN to find these best networks. Holdout verification testing with an unseen dataset was performed using the networks selected from the 15% prevalence dataset. This unseen data was set to a prevalence of 8.1% - the actual rate of preterm birth in Ontario (CIHI, 2012). This hold out verification testing was done to provide real world conditions to the model to give a more accurate estimate of results with new unseen data – a virtual prospective trial of future patients.

3.6.1.3 Stage 3: Decision Trees

3.6.1.3.1 Step 1: Preprocessing

5by2 cross validation

For additional verification, randomization, and to reduce bias on the dataset, a repeated 2-fold cross validation method (see Figure 3-1) was applied (Dietterich, 1998). Used by Gilchrist, three bash scripts were applied to the datasets (create_5by2_folds.sh, duplicate_names.sh and run_dt_5by2.sh), creating 5-by-2 fold cross validation sets (Gilchrist, 2012). The advantage of this, is that all of the data values can be used for both training and testing, making sure that key variables are not left out.

This works by first splitting the dataset (to be analyzed), into positive and negative cases (Step 1). Then the dataset is randomly split into equal sized training and

test sets. The first sets of positive and negative cases are grouped for training (Step 2 – A cases) and the second sets for testing (Step 2 – B cases). This allows an equal number of positive and negative cases to be selected in the train and test sets. The grouped train and test sets were then reversed and a new train set was created (formerly the test set) and new test set (formerly the train set)(Step 3). This process was repeated 5 times to create 10 separate training and test sets (Dietterich, 1998).

Figure 3-1 Steps to randomize a dataset using repeated 2-fold cross validation technique

3.6.1.3.2 Step 2: Software Execution

Using See5 to generate Decision Trees

The c5.0 algorithm was executed using an open source application known as See5 (it is also available as a commercial application). See5 requires the dataset to be divided into 3 distinct file types: names, data and test.

The names file contains the attributes and values. The attributes may be divided into explicitly defined attributes such as discrete, continuous, date, time etc. or implicitly defined attributes, which may be specified by a formula (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 a comma separated each value, 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. 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).

The scripts written by Gilchrist, format the resulting 5-by-2-fold train and test sets into the required data and test files to be read by See5, and also launch and run the application for these scripts. The names file has to be created to match the variables contained in the data being analyzed.

For the Parous related decision trees, each tree was made with a cost of misclassifying positive cases set at a value of 6.5. The Nulliparous related DTs were set at a misclassification cost of 6.0. This cost allowed the DT to emphasize classification of positive cases and therefore favouring the sensitivity seen in test results. These specific

values were determined to be ideal for use by Yu, as optimal with the data and See5, (Yu, 2009). Two methods for variable selection were used based on the DT outputs, to be used for the ANN Builder as inputs.

3.6.1.3.3 Step 3:Post Assessment – Results and Feature Selection

Results from executing See5 were stored in a text file. This file includes a list of main variables that were used to construct the classifier, including the percentage usage of the attributes and listed in order of highest importance. These files also output a confusion matrix. The confusion matrix provides additional details on the correct and incorrect classifications, and was used to evaluate various performance measures (Rulequest Research, 2012).

Two Methods of Feature Selection:

This thesis applies the initial method of feature selection used by Yu (2009), and introduces a new method of feature selection.

i) Method 1 Selection - Compromise:

The first method of feature selection outlined by Yu (2009), with the 5-by-2 cross validation method, assesses the results of the ten decision trees produced by See5, using performance measures, calculated from the confusion matrix output. To mirror the method used by Yu (2009), the tree with the best compromise, (smallest calculated difference - while optimizing sensitivity), between sensitivity and specificity was selected, and where little variation occurred, the number of variables used was factored (tree with smaller number of variables used became the determining factor). Once selected the variables in that tree were used and incorporated as features to build an ANN model for PTB.

ii) Method 2 Selection – Average Usage:

The idea was to come up with a way to select variables in the DT based on high average percent usage, but also automate the removal of variables without the constant intervention of a clinician.

We had our clinical advisor go through the variables and classify each into one of two categories: 1) include always (based on clinical logic), and 2) maybe include or exclude depending on percentage of usage by DT, based on including only a select number of variables for a given DT. For BORN data, 19 features were identified as include always (See Table 3.5).

Once all the ten DT models from See5 are generated, variables with their usages across the ten trees were taken and averaged. The variables that were classified as “include always”, regardless of their % average, were kept and the remainder of the variables were selected with the highest percentage of usage to give a select number of variables for a given DT. This accounts for the variation in model generation from the DT’s and cases.

Due to the large size of the BORN dataset, and limited testing capability of large datasets on the ANN Builder, this second methodology was originally developed to help reduce the size of dataset based on features. Time required to run the program, was a key factor and testing was performed as data was being preprocessed for ANN stages. Based on the testing results of BORN method 1, the largest number of variables assessed was capped at 45 variables. At the time this was the largest number of variables used and had a long run time, thus 45 variables was the cut-off number per dataset to be further analyzed in the ANN (note an exception in PRAMS) which, resulted in a total of 8 datasets to be analyzed with the ANN: 4 dataset categories x 2 methods.

Table 3-5 BORN Variables Inclusion Identification – 1, Include Always based on logic or 2, maybe include depending on % usage.

BORN Variables Inclusion for Method 2					
Variable	Inclusion	Variable	Inclusion	Variable	Inclusion
1 OBCOMP11	1	20 OBCOMP15	2	39 MATHP27	2
2 MULTGEST	1	21 LANGUAGE_up	2	40 INTBF	2
3 OBCOMP12	1	22 MATHP26	2	41 PREVCS	2
4 OBCOMP9	1	23 GENDER	2	42 MATHP18	2
5 OBCOMP7	1	24 PRENCLAS	2	43 MATHP_sub	2
6 OBCOMP4	1	25 REPASS	2	44 PCSNUM	2
7 PPRETERM	1	26 OBCOMP6	2	45 PARITY	2
8 OBCOMP8	1	27 FIRSTVIS	2	46 SMOKING	2
9 MATHP4	1	28 MATHP_ment	2	47 OBCOMP5	2
10 MATHP3	1	29 MATHP17	2	48 MATHP16	2
11 MATHP5	1	30 CONGAN_CARDIAC	2	49 OBCOMP14	2
12 CONGAN_GAS	1	31 CONGAN_OROFACIAL	2	50 MATHP19	2
13 OBCOMP3	1	32 CONGAN_GEN	2	51 PTERM	2
14 OBCOMP1	1	33 CONGAN_RES	2	52 MATHP2	2
15 OBCOMP2	1	34 CONGAN_CNS	2	53 CONGAN_CHR	2
16 MATWGTKG	1	35 CONGAN55	2	54 MATHP1	2
17 MATAGE	1	36 CONGAN_EYE	2	55 CONGAN52	2
18 OBCOMP13	1	37 CONGAN54	2		
19 BMI	1	38 CONGAN_MUS	2		

3.6.1.4 Stage 4: Artificial Neural Networks:

3.6.1.4.1 Step 1: Preprocessing

i) Identify Significant Variables

Once a list of significant variables was determined (as in section 3.5.1.3.3), datasets were created, and partitioned to match the variable list. This is done for each of the four sets, Parous, Nulliparous, Parous WO, Nulliparous WO, for each method (1 & 2). These datasets then had to be prepared for the ANN Builder.

ii) Normalize/Standardize

Data values are often normalized to account for large range differences that may occur between values, and to increase the efficiency of the ANN (Durai & Saro, 2006). This is used to help minimize the bias of one feature over another. Studies suggest that neural networks perform best when the values range between -1 and 1. In this study, 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):

$$x_i^n = \frac{x_i - \mu_i}{3\sigma_i}$$

where: x_i^n = Normalized value
 x_i^n = Value
 μ_i = Average
 σ_i = Standard deviation

A MATLAB script was developed to perform the normalization for the datasets.

iii) Handle Missing Values:

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

Database Formatting:

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.

Using the MIRG Case Based Reasoning System (CBRS)

The CBR system was used to fill in missing values. The KNN algorithm retrieves the closest match from the training sample for each missing value and replaces the missing value by accounting for the weights assigned to each input variable (Ennett 2003). This application was written 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 Interface (Figure 3-2) was designed using the Swing API (Cotea & Jiwani, 2003).

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.

Step 1: Set Up Connection

Data Source Name:

Username:

Password:

of Primary Keys:

Step 2: Make your Selection

Patient ID:

of matches:

Weights:

HBLOODP	LBLOODP	HHEARTR	LHEARTR	HRESPR	LTEMPF	Lf
100	100	90	50	0	100	0

Step 3: Send Request

Figure 3-2 Screenshot of the CBR Interface

iv) ANN Builder Data Preparation

Data Formatting:

Once the results file was output by the CBRs, it was converted to a comma, separated, value (csv) file format. For ANN Builder analysis the case number (CASELINK) has to be moved to the first column, and outcome should be in the last column. From here, the data needs to be split back into positive and negative cases, so that files can be split again into a 5-by-2 cross validation.

5by2 Cross-validation and Preparation files

As mentioned previously, the 5-by-2 cross validation method was used to obtain more reliable risk estimation models. Martirosyan (2014) adapted and modified scripts

created by Gilchrist (2012) to achieve a division into train, test and verification sets for ANNs. The process was as follows as outlined by Martirosyan (2014):

1. Data were separated into mortality and non-mortality sets.
2. Both the mortality and non-mortality sets were randomized and split into two equal sized sets (A and B)
3. The set A was split into A1 and A2 for both the mortality and non-mortality sets, leaving the set B as a whole set. Then the process was reversed: The set B was split into B1 and B2, and the set A was left as a whole set.
4. The mortality and non-mortality sets B, A1, A2 were merged to form the train, test and verification sets respectively, and were labeled as Set #1. Then, the sets A, B1, B2 were merged to form the second set, Set #2.

To create ten different train, test and verifications sets (Set #1 through Set #10), the steps 2 through 4 were repeated five times in total.

The division of the mortality and non-mortality sets into equal parts ensures that an equal number of positive and negative cases are trained, tested, and verified by the ANN models. The 5-by-2 cross validation method is advantageous since it allows all patient cases to be used during the training, not letting any case that lead to better predictions to be excluded (Dietterich, 1998), (Gilchrist, 2012). The 5-by-2 method is summarized in Figure 3-3.

Figure 3-3 5-by-2 Cross Validation Method for ANN

In the FANN library, the input files must follow certain formatting guidelines:
The first line in the input file must include the number of cases, the number of attributes

and the number of outputs. A line of inputs and a line of outputs for each case should follow. The formatting requirements are summarized in Table 3.6. Martirosyan (2014) wrote scripts that automatically formatted the 5-by-2 cross validation files.

Table 3-6 ANN Builder Formatting Requirements

ANN Builder Format		
Number of Cases	Number of input attributes	Number of outcomes
	Case i ⁿ inputs	
	Case i outputs	
ⁿ where n is number of cases		

3.6.1.4.2 Step 2: Software Execution and Results

Execution of the ANN Builder software to perform analysis on the datasets used the following settings: FAST mode with sub parameters, 3 hidden layers, thresholds of 25% sensitivity and 70% specificity for saving networks and 20 different hidden node values spread equally across the range of $2n+1$ hidden nodes.

FAST Mode tests approximately 0.5% of all possible combinations, excluding the number of hidden nodes, which depends on the data. Using the sub parameters option with fast mode adds 6 more parameters, resulting in more combinations to test. See section 2.9.4.1 for more details regarding the FANN and the modes. Figure 3-4 illustrates some of the training algorithms, activation functions and network parameters associated with FAST Mode applied to a BORN database sample. The full details of this sample can be found in Appendix C.

```

good_net_1
FANN_FLO_2.1
num_layers=3
learning_rate=0.500000
connection_rate=0.100000
network_type=0
learning_momentum=0.000000
training_algorithm=3
train_error_function=0
train_stop_function=0
cascade_output_change_fraction=0.010000
quickprop_decay=-0.000100
quickprop_mu=1.500000
rprop_increase_factor=1.200000
rprop_decrease_factor=0.500000
rprop_delta_min=0.000000
rprop_delta_max=50.000000
rprop_delta_zero=0.100000
cascade_output_stagnation_epochs=12
cascade_candidate_change_fraction=0.010000
cascade_candidate_stagnation_epochs=12
cascade_max_out_epochs=150
cascade_min_out_epochs=50
cascade_max_cand_epochs=150
cascade_min_cand_epochs=50
cascade_num_candidate_groups=2
bit_fail_limit=3.49999994039535522461e-01
cascade_candidate_limit=1.00000000000000000000e+03
cascade_weight_multiplier=4.00000005960464477539e-01
cascade_activation_functions_count=10
cascade_activation_functions=3 5 7 8 10 11 14 15 16 17
cascade_activation_steepnesses_count=4
cascade_activation_steepnesses=2.50000000000000000000e-01 5.00000000000000000000e-01 7.50000000000000000000e-01 10.00000000000000000000e-01
layer_sizes=46 77 2
scale_included=0
neurons (num_inputs, activation_function, activation_steepness)=(0, 0, 0.00000000000000000000e+00)
connections (connected_to_neuron, weight)=(45, -5.91836869716644287109e-01) (35, 1.076799482107162e-01)

```

Figure 3-4 Shows a screen shot of the details for a sample of the final options used for best networks for BORN Parous Method 2 Option 0, Good Network 1, at 15% prevalence.

Scripts written by Martirosyan (2014) were used to calculate the performance measures of the ANN Builder results. To simplify and automate the process of network selection from result files, a sorting script was also written.

ANN Model Selection Tool (Martirosyan 2014):

The script can be used to select the networks either by optimizing sensitivity or specificity. Once a selection option is made (e.g. optimizing for sensitivity) for each result file (10 sets after 5-by-2 cross validation), a series of sorting is done according to maximum sensitivity within verification sets and test sets. The best performing network is selected; the path to the network file, was parsed to extract the name of the network

file. After all 10 result files were sorted, the folder contained 10 selected networks. A selection according to the specificity worked in a similar way.

Options also exist for how to select for best type of sensitivity or specificity. For the purpose of this work, an option selection of 0 was selected to optimize by largest outcome of sensitivity.

Performance Measures Calculation Tool (Martirosyan 2014):

The MIRG ANN Builder can be used for the purpose of calculating further performance measures of selected networks.

The script written to calculate the performance measures uses the scores.csv file. First, it finds the number of positive and negative cases and then calculates the elements of the confusion matrix. From there, the Positive Predictive Value (PPV), Negative Predictive Value (NPV), Accuracy (ACC), Matthews Correlation Coefficient (MCC) and F1-Score are calculated (see Section 2.9 for equations). Afterwards, the ROC points and the Area Under Curve (AUC) are calculated. These steps were repeated for each of the ten result files, and all results saved into a classifier_statistics.csv file. As a last step, the means and standard deviations of these ten sets' performance measures are calculated and directed to the classifier_statistics_final.csv file.

Mean (Average):

$$\mu = \frac{\sum_{i=1}^N x_i}{N}$$

μ = average

N = total number of values

X_i = value

Standard Deviation:

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum (x_i - \mu)^2}{N}}$$

σ = standard deviation

σ^2 = variance

μ = average

N = total number of values

X_i = value

F1-score: uses both the precision and recall of the test, acting as a weighted average of both these measures

$$F1 = \frac{2 TP}{2TP + FP + FN}$$

The Mathews Correlation Coefficient (MCC): is a correlation coefficient between the observed and predicted classifications.

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

ANN Cluster System Use

Due to the limitations of computing power of a desktop computer, the ANN network generation for this thesis was performed using the ANN Builder tool deployed on SHARCNET, a High Performance Computing (HPC) platform shared by a network of 18 Canadian academic institutions (SHARCNET, 2015) (Inside HPC, 2015). SHARCNET was selected due to the large amount of computing power available and the compatibility of the software and libraries on the system.

HPC is the use of many high-end computing resources to aid with solving complex problems or to run computationally intensive workloads. In this case, the computer intensive workload was beyond that of tasks achievable on a desktop system.

Work performed for this thesis took approximately 54 CPU years of time on the SHARCNET cluster. That is, if data was analyzed by the ANN Builder on 1 CPU of 1 desktop machine, it would have taken approximately 54 years running 24 hours a day to complete.

Preliminary Results Collection

A selection of top results was pulled from the cluster's resulting networks, as there were millions of networks that resulted from the ANN Builder. This selection was based on a high sensitivity value, using the ANN Builder scripts to perform this task.

Once these preliminary results were pulled from the cluster system, a more thorough selection for best network was made using the ANN Builder Tool. Upon best network selection, the tool also outputs the resulting performance measures.

When the best networks with 15% prevalence had been determined, these networks were then used to perform a verification analysis on an unseen dataset with the true population prevalence of 8.1% preterm births.

3.6.1.5 Stage 5: Hold Out Verification Data:

Holdout verification testing with an unseen dataset was performed using the networks selected from the 15% prevalence dataset. This hold out verification testing was used to provide real world conditions to the model to give a more accurate estimate of results with new unseen data – a virtual prospective trial of future patients. In this case an 8.1% prevalence was set to match the Ontario preterm birth rate (CIHI, 2012).

3.6.1.5.1 Step 1: Preprocessing

This process followed the same steps as in section 3.5.1.4.1 i – iv (Identify Significant Variables, Normalize/Standardize, Handle Missing Values and ANN Builder Data Preparation), in preparation for use of the ANN tool. This was applied to the sets that were created to match verification data.

iv) ANN Builder Data Preparation:

The same steps were applied to the verification data, as in section 3.5.1.4.1 however here the 5by2 cross validation step was not required. Only the one set created was required. However, since previously the 5by2 cross validation method was applied and best networks were selected, 10 verification files were needed to assess these networks. Therefore 10 copies of the verification set were made for this.

3.6.1.5.2 Step 2: Software Execution and Results

ANN Builder – Verification Mode:

A special mode can be used to run the results of the 10 best-selected networks with the matching dataset for verification.

Results Scripts:

Similar scripts as before can be used to calculate the performance measure results for this verification set.

3.7 Second Objective

The second objective was to extend the DT-ANN hybrid approach and perform analysis on the new PRAMS database to create another prediction model (ANN) for preterm birth. Details of this objective are similar to the first objective:

- **2a) Extend the Model to four scenarios of cases**

- **2 b) Find minimal number of variables required while keeping high sensitivity and maintain specificity**
- **2 c) Introducing a 5by2 cross validation strategy**
- **2 d) Introduce and test the ANN Builder in place of the MIRG ANN**

Once the results were obtained, a comparative analysis was to be done, comparing New PRAMS data to Old PRAMS data, and with the BORN database. It also must be noted that a number of variables that had been previously collected in PRAMS were removed in the new PRAMS, or had a name change, or were collected differently from the previous version. Another factor that may have impact on the results between the older and newer datasets is a difference of opinion regarding variable selection and clinical acceptance of performance from our current clinical advisor, Dr. Bariciak, and the past advisor, Dr. Walker. The applied overall methodology follows the same steps as in Objective 1 with differences in number of variables and cases.

3.7.1 Accomplishing Objective 2 – Methodology

The Objective 2 Methodology follows the same steps as Objective 1 – See Section 3.5.1

3.7.1.1 Stage 1: Establish model and outcome of interest

Preterm birth is classified as gestational age at birth less than 37 complete weeks of less than 259 days. It is important to note that the PRAMS database records gestational age in days unlike the BORN database, which uses weeks.

Once again since this study was to include a generalized population, variables selected were identified as parous or nulliparous, this time applied to the PRAMS database. The group also wanted to observe the outcome of removing “clinically

obvious” variables from parous and nulliparous identified cases. Therefore four scenarios were to be modeled: Parous, Nulliparous, Parous without Obvious variables (ParousWO), and Nulliparous without Obvious variables (NulliparousWO). These were to be divided into four datasets to be analyzed.

3.7.1.2 Stage 2: Data Preprocessing

See Section 3.5.1.2.

3.7.1.2.1 Step 1: Cleaning Data and Handling Missing Values

See Section 3.5.1.2.1

3.7.1.2.2 Step 2: Feature Selection and Importance

One of the main tasks of analyzing a dataset for this thesis was to determine variables that are likely to be predictive. Variables that were not available before 23 weeks gestation were not considered. As mentioned previously, four scenarios were to be modeled: Parous, Nulliparous, Parous without Obvious variables (ParousWO), and Nulliparous without Obvious variables (NulliparousWO).

The Parous dataset originally contained 51 variables from the PRAMS database for the prediction of preterm birth. The list of variables can be seen below in Table 3.7.

The Nulliparous dataset originally contained 45 variables from the PRAMS database for the prediction of preterm birth. The list of variables can be seen below in Table 3.8.

The ParousWO dataset originally contained 40 variables from the PRAMS database for the prediction of preterm birth. The list of variables can be derived from the

list of Parous variables Table 3.7, but eliminating the obvious variables, listed in Table 3.9.

The NulliparousWO dataset originally contained 36 variables from the PRAMS database for the prediction of preterm birth. The list of variables can be derived from the list of Nulliparous variables Table 3.8, but eliminating the obvious variables, listed in Table 3.9.

Table 3-7 PRAMS Parous List of Variables for DT Input including Outcome Variable.

PRAMS Parous Pre DTs		
	Variable	Description
1	URB_RUR	URBAN/RURAL CATEGORY
2	MM_HBP	Hypertension?
3	MM_DIAB	Diabetes?
4	MARRIED	Marital Status
5	SEX	Gender of Infant
6	PLURAL	Plurality
7	MAT_ED	Maternal Education
8	PAT_ED	Paternal Education
9	MOMSMOKE	Did Mom Smoke?
10	MOMLBS	Maternal Weight Gain
11	OTH_TERM	Pregnancy History Other Terminations?
12	MAT_RACE	Maternal Race
13	PAT_RACE	Paternal Race
14	PRE_DIET	(a) Pre-pregnancy dieting
15	PRE_EXER	(b) Pre-pregnancy exercise 3+ days/wk.
16	PRE_RX	(c) Pre-pregnancy prescrip meds
17	PRE_MH	(f) Pre-pregnancy ck/tx for depression/a
18	PRE_DDS	(h) Pre-pregnancy teeth cleaned
19	VITAMIN	Multivitamin -- # of times /wk
20	BPG_DIAB	Diabetes before pregnancy
21	PREV_LB	Previous -- live birth
22	PREV6NBW	Previous -- normal birth weight
23	PREV_PRE	Previous -- premature
24	PREG_TRY	Preg -- trying
25	PG_GDB	Gestational diabetes- this pregnancy
26	MORB_BLD	a) Morbidity -- vag bleeding
27	MORB_KID	b) Morbidity -- kidney/bladder infect
28	MORB_NAU	c) Morbidity -- nausea
29	MORB_PLA	f) Morbidity -- placenta
30	MORB_LAB	g) Morbidity -- preterm labour
31	MORB_TRN	i) Morbidity -- Blood transfusion
32	MORB_CAR	j) Morbidity -- car crash injury
33	PAB6HUS	Abuse -- 12 mnths b4 preg, husb/part
34	PAD6HUS	Abuse -- dur preg, husb/part
35	INCOME5	Income -- 12 mnths bef, total income

36	LANGUAGE	Language -- English/Spanish
37	MOM_BMIG_QX_REV	VAR: MOM BMI GROUPE FROM QX REVISED
38	PREGHX	VAR: HX PREV LIVE BIRTHS
39	PNC_WKS	VAR: WEEKS 1ST PNC VISIT
40	SMK63L_A	VAR: YES/NO SMOKE LAST 3 MTH PREG
41	DRK63L_A	VAR: YES/NO DRINK LAST 3 MTH PREG
42	STRS_T_G	VAR: TOTAL # STRESSES, GROUPE
43	MAT_AGE_NAPHSIS	Maternal age grouped
44	LGA	Large for Gestational Age Based on 90th Percentile
45	SGA_10	Small for Gestational Age Based on 10th Percentile
46	FRACE_OTH	Paternal race - Other
47	FRACE_WHT	Paternal race - White
48	INFER_TR	Infertility treatment
49	MRACE_OTH	Maternal race - Other
50	MRACE_WHT	Maternal race - White
51	P_PRTERM	Previous preterm births
52	TERM	Outcome

Table 3-8 PRAMS Nulliparous List of Variables for DT Input including Outcome Variable.

