**AN INTEGRATED HYBRID DATA MINING SYSTEM FOR PRETERM BIRTH
RISK ASSESSMENT BASED ON A
'SEMANTIC WEB SERVICES FOR HEALTHCARE' FRAMEWORK**

by

Christina Anne Catley, M.A.Sc., B.Eng.

A thesis submitted to
The Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

IN ELECTRICAL ENGINEERING

Ottawa-Carleton Institute of Electrical and Computer Engineering
Department of Systems and Computer Engineering
Carleton University, Ottawa, Ontario, Canada

© Christina Anne Catley, 2007

Library and
Archives Canada

Bibliothèque et
Archives Canada

Published Heritage
Branch

Direction du
Patrimoine de l'édition

395 Wellington Street
Ottawa ON K1A 0N4
Canada

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

ISBN: 978-0-494-27090-5

Our file Notre référence

ISBN: 978-0-494-27090-5

NOTICE:

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

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

AVIS:

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

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

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

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

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

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

Canada

ABSTRACT

A re-engineered approach to the early prediction of preterm birth is presented as a complimentary technique to the current procedure of using costly and invasive clinical testing on high-risk maternal populations only. The preterm birth risk assessment tool is based on an integrated hybrid data mining screening system developed using a combination of: Artificial Neural Networks (ANNs), specifically a classification-based Multilayer Perceptron (MLP) ANN and a MLP ANN with risk stratification; case based reasoning; and a decision tree voting algorithm. In developing the integrated hybrid data mining system, the classification-based MLP ANN was employed as an outcome prediction tool for estimating preterm birth and high-risk preterm birth on a heterogeneous maternal population; risk estimations use obstetrical variables available to physicians prior to 23 weeks gestation. The trained hybrid system yields a sensitivity of 65% and a specificity of 84%, meeting the clinical criteria determined by collaborating physicians; the system is successfully validated on 9701 new patient cases.

The developed tool is part of a larger goal: to offer perinatal risk assessment to physicians, researchers, and patients via a 'Semantic Web service for healthcare' framework. The framework is based on Semantic Web technology; a diagnostic clinical decision support ontology for perinatal care is created, attaching meaning to all resources. Candidate Web services are defined and service interactions are specified through composition templates, the service integration architecture is proposed.

This is for my boys.

To my husband, Todd, you are ‘my rock’. I could never have completed this without your encouragement, support and belief in me.

And to our son, Malcolm, you are a delight and joy.

Thank you for bringing so much laughter to our lives
(as well as a massive incentive to finish this thesis so that
‘Mommy’ won’t always be working at her ‘om-uter’).

ACKNOWLEDGEMENTS

My deepest thanks to my co-supervisors, Dr. Monique Frize and Dr. Dorina Petriu. As well as being excellent supervisors, you are also amazing mentors. To Dr. Frize, thank you for encouraging me in every way over the past six years, I feel honored to have worked with you throughout the course of my graduate studies. To Dr. Petriu, thank you for being a sounding board for my thesis concerns and for always being available despite your hectic schedule.

I am grateful to Dr. Robin Walker for your clinical insights and knowledge of perinatal care; it was wonderful learning from you. Visiting the neonatal intensive care unit and seeing the babies who would benefit from this research made this problem very tangible.

Many thanks to my thesis committee: Dr. Tet Yeap, Dr. James Green, and Dr. Peter Liu, for your invaluable suggestions and constructive criticism.

I am very appreciative of the scholarships I have received from the Natural Sciences and Engineering Research Council of Canada, the Ontario Graduate Scholarships in Science and Technology program (and Dr. Petriu's financial contribution to this scholarship), Carleton University, and the Canadian Engineering Memorial Foundation.

Thank you a hundred times over to Bruce and Gail Catley for your hours of babysitting; you have been the most supportive parents and loving grandparents.