PRAMS Nulliparous Pre DTs		
	Variable	Description
1	URB_RUR	URBAN/RURAL CATEGORY
2	MM_HBP	Hypertension?
3	MM_DIAB	Diabetes?
4	MARRIED	Marital Status
5	SEX	Gender of Infant
6	PLURAL	Plurality
7	MAT_ED	Maternal Education
8	PAT_ED	Paternal Education
9	MOMSMOKE	Did Mom Smoke?
10	MOMLBS	Maternal Weight Gain
11	MAT_RACE	Maternal Race
12	PAT_RACE	Paternal Race
13	PRE_DIET	(a) Pre-pregnancy dieting
14	PRE_EXER	(b) Pre-pregnancy exercise 3+ days/wk
15	PRE_RX	(c) Pre-pregnancy prescrip meds
16	PRE_MH	(f) Pre-pregnancy ck/tx for depression/a
17	PRE_DDS	(h) Pre-pregnancy teeth cleaned
18	VITAMIN	Multivitamin -- # of times /wk
19	BPG_DIAB	Diabetes before pregnancy
20	PREG_TRY	Preg -- trying
21	PG_GDB	Gestational diabetes- this pregnancy
22	MORB_BLD	a) Morbidity -- vag bleeding
23	MORB_KID	b) Morbidity -- kidney/bladder infect
24	MORB_NAU	c) Morbidity -- nausea
25	MORB_PLA	f) Morbidity -- placenta
26	MORB_LAB	g) Morbidity -- preterm labor
27	MORB_TRN	i) Morbidity -- Blood transfusion
28	MORB_CAR	j) Morbidity -- car crash injury
29	PAB6HUS	Abuse -- 12 mnths b4 preg, husb/part
30	PAD6HUS	Abuse -- dur preg, husb/part

31	INCOME5	Income -- 12 mnths bef, total income
32	LANGUAGE	Language -- English/Spanish
33	MOM_BMIG_QX_REV	VAR: MOM BMI GROUPED FROM QX REVISED
34	PNC_WKS	VAR: WEEKS 1ST PNC VISIT
35	SMK63L_A	VAR: YES/NO SMOKE LAST 3 MTH PREG
36	DRK63L_A	VAR: YES/NO DRINK LAST 3 MTH PREG
37	STRS_T_G	VAR: TOTAL # STRESSES, GROUPED
38	MAT_AGE_NAPHSIS	Maternal age grouped
39	LGA	Large for Gestational Age Based on 90th Percentile
40	SGA_10	Small for Gestational Age Based on 10th Percentile
41	FRACE_OTH	Paternal race - Other
42	FRACE_WHT	Paternal race - White
43	INFER_TR	Infertility treatment
44	MRACE_OTH	Maternal race - Other
45	MRACE_WHT	Maternal race - White
46	TERM	Outcome

Table 3-9 PRAMS List of Obvious Variables removed from Parous and Nulliparous Datasets

PRAMS Obvious Variables		
	Variable	Description
1	MM_HBP	Hypertension?
2	MM_DIAB	Diabetes?
3	PLURAL	Plurality
4	BPG_DIAB	Diabetes before pregnancy
5	PREV_PRE	Previous – premature
6	PG_GDB	Gestational diabetes- this pregnancy
7	MORB_PLA	f) Morbidity – placenta
8	MORB_LAB	g) Morbidity - preterm labour
9	MORB_TRN	i) Morbidity - Blood transfusion
10	MORB_CAR	j) Morbidity - car crash injury
11	P_PRTERM	Previous preterm births

3.7.1.2.3 Step 3: Preparing Datasets

Parous dataset therefore had 51 variables and 65,996 numbers of cases. Nulliparous dataset therefore had 45 variables and 36,847 numbers of cases. ParousWO dataset therefore had 40 variables and 66,007 numbers of cases. NulliparousWO dataset therefore had 36 variables and 36,847 numbers of cases.

Table 3-10 Summary of datasets before DTs

PreDT PRAMS Post Script Processing				
Dataset	No. Features	Positive Cases	Negative Cases	Total Cases
Parous	52	21,452	44,544	65,996
Nulliparous	46	13,883	22,964	36,847
ParousWO	41	21,463	44,544	66,007
NulliparousWO	37	13,883	22,964	36,847

i) Hold Out Testing Data for Later Verification:

Once these datasets were established, a sample size of the most recently entered cases was removed from the dataset to be used later as unseen data for verification testing with ANN results (hold out testing). For PRAMS datasets, a group of 10,000 cases were selected and removed completely from the pool of cases for the final verification; half (5,000) of these selected cases needed to be positive (preterm) and the other half (5,000) of these cases needed negative (term). These cases were stored in a separate file. From the 10,000 cases set aside, cases would be randomly pulled out to create a verification set of 5,000 cases total, with a prevalence set to 8.1%. A selection size of 10,000 cases was set to ensure that there would be enough information to make a verification dataset with the size of 5,000 cases total (a quarter of the ANN Analysis Sample - 20,000 divided by 4 = 5,000 cases). There also needed to be enough cases to set prevalence of the verification set to match the rate of preterm birth in Ontario (8.1%) (CIHI, 2012).

ii) ANN Analysis Data:

From the remaining dataset, a sample of cases was selected for use with the ANN (data to train, test and verify), to select best networks. For PRAMS, a sample set of 20,000 cases was copied from the dataset and set aside for this initial ANN analysis. However, this data needs to have a prevalence of 15% for the ANN; for optimal training

and network selection, a higher than reported prevalence can be used to get best results (Ennett & Frize, 1998 and 2000). Therefore, 15% of 20,000 was 3,000 cases required for preterm and the remaining 17,000 cases for term (85%).

Using the filtering functions in Microsoft Excel, the 50,000 cases of sample data for ANN were filtered and selected based on cases with the least amount of missing values, and were cases that had been the most recently entered into the database.

The 20,000 case size sample was selected due to the PRAMS database having a significantly smaller number of cases compared to that of the BORN database.

iii) Data for Decision Tree Analysis

All of the data, (including the 20,000 from step ii) except for the hold out data (the 10,000 case removed for verification sets – step i) Before proceeding forward, another check for >50% missing variables and values was performed, to ensure there was enough data for proper analysis. Because a 15% prevalence was used for ANN analysis, the data for Decision Tree analysis was set to 15% as well. However with such a large difference in number of preterm cases to the number of term cases there was a need to adjust the number of cases so that the 15% prevalence could be met. The number of cases selected for each scenario (parous, nulliparous and without obvious) was optimized for DTs to analyze the most possible number of cases, with the 15% prevalence.

Ex. Simple math was applied to determine numeric values required (parous numbers used):

$$15\% = 6,800 \text{ preterm cases} / X \text{ total cases; want to determine } X.$$

$$X = 6,800 / 15\%$$

$$X = 45,300$$

Thus 45,300 total cases – 6,800 preterm cases = 38,500 term cases (the number of negative cases required to meet prevalence of positive cases).

Setting Prevalence:

Because training with high prevalence generates good resulting networks (Ennett & Frize 1998 and 2000), 15% was selected to train the ANN to find these best networks. Holdout verification testing with an unseen dataset was performed using the networks selected from the 15% prevalence dataset. This unseen data was set to a prevalence of 8.1% - the actual rate of preterm birth in Ontario (CIHI, 2012). This hold out verification testing was done to provide real world conditions to the model to give a more accurate estimate of results with new unseen data – a virtual prospective trial of future patients.

3.7.1.3 Stage 3: Decision Trees

See Section 3.5.1.3

3.7.1.3.1 Step 1: Preprocessing

See Section 3.5.1.3.1

3.7.1.3.2 Step 2: Software Execution

See Section 3.5.1.3.2

3.7.1.3.3 Step 3: Post Assessment – Results and Feature Selection

i) Method 1 Selection – Compromise:

See Section 3.5.1.3.3 i)

ii) Method 2 Selection – Average Usage:

We had our clinical advisor go through the variables and classify each into one of two categories: 1) include always (based on clinical logic), and 2) maybe include or

exclude depending on percent usage during DT construction. For PRAMS data 26 features if present were identified as include always. See Table 3.11.

3.7.1.4 Stage 4: Artificial Neural Networks:

See Section 3.5.1.4

3.7.1.4.1 Step 1: Preprocessing

See Section 3.5.1.4.1

3.7.1.4.2 Step 2: Software Execution and Results

See Section 3.5.1.4.2

Table 3-11PRAMS Variables Inclusion Identification – 1, Include Always based on logic or 2, maybe include depending on % usage.

PRAMS Variables Inclusion for Method 2					
Variable	Inclusion	Variable	Inclusion	Variable	Inclusion
1 MM_HBP	1	18 PAD6HUS	1	35 PRE_MH	2
2 MM_DIAB	1	19 MOM_BMIG_QX_REV	1	36 PRE_DDS	2
3 PLURAL	1	20 PREGHX	1	37 VITAMIN	2
4 MAT_ED	1	21 SMK63L_A	1	38 PREV_LB	2
5 MOMSMOKE	1	22 DRK63L_A	1	39 PREV6NBW	2
6 MOMLBS	1	23 MAT_AGE_NAPHSIS	1	40 PREG_TRY	2
7 PRE_DIET	1	24 LGA	1	41 MORB_NAU	2
8 PRE_EXER	1	25 SGA_10	1	42 MORB_TRN	2
9 BPG_DIAB	1	26 P_PRTERM	1	43 INCOME5	2
10 PREV_PRE	1	27 URB_RUR	2	44 LANGUAGE	2
11 PG_GDB	1	28 MARRIED	2	45 PNC_WKS	2
12 MORB_BLD	1	29 SEX	2	46 STRS_T_G	2
13 MORB_KID	1	30 PAT_ED	2	47 FRACE_OTH	2
14 MORB_PLA	1	31 OTH_TERM	2	48 FRACE_WHT	2
15 MORB_LAB	1	32 MAT_RACE	2	49 INFER_TR	2
16 MORB_CAR	1	33 PAT_RACE	2	50 MRACE_OTH	2
17 PAB6HUS	1	34 PRE_RX	2	51 MRACE_WHT	2

3.7.1.5 Stage 5: Hold Out Verification Data:

See Section 3.5.1.5.

3.7.1.5.1 Step 1: Preprocessing:

See Section 3.5.1.5.1

3.7.1.5.2 Step 2: Software Execution and Results

See Section 3.5.1.5.2

3.8 Comparative Analysis

Once all of the BORN and PRAMS results were obtained and performance measures were calculated a comparative analysis was made; this is discussed in Chapter 4.

The main performance measures used for comparison were Sensitivity, Specificity, AUC, LR+ and LR -. The aim was to meet or exceed past results that were considered clinically useful with sensitivity above 65.13% and specificity above 84.07%, which were met by Yu with her Parous dataset. Recent studies measuring LR+ and LR- values, are found to be clinical useful if $LR+ > 5$ and $LR- < 0.2$ (Honest et al. 2012).

Comparisons were made between the four scenarios within a database (Parous, Nulliparous, and the without Obvious). Individually they were compared between Method 1 and Method 2. The results between databases were also compared.

The number of variables selected was also a comparison factor as were past results, which used the old methodology.

4 Chapter: Results and Discussion

4.1 Prediction of Preterm Birth – BORN Database

4.1.1 Decision Tree Results

The BORN database was split into Parous, Nulliparous, Parous without Obvious, and Nulliparous without Obvious DT sets for the prediction of preterm birth. A C5.0 Based DT method was applied using a 5by2-fold cross validation method to generate ten train and ten test sets, ultimately ten DTs (one train and one test set per DT) per dataset (Section 3.5.1.3).

Each of the ten trees generated used different selections of variables from the original set to classify the data, resulting in varying test sensitivity and specificity values. Two methods for variable selection were used based on the DT outputs, to be used for the ANN Builder as inputs.

4.1.1.1 Method 1 Selection – Compromise:

For the Parous dataset a total of 80,000 cases were evaluated with 12,000 preterm and 68,000 term births. A total of 55 variables were used, and the prevalence (rate of preterm birth) of the datasets was set to 15%. Since 15% prevalence was used for ANN analysis for optimal results (Ennett & Frize 1998 and 2000), DTs were set to 15% prevalence as well. When the DT tools analyzed the BORN data and generated a tree to classify the patients, the tool selected a combination of variables out of the 55 inputs, to create the tree. For instance the tool selected 44 out of the 55 variables to create DT1a. Ten different trees were generated and of the 55 total input variables DT 1a selected 44, DT

1b selected 37, DT 2a selected 41, DT 2b selected 43, DT 3a selected 47, DT 3b selected 43, DT 4a selected 46, DT 4b selected 39, DT 5a selected 38, and DT 5b selected 44.

The sensitivities and specificities varied among the trees: DT 1a (75.8%, 70.8%), DT 1b (75.3%, 83.4%), DT 2a (74.7%, 83.9%), DT 2b (76.2%, 80.8%), DT 3a (77.9%, 73.2%), DT 3b (76.6%, 72.4%), DT 4a (78.3%, 64.8%), DT 4b (75.3%, 83.2%), DT 5a (73.9%, 87.1%) and DT 5b (76.8%, 75.7%). To mirror the method used by Yu (2009), the tree with the best compromise, (smallest calculated difference - while optimizing sensitivity), between sensitivity and specificity was selected, and where little variation occurred, the number of variables became the determining factor; the tree with the smaller number of variables was selected. Therefore the variables found in the attribute usage list for DT 2b were used for the ANN. DT results for Parous trees can be seen in Table 4.1. A similar selection method was used for the rest of the datasets (Nulliparous, ParousWO and NulliparousWO). The selected list of variables and attribute usage can be seen in Table 4.5 BORN Variable Lists Method 1.

Table 4-1BORN Train and Test Results Across Ten Generated Trees for Parous Data.

BORN Parous DT Results											
Decision Tree		1a	1b	2a	2b	3a	3b	4a	4b	5a	5b
Train	Sensitivity (%)	82.30%	79.25%	79.78%	80.25%	81.63%	82.43%	84.10%	78.02%	77.17%	80.55%
	Specificity (%)	85.55%	86.46%	85.97%	86.26%	85.15%	85.89%	84.25%	87.31%	87.88%	86.46%
	Accuracy	0.8507	0.8538	0.8504	0.8536	0.8462	0.8537	0.8423	0.8591	0.8627	0.8557
Test	Sensitivity (%)	75.77%	75.32%	74.68%	76.23%	77.85%	76.62%	78.30%	75.30%	73.85%	76.77%
	Specificity (%)	70.84%	83.42%	83.87%	80.78%	73.20%	72.39%	64.75%	83.22%	87.09%	75.66%
	Accuracy	0.7158	0.8220	0.8250	0.8010	0.7390	0.7303	0.6678	0.8203	0.8511	0.7582
# of Variables		44	37	41	43	47	43	46	39	38	44

For the Nulliparous dataset, a total of 66,500 cases were evaluated with 10,000 preterm and 56,500 term births. A total of 51 variables were used, and the prevalence of the datasets was set to 15%. Certain variables were not applicable to the nulliparous database: PPRTERM, PTERM, PREVCS & PCSNUM. Similar to the Parous DT results, the resulting sensitivities and specificities for the ten trees varied, as can be seen in Table 4.2. Applying the same selection method, the variables found in the attribute usage list for DT 3b were used for the ANN. The list of variables can be seen in Table 4.5: BORN Variable Lists Method 1.

Similar to the Parous and Nulliparous datasets, two more datasets removing clinically obvious variables (MULTGEST, PPRETERM, OBCOMP1, OBCOMP3, OBCOMP4, OBCOMP9, OBCOMP11, OBCOMP12) were evaluated with DTs. These variables are described in Chapter 3 in Table 3.3 For the ParousWO dataset a total of 80,000 cases were evaluated with 12,000 preterm and 68,000 term births. A total of 47 variables were used, and the prevalence of the datasets was set to 15%. For the NulliparousWO dataset a total of 73,000 cases were evaluated with 11,000 preterm and 62,000 term births. A total of 44 variables were used, and the prevalence of the datasets was set to 15%. The resulting sensitivities and specificities of the 10 trees generated, can be observed in Table 4.3 and Table 4.4; the variables found in attribute usage lists from DT 3b and DT 4b were used for the ANN, respectively. List of variables can be seen in Table 4.5: BORN Variable Lists Method 1

Table 4-2 BORN Train and Test Results Across Ten Generated Trees for Nulliparous Data.

BORN Nulliparous DT Results											
Decision Tree		1a	1b	2a	2b	3a	3b	4a	4b	5a	5b
Train	Sensitivity (%)	81.92%	82.60%	80.88%	82.52%	81.22%	79.72%	81.50%	82.46%	82.40%	82.66%
	Specificity (%)	87.60%	88.76%	88.47%	86.98%	88.20%	88.12%	87.12%	87.58%	88.18%	88.38%
	Accuracy	0.8675	0.8783	0.8733	0.8631	0.8715	0.8686	0.8627	0.8681	0.8731	0.8752
Test	Sensitivity (%)	77.46%	76.16%	75.80%	77.78%	76.12%	74.82%	76.54%	77.82%	77.22%	76.40%
	Specificity (%)	59.88%	55.78%	60.89%	53.48%	56.27%	65.70%	61.19%	59.24%	54.83%	56.96%
	Accuracy	0.6252	0.5884	0.6313	0.5714	0.5925	0.6707	0.6350	0.6204	0.5820	0.5989
# of Variables		35	39	37	40	39	38	41	38	40	39

Table 4-3 BORN Train and Test Results Across Ten Generated Trees for Parous without Obvious Data.

BORN Parous Without Obvious DT Results											
Decision Tree		1a	1b	2a	2b	3a	3b	4a	4b	5a	5b
Train	Sensitivity (%)	79.30%	79.37%	80.60%	82.12%	80.53%	78.63%	83.02%	81.03%	80.75%	83.37%
	Specificity (%)	72.59%	70.10%	70.74%	70.22%	74.13%	71.87%	68.99%	71.20%	71.34%	70.11%
	Accuracy	0.7360	0.7149	0.7222	0.7201	0.7509	0.7288	0.7109	0.7267	0.7276	0.7210
Test	Sensitivity (%)	69.22%	72.33%	71.72%	73.58%	70.63%	70.63%	74.67%	72.33%	71.47%	74.47%
	Specificity (%)	47.18%	45.36%	42.37%	42.04%	45.46%	50.47%	38.57%	44.22%	43.78%	38.69%
	Accuracy	0.5049	0.4941	0.4677	0.4678	0.4924	0.5350	0.4399	0.4844	0.4793	0.4406
# of Variables		38	40	38	39	40	36	39	37	38	39

Table 4-4 BORN Train and Test Results Across Ten Generated Trees for Nulliparous without Obvious Data.

BORN Nulliparous Without Obvious DT Results											
Decision Tree		1a	1b	2a	2b	3a	3b	4a	4b	5a	5b
Train	Sensitivity (%)	79.65%	75.84%	76.91%	78.09%	78.73%	79.45%	77.62%	78.27%	75.58%	75.62%
	Specificity (%)	73.37%	75.77%	78.16%	71.61%	73.69%	72.32%	72.23%	73.71%	76.47%	76.93%
	Accuracy	0.7432	0.7578	0.7798	0.7259	0.7445	0.7339	0.7304	0.7440	0.7634	0.7673
Test	Sensitivity (%)	72.55%	69.35%	69.51%	72.49%	70.04%	68.49%	69.25%	72.87%	67.24%	67.76%
	Specificity (%)	25.04%	30.25%	27.03%	25.35%	26.59%	25.12%	26.94%	27.53%	31.32%	29.31%
	Accuracy	0.3219	0.3614	0.3343	0.3246	0.3313	0.3166	0.3331	0.3436	0.3673	0.3510
# of Variables		34	32	34	32	34	32	34	34	32	35

Table 4-5BORN Selected DTs and Variable Lists Selected for ANN Analysis for Method 1.

BORN Variable Lists Method 1					
Decision Tree		Parous (2b)	Nulliparous (3b)	ParousWO (3b)	NulliparousWO (4b)
Test	Sensitivity (%)	76.23%	74.82%	70.63%	72.87%
	Specificity (%)	80.78%	65.70%	50.47%	27.53%
	Accuracy	0.8010	0.6707	0.5350	0.3436
# of Variables		43	38	36	34
Attribute % Usage	1	100% OBCOMP11	100% OBCOMP11	100% PTERM	100% OBCOMP8
	2	96% OBCOMP12	96% OBCOMP12	95% MATHP_sub	99% OBCOMP7
	3	94% MULTGEST	94% MULTGEST	92% OBCOMP8	99% OBCOMP5
	4	92% OBCOMP9	92% OBCOMP9	91% OBCOMP7	97% MATHP_sub
	5	92% MATHP_sub	90% MATHP4	90% OBCOMP5	95% OBCOMP14
	6	91% OBCOMP8	89% OBCOMP8	90% MATHP4	94% MATHP4
	7	91% PPRETERM	88% OBCOMP7	82% MATHP27	91% OBCOMP13
	8	90% OBCOMP7	88% OBCOMP4	82% MATAGE	90% MATHP27
	9	89% OBCOMP4	86% OBCOMP3	76% OBCOMP15	88% MATAGE
	10	89% MATHP4	85% OBCOMP1	74% MATHP3	82% LANGUAGE_up
	11	79% OBCOMP15	85% INTBF	68% INTBF	82% MATHP3
	12	74% MATHP3	81% MATHP_sub	68% LANGUAGE_up	79% INTBF
	13	74% MATAGE	80% MATHP3	62% REPASS	73% REPASS
	14	74% INTBF	77% OBCOMP15	60% SMOKING	66% FIRSTVIS
	15	73% OBCOMP3	71% OBCOMP13	56% FIRSTVIS	60% OBCOMP15
	16	65% MATHP18	70% OBCOMP14	56% MATHP5	56% PRENCLAS
	17	65% OBCOMP1	70% MATAGE	47% OBCOMP2	39% GENDER
	18	47% CONGAN_RES	69% LANGUAGE_up	46% PARITY	38% MATHP2
	19	18% MATHP_ment	50% OBCOMP5	46% OBCOMP13	37% OBCOMP2
	20	18% OBCOMP2	36% PRENCLAS	33% MATHP16	36% MATHP5
	21	17% MATHP5	36% CONGAN_GAS	33% MATHP_ment	34% SMOKING
	22	17% OBCOMP5	27% GENDER	24% GENDER	25% MATHP_ment
	23	16% PTERM	24% MATHP27	22% MATHP2	19% MATHP26
	24	15% LANGUAGE_up	21% MATHP26	17% PREVCS	14% CONGAN_RES
	25	14% MATHP27	20% FIRSTVIS	14% PRENCLAS	11% MATHP17
	26	8% FIRSTVIS	16% SMOKING	12% PCSNUM	10% CONGAN55
	27	8% SMOKING	12% MATHP2	11% CONGAN55	7% CONGAN_CHR
	28	8% OBCOMP14	8% CONGAN_CHR	6% MATHP26	7% MATHP19
	29	8% MATHP19	7% MATHP_ment	5% OBCOMP14	6% CONGAN_GAS
	30	7% OBCOMP13	6% MATHP5	4% OBCOMP6	5% MATHP1
	31	7% MATHP26	6% OBCOMP2	4% MATHP1	5% MATHP16
	32	6% GENDER	5% MATHP19	2% MATHP19	4% CONGAN_CARDIAC
	33	5% MATHP2	5% MATHP1	2% CONGAN_GAS	2% CONGAN_OROFACIAL
	34	4% MATHP17	4% MATHP16	2% CONGAN_CHR	2% CONGAN_GEN
	35	4% REPASS	3% CONGAN_MUS	1% CONGAN_EYE	
	36	3% PARITY	3% CONGAN55	1% CONGAN_GEN	
	37	3% PREVCS	3% CONGAN_CNS		
	38	3% PRENCLAS	1% REPASS		
	39	3% CONGAN_GAS			
	40	3% CONGAN55			
	41	2% PCSNUM			
	42	2% CONGAN_EYE			
	43	2% CONGAN_CNS			

4.1.1.2 Method 2 Selection - Average Usage

Section 3.5.1.3.3 ii) outlined a second method of variable selection introduced in this work. To make use of another aspect of DT output, a method was developed to select variables in the DT based on high average percent usage; this enables automated removal of variables without the constant intervention of a clinician.

Variables were identified as fitting into one of two categories: 1) include always (based on clinical logic), and 2) include or exclude depending on percent usage. For BORN data, 19 features, if present, were identified as ‘to be included always’.

Once all ten DT models from See5 were generated, variables with their usage across the ten trees were taken and averaged. A list of variables was then selected from those that were classified as “include always” regardless of their % average, and the rest of the variables were selected by “Top” % usages to include 45 variables in total, from the “maybe” classification.

The ANN Builder had previously only run small datasets with a small number of cases and variables (largest set used: 603 cases and 18 variables). Therefore the limitations and run times for larger datasets was unknown until we attempted to run BORN data using the ANN Builder tool on desktop machines in the MIRG lab. BORN datasets were significantly larger (50,000 cases 43 variables) than previously used datasets.

Initial testing was performed on desktop CPUs in the MIRG lab to make estimations for the length of time the ANN Builder would take to run an analysis to completion for different sized sets. However it was found that to handle larger datasets, more processing power was required in order to run data analysis to completion within reasonable amount of time. For example with FAST mode, 10 5by2 datasets with 47 variables and 50,000 cases would take approximately 634 hours to run to completion, approximately 63.4 hours a set. Additionally, the more variables and cases that were assessed, the more time exponentially was required to run to completion. More testing values from larger datasets using FAST mode can be seen in Appendix A. Changing modes and parameter

settings also increased processing time as reported by Martiryosan (2014): using a dataset with 603 cases and 18 variables in FAST mode took 1 hour to run to completion, FAST with sub parameters took 10 hours, MEDIUM took about a week, and MEDIUM with sub parameters took about 25 days. To access more processing power the ANN Builder was then set up to work on SHARCNET system instead of desktop machines.

This second methodology was originally developed to help reduce the size of the dataset based on features. Based on the long run times of BORN method 1 results, a cut off number of up to 45 variables was selected and applied per dataset for method 2, to be further analyzed in the ANN. There was the possibility to run the ANN with more variables, however the duration would be exponentially longer. Table 4.6 shows the selection lists for BORN Variables for the 4 datasets and their averaged percent usage across 10 trees.

Table 4-6 BORN Variables Selected for ANN Analysis for Method 2. Highlighted variables indicate the ones that are ‘Keep Always’.

BORN Parous Variables Method 2			BORN Nulliparous Variables Method 2			BORN Parous WO Variables Method 2			BORN Nulliparous WO Variables Method 2		
Variables		Average % Usage	Variable		Average % Usage	Variables		Average % Usage	Variables		Average % Usage
1	OBCOMP11	100	1	OBCOMP11	100	1	PTERM	100	1	OBCOMP8	100
2	OBCOMP12	95.9	2	OBCOMP12	96	2	MATHP_sub	95.8	2	OBCOMP5	98.8
3	MULTGEST	94	3	MULTGEST	94	3	OBCOMP8	92.3	3	MATHP4	97.7
4	OBCOMP9	92	4	OBCOMP9	91.9	4	OBCOMP7	91.2	4	OBCOMP14	92.5
5	OBCOMP8	90.7	5	OBCOMP7	88.8	5	OBCOMP5	90.9	5	OBCOMP7	97.5
6	PPRETERM	90.6	6	OBCOMP8	88.9	6	MATHP4	90.2	6	MATHP_sub	83.3
7	OBCOMP7	90.2	7	OBCOMP4	88.2	7	MATHP27	81.8	7	MATHP3	90.5
8	MATHP_sub	82.4	8	MATHP_sub	82.3	8	MATHP3	78.8	8	MATHP27	89.2
9	OBCOMP4	89	9	MATHP4	87.2	9	OBCOMP15	72.3	9	INTBF	86.7
10	MATHP4	87.7	10	OBCOMP1	63.7	10	INTBF	73.9	10	LANGUAGE_up	81.4
11	OBCOMP15	81.1	11	OBCOMP5	72.7	11	LANGUAGE_up	68.3	11	MATAGE	86.9
12	OBCOMP3	78	12	OBCOMP14	40.9	12	MATAGE	78.5	12	REPASS	74.1
13	OBCOMP1	63.5	13	OBCOMP3	82.7	13	MATHP19	25.1	13	OBCOMP15	73.6
14	INTBF	76	14	INTBF	84.3	14	REPASS	57.8	14	PRENCLAS	59.6
15	OBCOMP5	48.6	15	OBCOMP15	66	15	MATHP5	37.8	15	SMOKING	44.5
16	MATAGE	56.7	16	MATHP19	23.4	16	PARITY	58	16	GENDER	52.9
17	OBCOMP2	25.5	17	OBCOMP2	60.7	17	FIRSTVIS	48.5	17	MATHP17	30.7
18	FIRSTVIS	22.4	18	MATAGE	68.3	18	OBCOMP13	43.2	18	OBCOMP2	36.3
19	MATHP19	21.1	19	FIRSTVIS	34.5	19	GENDER	32.4	19	MATHP_ment	34.7
20	MATHP27	28.4	20	LANGUAGE_up	72.3	20	SMOKING	48.4	20	FIRSTVIS	55.2
21	MATHP5	29.3	21	GENDER	32.2	21	MATHP2	29.8	21	MATHP16	22.4
22	MATHP18	37.3	22	CONGAN55	13.3	22	MATHP16	17.2	22	OBCOMP13	66.2
23	MATHP17	17.2	23	PRENCLAS	46.7	23	OBCOMP2	41.8	23	MATHP26	31.6
24	MATHP26	9	24	MATHP_ment	22.6	24	PREVCS	27.5	24	MATHP2	34.3
25	MATHP2	10.2	25	MATHP2	18.2	25	MATHP_ment	37.9	25	MATHP5	32
26	GENDER	11.4	26	MATHP27	34.4	26	CONGAN55	13.1	26	MATHP19	37.4
27	PTERM	22.7	27	CONGAN_CNS	3.8	27	MATHP26	25.3	27	CONGAN55	13.2
28	MATHP_ment	22.5	28	MATHP5	10.4	28	CONGAN_RES	7.1	28	CONGAN_MUS	3.1
29	CONGAN55	9.6	29	OBCOMP13	42.5	29	PCSNUM	16	29	CONGAN_GAS	2.5
30	LANGUAGE_up	24.9	30	SMOKING	25.5	30	PRENCLAS	19.8	30	CONGAN_CHR	3.4
31	SMOKING	15.7	31	CONGAN_CHR	7.9	31	CONGAN_CNS	5.6	31	MATHP1	3.2
32	MATHP3	48.1	32	CONGAN_OROFACIAL	2.9	32	CONGAN_GEN	3.4	32	CONGAN_OROFACIAL	2.4
33	PREVCS	9.7	33	REPASS	16.7	33	CONGAN_CHR	1	33	CONGAN_GEN	3.9
34	CONGAN_CHR	4.2	34	CONGAN_GAS	7.5	34	CONGAN_MUS	1.2	34	MATHP18	0.2
35	CONGAN_RES	11.2	35	MATHP26	16.7	35	OBCOMP14	15.6	35	CONGAN_CNS	3
36	MATHP16	5.2	36	MATH3	64.9	36	CONGAN_GAS	3.8	36	CONGAN_CARDIAC	0.7
37	OBCOMP14	17.4	37	MATH16	11.8	37	CONGAN_EYE	1.3	37	CONGAN_RES	9.4
38	PARITY	5.7	38	MATH17	13.8	38	OBCOMP6	3.9			
39	PRENCLAS	4.1	39	MATH18	17.5	39	MATHP17	25.3			
40	CONGAN_GAS	4.9	40	CONGAN_MUS	5	40	MATHP1	4.5			
41	PCSNUM	5.4	41	CONGAN_CARDIAC	1.7	41	CONGAN_CARDIAC	1.9			
42	OBCOMP13	9.9	42	MATHP1	1.5	42	CONGAN_OROFACIAL	2.2			
43	REPASS	11.7	43	CONGAN_EYE	1.7	43	CONGAN54	0.1			
44	CONGAN_CNS	4.2	44	CONGAN_GEN	1	44	MATHP18	2.7			
45	CONGAN_CARDIAC	2.5									

4.1.2 ANN Results

Table 4.7 shows the verification performance results using the ANN Builder classification for the BORN database sets: Parous, Nulliparous, ParousWO, and NulliparousWO with both variable selection methods (1- by compromise, 2- by averaged % usage). A sample of the original datasets was used to perform the ANN classifications; each set used 50,000 cases with a 15% prevalence of positive cases. As

mentioned before, for optimal training and network selection, a higher than reported prevalence can be used to get best results, thus 15% prevalence was used (Ennett & Frize 1998 and 2000).

For the BORN datasets, overall Method 2 appears to perform better than Method 1, with the best performing sets being Parous and Nulliparous for both methods. Using method 2 for the parous sets gives a sensitivity of 77.67 %, specificity of 82.14%, and AUC of 0.8462. Method 2 for nulliparous sets gives sensitivity of 79.31%, specificity of 77.88% and AUC of 0.8622.

To be considered as clinically useful a Sensitivity of 65% and Specificity of 85% must be met as recommended by our clinician partner, reported by Gunaratnam (2013). The sensitivity values do meet the requirement, however, collectively these values do not meet the performance measures reported.

Additionally, the positive and negative likelihood ratios were calculated from the sensitivity and specificity values, as shown in Table 4.8 and they also do not meet the criteria set out in (Honest et al., 2012), which uses LR+ of >5 and LR- of <0.2 as a measure of clinical usefulness.

ANN analysis was performed using Fast Mode with Sub Parameters settings on the SHARCNET cluster system. A number of options for final results existed for selecting best networks for performance based on sensitivity and specificity results (described section 3.5.1.4.2). The results shown in this section are based on selecting option 0; selecting the best resulting sensitivity as a default. Other options exist (1-10), these other options perform selection and optimization differently while considering specificity while option 0 does not. Work performed for this thesis took approximately

54 CPU years of time on the SHARCNET cluster. That is, if data was analyzed by the ANN Builder on 1 CPU of 1 desktop machine, it would have taken approximately 54 years running 24 hours a day to complete.

In order to select the best networks a full analysis was performed with 15% data, requiring training, testing and verification analysis. Best networks were selected from the 15% verification results. These results are shown in the following Tables 4.7 and 4.8. The final results are assessed with the 8.1% verification data analysis with these selected networks, in the next section (see Section 4.1.3)

Table 4-7 Verification Results at 15% Prevalence Data – ANN Outcome.

BORN DT- ANN Results with 15% Prevalence Datasets													
Dataset	Method	PPV	stdev	NPV	stdev	ACC	stdev	MCC	stdev	F1score	stdev	AUC	stdev
Parous	1	0.1489	0.0296	0.8442	0.0485	0.4458	0.1840	-0.0010	0.0679	0.2285	0.0561	0.4992	0.0675
Nulliparous	1	0.5586	0.2013	0.9587	0.0048	0.8463	0.0766	0.5732	0.1408	0.6322	0.1205	0.8809	0.0359
ParousWO	1	0.1602	0.0507	0.8407	0.0433	0.4459	0.1858	-0.0012	0.0817	0.2313	0.0420	0.4954	0.0770
NulliparousWO	1	0.1560	0.0140	0.8622	0.0395	0.4099	0.1106	0.0223	0.0607	0.2495	0.0252	0.5116	0.0433
Parous	2	0.4554	0.0977	0.9543	0.0060	0.8147	0.0548	0.4936	0.0810	0.5668	0.0722	0.8462	0.0358
Nulliparous	2	0.4135	0.1311	0.9552	0.0049	0.7810	0.0574	0.4563	0.0975	0.5319	0.0838	0.8622	0.0214
ParousWO	2	0.4864	0.1621	0.9334	0.0082	0.8174	0.0774	0.4567	0.1262	0.5417	0.1012	0.8101	0.0412
NulliparousWO	2	0.7147	0.2726	0.9206	0.0068	0.8654	0.0856	0.5416	0.1791	0.5908	0.1283	0.8055	0.0446

Table 4-8 Verification Results at 15% Prevalence Data – Calculated LR values.

BORN ANN Results using 15% Prevalence (Option 0)										
Dataset	No. Variables	Method	Sensitivity (%)	stdev	Specificity (%)	stdev	AUC	stdev	LR +	LR -
Parous	43	1	58.48%	0.2636	42.12%	0.2606	0.4992	0.0675	1.01	0.99
Nulliparous	38	1	79.15%	0.0318	85.59%	0.0939	0.8809	0.0359	5.49	0.24
ParousWO	40	1	57.16%	0.2508	42.38%	0.2595	0.4954	0.0770	0.99	1.01
NulliparousWO	34	1	66.20%	0.1694	36.55%	0.1564	0.5116	0.0433	1.04	0.92
Parous	45	2	77.67%	0.0391	82.14%	0.0681	0.8462	0.0358	4.35	0.27
Nulliparous	44	2	79.31%	0.0277	77.88%	0.0702	0.8622	0.0214	3.59	0.27
ParousWO	44	2	66.05%	0.0436	84.51%	0.0939	0.8101	0.0412	4.26	0.40
NulliparousWO	37	2	55.38%	0.0265	92.04%	0.1038	0.8055	0.0446	6.96	0.48

4.1.3 Hold Out Verification Results

Hold out verification analysis was performed with an unseen dataset set to a prevalence of 8.1%. This prevalence was set to match the “real life” rate of preterm birth in Ontario, (CIHI, 2012). Recall that hold out verification testing is to provide real world conditions to the model to give a more accurate estimate of results with new unseen data – a virtual prospective trial of future patients. This verification analysis is performed using the networks selected using the 15% prevalence data selected in the previous section (Section 4.1.2) and as described in Chapter 3 (Section 3.5.1.4).

Table 4.9 shows the verification performance results using the ANN Builder verification mode for the BORN database verification sets: Parous, Nulliparous, ParousWO, and NulliparousWO with both variable selection methods (1- by compromise, 2- by averaged % usage). A sample size of the original datasets was used to perform the ANN classifications; each set used 12,500 cases with an 8.1% prevalence of positive cases.

For the BORN verification sets, the outcomes of results from both methods for each set were similar to one another (with Method 2 performing only slightly better than Method 1 overall). To be considered as clinically useful a Sensitivity of 65% and Specificity of 85% must be met as recommended by our clinician partner, reported by Gunaratnam (2013). The specificity values do meet the requirement, however, collectively these values do not meet the performance measures reported.

Additionally, the positive and negative likelihood ratios were calculated from the sensitivity and specificity values, as shown in Table 4.10, and they also do not meet the

criteria set out in (Honest et al., 2012), which uses LR+ of >5 and LR- of <0.2 as a measure of clinical usefulness.

Table 4-9 Verification Results at 8.1% Prevalence Unseen Data

BORN DT- ANN Results with 8.1% Prevalence Verification Data													
Dataset	Method	PPV	stdev	NPV	stdev	ACC	stdev	MCC	stdev	F1score	stdev	AUC	stdev
Parous	1	0.4410	0.2749	0.9539	0.0140	0.8457	0.0987	0.3461	0.1300	0.3692	0.1393	0.7529	0.0794
Nulliparous	1	0.5249	0.3631	0.9531	0.0181	0.7826	0.1925	0.3376	0.1447	0.3259	0.1202	0.7495	0.0972
ParousWO	1	0.2698	0.1996	0.9267	0.0399	0.6737	0.3015	0.1802	0.1432	0.2590	0.0777	0.6456	0.0998
NulliparousWO	1	0.1945	0.0776	0.9442	0.0143	0.7248	0.2215	0.1904	0.1031	0.2648	0.0788	0.6644	0.0775
Parous	2	0.6041	0.2446	0.9554	0.0073	0.8828	0.1382	0.4722	0.1354	0.4919	0.1190	0.7721	0.0456
Nulliparous	2	0.6795	0.2386	0.9596	0.0096	0.9204	0.0508	0.5416	0.1161	0.5494	0.1165	0.7970	0.0324
ParousWO	2	0.3082	0.2440	0.9099	0.0722	0.7146	0.2789	0.1717	0.1457	0.2383	0.0727	0.6341	0.0701
NulliparousWO	2	0.2800	0.1374	0.9456	0.0160	0.7130	0.2846	0.2414	0.1249	0.3055	0.0984	0.7121	0.0341

Table 4-10 Verification Results at 8.1% Prevalence Unseen Data – Calculated LR values

BORN Verification Results using 8.1% Prevalence (Option 0)										
Dataset	No. Variables	Method	Sensitivity (%)	stdev	Specificity (%)	stdev	AUC	stdev	LR +	LR -
Parous	43	1	50.52%	0.2005	87.57%	0.1198	0.7529	0.0794	4.06	0.57
Nulliparous	38	1	52.04%	0.2861	80.57%	0.2297	0.7495	0.0972	2.68	0.60
ParousWO	40	1	50.80%	0.2340	68.83%	0.3476	0.6456	0.0998	1.63	0.71
NulliparousWO	34	1	52.57%	0.2105	74.23%	0.2563	0.6644	0.0775	2.04	0.64
Parous	45	2	50.53%	0.1328	91.61%	0.1604	0.7721	0.0456	6.02	0.54
Nulliparous	44	2	53.96%	0.1306	95.40%	0.0620	0.7970	0.0324	11.72	0.48
ParousWO	44	2	41.43%	0.2554	74.11%	0.3250	0.6341	0.0701	1.60	0.79
NulliparousWO	37	2	55.25%	0.2308	72.72%	0.3282	0.7121	0.0341	2.03	0.62

4.2 Prediction of Preterm Birth – PRAMS Database

4.2.1 Decision Tree Results

The PRAMS database was split into Parous, Nulliparous, Parous without Obvious, and Nulliparous without Obvious DT sets for the prediction of preterm birth. A C5.0 Based DT method was applied using a repeated 5by2-fold cross validation method to

generate ten train and ten test sets, ultimately ten DTs (one train and one test set per DT) per dataset (Section 3.6.1.3).

Each of the ten trees generated used different selections of variables from the original set to classify the data, resulting in varying test sensitivity and specificity values. Two methods for variable selection were used based on the DT outputs, to be used for the ANN Builder as inputs.

4.2.1.1 Method 1 Selection - Compromise

For the Parous dataset a total of 45,300 cases were evaluated with 6,800 preterm and 38,500, term births. A total of 51 variables were used, and the prevalence of the datasets was 15%. Since 15% prevalence was used for ANN analysis for optimal results (Ennett & Frize 1998 and 2000), DTs were set to 15% prevalence as well. When the DT tools analyzed the BORN data and generated a tree to classify the patients, the tool selected a combination of variables out of the 51 inputs, to create the tree. For instance the tool selected 45 out of the 51 variables to create DT1a. Ten different trees were generated and of the 55 total input variables DT 1a selected 45 variables, DT 1b selected 49, DT 2a selected 46, DT 2b selected 48, DT 3a selected 44, DT 3b selected 46, DT 4a selected 45, DT 4b selected 47, DT 5a selected 48, and DT 5b selected 50.

The sensitivities and specificities varied among the trees: DT 1a (44.7%, 83.6%), DT 1b (40.6%, 84.0%), DT 2a (45.9%, 83.8%), DT 2b (45.5%, 84.0%), DT 3a (43.9%, 83.6%), DT 3b (39.1%, 84.7%), DT 4a (39.7%, 86.5%), DT 4b (42.9%, 82.2%), DT 5a (45.1%, 84.5%) and DT 5b (42.8%, 83.5%). To mirror the method used by Yu (2009), the tree with the best compromise, (smallest calculated difference - while optimizing sensitivity), between sensitivity and specificity was selected, and where little variation

occurred the number of variables became the determining factor; the tree with the smaller number of variables was selected to generate. Therefore the variables found in the attribute usage list for DT 2b were used for the ANN. DT results for Parous trees can be seen in Table 4.11. Similar selection method was used for the rest of the datasets (Nulliparous, ParousWO and NulliparousWO). The selected list of variables and attribute usage can be seen in Table 4.15 PRAMS Variable Lists Method 1.

Table 4-11PRAMS Train and Test Results Across Ten Generated Trees for Parous Data.

PRAMS Parous DT Results											
Decision Tree		1a	1b	2a	2b	3a	3b	4a	4b	5a	5b
Train	Sensitivity (%)	99.71%	99.56%	99.76%	99.79%	99.56%	99.79%	99.56%	99.71%	99.62%	99.59%
	Specificity (%)	98.28%	98.78%	98.92%	98.03%	98.05%	98.94%	98.16%	98.17%	98.52%	99.14%
	Accuracy	0.9849	0.9890	0.9905	0.9829	0.9827	0.9907	0.9837	0.9840	0.9869	0.9921
Test	Sensitivity (%)	44.74%	40.59%	45.85%	45.53%	43.94%	39.15%	39.71%	42.88%	45.06%	42.79%
	Specificity (%)	83.59%	84.06%	83.79%	84.02%	83.52%	84.74%	86.46%	82.19%	84.50%	83.53%
	Accuracy	0.7776	0.7753	0.7809	0.7824	0.7758	0.7790	0.7944	0.7629	0.7858	0.7741
# of Variables		45	49	46	48	44	46	45	47	48	50

For the Nulliparous dataset, a total of 21,050 cases were evaluated with 3,150 preterm and 17,900 term births. A total of 45 variables were used, and the prevalence of the datasets was 15%. Certain variables were not applicable to the nulliparous database: OTH_TERM, PREV_LB, PREV6NBW, PREV_PRE, P_PRTERM. Similar to the Parous DT results, the resulting sensitivities and specificities for the ten trees varied, as can be seen in Table 4.12. Applying the same selection method, the variables found in the attribute usage list for DT 4b were used for the ANN. List of variables can be seen in Table 4.15: PRAMS Variable Lists Method 1.

Similar to the Parous and Nulliparous datasets, two more datasets removing clinically obvious variables (MM_HBP, MM_DIAB, PLURAL, BPG_DIAB, PREV_PRE,

PG_GDB, MORB_PLA, MORB_LAB, MORB_TRN, MORB_CAR, P_PRTERM) were evaluated with DTs. These variables are described in Chapter 3 in Table 3.9. For the ParousWO dataset a total of 45,300 cases were evaluated with 6,800 preterm and 38,500 term births. A total of 40 variables were used, and the prevalence of the datasets was 15%. For the NulliparousWO dataset a total of 21,050 cases were evaluated with 3,150 preterm and 17,900 term births. A total of 36 variables were used, and the prevalence of the datasets was 15%. The resulting sensitivities and specificities, of the 10 trees generated, can be observed in Table 4.13 and Table 4.14; the variables found in attribute usage lists from DT 5b and DT 3b were used for the ANN, respectively. List of variables can be seen in Table 4.15: BORN Variable Lists Method 1.

Table 4-12 PRAMS Train and Test Results Across Ten Generated Trees for Nulliparous Data.

PRAMS Nulliparous DT Results											
Decision Tree		1a	1b	2a	2b	3a	3b	4a	4b	5a	5b
Train	Sensitivity (%)	99.43%	99.05%	99.68%	98.92%	99.43%	99.30%	99.37%	99.30%	99.37%	99.30%
	Specificity (%)	97.47%	97.94%	98.69%	97.74%	97.66%	98.16%	98.64%	98.16%	97.28%	97.71%
	Accuracy	0.9777	0.9811	0.9884	0.9792	0.9793	0.9833	0.9875	0.9833	0.9760	0.9795
Test	Sensitivity (%)	47.30%	44.25%	40.83%	47.94%	36.44%	47.43%	39.43%	48.95%	44.70%	45.02%
	Specificity (%)	82.93%	84.37%	81.80%	79.31%	80.19%	81.04%	83.31%	81.46%	82.16%	82.19%
	Accuracy	0.7760	0.7837	0.7567	0.7461	0.7364	0.7601	0.7674	0.7660	0.7655	0.7663
# of Variables		41	42	40	41	38	41	41	41	43	37

Table 4-13 PRAMS Train and Test Results Across Ten Generated Trees for Parous without Obvious Data.

PRAMS Parous Without Obvious DT Results											
Decision Tree		1a	1b	2a	2b	3a	3b	4a	4b	5a	5b
Train	Sensitivity (%)	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
	Specificity (%)	97.89%	98.22%	98.00%	97.41%	97.06%	98.30%	97.85%	97.97%	97.56%	97.90%
	Accuracy	0.9820	0.9849	0.9830	0.9780	0.9750	0.9855	0.9818	0.9828	0.9793	0.9821
Test	Sensitivity (%)	38.35%	39.29%	38.38%	36.26%	38.50%	36.41%	38.71%	36.47%	37.62%	42.53%
	Specificity (%)	82.03%	79.58%	80.61%	80.16%	83.10%	83.35%	79.39%	83.34%	82.30%	80.82%
	Accuracy	0.7547	0.7354	0.7427	0.7357	0.7641	0.7630	0.7328	0.7630	0.7559	0.7507
# of Variables		37	37	37	36	38	37	38	35	37	38

Table 4-14PRAMS Train and Test Results Across Ten Generated Trees for Nulliparous without Obvious Data.

PRAMS Nulliparous Without Obvious DT Results											
Decision Tree		1a	1b	2a	2b	3a	3b	4a	4b	5a	5b
Train	Sensitivity (%)	98.79%	99.37%	98.86%	98.92%	98.92%	98.92%	99.75%	99.37%	99.17%	99.11%
	Specificity (%)	97.35%	97.12%	97.78%	95.64%	97.75%	96.69%	97.26%	98.23%	97.51%	97.17%
	Accuracy	0.9757	0.9745	0.9794	0.9613	0.9793	0.9703	0.9763	0.9840	0.9776	0.9746
Test	Sensitivity (%)	31.11%	36.38%	35.17%	34.98%	28.57%	37.14%	34.67%	36.38%	32.83%	29.14%
	Specificity (%)	80.65%	79.16%	77.83%	78.70%	79.89%	80.69%	79.01%	79.34%	81.41%	78.59%
	Accuracy	0.7324	0.7276	0.7145	0.7216	0.7221	0.7418	0.7237	0.7291	0.7414	0.7119
# of Variables		33	32	33	32	35	32	35	32	31	34

Table 4-15PRAMS Selected DTs and Variable Lists Selected for ANN Analysis for Method 1.

PRAMS Variable Lists Method 1					
Decision Tree		Parous (2b)	Nulliparous (4b)	ParousWO (4b)	NulliparousWO (3b)
Test	Sensitivity (%)	45.53%	48.95%	42.53%	37.14%
	Specificity (%)	84.02%	81.46%	80.82%	80.69%
	Accuracy	0.7824	0.7660	0.7507	0.7418
# of Variables		48	41	38	32
Attribute % Usage	1	100% PREV_LB	100% PLURAL	97% LGA	97% LGA
	2	98% MORB_LAB	97% MORB_LAB	90% MAT_RACE	85% MOMLBS
	3	89% LGA	97% LGA	79% MOMLBS	81% MORB_BLD
	4	73% MM_HBP	71% MAT_RACE	73% MORB_BLD	80% MAT_RACE
	5	71% OTH_TERM	62% PNC_WKS	68% PNC_WKS	56% PNC_WKS
	6	62% MORB_PLA	62% PAB6HUS	56% PREGHX	44% FRACE_WHT
	7	62% BPG_DIAB	61% MORB_TRN	50% OTH_TERM	43% PAB6HUS
	8	61% PREV_PRE	61% MM_HBP	31% SMK63L_A	38% PAD6HUS
	9	52% SGA_10	60% MOMLBS	29% PAD6HUS	30% VITAMIN
	10	48% PNC_WKS	52% MORB_BLD	23% PRE_MH	28% SEX
	11	47% MORB_TRN	47% DRK63L_A	21% SGA_10	24% PRE_MH
	12	42% MOMLBS	41% MORB_CAR	20% MOM_BMIG_QX_REV	23% LANGUAGE
	13	41% MORB_BLD	28% MAT_AGE_NAPHSIS	20% MAT_AGE_NAPHSIS	22% MAT_ED
	14	38% MORB_CAR	25% MORB_PLA	16% PAB6HUS	19% MOMSMOKE
	15	36% MAT_AGE_NAPHSIS	24% BPG_DIAB	15% INCOME5	17% PRE_RX
	16	30% PAD6HUS	22% PRE_DIET	13% PREV6NBW	14% INCOME5
	17	28% MORB_KID	17% VITAMIN	12% MORB_KID	14% MOM_BMIG_QX_REV
	18	26% PRE_RX	14% FRACE_WHT	11% PAT_RACE	14% DRK63L_A
	19	26% MM_DIAB	14% PRE_MH	10% VITAMIN	10% SGA_10
	20	25% PLURAL	13% PAD6HUS	8% PRE_RX	9% SMK63L_A
	21	23% PREV6NBW	13% MAT_ED	7% MORB_NAU	9% PRE_EXER
	22	22% MAT_RACE	12% INCOME5	7% PAT_ED	9% PAT_RACE
	23	21% MOMSMOKE	9% LANGUAGE	6% PRE_EXER	6% STRS_T_G
	24	19% PAB6HUS	9% MARRIED	5% DRK63L_A	6% PRE_DDS
	25	14% PG_GDB	8% MORB_NAU	5% INFER_TR	6% MARRIED
	26	12% DRK63L_A	8% PREG_TRY	5% PRE_DIET	6% MORB_KID
	27	9% LANGUAGE	7% PRE_RX	5% MARRIED	4% PRE_DIET
	28	9% SEX	6% PG_GDB	5% MOMSMOKE	4% PREG_TRY
	29	8% PREG_TRY	6% MORB_KID	5% SEX	2% PAT_ED
	30	8% FRACE_WHT	6% MOM_BMIG_QX_REV	4% STRS_T_G	2% MORB_NAU
	31	7% STRS_T_G	5% PRE_DDS	4% URB_RUR	2% MRACE_WHT
	32	7% PAT_ED	5% SGA_10	3% LANGUAGE	2% URB_RUR
	33	6% PAT_RACE	5% MM_DIAB	3% PREV_LB	
	34	5% VITAMIN	4% PRE_EXER	3% FRACE_OTH	
	35	5% SMK63L_A	3% PAT_RACE	3% PREG_TRY	
	36	5% MARRIED	3% SMK63L_A	2% PRE_DDS	
	37	5% P_PRTERM	3% MOMSMOKE	1% MAT_ED	
	38	4% MOM_BMIG_QX_REV	3% URB_RUR	1% MRACE_WHT	
	39	4% INCOME5	2% SEX		
	40	4% PRE_DIET	1% PAT_ED		
	41	4% PRE_DDS	1% STRS_T_G		
	42	3% INFER_TR			
	43	3% URB_RUR			
	44	3% PRE_EXER			
	45	2% MORB_NAU			
	46	2% MAT_ED			
	47	2% PRE_MH			
	48	1% MRACE_WHT			

4.2.1.2 Method 2 Selection – Average Usage

Section 3.6.1.3.3 ii) outlined a second method of variable selection introduced in this work. To make use of another aspect of DT output: a way to select variables in the DT, based on high average percent usage, and automate the removal of variables without the constant intervention of a clinician.

Variables were identified into one of two categories: 1) include always (based on clinical logic), and 2) maybe include or exclude depending on percent usage. For PRAMS data 26 features if present were identified as include always.

Once all ten DT models from See5 were generated, variables with their usages across the ten trees were taken and averaged. A list of variables was then selected from those are classified as “include always” regardless of their % average, and the rest of the variables selected by “Top” % usages, from the “maybe” classification.

To remain consistent with BORN, for comparison purposes, the “maximum” number of 45 variables was selected and applied per dataset for method 2, to be further analyzed in the ANN. There is the possibility to run the ANN with more variables, however this would take an exponentially longer run time. Table 4.16 shows the selection lists for BORN Variables for the 4 datasets and their averaged percent usage across 10 trees.

Table 4-16PRAMS Variables Selected for ANN Analysis for Method 2. Highlighted variables indicate the ones that are ‘Keep Always’.

PRAMS Parous Variables Method 2			PRAMS Nulliparous Variables Method 2			PRAMS Parous WO Variables Method 2			PRAMS Nulliparous WO Variables Method 2		
Variables	Average % Usage		Variables	Average % Usage		Variables	Average % Usage		Variables	Average % Usage	
1	PREV_LB	100	1	PLURAL	100	1	LGA	97.4	1	MOMLBS	86.9
2	MORB_LAB	90.5	2	MORB_LAB	97.2	2	MAT_RACE	87.7	2	MAT_RACE	79
3	LGA	87.5	3	LGA	92.3	3	OTH_TERM	58.2	3	LGA	95
4	OTH_TERM	57.3	4	BPG_DIAB	37.8	4	MORB_BLD	69.9	4	PNC_WKS	62.7
5	MORB_PLA	65.9	5	MORB_TRN	35.8	5	MOMLBS	75.5	5	LANGUAGE	35.2
6	PNC_WKS	53.6	6	PAB6HUS	31.6	6	PNC_WKS	63.7	6	MORB_BLD	51.1
7	MOMLBS	62.4	7	MAT_RACE	64.4	7	LANGUAGE	18	7	DRK63L_A	35.1
8	MM_HBP	61.9	8	MOMLBS	73.6	8	DRK63L_A	28.2	8	INCOME5	24.8
9	MAT_RACE	55.6	9	MM_HBP	53.4	9	PREGHX	46.7	9	FRACE_OTH	4.7
10	PREV_PRE	47.1	10	MORB_PLA	46.7	10	PREV6NBW	14.8	10	FRACE_WHT	24.3
11	MM_DIAB	24.5	11	MM_DIAB	35.3	11	PRE_RX	11.4	11	PAT_RACE	9.1
12	SGA_10	20.9	12	MORB_CAR	26.1	12	PREV_LB	11.1	12	SEX	13.5
13	MORB_CAR	29.2	13	PNC_WKS	64.2	13	PAB6HUS	21.2	13	MAT_AGE_NAPHSIS	10.4
14	PG_GDB	15.4	14	MOM_BMIG_QX_REV	9.9	14	MARRIED	11.9	14	PRE_RX	12.2
15	PAD6HUS	18.6	15	PAD6HUS	32.3	15	MOMSMOKE	17.6	15	PAB6HUS	18.8
16	PLURAL	37.1	16	PG_GDB	19	16	SGA_10	18.8	16	MAT_ED	13.3
17	MOMSMOKE	15.9	17	PAT_RACE	6.5	17	MOM_BMIG_QX_REV	17	17	PRE_DIET	7.9
18	MORB_TRN	44.2	18	DRK63L_A	21.9	18	VITAMIN	12.5	18	INFER_TR	1.5
19	PRE_RX	12.3	19	MORB_BLD	30	19	PRE_EXER	9.4	19	PREG_TRY	5.1
20	SMK63L_A	6.6	20	PRE_EXER	9.4	20	SMK63L_A	15.7	20	PRE_MH	14.1
21	MORB_BLD	25.3	21	PRE_DIET	12.2	21	FRACE_WHT	6	21	MARRIED	6.3
22	BPG_DIAB	39	22	FRACE_WHT	12.2	22	MORB_NAU	8.8	22	STRS_T_G	9.6
23	DRK63L_A	19.6	23	INCOME5	11.2	23	MAT_ED	5.9	23	MORB_KID	8.6
24	VITAMIN	11	24	LANGUAGE	27.7	24	INCOME5	14	24	SMK63L_A	11.2
25	LANGUAGE	15.3	25	PRE_RX	13.7	25	MAT_AGE_NAPHSIS	9	25	SGA_10	10.2
26	MARRIED	7.8	26	PRE_MH	9.3	26	PRE_MH	14.3	26	MOM_BMIG_QX_REV	9.8
27	PRE_EXER	6.3	27	SGA_10	7.7	27	MORB_KID	11.4	27	VITAMIN	15.9
28	INCOME5	6.2	28	MOMSMOKE	12	28	SEX	10	28	MOMSMOKE	19.1
29	PRE_DIET	6.5	29	MORB_NAU	8.6	29	PREG_TRY	12.4	29	PRE_EXER	8.2
30	PAB6HUS	15.5	30	MAT_ED	8.2	30	PAT_RACE	8.9	30	URB_RUR	2
31	PREG_TRY	12.8	31	VITAMIN	5.4	31	STRS_T_G	4.6	31	MORB_NAU	7.5
32	PRE_MH	8.9	32	SMK63L_A	7.9	32	PRE_DDS	9.5	32	MRACE_OTH	0.8
33	PRE_DDS	5.3	33	PRE_DDS	5.5	33	PRE_DIET	7.9	33	PRE_DDS	5.9
34	SEX	7.5	34	STRS_T_G	5.2	34	INFER_TR	1.8	34	PAD6HUS	34.1
35	MAT_AGE_NAPHSIS	10.6	35	MARRIED	5.4	35	PAT_ED	8.4	35	PAT_ED	3.3
36	PREV6NBW	21.5	36	PREG_TRY	5.1	36	URB_RUR	2.5	36	MRACE_WHT	1.2
37	FRACE_WHT	4.8	37	MAT_AGE_NAPHSIS	14.8	37	MRACE_WHT	0.3			
38	PAT_ED	4.1	38	SEX	8.6	38	PAD6HUS	20.5			
39	STRS_T_G	4.3	39	URB_RUR	3.1	39	FRACE_OTH	2.5			
40	MORB_NAU	4.3	40	MORB_KID	5.6	40	MRACE_OTH	0.3			
41	MOM_BMIG_QX_REV	8.9	41	FRACE_OTH	1.1						
42	MAT_ED	4.2	42	INFER_TR	2.3						
43	MORB_KID	10.8	43	PAT_ED	1.4						
44	PREGHX	7.8	44	MRACE_WHT	0.4						
45	P_PRTERM	3.9	45	MRACE_OTH	1.6						

4.2.2 ANN Results

Table 4.17 shows the verification performance results using the ANN Builder classification for the PRAMS database sets: Parous, Nulliparous, ParousWO, and NulliparousWO with both variable selection methods (1- by compromise, 2- by averaged % usage). A sample size of the original datasets were used to perform the ANN classifications, each set used 20,000 cases with a 15% prevalence for positive cases. As mentioned before, for optimal training and network selection, a higher than reported

prevalence can be used to get best results, thus 15% prevalence was used (Ennett & Frize, 1998 and 2000).

For the PRAMS datasets, all sets using both methods performed close to perfect, 100%. All sensitivity and specificity values showed a near perfect performance. To be considered as clinically useful a Sensitivity of 65% and Specificity of 85% must be met as recommended by our clinician partner, reported by Gunaratnam (2013). Collectively these values meet and surpass the performance measures reported.

Additionally, the positive and negative likelihood ratios were calculated from the sensitivity and specificity values, as shown in Table 4.18 and they also appear to meet the criteria set out in (Honest et al., 2012), which uses LR+ of >5 and LR- of <0.2 as a measure of clinical usefulness.

ANN analysis was performed using Fast Mode with Sub Parameters On settings on the SHARCNET cluster system. A number of options for final results existed for selecting best networks for performance based on sensitivity and specificity results (described section 3.5.1.4.2). The results shown in this section are based on selecting option 0; selecting the best resulting sensitivity as a default. Other options exist (1-10), these other options perform selection and optimization differently while considering specificity while option 0 does not. The total ANN analysis work took 54 CPU years of processing time.

These results appear to meet and surpass clinical usefulness acceptability reported by Gunaratnam (2013), but at a closer look at Tables 4.17 and 4.18, prediction is almost perfect. Having found this unexpected, an investigation was performed to determine if a tool or calculation error occurred in data processing at any point. It was found that

multiple duplicates of patient data was found in the 15% prevalence train, test, and verifying datasets. Further investigation was then performed on both BORN and PRAMS datasets; findings and implications are discussed in section 4.4 and the detailed results of the investigation can be found in Appendix B.

In order to select the best networks a full analysis was performed with 15% data, requiring training, testing and verification analysis. Best networks were selected from the 15% verification results. These results are shown in the following Tables 4.17 and 4.18. The final results are assessed with the 8.1% verification data analysis with these selected networks, in the next section (see Section 4.2.3).

Table 4-17 Verification Results at 15% Prevalence Data – ANN Outcome.

PRAMS DT-ANN Results with 15% Prevalence Datasets													
Dataset	Method	PPV	stdev	NPV	stdev	ACC	stdev	MCC	stdev	F1score	stdev	AUC	stdev
Parous	1	0.9797	0.0069	0.9982	0.0007	0.9954	0.0012	0.9821	0.0045	0.9848	0.0038	0.9948	0.0032
Nulliparous	1	0.9984	0.0047	1.0000	0.0001	0.9997	0.0008	0.9989	0.0033	0.9991	0.0028	0.9996	0.0012
ParousWO	1	0.9621	0.0107	0.9976	0.0007	0.9921	0.0021	0.9696	0.0080	0.9742	0.0068	0.9909	0.0029
NulliparousWO	1	0.9983	0.0029	0.9999	0.0002	0.9997	0.0006	0.9987	0.0024	0.9989	0.0020	0.9996	0.0008
Parous	2	0.9767	0.0108	0.9983	0.0008	0.9950	0.0021	0.9804	0.0082	0.9833	0.0070	0.9930	0.0045
Nulliparous	2	0.9992	0.0024	0.9999	0.0004	0.9998	0.0007	0.9991	0.0026	0.9993	0.0022	0.9995	0.0014
ParousWO	2	0.9691	0.0108	0.9980	0.0009	0.9935	0.0021	0.9750	0.0078	0.9787	0.0067	0.9922	0.0032
NulliparousWO	2	0.9976	0.0059	0.9999	0.0003	0.9995	0.0011	0.9981	0.0043	0.9984	0.0036	0.9994	0.0013

Table 4-18 Verification Results at 15% Prevalence Data – Calculated LR values.

PRAMS ANN Results using 15% Prevalence (Option 0)										
Dataset	No. Variables	Method	Sensitivity (%)	stdev	Specificity (%)	stdev	AUC	stdev	LR +	LR -
Parous	48	1	98.99%	0.004	99.64%	0.0013	0.9948	0.0032	273.14	0.01
Nulliparous	41	1	99.97%	0.001	99.97%	0.0008	0.9996	0.0012	3545.15	0.00
ParousWO	38	1	98.65%	0.004	99.31%	0.0020	0.9909	0.0029	143.58	0.01
NulliparousWO	32	1	99.95%	0.001	99.97%	0.0005	0.9996	0.0008	3266.23	0.00
Parous	45	2	99.01%	0.005	99.58%	0.0020	0.9930	0.0045	236.42	0.01
Nulliparous	45	2	99.93%	0.002	99.99%	0.0004	0.9995	0.0014	7087.47	0.00
ParousWO	40	2	98.85%	0.005	99.44%	0.0020	0.9922	0.0032	177.25	0.01
NulliparousWO	36	2	99.92%	0.002	99.96%	0.0011	0.9994	0.0013	2356.60	0.00

4.2.3 Hold Out Verification Results

Hold out verification analysis was performed with an unseen dataset set to a prevalence of 8.1%. This prevalence was set to match the “real life” rate of preterm birth in Ontario, (CIHI, 2012). Recall that hold out verification testing is to provide real world conditions to the model to give a more accurate estimate of results with new unseen data – a virtual prospective trial of future patients. This verification analysis is performed using the networks selected using the 15% prevalence data selected in the previous section (Section 4.2.2) and as described in Chapter 3 (Section 3.6.1.4).

Table 4.19 shows the verification performance results using the ANN Builder verification mode for the BORN database verification sets: Parous, Nulliparous, ParousWO, and NulliparousWO with both variable selection methods (1- by compromise, 2- by averaged % usage). A sample size of the original datasets was used to perform the ANN classifications; each set used 5,000 cases with an 8.1% prevalence of positive cases.

For the PRAMS verification sets, the outcomes of results from both methods of each set were similar to one another (with Method 2 performing only slightly better than Method 1). To be considered as clinically useful a Sensitivity of 65% and Specificity of 85% must be met as recommended by our clinician partner, reported by Gunaratnam (2013). The specificity values do meet the requirement, however, collectively these values do not meet the performance measures reported.

Additionally, the positive and negative likelihood ratios were calculated from the sensitivity and specificity values, as shown in Table 4.20 and they also do not meet the

criteria set out in (Honest et al., 2012), which uses LR+ of >5 and LR- of <0.2 as a measure of clinical usefulness.

It was interesting to note is that despite the superior results achieved with the 15% prevalence data sets, these networks do not perform as well for the lower prevalence of 8.1%. An investigation was performed (more details in discussion Section 4.4), observing that many identical cases were found in the 15% prevalence train, test and verification sets. Over fitting appeared to be occurring affecting network selection. It was found that more unique cases (less identical) occurred in the 8.1% data sets, which explains the poorer performance using the same networks.

Table 4-19 Verification Results at 8.1% Prevalence Unseen Data

PRAMS DT-ANN Results with 8.1% Prevalence Verification Data													
Dataset	Method	PPV	stdev	NPV	stdev	ACC	stdev	MCC	stdev	F1score	stdev	AUC	stdev
Parous	1	0.1714	0.0510	0.9607	0.0060	0.6380	0.2084	0.2064	0.0669	0.2640	0.0571	0.7298	0.0442
Nulliparous	1	0.5779	0.1698	0.9466	0.0021	0.9183	0.0338	0.4228	0.0947	0.4472	0.0710	0.6994	0.0283
ParousWO	1	0.2136	0.0310	0.9411	0.0018	0.8311	0.0326	0.1970	0.0275	0.2711	0.0237	0.6522	0.0131
NulliparousWO	1	0.6004	0.0759	0.9489	0.0020	0.9292	0.0081	0.4559	0.0472	0.4816	0.0381	0.7208	0.0115
Parous	2	0.1704	0.0499	0.9599	0.0082	0.6499	0.1858	0.2060	0.0701	0.2625	0.0584	0.7256	0.0528
Nulliparous	2	0.4427	0.1274	0.9473	0.0035	0.9017	0.0308	0.3651	0.0806	0.4105	0.0664	0.7064	0.0164
ParousWO	2	0.3082	0.2440	0.9099	0.0722	0.7146	0.2789	0.1717	0.1457	0.2383	0.0727	0.6341	0.0701
NulliparousWO	2	0.4558	0.1533	0.9494	0.0034	0.9003	0.0341	0.3812	0.0793	0.4231	0.0626	0.7166	0.0162

Table 4-20 Verification Results at 8.1% Prevalence Unseen Data – Calculated LR values

PRAMS Verification Results using 8.1% Prevalence (Option 0)										
Dataset	No. Variables	Method	Sensitivity (%)	stdev	Specificity (%)	stdev	AUC	stdev	LR +	LR -
Parous	48	1	69.21%	0.1443	63.33%	0.2391	0.7298	0.0442	1.89	0.49
Nulliparous	41	1	38.22%	0.0163	96.55%	0.0372	0.6994	0.0283	11.09	0.64
ParousWO	38	1	38.20%	0.0358	87.07%	0.0380	0.6522	0.0131	2.95	0.71
NulliparousWO	32	1	40.40%	0.0220	97.55%	0.0075	0.7209	0.0115	16.45	0.61
Parous	45	2	68.15%	0.1496	64.71%	0.2136	0.72560	0.0528	1.93	0.49
Nulliparous	45	2	40.35%	0.0417	94.57%	0.0343	0.7064	0.0164	7.42	0.63
ParousWO	40	2	38.35%	0.1170	82.02%	0.0937	0.6265	0.0594	2.13	0.75
NulliparousWO	36	2	42.99%	0.0552	94.17%	0.0405	0.7166	0.0162	7.38	0.61

4.3 Variable Comparison

In this section, variables across databases were compared. For ease of comparison, we first look at PRAMS, with the Old PRAMS variables (used by Yu, 2009). Of note, a number of the variables between the datasets changed in their name, collection, and representation, making it challenging to do a simple comparison. Table 4.21 and 4.22 list variables used in the Old PRAMS ANN analysis and indicate which variables are similar to those used in the recent PRAMS ANN analysis. An “X” indicates that the variable is the same or very similar to one used currently and a “~” indicates a variable that may be a similar variation used, or changed in the new database, otherwise it is a variable that is not used in the new set. In the Parous list, 9 variables are the same or similar and 6 are identified as having a variation in the Newer PRAMS variables. For nulliparous, there are 7 and 5.

Table 4-21 Old PRAMS Parous Variable Used on MIRC ANN – Identifying Variables Similar to New PRAMS Variables Used with ANN Builder.

Old PRAMS Parous Variables - used with MIRC ANN			
	Variable Name	Variable Label	
1	PLURAL	plurality single, twin, triplet (specify)	X
2	MAT_AGE	mat age	X
3	MOM_CIG_later	mat smoking: no of cigarettes smoked in last 3 months of pregnancy	
4	DRK4_3L	mat drinking: avg no. of drinks in last 3 months of pregnancy	
5	BMI	body mass index just before pregnancy	X
6	MORB_PLA	morbidity: problems with placenta	X
7	PREV_PRE	previous no of children	X
8	MM_HBP	hypertension (prepregnancy, gestation, eclampsia)	X
9	MORB_BLD	vaginal bleeding during pregnancy	X
10	MORB_CRV	morbidity: cervix had to be sewn shut (incompetent cervix, cerclage)	
11	BLACK	mat race black	~
12	HAWAII	mat race Hawaii	~
13	PREV_LBW	previous low birthweight	
14	STRS_DV3	stress: separation or divorce from husband or partner	~
15	MOMSMOKE	mat smoking	X
16	STRS_JL3		~
17	STRS_BIL	stress: bills that could not be paid	~
18	MM_DIAB	mat diabetes (prepregnancy & gestational)	X
19	STRS_JOB	stress: husband or partner lost job	~
X	same or very similar value		
~	different but close variation of similar value		

Table 4-22 Old PRAMS Nulliparous Variable Used on MIRG ANN – Identifying Variables Similar to New PRAMS Variables Used with ANN Builder.

Old PRAMS Nulliparous Variables - used with MIRG ANN				
Variable Name		Variable Label		
1	PLURAL	plurality single, twin, triplet (specify)		X
2	MAT_AGE	mat age		X
3	OTH_TERM	history of other terminations		
4	MORB_PLA	morbidity: problems with placenta		X
5	HAWAII	mat race Hawaii		~
6	MM_HBP	hypertension (prepregnancy, gestation, eclampsia)		X
7	MORB_CRV	morbidity: cervix had to be sewn shut (incompetent cervix, cerclage		
8	MORB_BLD	vaginal bleeding during pregnancy		X
9	BLACK	mat race black		~
10	MM_NOMD	no medical risk factor		
11	PACIFIC	mat race pacific		~
12	INCOME	Income		X
13	STRS_WRK	stress: mother lost job		~
14	STRS_JOB	stress: husband or partner lost job		~
15	MORB_KID	morbidity: kidney or bladder (urinary tract) infection		X
16	SMK2YRS	mat smoking: at least 100 cigarettes in the past 2 years		
X same or very similar value				
~ different but close variation of similar value				

The following tables examine variables used across the BORN and PRAMS databases in Parous and Nulliparous sets. These two databases are collected very differently and use many different labels. An attempt at identifying matching or similar variables was made. Table 4.23 and 4.24 list variables used for the ANN analysis, with a column identifying similar or possibly overlapping variable in the PRAMS ANN analysis list. Tables 4.25 and 4.26 do the same in reverse, listing PRAMS variables and identifying variables that were similar in the BORN variable list. Between BORN and PRAMS there are 22 variables that seemingly overlapped in the Parous list, and 18 in the Nulliparous list.

Table 4-23BORN Parous Variable List from M1 and M2 – Identifying Overlapping Variables in the PRAMS dataset used in M1 & M2.

BORN Parous Variables Used for ANN Methods 1 & 2		
Variable	Description	Similar/Overlapping Variable
1 MULTGEST	Multiple gestation	PLURAL
2 MATAGE	Mother's age (years)	MAT_AGE_NAPHSIS
3 PPRETERM	Number previous preterm babies	PREV_PRE
4 PTERM	Number previous term babies	
5 PARITY	Parity	OTH_TERM/PREV_LB
6 PREVCS	Previous C/S	
7 GENDER	Baby's sex	SEX
8 INTBF	Intention to breastfeed	
9 SMOKING	Smoking	MOMSMOKE
10 MATHP2	Maternal health problem - Thyroid disease	
11 MATHP3	Maternal health problem - Chronic hypertension	MM_HBP
12 MATHP4	Maternal health problem - Diabetes insulin	BPG_DIAB/MM_DIAB
13 MATHP5	Maternal health problem - Diabetes non-insulin	BPG_DIAB/MM_DIAB
14 MATHP16	Maternal health problem - Heart disease	
15 MATHP17	Maternal health problem - Hepatitis B	
16 MATHP18	Maternal health problem – HIV	
17 MATHP19	Maternal health problem – Lupus	
18 MATHP26	Maternal health problem - Thyroid disease	
19 MATHP27	Maternal health problem – Other	
20 OBCOMP1	Obstetrical complications – Eclampsia	
21 OBCOMP2	Obstetrical complications - Gestational diabetes	PG_GDB
22 OBCOMP3	Obstetrical complications - Hypertension (gestational or transient)	MM_HBP
23 OBCOMP4	Obstetrical complications – IUGR/SGA	SGA_10
24 OBCOMP5	Obstetrical complications – LGA	LGA
25 OBCOMP7	Obstetrical complications - Placenta previa	MORB_PLA
26 OBCOMP8	Obstetrical complications - Placental abruption	MORB_PLA
27 OBCOMP9	Obstetrical complications – Pre-eclampsia	
28 OBCOMP11	Obstetrical complications - Preterm labour	MORB_LAB
29 OBCOMP12	Obstetrical complications - Preterm premature	
30 OBCOMP13	Obstetrical complications – UTI	MORB_KID
31 OBCOMP14	Obstetrical complications - Other cervical/vaginal	MORB_BLD
32 OBCOMP15	Obstetrical complications – Other	
33 REPASS	Reproductive assistance	INFER_TR
34 FIRSTVIS	First trimester visit	
35 CONGAN55	Congenital anomalies - Anomalies unclassified elsewhere - Other syndromes	
36 PRENCLAS	Prenatal classes	PNC_WKS
37 PCSNUM	No. of previous C/S's	
38 LANGUAGE_up	Mother's primary language	LANGUAGE
39 MATHP_sub	Maternal health problem - Substance use	SMK63L_A/DRK63L_A
40 MATHP_ment	Maternal health problem - Mental health	
41 CONGAN_CNS	Congenital anomalies - Central Nervous System	
42 CONGAN_RES	Congenital anomalies - Respiratory system	
43 CONGAN_GAS	Congenital anomalies – Gastrointestinal	
44 CONGAN_CHR	Congenital anomalies – Chromosomal	
45 CONGAN_CARDIAC	Congenital anomalies – Cardiac	
46 CONGAN_EYE	Congenital anomalies – Eye/Ear	

Table 4-24 BORN Nulliparous Variable List from M1 and M2 – Identifying Overlapping Variables in the PRAMS dataset used in M1 & M2.

BORN Nulliparous Variables Used for ANN Methods 1 & 2		
Variable	Description	Similar/Overlapping Variable
1 MULTGEST	Multiple gestation	PLURAL
2 MATAGE	Mother's age (years)	MAT_AGE_NAPHSIS
3 GENDER	Baby's sex	SEX
4 INTBF	Intention to breastfeed	
5 SMOKING	Smoking	MOMSMOKE
6 MATHP2	Maternal health problem - Thyroid disease	
7 MATHP3	Maternal health problem - Chronic hypertension	MM_HBP
8 MATHP4	Maternal health problem - Diabetes insulin	BPG_DIAB/MM_DIAB
9 MATHP5	Maternal health problem - Diabetes non-insulin	BPG_DIAB/MM_DIAB
10 MATHP16	Maternal health problem - Heart disease	
11 MATHP18	Maternal health problem – HIV	
12 MATHP19	Maternal health problem – Lupus	
13 MATHP26	Maternal health problem - Thyroid disease	
14 MATHP27	Maternal health problem – Other	
15 OBCOMP1	Obstetrical complications – Eclampsia	
16 OBCOMP2	Obstetrical complications - Gestational diabetes	PG_GDB
17 OBCOMP3	Obstetrical complications - Hypertension (gestational or transient)	MM_HBP
18 OBCOMP4	Obstetrical complications – IUGR/SGA	SGA_10
19 OBCOMP5	Obstetrical complications – LGA	LGA
20 OBCOMP7	Obstetrical complications - Placenta previa	MORB_PLA
21 OBCOMP8	Obstetrical complications - Placental abruption	MORB_PLA
22 OBCOMP9	Obstetrical complications – Pre-eclampsia	
23 OBCOMP11	Obstetrical complications - Preterm labour	MORB_LAB
24 OBCOMP12	Obstetrical complications - Preterm premature	
25 OBCOMP13	Obstetrical complications – UTI	MORB_KID
26 OBCOMP15	Obstetrical complications – Other	
27 REPASS	Reproductive assistance	INFER_TR
28 FIRSTVIS	First trimester visit	
29 CONGAN55	Congenital anomalies - Anomalies unclassified elsewhere - Other syndromes	
30 PRENCAS	Prenatal classes	
31 LANGUAGE_up	Mother's primary language	LANGUAGE
32 MATHP_sub	Maternal health problem - Substance use	SMK63L_A/DRK63L_A
33 MATHP_ment	Maternal health problem - Mental health	
34 CONGAN_CNS	Congenital anomalies - Central Nervous System	
35 CONGAN_EYE	Congenital anomalies – Eye/Ear	
36 CONGAN_OROFACIAL	Congenital anomalies – Orofacial	
37 CONGAN_CARDIAC	Congenital anomalies – Cardiac	
38 CONGAN_GAS	Congenital anomalies – Gastrointestinal	
39 CONGAN_CHR	Congenital anomalies – Chromosomal	
40 OBCOMP14	Obstetrical complications - Other cervical/vaginal	
41 MATHP1	Maternal health problem - Alcohol dependence syndrome/Alcoholism	
42 MATHP17	Maternal health problem - Hepatitis B	
43 CONGAN_MUS	Congenital anomalies – Musculoskeletal	
44 CONGAN_GEN	Congenital anomalies – Genitourinary	

Table 4-25 PRAMS Parous Variable List from M1 and M2 – Identifying Overlapping Variables in the BORN dataset used in M1 & M2.

PRAMS Parous Variables Used for ANN Methods 1 & 2			
	Variable	Description	Similar/Overlapping Variable
1	URB_RUR	URBAN/RURAL CATEGORY	
2	MM_HBP	Hypertension?	MATHP/OBCOMP3
3	MM_DIAB	Diabetes?	MATHP4/MATHP5
4	MARRIED	Marital Status	
5	SEX	Gender of Infant	GENDER
6	PLURAL	Plurality	MULTGEST
7	MAT_ED	Maternal Education	
8	PAT_ED	Paternal Education	
9	MOMSMOKE	Did Mom Smoke?	SMOKING
10	MOMLBS	Maternal Weight Gain	
11	OTH_TERM	Pregnancy History Other Terminations?	PARITY
12	MAT_RACE	Maternal Race	
13	PAT_RACE	Paternal Race	
14	PRE_DIET	(a) Pre-pregnancy dieting	
15	PRE_EXER	(b) Pre-pregnancy exercise 3+ days/wk	
16	PRE_RX	(c) Pre-pregnancy prescrip meds	
17	PRE_MH	(f) Pre-pregnancy ck/tx for depression/a	
18	PRE_DDS	(h) Pre-pregnancy teeth cleaned	
19	VITAMIN	Multivitamin -- # of times /wk	
20	BPG_DIAB	Diabetes before pregnancy	MATHP4/MATHP5
21	PREV_LB	Previous -- live birth	PARITY
22	PREV6NBW	Previous -- normal birth weight	
23	PREV_PRE	Previous -- premature	PPRETERM
24	PREG_TRY	Preg -- trying	
25	PG_GDB	Gestational diabetes- this pregnancy	OBCOMP2
26	MORB_BLD	a) Morbidity -- vag bleeding	OBCOMP14
27	MORB_KID	b) Morbidity -- kidney/bladder infect	OBCOMP13
28	MORB_NAU	c) Morbidity -- nausea	
29	MORB_PLA	f) Morbidity -- placenta	OBCOMP7/OBCOMP8
30	MORB_LAB	g) Morbidity -- preterm labor	OBCOMP11
31	MORB_TRN	i) Morbidity -- Blood transfusion	
32	MORB_CAR	j) Morbidity -- car crash injury	
33	PAB6HUS	Abuse -- 12 mnths b4 preg, husb/part	
34	PAD6HUS	Abuse -- dur preg, husb/part	
35	INCOME5	Income -- 12 mnths bef, total income	
36	LANGUAGE	Language -- English/Spanish	LANGUAGE_up
37	MOM_BMIG_QX_REV	VAR: MOM BMI GROUPED FROM QX REVISED	
38	PREGHX	VAR: HX PREV LIVE BIRTHS	
39	PNC_WKS	VAR: WEEKS 1ST PNC VISIT	PRENCLAS
40	SMK63L_A	VAR: YES/NO SMOKE LAST 3 MTH PREG	MATHP_sub
41	DRK63L_A	VAR: YES/NO DRINK LAST 3 MTH PREG	MATHP_sub
42	STRS_T_G	VAR: TOTAL # STRESSES, GROUPED	
43	MAT_AGE_NAPHSIS	Maternal age grouped	MATAGE
44	LGA	Large for Gestational Age Based on 90th Percentile	OBCOMP5
45	SGA_10	Small for Gestational Age Based on 10th Percentile	OBCOMP4
46	FRACE_WHT	Paternal race - White	
47	INFER_TR	Infertility treatment	REPASS
48	MRACE_WHT	Maternal race - White	
49	P_PRTERM	Previous Preterm Births	

Table 4-26 PRAMS Nulliparous Variable List from M1 and M2 – Identifying Overlapping Variables in the BORN dataset used in M1 & M2.

PRAMS Nulliparous Variables Used for ANN Methods 1 & 2			
	Variable	Description	Similar/Overlapping Variable
1	URB_RUR	URBAN/RURAL CATEGORY	
2	MM_HBP	Hypertension?	MATHP/OBCOMP3
3	MM_DIAB	Diabetes?	MATHP4/MATHP5
4	MARRIED	Marital Status	
5	SEX	Gender of Infant	GENDER
6	PLURAL	Plurality	MULTGEST
7	MAT_ED	Maternal Education	
8	PAT_ED	Paternal Education	
9	MOMSMOKE	Did Mom Smoke?	SMOKING
10	MOMLBS	Maternal Weight Gain	
11	MAT_RACE	Maternal Race	
12	PAT_RACE	Paternal Race	
13	PRE_DIET	(a) Pre-pregnancy dieting	
14	PRE_EXER	(b) Pre-pregnancy exercise 3+ days/wk	
15	PRE_RX	(c) Pre-pregnancy prescrip meds	
16	PRE_MH	(f) Pre-pregnancy ck/tx for depression/a	
17	PRE_DDS	(h) Pre-pregnancy teeth cleaned	
18	VITAMIN	Multivitamin -- # of times /wk	
19	BPG_DIAB	Diabetes before pregnancy	MATHP4/MATHP5
20	PREG_TRY	Preg -- trying	
21	PG_GDB	Gestational diabetes- this pregnancy	OBCOMP2
22	MORB_BLD	a) Morbidity -- vag bleeding	OBCOMP14
23	MORB_KID	b) Morbidity -- kidney/bladder infect	OBCOMP13
24	MORB_NAU	c) Morbidity -- nausea	
25	MORB_PLA	f) Morbidity -- placenta	OBCOMP7/OBCOMP8
26	MORB_LAB	g) Morbidity -- preterm labor	OBCOMP11
27	MORB_TRN	i) Morbidity -- Blood transfusion	
28	MORB_CAR	j) Morbidity -- car crash injury	
29	PAB6HUS	Abuse -- 12 mnths b4 preg, husb/part	
30	PAD6HUS	Abuse -- dur preg, husb/part	
31	INCOME5	Income -- 12 mnths bef, total income	
32	LANGUAGE	Language -- English/Spanish	LANGUAGE_up
33	MOM_BMIQ_QX_REV	VAR: MOM BMI GROUPED FROM QX REVISED	
34	PNC_WKS	VAR: WEEKS 1ST PNC VISIT	PRENCLAS
35	SMK63L_A	VAR: YES/NO SMOKE LAST 3 MTH PREG	MATHP_sub
36	DRK63L_A	VAR: YES/NO DRINK LAST 3 MTH PREG	MATHP_sub
37	STRS_T_G	VAR: TOTAL # STRESSES, GROUPED	
38	MAT_AGE_NAPHSIS	Maternal age grouped	MATAGE
39	LGA	Large for Gestational Age Based on 90th Percentile	OBCOMP5
40	SGA_10	Small for Gestational Age Based on 10th Percentile	OBCOMP4
41	FRACE_WHT	Paternal race - White	
42	INFER_TR	Infertility treatment	REPASS
43	MRACE_OTH	Maternal race - Other	
44	FRACE_OTH	Paternal race - Other	
45	MRACE_WHT	Maternal race - White	

4.4 Discussion

The main performance measures used were Sensitivity, Specificity, AUC, LR+ and LR –. The aim was to meet or exceed clinically acceptable values (sensitivity of 65% and specificity of 85%) and compare to values achieved by Yu (2009) and Catley (2007). Past results that were achieved by Yu (2009), with Parous data, were a sensitivity of 65.13% and specificity of 84.07%, and for Catley (2007), as sensitivity of 64% and

specificity of 84%. Recent studies measuring LR+ and LR- values, are found to be clinical useful if LR+ >5 and LR - <0.2 (Honest et al. 2012).

The performance of the BORN database using the ANN Builder using FAST mode with sub parameters, does not meet the clinically useful performance measures. The best results in the BORN analysis were the Parous and Nulliparous sets using Method 2 at 15% prevalence and the verification 8.1% prevalence (real world conditions). For Parous at 15%: Sensitivity (77.67%), Specificity (82.14%), and AUC (0.8462). For Nulliparous at 15%: Sensitivity (79.31%), Specificity (77.88%), and AUC (0.8622). For Parous at 8.1%: Sensitivity (50.53%), Specificity (53.96%), and AUC (0.7721). For Nulliparous at 8.1%: Sensitivity (91.61%), Specificity (95.40%), and AUC (0.7970). BORN also used a significantly larger number of variables (34-45) than were used in the Old PRAMS analysis (16-19, further split to 26), however the larger set is comparable to the number of variables (48) first evaluated by Catley (2007) with the Old PRAMS database, suggesting further refinement of variables may result in better results.

The calculated likelihood ratios for all 8 scenarios at the 8.1% prevalence also do not meet clinical performance expectations (LR+ >5 and LR- <0.2). Method 1 LR+ and LR- values: Parous (4.06, 0.57), Nulliparous (2.68, 0.60), ParousWO (1.63, 0.60), NulliparousWO (2.04, 0.64). Method 2 LR+ and LR- values: Parous (6.02, 0.54), Nulliparous (11.72, 0.48), ParousWO (1.60, 0.79), NulliparousWO (2.03, 0.62).

These do not meet or exceed the Old PRAMS Sensitivity, Specificity and AUC for Parous (65.13%, 84.07%, and 0.8195), and Nulliparous (61.08%, 71.14% and 0.7195) (Yu, 2009).

The performance of the PRAMS database using the ANN Builder using FAST mode with sub parameters, does not meet the clinical performance measures for the verification set results. The best results in the PRAMS analysis were the Parous and Nulliparous sets using Method 2 the verification 8.1% prevalence (real world conditions). For Parous at 8.1%: Sensitivity (68.15%), Specificity (64.71%), and AUC (0.7256). For Nulliparous at 8.1%: Sensitivity (40.35%), Specificity (94.57%), and AUC (0.7064).

The calculated likelihood ratios for all 8 scenarios at the 8.1% prevalence also do not meet clinical performance expectations ($LR+ > 5$ and $LR- < 0.2$). Method 1 $LR+$ and $LR-$ values: Parous (1.89, 0.49), Nulliparous (11.09, 0.64), ParousWO (2.95, 0.71), NulliparousWO (16.45, 0.61). Method 2 $LR+$ and $LR-$ values: Parous (1.93, 0.49), Nulliparous (7.42, 0.63), ParousWO (2.13, 0.75), NulliparousWO (7.38, 0.61). PRAMS also used a significantly larger number of variables (32-48) than were used in the Old PRAMS analysis (16-19, further split to 26), however the larger set is comparable to the number of variables (48), first evaluated by Catley (2007) with the Old PRAMS database, suggesting further refinement of variables may result in better performance. It can also be observed that between applied methods of variable selection, Method 2 results are higher than Method 1, however method 1 results were often close to matching method 2 values.

With 15% prevalence of PTB, the PRAMS datasets appear to meet the performance measures expected by clinicians. Sensitivity, Specificity and AUC for results for Method 2: Parous Sensitivity (99.01%), Specificity (99.58%), and AUC (0.9930) and Nulliparous Sensitivity (99.93%), Specificity (99.99%), and AUC (0.9995). These meet and exceed the Old PRAMS Sensitivity, Specificity and AUC for Parous (65.13%, 84.07%, 0.8195) and Nulliparous (61.08%, 71.14%, 0.7195). But this does not

represent the real world condition and thus cannot be used as a clinical tool until a lower prevalence performs acceptably. . The high performance was also the result of multiple identical cases in the training and testing datasets.

As was seen in the PRAMS datasets with 15% prevalence, testing data was performing near perfect. An investigation was done to verify the validity of these unexpected results. It was observed that many similar or identical cases were found in the 15% prevalence train, test and verification datasets. To ensure the issue was not with tools or calculations made during the experiments, the datasets were checked at major steps throughout the methodology for “uniqueness”, using the UNIX tools ‘sort’ and ‘uniq’ to display unique cases within a set. Data was checked with the original datasets, at case division before DTs (splitting into Parous, Nulliparous and without Obvious cases), after DTs just before normalization, after data normalization before CBRS tool used, after CBRS processing just before ANN analysis and the verification sets before ANN. Recorded details can be found in Appendix B which shows the significance of the discovery of these similar cases.

For instance the Original BORN database contained 226 variables and 669,134 cases with nearly 100% unique cases; however, for PRAMS, there were 372 variables and 109,319 cases, but only 9% were unique. Each of these cases has unique identifiers.

It is probable that, as we reduced the number of variables observed, we limited the factor of uniqueness amongst these cases. As well, it is likely that there is a factor in relation to type of variables that are assessed: the PRAMS database has several variables with binary options (yes or no, 1 or 0) whereas BORN databases tend to have more

options per variable (continuous type data). These factors were not previously considered.

Thus, when looking at PRAMS results, near perfect predictions were made for 15% prevalence dataset, as there is little unique data in dataset. For example, the Parous Method 1 dataset had about 13% unique cases (nulliparous had 12%), before applying the 5by2 cross validation method. So, a lot of duplicate data was randomly assigned in train, test and verification sets, which would make it “easy” for ANN to predict as it has seen these cases before.

In the 8.1% prevalence verification data, the unseen data from hold out testing (refer to Section 3.5.1.2.3), there were 83% unique cases for Parous and 49% unique cases for Nulliparous, which is higher than in the 15% sets. It seemed like the model had more difficulty time with predicting cases with the 8.1% data since there is more unique data than the 15% sets, which contained more identical data. The ANN seems to be over-fitting data during the training so it cannot generalize with new unseen cases.

This observation highlights the importance of using unseen data with real world prevalence as final verification. Again, these factors were not previously considered or applied in training, and would be important to consider in the future.

With the close results of method 1 and method 2 occurring in both BORN and PRAMS data, this suggests that feature selection could be further reduced in size, and further exploration of feature selection filtering may be assessed while being able to maintain high sensitivity, specificity, AUC, and meeting the LR thresholds.

Additionally, it was expected that removing clinically obvious variables from datasets would perform much worse than if kept, which does appear to occur in BORN

and PRAMS in both methods. For example, BORN method 2 verification results of Parous compared to Parous without obvious variables have close sensitivity (50.53%, 41.43%), specificity (91.61%, 74.11%) and AUCs (0.9554, 0.9099).

The interest in observing variable sets that include or exclude clinically ‘obvious’ variables already known to obstetricians for predicting preterm birth is an attempt to illustrate the added utility a prediction tool such as this one would add to information that is already known by the medical team managing the pregnancy. Our results further support findings in the literature suggesting that it is difficult to find a tool that can accurately predict risk of preterm birth in a truly generalized, low risk population at this time (Honest et al., 2012).

At present the BORN database sets appear to contain more unique data and at the validation stage, BORN may perform better than PRAMS. This is left to be seen.

In addition, it is a challenge to make a concrete assessment of the new method and analysis compared to the old as several significant modifications to the methodology were made: use of a new ANN tool, and a more rigorously structured approach was developed with the introduction of: 5by2 cross validation, hold out testing of unseen data, setting high prevalences to select best networks (prediction models), and verifying these networks with unseen data at a prevalence equivalent to real world conditions.

It is also important to note that the results of this thesis work were evaluated with the ANN Builder using “FAST” mode with sub parameters (see Section 3.5.1.4.2.). Recall that ANNs are trained with data, which are applied to classify new and unknown cases, and test for reliability (see section 2.9.4.1 for details on the FANN and the modes). This work performed an analysis of datasets using “Fast Mode”. This mode

approximately tests 0.5% of all possible combinations, excluding the number of hidden nodes.

Run time is highly dependent on the number of cases and the number of variables being evaluated. Testing results for FAST mode can be seen in Appendix A for approximate time estimates and specifications of the machines they were run on. For example, one 5by2 set evaluating 50,000 cases, with 47 variables on a machine in the MIRG lab could take approximately 63.4 hours to complete in FAST mode. While FAST mode only analyzes about 0.5% of all network combination possibilities MEDIUM mode would evaluate approximately 7% of all options and SLOW mode runs through all possible combinations. Using sub parameters and other functions of the ANN Builder increase the number of possibilities, however this also increases the processing and assessment time required. FAST mode with sub parameters assesses about 3% of all combinations. Original testing datasets were run on desktop machines in the MIRG lab, but proved to not be enough processing power. Thus the ANN Builder was adapted to run on the SHARCNET cluster system. However, by then, time for assessment became limited even with more processing power, thus would be useful to perform further analysis in the future with the other modes. Additional computational and processing power would help decrease run time.

These results are reported using the option of optimizing sensitivity, but the ANN Builder is flexible and the optimization of specificity can also be obtained. In a clinical setting, the ability to optimize one or both would be of benefit allowing results to be tailored to the physician's intended action for the result.

From a clinical perspective, the decision to prioritize sensitivity versus specificity depends on what may be done with the information obtained/observed (how it will be used) and what are the costs to the patient and healthcare system of any changes to clinical management based on this information.

According to our clinician partner, when considering scenarios or conditions that would require extreme interventions, such as deciding that a woman at high risk of preterm birth should travel to stay near a centre with facilities for providing high level neonatal intensive care, one would want to use a model that has a high specificity to minimize the false positives given the high cost of the intervention (moving a patient to a new location for an extended period of time). Not only would this model require a high specificity, but clinicians would also request a high PPV (positive predictive value, where a positive prediction rules-in those who will truly deliver early). PPV values in this study resulted in low values below 0.5 other than in BORN 8.1% method 2 results for Parous (0.6041), Nulliparous (0.6795) and the PRAMS 15% results all around 0.9 (as expected with near perfect predictions). Without having a high sensitivity as well, some women who will also deliver early would be missed. However, for the Ottawa area or a similar larger center, one might want to use a system that has a high sensitivity to identify all women at risk and then watch them more closely at each prenatal clinic visit until more evidence for risk of preterm birth unfolds (lower cost intervention). Without a high specificity as well, some women would be labeled as high risk for PTB who actually are not (high false positive rate). This approach is acceptable if the information used does not lead to making a high cost/high stakes change to patient management based on the prediction.

A systematic review from 2012 looked at published studies of PTB prediction tests in the literature and compared the accuracy of the various tests. They concluded that none of the tests have both a high sensitivity and specificity (they report positive and negative likelihood ratio calculated from sensitivity and specificity), and that no single test is useful for all PTB predictions (Honest et al., 2012). This includes fetal fibronectin testing which is considered a benchmark test. It is suggested that various tests need to be used together in conjunction with other information depending on the patient and the reason for the prediction.

Thus in a clinical setting, our clinical partner suggested that using models similar to those obtained from the ANN Builder, a tool could be developed and used to flag women who might be at high risk for PTB. It would then notify the obstetrician to watch certain patients more closely for other indicators of preterm birth, using the tool as an adjunct to other tests or findings.

5 Chapter: Concluding Remarks, Contributions to Knowledge and Future Work

5.1 Conclusion

An integrated data mining method has been improved and applied to two recent databases with the intent to improve the prediction of a medical problem, preterm birth, based on information collected by 23 weeks gestation in an obstetrical environment. In an attempt to help predict the occurrence of this pregnancy complication, this thesis work has raised questions that require further examination and assessment. These findings are presented in the conclusion, contribution to knowledge and future work portions of this chapter.

5.1.1 Accomplished Objectives

The first objective was to evaluate a preterm birth prediction ANN model by applying a DT-ANN hybrid methodology with the unseen BORN database, with the intent of matching or improving previous classification results obtained with the old PRAMS database. In order to accomplish the first objective, a new rigorous method was developed and applied: a 5by2 cross validation method was introduced in order to reduce bias on the results, a new feature selection method was designed and analyzed, specifying training and verification prevalence were incorporated into the methodology, and a new ANN tool was used. This was first assessed with the BORN database. The DT-ANN hybrid produced preliminary results for a general population. The second objective was to apply the same DT-ANN methodology to create an ANN prediction model of preterm birth using an updated PRAMS database, (a previous version was used in prior studies),

and observe classification results. It was found that the ANN results using the unseen verification data in both the BORN and PRAMS databases do not surpass the previous results of the DT-ANN hybrid application and systems that previous studies showed. We also attempted to determine if this prediction system could add more to clinical information about PTB risk already known to medical teams, by using four scenarios of cases within each of the databases: Parous, Nulliparous, Parous without Obvious variables (ParousWO) and Nulliparous without Obvious variables (NulliparousWO). Two Methods of feature selection were utilized for each scenario. The DT-ANN Hybrid produced network selection results using train, test and verification data at 15% prevalence. Due to limited time and processing power available, only a sample of all possible network options were analyzed. To verify these results, an unseen dataset with an 8.1% prevalence, representing real world conditions – Ontario preterm birth rate, was assessed with the best performing networks using data with 15% prevalence. This thesis work has led to the following:

1. An integrated hybrid model was created for both mothers who have had previous children (Parous) and for mothers giving birth to their first child (Nulliparous). This DT-ANN hybrid system was able to predict the complex heterogeneous, multi-factor problem of preterm birth in the 4 scenarios (Parous, Nulliparous, ParousWO, and NulliparousWO), using two methods of feature selection. This has not been previously performed using the BORN database. This was also applied to a new PRAMS database (2009-2011) containing unseen data and new variables. The BORN database consists of data collected from Ontario, Canada, while the PRAMS data is collected from the USA.

- The best results in the BORN analysis were the Parous and Nulliparous sets using Method 2 (feature selection by average % usage) for both the results at 15% prevalence and the verification 8.1% prevalence (real world conditions). For Parous at 15%: Sensitivity (77.67%), Specificity (82.14%), and AUC (0.8462). For Nulliparous at 15%: Sensitivity (79.31%), Specificity (77.88%), and AUC (0.8622). For Parous at 8.1%: Sensitivity (50.53%), Specificity (53.96%), and AUC (0.7721). For Nulliparous at 8.1%: Sensitivity (91.61%), Specificity (95.40%), and AUC (0.7970). LR+ and LR- values, method 2: Parous (6.02, 0.54), Nulliparous (11.72, 0.48). In BORN these do not meet or exceed the clinical usefulness criteria sensitivity of 65% and specificity of 85%, nor the Old PRAMS Parous and Nulliparous values for Sensitivity (65.13%, 61.08%), Specificity (84.07%, 71.14%) and AUC (0.8195, 0.7195). Old PRAMS also used more than the 8.1% prevalence (sets ranged 15.9% to 17.9% prevalence) and a different methodology was used so it is not directly comparable. The calculated likelihood ratios for all 8 scenarios at the 8.1% prevalence also do not meet clinical performance expectations (LR+ >5 and LR- <0.2).
- The best results in the PRAMS analysis were the Parous and Nulliparous sets using Method 2 the verification 8.1% prevalence (real world conditions). For Parous at 8.1%: Sensitivity (68.15%), Specificity (64.71%), and AUC (0.7256). For Nulliparous at 8.1%: Sensitivity (40.35%), Specificity (94.57%), and AUC (0.7064). Old PRAMS also

used more than the 8.1% prevalence (sets ranged 15.9% to 17.9% prevalence) and a different methodology was used so it is not directly comparable. With 15% prevalence of PTB, the PRAMS datasets appear to meet the performance measures expected by clinicians. The Method 2 LR+ and LR- values: Parous (1.93, 0.49), Nulliparous (7.42, 0.63). The calculated likelihood ratios for all scenarios at the 8.1% prevalence also do not meet clinical performance expectations (LR+ >5 and LR- <0.2).

- The positive predictive value (PPV) is the likelihood that the positive prediction is correct. Previous models by Yu (0.43) resulted in a low PPV for preterm birth. PPV values were found to be higher in this work in BORN Method 2 scenarios (BORN P method 2 = 0.6041, NP M2 = 0.6795).
2. Improvements to past models were applied and tested: two variable selection methods used, a method to reduce bias added (5by2 cross validation), and a new ANN tool introduced. The consideration of dataset prevalence being relevant was also applied.
 3. The DTs identified significant contributing variables for predicting preterm birth in both the BORN and PRAMS databases.
 - In BORN: for Parous DTs, at total of 53 variables were submitted to the input and the analysis resulted in 43 and 45 variables being selected for methods 1 and 2 respectively; for Nulliparous DTs, a total of 49 variables was input and resulted in 38 and 44 variables for methods 1 and 2 respectively; for ParousWO DTs, at total of 45 variables was input and

resulted in 36 and 44 variables for methods 1 and 2 respectively; and for NulliparousWO DTs, a total of 42 variables was input and resulted in 34 and 37 variables for methods 1 and 2 respectively.

- In PRAMS: for Parous DTs, at total of 51 variables were input and resulted in 48 and 45 variables for methods 1 and 2 respectively; for Nulliparous DTs, a total of 45 variables was input and resulted in 41 and 45 variables for methods 1 and 2 respectively; for ParousWO DTs, at total of 40 variables was input and resulted in 36 and 40 variables for methods 1 and 2 respectively; and for NulliparousWO DTs, a total of 36 variables was input and resulted in 32 and 36 variables for methods 1 and 2 respectively.

4. Within the 10 generated trees per scenario, some variables were used consistently throughout the generated trees. The difference in consistency highlights the complexity and difficulties encountered when attempting to predict preterm birth.

- Across the 10 generated DTs in BORN: for Parous 52 (of 53) were found in at least one of the trees and 26 of these variables were used in all 10 trees; for Nulliparous 44 (of 49) were found in at least one of the trees and 29 of these variables were used in all 10 trees; for ParousWO 44 (of 45) were found in at least one of the trees and 30 of these variables were used in all 10 trees; for Nulliparous 37 (of 42) were used found in at least one of the trees and 28 of these variables were used in all 10 trees.
- Across the 10 generated DTs in PRAMS: for Parous all 51 variables were used to generate the trees and 41 of these variables were used in all 10

trees; for Nulliparous all 45 variables were used to generate the trees and 30 of these variables were used in all 10 trees; for ParousWO all 40 variables were used to generate the trees and 32 of these variables were used in all 10 trees; for Nulliparous all 36 variables were used to generate the trees and 24 of these variables were used in all 10 trees.

5. Two methods of feature selection were utilized and compared. This is the first time our research group has considered using average % usage of variables across 10 generated DTs for feature selection. M2 appeared to be a “stronger” method of feature selection, producing higher performance results than M1 for both BORN and PRAMS data analysis. It could be that selection of M2 is more refined because the focus of the variables assessed are based on how frequently they are used across the 10 generated trees from 5by2 cross validation. A question remains regarding the impact of finding identical cases within a database.
6. Preliminary results with the BORN database for “without obvious” variable sets indicated poorer performance compared with obvious” variable sets. This was expected due to the elimination of key identity variable types in the database.
7. Preliminary results in the PRAMS database for “without obvious” variables sets indicate close performance to “with obvious” variables, but still had a poorer performance. This was expected due to knowledge of elimination of key identity variables types in the database.
8. BORN used a significantly larger number of variables (34-45) than were used in the Old PRAMS analysis (16-19, further split to 26), however the larger set is comparable to the number of variables (48), first evaluated by Catley (2007) with

the Old PRAMS database, suggesting further refinement of variables may be performed and may result in better overall results.

9. PRAMS also used a significantly larger number of variables (32-48) than were used in the Old PRAMS analysis (16-19, further split to 26), however the larger set is comparable to the number of variables (48), first evaluated by Catley (2007) with the Old PRAMS database, suggesting further refinement of variables may be performed and may result in better overall results.
10. The BORN database contains much more data than has ever been previously applied to this DT-ANN hybrid system. Therefore, limitations of existing MIRC tools (CBRS, ANN Builder, Desktop Machines) were tested. The CBRS system could handle up to 50,000 cases and 49 variables to run at a reasonable time (~2 days, up to 4). The ANN Builder was first limited by the CPU power of the desktop machines in the MIRC lab which would take up to a week to run data just in FAST mode without sub parameters (most basic settings). The processing was then run on the SHARCNET cluster system for more processing power, and with the sub parameters setting, ran the equivalent of 54 CPU years (running a desktop CPU for 54 years 24/7), in only a few weeks.
11. The FAST mode setting with sub parameters were used with the ANN, which meant only a small sample of all possible networks could be analyzed (approximately 3% of all possibilities), so there is great potential for better performance. This work optimized sensitivity, however the ANN builder has the ability to optimize specificity.

12. Preliminary results in the PRAMS database verification sets show poorer performance overall compared to the to BORN database. This may be attributed to identical cases within the datasets. The PRAMS ANN 15% prevalence data produced near perfect prediction results, which raised questions. An investigation was performed and it was found that there were many identical cases within datasets. The Original BORN database provided contained nearly 100% unique cases however, for PRAMS 9% of cases were unique. So when looking at PRAMS datasets, for example Parous Method 1, there are little unique data in the dataset. The Parous Method 1 dataset had about 13% unique cases and nulliparous had 12%; a lot of duplicate data was randomly assigned in train, test and verification sets, which would make it “easy” for the ANN to predict as it has seen these cases before. In the PRAMS 8.1% prevalence verification data, which was removed early in the process, there were 83% unique cases for Parous and 49% unique cases for Nulliparous. It appears that the new model had a difficult time making predictions with the 8.1% data since it was more unique unlike the 15% data, which was much less unique (many identical cases). The ANN is over-fitting data in the training so then it cannot generalize with new unseen cases. These findings and large change to the methodology make it difficult to truly assess the results of this work and compare them with previous results.
13. The intent of this thesis was to generate performance results of an improved and refined DT-ANN analysis for the prediction of preterm birth and to compare these results to past performances: Catley and Yu. Though the new results barely reached matching values of Catley and Yu, primarily due to the discovery of

identical cases occurring within the data, it could also be said that the comparison may have been difficult otherwise. Catley did not assess Nulliparous cases; Nicole did however with a lot less variables used. Their similar methodology is now significantly less rigorous than the one that we developed and applied for this thesis work (adding the consideration and importance of prevalence when training, testing and verifying, the importance of hold out testing with unseen data at real world conditions). Therefore identifying specific type of data that should be collected by PRAMS and BORN programs is not possible at present.

5.2 Contributions to Knowledge

1. This work demonstrated that the DT- ANN hybrid system could be used to apply the same methodology and tools to analyze a similar problem (predicting preterm birth in this case), on different populations of data (PRAMS data from the USA and BORN data from Ontario). To our knowledge it is the first time an integrated DT-ANN system has been applied to predict preterm birth using the BORN database.
2. A new and rigorous methodology was applied for modeling the prediction of preterm birth (with an ANN) using information obtained before 23 weeks gestation, which can now be applied and tested with other studies.
3. Found that predictions with DT-ANN with BORN data perform better than with PRAMS data; however, this may be due to the impact of the low amount of unique cases used for training analysis.

4. Method 2 of feature selection was developed and introduced into the methodology. It appears to be a stronger method, as the selection of features is more refined because the variables selected are based on how strongly and commonly they are used across the 10 generated trees from 5by2 cross validation. However, a final conclusion cannot be drawn due to the findings of the impact of identical cases within a database.
5. Real world data is often difficult to obtain for research especially medical data due to concerns with the risk of patient identification. What enhances this thesis work is MIRG's access to real world data from the BORN and PRAMS databases.
6. One of the key findings of this thesis work was the identification of multiple identical cases within the databases, and ultimately the data sets trained, tested and verified. Because of the large amount of duplicates, these cases were randomly assigned in train, test and verification sets, which would make it "easy" for ANN to predict as it has seen these cases before. It was observed in the 8.1% PRAMS verification analysis that the model had a more difficult time with prediction cases when there was more "unique" data within a set. Therefore the ANN was over-fitting data in training so then it could not generalize with new unseen cases, which may have occurred throughout the data assessed in this work. This observation also highlights the importance of using unseen data with real world prevalence as final verification. These factors were not previously considered or applied in training, and would be important to consider in the future.

5.3 Future Work

1. Further investigation required with other ANN Builder modes (Medium, Slow), and settings (thresholds, sub parameters, hidden nodes, layers) to determine if better results can be obtained; the possibility exists. Analyzing these settings allow for more networks to be observed with potentially better resulting performance as opposed to only about 3%.
2. Further work is needed to improve all scenario outcomes. These preliminary results do not meet clinical expectations of past set values from clinical partner. As well we were unable to properly draw conclusions from these results with past ones due to the differences in approach, and prevalence consideration. Hold out testing added to this difficulty, as it was not previously part of the methodology.
3. Parous and Nulliparous results do not meet clinical standards for likelihood ratio (LR) assessment, in the hold out test. This requires further exploration to improve results.
4. Further work is needed to attempt to reduce the number of variables used. This work presented preliminary results assessing new tools and making minor improvements. A much larger dataset than previously used, resulted in more effort spent on the preprocessing and processing of the data. As a consequence less time was available to focus on variable reduction (pare down). Finding the minimal number of input variables required is important to optimize a physician's time and reduce the complexity of the problem. An optimal set of input variables needs to be determined that still performs well to make accurate predictions. Also

more complexity creates more challenges and increases the time for modeling with an ANN.

5. Identifying specific type of data that should be collected by PRAMS and BORN programs is not possible at present. Future work and discovery need to go into this, after conclusions about uniqueness is within data is addressed.
6. Further investigate balance of too little and too many variables and a way to identify specific key variables, as well as examine variables used across databases.
7. Further investigation is needed to identify the impact of having unique or non-unique cases present in the train, test and verification sets for ANN analysis, and if there is an effect on the networks (models) selected.
8. Apply the methodology to the datasets of other perinatal databases on a national, international or global level to assess the DT-ANN algorithm on different perinatal environments.
9. Apply the methodology of DT – ANN hybrid method to create and evaluate models of other medical conditions. The design implemented was intended to be generic and applicable to other scenarios.
10. Within the BORN database, many important variables were restricted from inclusion in the dataset due to the fact that these variables increased the risk of identifying individual patient health information. These variables included information on marital status, family size, education level, employment status, ethnic origin, and income. There needs to be a way to evaluate the inclusion of these variables without using values that closely link to patient identification.

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

11. Implement a clinically accurate tool, once developed, with the proposed Physician Mothers Support (PMODS) System in a clinical setting, and perform a usability study to determine the usefulness and ease of use of such a tool. Determine the physician views and any area of concern such as: security, privacy, ethical dilemmas etc. PMODS has been designed with Drupal based on Physicians – Mother Decision Support (PPADS) model; data needs to be populated into the model.

5.4 Final Statement

Preliminary results of the new integrated hybrid system utilizing the ANN Builder provided useful insight for improving the process of data analysis when estimating preterm birth. Although the aim was originally to compare these results with Catley (2007) and Yu (2009), a different approach was developed and applied, limitations of the tools were tested, hold out testing for real world conditions was carried-out, and the impact of identical cases was observed and should be taken into future consideration in future work. Though these results are not clinically useful now, this work has further contributed to knowledge about how machine-learning techniques can be integrated into database analysis to predict clinically important outcomes. The hope is that one-day a system such as this will be put to clinical use to help improve healthcare outcomes.

References

Allen N. et al. (2002). Preterm Birth: Making a Difference. Best Start 2002. Available at http://www.beststart.org/resources/rep_health/pdf/Preterm_mstr.pdf Accessed July 9th, 2015

Ambrosius, W., *Topics in Biostatistics (Methods in Molecular Biology)* Humana Press, 1st ed., 2007.

Angus D., Shorr A., White A., Dremsizov T., Schmitz R., and Kelley M., 2006. “Critical Care Delivery in the United States: Distribution of Services and Compliance with Leapfrog Recommendations,” *Critical Care Medicine*, vol. 34, pp. 311–314, 2006.

Basso O., Baird D. Infertility and preterm delivery, birthweight, and caesarean section: A study within the Danish National Birth Cohort. *Human Reproduction*. 2003;18(11):2478–2484.

Behrman R.E., Butler A.S. (Eds.), 2007. Preterm birth: causes, consequences, and prevention/Committee on Understanding Premature Birth and Assuring Healthy Outcomes, Board on Health Sciences Policy National Academies Press, Washington (2007). Available at <http://www.ncbi.nlm.nih.gov/books/NBK11388/> Accessed July 9th, 2015.

Berghella V., Hayes E., Visintine J., and Baxter J., “Fetal Fibronectin Testing for Reducing the Risk of Preterm Birth (Review),” *Cochrane Database Syst Rev.*, vol. 8, October 2008.

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*. Available at <http://www.cdc.gov/reproductivehealth/ProductsPubs/DatatoAction/pdf/birout2.pdf> Accessed July 8th, 2015.

Blount, M., Ebling, M. R., Eklund, J. M., James, A. G., McGregor, C., Percival, N., and Sow, D. (2010). Real-time analysis for intensive care: development and deployment of the Artemis analytic system. *Engineering in Medicine and Biology Magazine, IEEE*, 29 (2), 110-118.

BORN Ontario. (2013). BORN Bulletin. 1(1). Available at <https://www.bornontario.ca/assets/documents/bornbulletin/BORN%20Ontario%20Bulletin%20-%20BORN%20Reporting.pdf> Accessed July 9th, 2015

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.

Centers for Disease Control and Prevention, "Pregnancy Risk Assessment Monitoring System (PRAMS): Home," May 5th 2009. Available at <http://www.cdc.gov/prams/>

Accessed on July 9th, 2015.

Center for Disease Control and Prevention, 2014. “Premature Birth” Available at <http://www.cdc.gov/features/prematurebirth/> Accessed July 9th, 2015.

Chandiramani M., Tribe R., and Shennan A., “Preterm labour and prematurity,” *Obstetrics, Gynaecology and Reproductive Medicine–Elsevier*, pp. 232–237, 2007.

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

Canadian Institute for Health Information, 2009. “Too Early, Too Small: A Profile of Small Babies Across Canada.” Available at http://secure.cihi.ca/cihiweb/dispPage.jsp?cw_page=AR_1106_E Accessed July 9th, 2015.

Canadian Institute for Health Information, 2012. Highlights of 2010–2011 Selected Indicators Describing the Birthing Process in Canada, Available at https://secure.cihi.ca/free_products/Childbirth_Highlights_2010-11_EN.pdf Accessed July 9th, 2015.

Concise Medical Dictionary. “cervical cerclage”, Oxford University Press 2007. Carleton University. Available at <http://www.oxfordreference.com/views/ENTRY.html?subview=Main&entry=t60.e1665> Accessed July 9th, 2015

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

DeFranco EA, Lewis DF, Odibo AO. Improving the screening accuracy for preterm labor: is the combination of fetal fibronectin and cervical length in symptomatic patients a useful predictor of preterm birth? A systematic review. *Am J Obstet Gynecol* 2013;208:233.e1-6.

Dietterich, T. G. (1998). Approximate statistical tests for comparing supervised classification learning algorithms. *Neural Computation* 10 (7), 1895 -1924

Duda R., Hart P., and Stork D., *Pattern Classification*. New York, NY: John Wiley & Sons, Inc., 2nd ed., 2001.

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*. Available at April 12, 2012 from <http://www.waset.org/journals/waset/v17/v17-12.pdf> Accessed July 9th, 2015.

Ennett CM, Frize M. (1998) An investigation into the strengths and limitations of artificial neural networks: an application to an adult ICU patient database. *Proc AMLA Symp* 1998:998.

Ennett CM, Frize M. (2000) Selective sampling to overcome skewed a priori probabilities with neural networks. *Proc AMLA Symp* 2000;(20 Suppl):225-9.

Ennett, C. M., Frize, M., and Scales, N. (2002, October). Logarithmic-sensitivity index as a stopping criterion for neural networks. *In Proceedings of 24th Annual Conference of the Engineering in Medicine and Biology and the Annual Fall Meeting of the Biomedical Engineering Society*, pp 23-26.

Ennett, C. M., & Frize, M. (2003, September). Validation of a hybrid approach for imputing missing data. *In Engineering in Medicine and Biology Society, 2003. Proceedings of the 25th Annual International Conference of the IEEE* (Vol. 2, pp. 1268 – 1271).

Ennett C., Frize M., and Charette E., “Improvement and automation of artificial neural networks to estimate medical outcomes,” *Medical Engineering & Physics — Elsevier*, vol. 26, pp. 321–328, 2004.

Fast Artificial Neural Network library documentation (FANN). Available at: <http://leenissen.dk/fann/wp>. Accessed July 9th, 2015.

Fawcett T., (2006)“An Introduction to ROC Analysis,” *Pattern Recognition Letters—Elsevier*, pp. 861–87, 2006.

ffNTest.com, 2010. Available at <http://www.ffntest.com> Accessed July 9th, 2015

Frize, M., Weyand, S., Bariciak, E., Gilchrist, J., Dunn, S., and Tozer, S. (2011, May). Combined Physician-Parent Decision Support tool for the neonatal intensive care unit. In *Medical Measurements and Applications Proceedings (MeMeA), 2011 IEEE International Workshop on*, pp 59-64.

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, Bariciak E, and Gilchrist J (2013) PPADS: Physician- Parent Decision-Support for Neonatal Intensive Care, MEDINFO Proc. vol. 192, World Congress on Medical and Health Informatics, Copenhagen, Denmark, 2013, pp 23-27.

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

Goldenberg, RL and Rouse, DJ.(1998) The prevention of premature birth. *N Engl J Med.* 1998; 339: 313–320

Goldenberg RL, Iams JD, Mercer BM, Meis PJ, Moawad A, Das A, Miodovnik M, VanDorsten PJ, Caritis SN, Thurnau G, Dombrowski MP. (2001) The Preterm Prediction Study: Toward a multiple-marker test for spontaneous preterm birth. *American Journal of Obstetrics and Gynecology.* 2001;185(3):643–651.

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

Gomez,R., Romero R., Nien, J., Chaiworapongsa, T., Medina, L., Kim, Y., Yoon, B., Carstens, M., Espinoza, J., Iams, J., and Gonzalex, R., (2005) “A short cervix in women with preterm labor and intact membranes: A risk factor for microbial invasion of the amniotic cavity,” *American Journal of Obstetrics and Gynecology*, pp. 678–689, 2005.

Gorthi, A., Firtion, C., Vepa, J. (2009), "Automated risk assessment tool for pregnancy care," *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE* , vol., no., pp.6222-6225, 3-6 Sept. 2009 Available at:

<http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5334644&isnumber=5332379>

Accessed July 9th, 2015

Gullo A. and Lumb P., 2005. "Scoring Systems and Outcome," in *Intensive and Critical Care Medicine: Reflections, Recommendations and Perspectives*, pp. 117–131, Italy: Springer-Verlag Italia, 2005.

Gunaratnam, M. (2013), 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. M.A.Sc. diss. Carleton University.

Hassen S., Romero R., Berry S., Dang K., Blackwell S., Treadwell M., and Wolfe H., (2000) "Patients with an ultrasonographic cervical length ≤ 15 mm have nearly a 50% risk of early spontaneous preterm delivery," *Am J Obstet Gynecol–Elsevier*, vol. 182, pp. 1458–1467, 2000.

Hibbard J., Tart M., and Moawad, 2000. "Cervical Length at 16—22 Weeks' Gestation and Risk for Preterm Delivery," *The American A. Journal of Obstetricians and Gynecologists*, vol. 96, pp. 972–978, 2000.

HL7 (2015). Health Level Seven International. Available at <http://www.hl7.org>
Accessed July 9th, 2015

Honest, H., Bachmann, L., Sundaram, L., Gupta, J., Kleijnen, J., and Khan, K., (2004)
“Review. The accuracy of risk scores in predicting preterm birth—a systematic review,”
Journal of Obstetrics and Gynaecology, vol. 24, pp. 1343–359, 2004.

Honest H, Bachmann LM, Gupta JK, Kleijnen J, Khan KS. Accuracy of cervicovaginal
fetal fibronectin test in predicting risk of spontaneous preterm birth: systematic review.
BMJ 2002;325:301-301 Available at:

[http://ac.els-cdn.com/S0002937813009459/1-s2.0-S0002937813009459-](http://ac.els-cdn.com/S0002937813009459/1-s2.0-S0002937813009459-main.pdf?_tid=39a674c8-5ad9-11e4-92fd-00000aacb361&acdnat=1414085155_2031757bb92ef4576fb63b22a2a322ba)
[main.pdf?_tid=39a674c8-5ad9-11e4-92fd-](http://ac.els-cdn.com/S0002937813009459-main.pdf?_tid=39a674c8-5ad9-11e4-92fd-00000aacb361&acdnat=1414085155_2031757bb92ef4576fb63b22a2a322ba)
[00000aacb361&acdnat=1414085155_2031757bb92ef4576fb63b22a2a322ba](http://ac.els-cdn.com/S0002937813009459-main.pdf?_tid=39a674c8-5ad9-11e4-92fd-00000aacb361&acdnat=1414085155_2031757bb92ef4576fb63b22a2a322ba)

Accessed July 9th, 2015

Honest H., Hyde C.J., Khan K.S., 2012. Prediction of spontaneous preterm birth: no good
test for predicting a spontaneous preterm birth. *Curr Opin Obstet Gynecol* 2012; 24: 422–
433. Available at <http://www.ncbi.nlm.nih.gov/pubmed/23099810> Accessed July 9th,
2015.

Henriksen TB, Baird DD, Olsen J, Hedegaard M, Secher NJ, Wilcox AJ. (1997) Time to
pregnancy and preterm delivery. *Obstetrics and Gynecology*. 1997;89(4):594–599.

IBM InfoSphere. (2010). Leveraging key data to provide proactive patient care. Available
at [http://www-01.ibm.com/software/success/cssdb.nsf/CS/SSAO-](http://www-01.ibm.com/software/success/cssdb.nsf/CS/SSAO-8BQ2D3?OpenDocument&Site=corp&cty=en_us)
[8BQ2D3?OpenDocument&Site=corp&cty=en_us](http://www-01.ibm.com/software/success/cssdb.nsf/CS/SSAO-8BQ2D3?OpenDocument&Site=corp&cty=en_us) Accessed July 10th, 2015.

Incerti M., Ghidini A., Korker V., and Pezzullo J., 2007. "Performance of cervicovaginal fetal fibronectin in a community hospital setting," *Arch Gynecol Obstet*, vol. 275, pp. 347–351, 2007.

Inside HPC, 2015. Available at <http://insidehpc.com/hpc-basic-training/what-is-hpc/>
Accessed July 9th, 2015

Janghel, R.R., Shukla, A., Tiwari, R., Tiwari, P., 2009, "Clinical Decision Support System for Fetal Delivery Using Artificial Neural Network," *New Trends in Information and Service Science, 2009. NISS '09. International Conference on* , vol., no., pp.1070-1075, June 30 2009-July 2 2009 Available at
<http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5260650&isnumber=5260409>
Accessed July 9th, 2015

Jha, G.K. Artificial Neural Networks. (2011). Available at
http://www.iasri.res.in/ebook/EB_SMAR/e-book_pdf%20files/Manual%20IV/3-ANN.pdf Accessed July 8th, 2015

Joffe M, Li Z. Association of time to pregnancy and the outcome of pregnancy. *Fertility and Sterility*. 1994;62(1):71–75.

Jordan E.T., 2013 "Moms-To-Be: What You Should Know About Cervical Length". National Healthy Mothers, Healthy Babies Coalition (HMHB). 2012. Available at:
<http://www.hmhb.org/2013/02/moms-to-be-cervical-length/> Accessed July 9th, 2015

Joseph K., Kramer M., Marcoux S., Ohlsson A., Wu-Wen S., and Allen R., and Platt A., 1998. “Determinants of Preterm Birth Rates in Canada from 1981 through 1983 and from 1992 through 1994,” *The New England Journal of Medicine*, vol. 339, pp. 1434–1439, November 12 1998.

Kaminski, M., Goujard, J., and Rumeau-Rouquette, C. 1973. Prediction of low birthweight and prematurity by a multiple regression analysis with maternal characteristics known since the beginning of pregnancy. *Int. J. Epidemiol.* 2:195-204

Kantardzic M. (2012), Data Mining Concepts, Models, Methods, and Algorithms. Hoboken, NJ: *The Institute of Electrical and Electronics Engineers Press* — John Wiley & Sons, 2002.

Kassel, K.S., 2015. Risk Factors for Preterm Labor and Delivery. Available at <http://health.cvs.com/GetContent.aspx?token=f75979d3-9c7c-4b16-af56-3e122a3f19e3&chunkiid=101358> Accessed July 9th, 2015.

Klosgenand W., Zytkow J., “Decision-Tree Discovery,” in *Handbook of Data Mining and Knowledge Discovery*, pp. 267–276, New York, NY: Oxford University Press, 2002.

section author: R. Kohavi and Quinlan.

Kluge E.-H., 2007. “Resource Allocation in Healthcare: Implications of Models of Medicine as a Profession,” *Medscape*, vol. 9, p. 57, 2007.

Kocis, K.C. and Kocis, D.J. (2012) U.S. Patent Application 12/721,010.

Kumar P. and Suresh G., 2008. “Complications After Preterm Birth: An Overview for

Emergency Physicians,” *Clinical Pediatric Emergency Medicine*, pp. 191–199, 2008.

Latendresse, G. 2009, “The Interaction Between Chronic Stress and Pregnancy: Preterm Birth from A Biobehavioral Perspective,” *American College of Nurse-Midwives*, pp. 8–17, 2009.

Leiticha H. and Kaider A., 2003. “Fetal fibronectin—how useful is it in the prediction of preterm birth?,” *BJOG: an International Journal of Obstetrics and Gynaecology*, vol. 110, pp. 66–77, 2003.

R. Li and Z. Wang, “Mining classification rules using rough sets and neural networks,” *European Journal of Operational Research*, pp. 439–448, 2004.

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

Lim G., Tracey J., Boom N., Karmakar S., Wang J., Berthelot J.M. and Heick C. 2009. CIHI Survey: Hospital Costs for Preterm and Small-for-Gestational Age Babies in Canada. *Healthcare Quarterly*, 12(4) September 2009: 20-24. Available at <http://www.longwoods.com/content/21121> Accessed July 9th, 2015

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

Nissen S. (2007), Large Scale Reinforcement Learning using Q- SARSA(λ) and Cascading Neural Networks. Master’s thesis, Department of Computer Science, University of Copenhagen, Denmark.

Maner W. and. Garfield R E., “Identification of Human Term and Preterm Labor using Artificial Neural Networks on Uterine Electromyography Data,” *Annals of Biomedical Engineering*, vol. 35, pp. 465–473, March 2007.

March of Dimes, PMNCH, Save the Children, WHO. Born Too Soon: The Global Action Report on Preterm Birth. Eds CP Howson, MV Kinney, JE Lawn. World Health Organization. Geneva, 2012.

March of Dimes Foundation. 2015. Available at <http://www.marchofdimes.org/> Accessed on July 9th, 2015.

Martin J.A., Hamilton B.E., Osterman M.J.K., 2013. Births: Final data for 2012. National vital statistics reports; vol 62 no 9. Hyattsville, MD: National Center for Health Statistics. 2013.

Martirosyan, H. (2014). Integration of Neonatal Mortality Prediction Models into a Clinical Decision Support System. M.A.Sc. diss. Carleton University.

Martirosyan, H., Frize, M., Ong, D.E., Gilchrist, J., Bariciak, E., 2015. A Decision-Support System for Expecting Mothers and Obstetricians. Springer International Publishing. January 2015. *6th European Conference of the International Federation for Medical and Biological Engineering IFMBE Proceedings* Volume 45, 2015, pp 703-706

McParland P., Jones G., and Taylor D., 2004. “Preterm Labour and Prematurity,” *Current Obstetrics & Gynaecology Elsevier*, vol. 14, pp. 309–319, 2004.

Mercer B., Goldenberg R., Das A., Moawad A., Iams J., Meis P., Copper R., Johnson F., Thom E., McNellis D., Miodovnik M., Menard M., Caritis S., Thurnau G., Bottoms S., and Roberts J. 1996, “The Preterm Prediction study: A Clinical Risk Assessment System,” *American Journal Of Obstetrical & Gynecology*, vol. 174, pp. 1885– 95, 1996.

S. Mitra S. and Acharya, T. (2003), *Data Mining Multimedia, Soft Computing and Bioinformatic*. Hoboken, NJ: John Wiley & Sons, Inc., 2003.

Moorman, J. R., Rusin, C. E., Lee, H., Guin, L. E., Clark, M. T., Delos, J. B., and Lake, D. E. (2011, August). Predictive monitoring for early detection of subacute potentially catastrophic illnesses in critical care. *In Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE* (pp. 5515- 5518)

Moorman, J. R., Carlo, W. A., Kattwinkel, J., Schelonka, R. L., Porcelli, P. J., Navarrete, C. T., and Michael O’Shea, T. (2011). Mortality reduction by heart rate characteristic monitoring in very low birth weight neonates: a randomized trial. *The Journal of pediatrics*, 159 (6), 900-906.

Ontario’s maternal newborn and early child development resource centre. (2012). *Preterm Labour Signs & Symptom*. Available at http://www.beststart.org/resources/rep_health/preterm/Preterm_English_2012.pdf Accessed July 9th, 2015.

Oxford University Press 2007. Carleton University. “preterm birth”, “Concise Medical Dictionary.” Available at:

<http://www.oxfordreference.com/views/ENTRY.html?subview=Main&entry=t60.e8177>

Accessed July 2015

Pan, A. (2002), “MIRG ANN Guide,” 2002. University of Ottawa — School of Information Technology and Engineering.

Peleg, M., and Tu, S. (2011). Decision Support, Knowledge Representation and Management in Medicine. Stanford Centre for Biomedical Informatics Research. Available at http://bmir.stanford.edu/file_asset/index.php/1009/SMI-2006-1088.pdf Accessed July 9th 2015.

Peng, W., et al. (2006). An Implementation of ID3 – Decision Tree Learning Algorithm. *University of New South Wales*. Available at <http://web.arch.usyd.edu.au/~wpeng/DecisionTree2.pdf> Accessed July 9th 2015

Petrini J., Dias T., McCormick M., Massolo M., and Green N., 2008. “Increased Risk of Adverse Neurological Development for Late Preterm Infants,” *The Journal of Pediatrics*, pp. 1–11, 2008.

Pingili R., Bamigboye V., and Jegede K., 2008. “Multiple pregnancy: a blessing or a curse?(Report),” *The Internet Journal of Gynecology and Obstetrics*, vol. 84, July 2008.— University of Ottawa.

Preventing Preterm Birth, 2015. “Preventing Preterm Birth”. Keep’Em Cookin’ (blog). 2015 Available at:

<http://www.keepemcookin.com/educate-yourself/prevention-treatment/preventing-preterm-birth/> Accessed July 9th, 2015.

REALTROMIS, Inc. – Product Pipeline Analysis. GlobalData. (2013). Available at <http://www.marketresearch.com/GlobalData-v3648/REALTROMINS-Product-Pipeline-7295746/> Accessed July 9th 2015.

Rochman, B. 2012. The Cost of Premature Birth: For One Family, More than \$2 Million. Available at <http://healthland.time.com/2012/05/02/the-cost-of-premature-birth-for-one-u-s-family-it-was-more-than-2-million/> Accessed July 9th, 2015

Romero R, Oyarzun E, Mazor M, Sirtori M, Hobbins JC, Bracken M. (1989) Meta-analysis of the relationship between asymptomatic bacteriuria and preterm delivery/low birth weight. *Obstetrics and Gynecology*. 1989;73(4):576–582.

Rulequest Research. (2010). Is See5/C5.0 better than C4.5? Available at <http://rulequest.com/see5-comparison.html> Accessed July 9th, 2015.

Rulequest Research. (2012). See5: An Informal Tutorial. Available at <http://www.rulequest.com/see5-win.html#.names> Accessed July 2015

Rybachynski, 2005) Rybachynski, D. (2005). Design of an Artificial Neural Network Research Framework to Enhance the Development of Clinical Prediction Models. Master's Thesis, University of Ottawa, Ottawa, Ontario, Canada.

Segal S. and Norwitz E.,(2004) “Chapter 20– preterm labour,” in *Anaesthetic and Obstetric Management of High-Risk Pregnancy*, Springer- Verlag New York, third ed., 2004. Available at <http://lib.myilibrary.com.proxy.library.carleton.ca/Browse/open.asp?ID=18890> Accessed July 9th, 2015

SHARCNET, 2015. Available at <https://www.sharcnet.ca/my/about> Accessed July 9th, 2015

Statistics Canada (2014). Births, estimates, by province and territory. Available at <http://www.statcan.gc.ca/tables-tableaux/sum-som/l01/cst01/demo04a-eng.htm> Accessed July 9th, 2015

Srihari, S. (2007). Introduction to Pattern Recognition. *Department of Computer Science and Engineering, University of Buffalo*. Available at <http://www.cedar.buffalo.edu/~srihari/CSE555/> Accessed July 10th 2015

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.

H. Tan, S. Wu Wen, X. Kuan Chen, K. Demissie, and M. Walker, “Early prediction of preterm birth for singleton, twin, and triplet pregnancies,” *European Journal of Obstetrics & Gynecology and Reproductive Biology–Elsevier*, vol. 131, pp. 132– 137, 2007.

Tan, P., Steinbach, M., and Kumar, V., *Introduction to Data Mining*. Boston, MA: Pearson Education Inc, 2006.

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*. Available at <http://smj.sma.org.sg/4903/4903a1.pdf> Accessed July 9th, 2015.

The Canadian Neonatal Network. 2009. Available at <http://www.canadianneonatalnetwork.org/portal/> Accessed on July 9th, 2015.

The Mothers Program. 2013. BORN – Better Outcomes Registry & Network. Available at <http://www.themothersprogram.ca/resources-and-information/born> Accessed on July 9th, 2015.

Thommandram, A., Eklund, J. M., McGregor, C., Pugh, J. E., & James, A. G. (2014, June). A Rule-Based Temporal Analysis Method for Online Health Analytics and Its Application for Real-Time Detection of Neonatal Spells. *In Big Data (BigData Congress), 2014 IEEE International Congress on* (pp. 470-477)

Tomashek K., Shapiro-Mendoza C., Davidoff M., and Petrini J., 2007. “Differences in Mortality between Late-Preterm and Term Singleton Infants in the United States, 1995—2002,” *Mosby, Inc*, pp. 450–456, 2007.

Vidakovic, B. 2011. *Statistics for Bioengineering Sciences with MATLAB and WinBUGS Support*. New York: Springer.

Weyand, S., Frize, M., Bariciak, E., Dunn, S. and S. Tozer. Combined Physician-Parent Decision Support Tool for the Neonatal Intensive Care Unit. *Proc. MeMeA 2011, Bari, Italy*.

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*.

Women's Imaging Services, Obstetric & Gynaecology Ultrasound, 2012. Available at: <http://www.womensimaging services.com.au/obstetric-ultrasound/cervical-length-assessment.aspx> Accessed July 9th, 2015

Yost NP, Owen J, Berghella V, MacPherson C, Swain M, Dildy GA III, Miodovnik M, Langer O, Sibai B. (2004) Second-trimester cervical sonography: Features other than cervical length to predict spontaneous preterm birth. *Obstetrics and Gynecology*. 2004;103(3):457–462.

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.

Appendices

Appendix A – Desktop CPU Testing with ANN Builder

Appendix B – BORN and PRAMS Investigation Into Identical Data within Databases

Appendix C – Sample of ANN Final Options

Appendix A - Desktop CPU Testing with ANN Builder

Before the Cluster System with ANN Builder on SHARCNET was set up. Data was going to be originally run on MIRG desktop machines. Shows the specifications of the CPUs used to perform ANN Builder testing analysis. See the run times and information on the data used for the test runs. BORN and PRAMS data was used.

Table A-1 Desktop Machine CPU Specifications

Machine	Specifications
BME – 13	Intel Core i7 3770 (3.4GHz) processor, 8GB DDR3 RAM (800 MHz), Windows 7 64bit.
BME – 12	Intel Core i5 760 (2.80GHz) processor, 8GB RAM (), Windows 7 64bit.
BME – 11	Intel Core i5 76 (2.80 GHz) processor, 8GB RAM (), Windows 7 64bit.
BME – 10	Intel Core 2 Duo E8400 (3.00 GHz) processor, 4GB RAM (), Windows Vista 32bit.
BME – 14	Intel Core i7-3770 (3.4GHz) processor, 8 GB RAM (), Windows 7 64bit

Table A-2 ANN Builder Run Information Using Sample BORN Data

BORN ANN Builder Dataset Analysis Run Information								
	Parous M1	Nulliparous M1	ParousWO M1	NulliparousWO M1	Parous M2	Nulliparous M2	ParousWO M2	NulliparousWO M2
No. Features	47	40	37	30	30	30	30	30
No. Cases	50,000	50,000	50,000	50,000	50,000	50,000	50,000	50,000
Machine Used	BME 12	BME 14	BME 14	BME 14	BME 14	BME 13	BME 13	BME 11
No. Sets	10	10	10	10	10	10	10	10
Total Hours	634	185	186	160	161	174	170	260
Hours Per Set	63.4	18.5	18.6	16	16.1	17.4	17	26

*M1 = Method 1, M2 = Method 2

Table A-3 ANN Builder Run Information Using Sample PRAMS Data

PRAMS ANN Builder Dataset Analysis Run Information								
	Parous M1	Nulliparous M1	ParousWO M1	NulliparousWO M1	Parous M2	Nulliparous M2	ParousWO M2	NulliparousWO M2
No. Features	48	43	39	35	30	30	30	30
No. Cases	10,000	10,000	10,000	10,000	10,000	10,000	10,000	10,000
Machine Used	BME 14	BME 14	BME 11	BME 14	BME 14	BME 13	BME 14	BME 13
No. Sets	10	10	10	10	10	10	10	10
Total Hours	73	36	55	51	32	40	33	76
Hours Per Set	7.3	3.6	5.5	5.1	3.2	4	3.3	7.6

*M1 = Method 1, M2 = Method 2

Table A-4 ANN Builder Run Information Using Sample Old PRAMS Data

OLD PRAMS ANN Builder Dataset Analysis Run Information		
	10 validation	5by2
No. Features	26	26
No. Cases	14,124	14,124
Machine Used	BME 13/14	BME 13
No. Sets	10	10
Total Hours	11	15
Hours Per Set	1.1	1.5

Appendix B - BORN and PRAMS Investigation Into Identical Data within Databases

Investigation Data:

Originally it was observed that there was several incidences of duplicate patient data were found in the 15% prevalence train, test and verification datasets. An investigation went into observing if there was a potential problem. Initial thoughts were that the issues might have come from the CBRS stage, or with the normalizing of the data prior to CBRS.

To ensure that these duplications were not an issue with tools or calculations made for processing, the datasets were checked at major steps throughout the methodology for “uniqueness”, using the UNIX tools ‘sort’ and ‘uniq’ to display unique cases within a set. Data was checked with the original datasets, at case division before DTs (splitting into Parous, Nulliparous and without Obvious cases), after DTs just before normalization, after data normalization before CBRS tool used, after CBRS processing just before ANN analysis and the verification sets before ANN.

The following tables are the observations found for BORN and PRAMS. The number of Unique Cases for each set was identified, and compared to the total. The % of uniqueness within a set was calculated.

B.1 Investigation Results – BORN Data

Table B-1 Uniqueness in the Original Untouched Data Set

Original Data Set (226)	
Total Cases	669,134
Unique Cases	668,953
% Unique Cases	99.97%

Table B-2 Uniqueness within Cases to be Investigated

Dataset Numbers After Case Type Division (by Variable) Prior to Data Cleaning				
	Parous (56)	Nulliparous (52)	ParousWO (48)	NulliparousWO (45)
Total Cases	669,134	669,134	669,134	669,134
Unique Cases	447,373	417,378	430,814	430,814
% Unique Cases	66.86%	62.38%	64.38%	64.38%

Table B-3 Uniqueness before Data Normalization

Dataset Numbers Before Normalized				
	Parous	Nulliparous	ParousWO	NulliparousWO
Total Cases	50,000	50,000	50,000	50,000
Unique Cases	27,671	20,853	25,995	17,860
% Unique Cases	55.34%	41.71%	51.99%	35.72%

Table B-4 Uniqueness after Data Normalization

Normalized Dataset Numbers Before CBRS Run				
	Parous	Nulliparous	ParousWO	NulliparousWO
Total Cases	50,000	50,000	50,000	50,000
Unique Cases	27,671	20,853	25,995	17,860
% Unique Cases	55.34%	41.71%	51.99%	35.72%

Table B-5 Uniqueness in Data Prior to ANN analysis for Method 1

Dataset Numbers After CBRS and Prepared Before Running with ANN - For Method 1				
	Parous (43)	Nulliparous (38)	ParousWO (40)	NulliparousWO (34)
Mortality				
Total Cases	7,500	7,500	7,500	7,500
Unique Cases	7,301	6,465	6,409	4,830
% Unique Cases	97.35%	86.20%	85.45%	64.40%
Nonmortality				
Total Cases	42,500	42,500	42,500	42,500
Unique Cases	20,043	13,756	19,124	12,402
% Unique Cases	47.16%	32.37%	45.00%	29.18%

Table B-6 Uniqueness of Hold Out Verification Data for Method 1

Dataset Numbers of the prepared Verification Sets - For Method 1				
	Parous (43)	Nulliparous (38)	ParousWO (40)	NulliparousWO (34)
Total Cases	12,500	12,500	12,500	12,500
Unique Cases	9,064	7,549	8,878	7,058
% Unique Cases	72.51%	60.39%	71.02%	56.46%

B.2 Investigation Results – PRAMS Data

Table B-7 Uniqueness in the Original Untouched Data Set

Original Data Set (372)	
Total Cases	109,319
Unique Cases	10,000
% Unique Cases	9.15%

Table B-8 Uniqueness within Cases to be Investigated

Dataset Numbers After Case Type Division (by Variable) Prior to Data Cleaning				
	Parous (52)	Nulliparous (46)	ParousWO (41)	NulliparousWO (37)
Total Cases	70,880	38,439	70,880	38,439
Unique Cases	6,477	3,523	6,477	3,523
% Unique Cases	9.14%	9.17%	9.14%	9.17%

Table B-9 Uniqueness before Data Normalization

Dataset Numbers Before Normalized				
	Parous	Nulliparous	ParousWO	NulliparousWO
Total Cases	20,000	20,000	20,000	20,000
Unique Cases	2,684	2,420	2,858	2,438
% Unique Cases	13.42%	12.10%	14.29%	12.19%

Table B-10 Uniqueness after Data Normalization

Normalized Dataset Numbers Before CBRS Run				
	Parous	Nulliparous	ParousWO	NulliparousWO
Total Cases	20,000	20,000	20,000	20,000
Unique Cases	2,684	2,420	2,858	2,438
% Unique Cases	13.42%	12.10%	14.29%	12.19%

Table B-11 Uniqueness in Data Prior to ANN analysis for Method 1

Dataset Numbers After CBRS and Prepared Before Running with ANN - For Method 1				
	Parous (48)	Nulliparous (41)	ParousWO (38)	NulliparousWO (32)
Mortality				
Total Cases	3,000	3,000	3,000	3,000
Unique Cases	566	430	576	436
% Unique Cases	18.87%	14.33%	19.20%	14.53%
Nonmortality				
Total Cases	17,000	17,000	17,000	17,000
Unique Cases	2,118	1,990	2,282	2,002
% Unique Cases	12.46%	11.71%	13.42%	11.78%

Table B-12 Uniqueness of Hold Out Verification Data for Method 1

Dataset Numbers of the prepared Verification Sets - For Method 1				
	Parous (48)	Nulliparous (41)	ParousWO (38)	NulliparousWO (32)
Total Cases	5,000	5,000	5,000	5,000
Unique Cases	4,162	2,433	4,147	2,446
% Unique Cases	83.24%	48.66%	82.94%	48.92%

Appendix C - Sample of ANN Final Options

A detailed sample of the final FANN library options used for best networks for

BORN Parous Method 2 Option 0, Good Network 1, at 15% prevalence.

[illegible]

[illegible]

6.52309894561767578125e-01) (13, 7.10850894451141357422e-01) (38, -5.86697697639465332031e-01) (40, 2.14233249425888061523e-01) (36, 3.07997196912765502930e-01) (45, -1.93622976541519165039e-01) (39, 7.06248700618743896484e-01) (20, -4.81044143438339233398e-01) (4, 4.13434579968452453613e-02) (23, -7.24720001220703125000e-01) (45, 2.82841175794601440430e-01) (32, -2.63646453619003295898e-01) (24, 1.36788785457611083984e+00) (36, 3.03051501512527465820e-01) (6, -7.04784333705902099609e-01) (2, 8.89054685831069946289e-02) (45, -8.64927113056182861328e-01) (40, 2.92503803968429565430e-01) (3, -7.80240774154663085938e-01) (35, 2.74938017129898071289e-01) (9, 3.57987612485885620117e-01) (45, -2.13164776563644409180e-01) (41, -3.82343292236328125000e-01) (9, 8.03279131650924682617e-02) (3, 2.01036527752876281738e-01) (37, -6.76640644669532775879e-02) (14, -2.54275679588317871094e-01) (45, 3.55240076780319213867e-01) (11, -5.90291202068328857422e-01) (24, -4.21974986791610717773e-01) (4, 5.99358975887298583984e-01) (40, -4.83500003814697265625e-01) (45, -1.90794020891189575195e-01) (35, -9.86554145812988281250e-01) (11, 2.97937065362930297852e-01) (26, 7.23942816257476806641e-01) (18, -2.03573647886514663696e-02) (31, 7.10905849933624267578e-01) (45, -4.92818981409072875977e-01) (13, 5.15494167804718017578e-01) (6, 5.11064529418945312500e-01) (38, 1.10907530784606933594e+00) (2, -5.30992627143859863281e-01) (45, 3.48044008016586303711e-01) (19, -7.11613297462463378906e-02) (30, 3.88662964105606079102e-02) (37, 6.97033047676086425781e-01) (15, 4.25050795078277587891e-01) (10, -3.60575199127197265625e-01) (45, -3.93108606338500976562e-01) (9, 3.19086350500583648682e-02) (32, -6.76210463047027587891e-01) (22, -2.24504053592681884766e-01) (16, 1.95334896445274353027e-01) (45, 1.26694783568382263184e-01) (1, 9.14594709873199462891e-01) (22, 1.57739490270614624023e-01) (25, -3.24844866991043090820e-01) (2, 9.43001508712768554688e-01) (19, 6.97212159633636474609e-01) (45, -7.2587704658508309731445e-01) (44, 2.9758706688809204102e-01) (24, 1.4637742936611755371e-01) (15, -1.93836875259876251221e-02) (29, 3.13579559326171875000e-01) (45, -3.62046137452125549316e-02) (10, -1.98383688926696777344e-01) (28, -1.83008217811584472656e+00) (6, -4.61158901453018188477e-01) (33, -6.71582520008087158203e-02) (1, -7.98896253108978271484e-01) (45, -4.94721144437789916992e-01) (36, -3.35715323686599731445e-01) (31, -2.71966040134429931641e-01) (39, 7.95894920825958251953e-01) (41, 5.56004405021667480469e-01) (45, -5.85279881954193115234e-01) (28, -2.42134124040603637695e-01) (37, 8.25267881155014038086e-02) (25, -3.83086323738098144531e-01) (34, 5.60741782188415527344e-01) (40, -6.43863022327423095703e-01) (45, -1.60481810569763183594e-01) (15, -4.47882831096649169922e-01) (13, 6.437526140271004938477e-01) (20, -6.56014978885650634766e-01) (37, -7.95946657657623291016e-01) (45, -7.58640468120574951172e-01) (9, -1.80453404784202575684e-01) (17, 2.52495467662811279297e-01) (36, 9.60726812481880187988e-02) (42, 6.35948851704597473145e-02) (25, 5.81395387649536132812e-01) (45, -3.69800031185150146484e-01) (37, -7.37512528896331787109e-01) (7, -4.43337380886077880859e-01) (26, -4.27102036774158477783e-02) (43, -1.70070633292198181152e-01) (45, -1.55916392803192138672e-01) (14, 3.34274649620056152344e-01) (32, 2.59248882532119750977e-01) (7, 1.34337440133094787598e-01) (33, 3.90444546937942504883e-02) (0, -2.53755172729492187500e+02) (45, 7.21870779991149902344e-01) (14, -2.12242186069488525391e-01) (36, 4.73426759243011474609e-01) (8, 3.10583800077438354492e-01) (28, -7.05590677261352539062e+00) (45, 3.52847903966903686523e-01) (25, -3.66873472929000854492e-01) (3, -6.777526140271004938477e-01) (2, 2.50548124313354492188e-01) (40, 8.28312337398529052734e-02) (9, -2.43933927267789840698e-02) (45, -2.30653777718544006348e-01) (14, 3.84358391165733337402e-02) (42, -3.15141975879669189453e-01) (40, 6.24739825725555419922e-02) (13, 2.68275707960128784180e-01) (45, -5.29718518257141113281e-01) (28, 2.06142067909240722656e+00) (14, -4.89859670400619506836e-01) (23, 3.48314717411994934082e-02) (8, 6.64300799369812011719e-01) (5, -4.83694136142730712891e-01) (45, 7.04589545726776123047e-01) (24, 2.95213103294372558594e-01) (42, 1.45440220832824707031e-01) (8, -5.36202311515808105469e-01) (25, -4.31153297424316406250e-01) (45, -2.72818148136138916016e-01) (5, -2.10462450981140136719e-01) (12, 2.61485278606414794922e-01) (20, -8.88308361172676086426e-02) (8, -3.61437164247335980225e-02) (17, -7.78891980648040771484e-01) (45, -3.07752937078475952148e-01) (20, -5.84121882915496826172e-01) (31, -1.48307383060455322266e-01) (10, -6.84699416160583496094e-01) (6, -4.14069563150405883789e-01) (45, -9.43567082285881042480e-02) (17, 2.20539078116416931152e-01) (19, 3.20409148931503295898e-01) (42, 3.56686979532241821289e-01) (22, -5.81643581390380859375e-01) (4, -4.74709033966064453125e-01) (45, 6.04363739490509033203e-01) (6, -2.71579381078481674194e-02) (24, 5.34550964832305908203e-01) (10, -6.94083154201507568359e-01) (31, -1.79142683744430541992e-01) (45, -9.80472266674041748047e-01) (11, 5.80342952162027359009e-03) (18, 9.05638933181762695312e-02) (23, -1.03826534748077392578e+00) (36, -5.33062126487493515015e-03) (39, -2.73540914058685302734e-01) (45, 5.33074557781219482422e-01) (1, 2.44035050272941589355e-01) (38, 1.47469311952590942383e-01) (35, -7.94031679630279541016e-01) (6, 5.53449213504791259766e-01) (45, 8.04663971066474914551e-02) (21, 2.77097642421722412109e-01) (38, -3.31441760063171386719e-01) (12, 3.40657174587249755859e-01) (39, -3.59089940786361694336e-01) (11, -3.49457412958145141602e-01) (45, -4.85512226819992065430e-01) (19, -3.62256467342376708984e-01) (15, -1.02496901527047157288e-02) (39, 7.93707445263862609863e-02) (7, 6.41797423362731933594e-01) (45, -2.75652080774307250977e-01) (7, -3.07076424360275268555e-01) (9, 7.70045638084411621094e-01) (24, -1.55674800276756286621e-01) (12, -3.43179404735565185547e-01) (23, -5.98267875611782073975e-02) (45, -2.23697200417518615723e-01) (32, 5.47695495188236236572e-02) (25, 1.32391691207885742188e-01) (8, -5.16442358493804931641e-01) (43, 2.13376507163047790527e-01) (45, -3.58897536993062733398e-01) (33, 7.24103972315788269043e-02) (32, -1.09263002872467041016e-01) (27, -2.52655267715454101562e-01) (44, -9.31113719940185546875e-01) (37, 1.21778296306729316711e-02) (45, 1.08463428914546696553e-01) (12, -1.64570778608322143555e-01) (43, 8.24429839849472045898e-02) (29, -5.02716183662414550781e-01) (30, -6.60514354705810546875e-01) (45, 2.27327495813369750977e-01) (8, 6.80653035640716552734e-01) (9, -3.80972146987915039062e-01) (15, 7.21845924854278564453e-01) (23, -7.25298702716827392578e-01) (19, 5.89102581143379211426e-02) (45, -8.08840468525886535645e-02) (13, 3.02227467298507690430e-01) (38, -5.65992712974548339844e-01) (43, 2.66797155141830444336e-01) (22, 1.16338157653808593750e+00) (45, -1.07813882827758789062e+00) (0, 3.47582672119140625000e+02) (24,

4.93528723716735839844e-01) (37, 7.90841460227966308594e-01) (43, -1.50000000000000000000e+03) (35, -
 6.31449103355407714844e-01) (45, 9.12331491708755493164e-02) (34, -1.35460898280143737793e-01) (27, -
 1.77593100070953369141e+00) (14, 3.37365776300430297852e-01) (35, -1.36938700452446937561e-02) (45,
 3.89007389545440673828e-01) (3, 4.08339977264404296875e-01) (11, -1.96274101734161376953e-01) (40, -
 1.92236408591270446777e-01) (6, -2.80875951051712036133e-01) (19, 1.97745159268379211426e-01) (45,
 6.42069697380065917969e-01) (33, 2.02794298529624938965e-02) (15, 4.29230779409408569336e-01) (8,
 2.71205365657806396484e-01) (6, -1.56814888119697570801e-01) (45, -2.67704665660858154297e-01) (18,
 2.61477470397949218750e-01) (11, -1.17738649249076843262e-01) (38, 4.73922863602638244629e-02) (2, -
 4.13558363914489746094e-01) (1, -6.77269220352172851562e-01) (45, 4.68400597572326660156e-01) (4,
 6.11781358718872070312e-01) (26, 1.04771673679351806641e+00) (19, 4.80123072862625122070e-01) (41,
 6.42586350440979003906e-01) (45, -6.64172232151031494141e-01) (15, -5.17639219760894775391e-01) (10, -
 2.14612126350402832031e-01) (17, 2.81184047460556030273e-01) (23, 5.05749046802520751953e-01) (34,
 6.73704028129577636719e-01) (45, -6.11167788505554199219e-01) (2, 7.82987236976623535156e-01) (36, -
 5.84005892276763916016e-01) (7, 5.59560298919677734375e-01) (8, -1.42325431108474731445e-01) (45, -
 1.70350134372711181641e-01) (23, -3.82569670677185058594e-01) (4, 1.37728407979011535645e-01) (32, -
 2.15565562248229980469e-01) (35, -2.86606550216674804688e-01) (15, -6.02677166461944580078e-01) (45,
 3.39699864387512207031e-01) (36, 5.70388913154602050781e-01) (16, 2.23759841173887252808e-02) (1,
 4.51828837394714355469e-01) (12, 4.03671450912952423096e-02) (45, 6.43691003322601318359e-01) (33, -
 1.38002187013626098633e-01) (41, 3.42473655939102172852e-01) (38, 8.17721247673034667969e-01) (11, -
 1.40160262584686279297e-01) (17, 2.04418063163757324219e-01) (122, 3.20001554489135742188e+00) (46,
 1.43006950616836547852e-01) (47, 1.54054179787635803223e-01) (48, -3.59923601150512695312e-01) (49,
 3.58579397201538085938e-01) (50, 7.14190721511840820312e-01) (51, 4.72858250141143798828e-01) (52,
 1.24634966254234313965e-01) (53, 3.10299992561340332031e-01) (54, -6.57229125499725341797e-01) (55,
 1.94391570985317230225e-02) (56, 2.27676913142204284668e-01) (57, 9.70135703682899475098e-02) (58,
 1.28397434949874877930e-01) (59, -2.52493411302566528320e-01) (60, -3.43879133462905883789e-01) (61, -
 3.91250252723693847656e-02) (62, -8.58102142810821533203e-01) (63, -3.11073392629623413086e-01) (64, -
 2.59148597717285156250e-01) (65, -4.48308318853378295898e-01) (66, -4.31667506694793701172e-01) (67, -
 2.35764726996421813965e-01) (68, 2.81650662422180175781e-01) (69, 4.26426045596599578857e-02) (70, -
 3.46176743507385253906e-01) (71, -1.64830178022384643555e-01) (72, -6.25177547335624694824e-02) (73,
 6.87909349799156188965e-02) (74, 1.20926328003406524658e-01) (75, 4.97217714786529541016e-01) (76,
 5.18955945968627929688e-01) (77, -6.08848750591278076172e-01) (78, -5.24905286729335784912e-02) (79,
 2.64033734798431396484e-01) (80, -4.73694175481796264648e-01) (81, -1.83808520436286926270e-01) (82,
 3.86276282370090484619e-02) (83, 4.44766759872436523438e-01) (84, 1.23103305697441101074e-01) (85, -
 3.18655818700790405273e-01) (86, -4.33914840221405029297e-01) (87, -1.57840237021446228027e-01) (88, -
 2.74147535674273967743e-03) (89, 5.60027778148651123047e-01) (90, 3.00918556749820709229e-02) (91,
 5.91206133365631103516e-01) (92, -5.78729152679443359375e-01) (93, -2.10507065057754516602e-01) (94, -
 1.36310786008834838867e-01) (95, 2.59455680847167968750e-01) (96, 5.04221081733703613281e-01) (97,
 3.75474125146865844727e-01) (98, 2.04642459750175476074e-01) (99, 8.38201567530632019043e-02) (100,
 1.40254765748977661133e-01) (101, 2.11245298385620117188e+00) (102, 2.28825956583023071289e-01) (103,
 4.52963374555110931396e-02) (104, 2.68477231264114379883e-01) (105, -1.72276079654693603516e-01) (106, -
 6.99033141136169433594e-01) (107, -3.31750452518463134766e-01) (108, 3.24707478284835815430e-04) (109,
 1.66931852698326110840e-01) (110, 4.08844023942947387695e-01) (111, 1.18498861789703369141e+00) (112, -
 9.09749746322631835938e-01) (113, -3.31096738576889038086e-01) (114, -1.78518705070018768311e-02) (115,
 1.01937755942344665527e-01) (116, 4.48576688766479492188e-01) (117, 1.59204438328742980957e-01) (118,
 3.94143849611282348633e-01) (119, -1.35678857564926147461e-01) (120, -1.49015545845031738281e-01) (121, -
 2.60443478822708129883e-01)