TABLE OF CONTENTS

ABSTRACT	II
ACKNOWLEDGEMENTS	IV
TABLE OF CONTENTS	V
LIST OF TABLES	X
LIST OF FIGURES	XII
GLOSSARY OF TERMS AND ACRONYMS	XIV
CHAPTER 1. INTRODUCTION	1
1.1 MOTIVATION	1
1.2 PROBLEM STATEMENT	3
1.2.1 PROBLEM #1: DEVELOPING AN INTEGRATED HYBRID DATA MINING SCREENING SYSTEM FOR PTB RISK ASSESSMENT	3
1.2.2 PROBLEM #2: DEVELOPING A SEMANTIC WEB SERVICES-BASED FRAMEWORK FOR PERINATAL CLINICAL DECISION SUPPORT	4
1.2.3 COHESIVE DESIGN OVERVIEW: LINKING PROBLEM #1 AND PROBLEM #2	6
1.3 HYPOTHESIS	9
1.4 SIGNIFICANCE OF THE RESEARCH	10
1.5 THESIS OUTLINE	11
CHAPTER 2. OVERVIEW OF RELEVANT MEDICAL AND TECHNOLOGICAL CONCEPTS	13
2.1 EARLY PREDICTION OF PRETERM BIRTH	13
2.1.1 RISK MODEL FOR PRETERM BIRTH	14
2.1.1.1 Higher Order Births	15
2.1.1.2 Previous Preterm Birth	15
2.1.1.3 Uterine Abnormalities	16
2.1.1.4 Other Risk Factors	16
2.2 ARTIFICIAL INTELLIGENCE AND DATA MINING IN MEDICINE	17
2.2.1 ARTIFICIAL NEURAL NETWORKS (ANN)	18
2.2.1.1 ANN Architecture	19
2.2.1.2 ANN Learning	21
2.2.1.3 Training, Testing and Verification Sets	22
2.2.1.4 Advantages of Neural Networks	23
2.2.1.5 Prior Uses of ANN within MIRG	24
2.2.2 ANN RESEARCH FRAMEWORK	25
2.2.3 ADDING RISK STRATIFICATION TO THE MLP ANN	26
2.2.3.1 Summary of Zhou and Frize's Approach Combining Maximum Likelihood Estimation and Neural Network Learning	28
2.2.4 CASE-BASED REASONING (CBR) – K-NEAREST NEIGHBOURS (KNN) ALGORITHM	29

	vi
2.2.5 DECISION TREE	31
2.2.6 PARENTS DECISION SUPPORT (PADS)	31
2.3 COLLABORATIVE DATA MINING METHODS	32
2.3.1 PARALLEL SYSTEMS	32
2.3.2 HYBRID (SERIES) SYSTEMS	33
2.4 SEMANTIC WEB SERVICES FOR HEALTHCARE – OVERVIEW OF KEY CONCEPTS	34
2.4.1 WEB SERVICES	35
2.4.1.1 Web Services – Overview	35
2.4.1.2 Web Services – Detailed Discussion	37
2.4.1.2.1 XML	39
2.4.1.2.2 SOAP	39
2.4.2 THE SEMANTIC WEB	40
2.4.2.1 Semantic Web – Overview	40
2.4.2.2 Semantic Web – Detailed Discussion	41
2.4.2.2.1 Knowledge Representation	42
2.4.2.2.2 Ontologies	42
2.4.2.2.3 Agents	43
2.4.3 SEMANTIC WEB SERVICES	44
CHAPTER 3. CRITICAL ANALYSIS	45
3.1 INTRODUCTION	45
3.2 EVIDENCE THAT PROBLEM #1 HAS NOT PREVIOUSLY BEEN SOLVED	45
3.2.1 LIMITATIONS OF CURRENT PREDICTION APPROACHES	45
3.2.1.1 Risk-Scoring Models	46
3.2.1.2 Invasive Diagnostics	47
3.2.1.3 Information Technology-Based Approaches	50
3.2.1.3.1 Lockwood and Kuczynski	51
3.2.1.3.2 Goodwin and Mayer	52
3.2.1.3.3 Van Dyne, Woolery, Gryzmala-Busse, Tsatsoulis	53
3.2.2 OTHER HYBRID SYSTEMS USED IN CLINICAL OUTCOME PREDICTION	56
3.2.2.1 Hybrid Systems Based on Combining ANNs and CBRs	56
3.2.2.2 Hybrid Systems Based on Combining (ANNs (and/or) CBRs) with Decision Trees	58
3.2.2.2.1 Decision Tree -> Neural Network Combinations	58
3.2.2.2.2 Decision Tree <-> Neural network Combinations	59
3.2.2.2.3 Neural Network -> Decision Tree Combinations	61
3.2.2.2.4 Decision Tree -> kNN	62
3.3 VALUE OF DEVELOPING A NEW HYBRID DATA MINING APPROACH FOR PREDICTING PTB	62
3.3.1 OTHER APPLICATIONS	63
3.4 EVIDENCE THAT PROBLEM #2 HAS NOT PREVIOUSLY BEEN SOLVED	64
3.4.1 OVERVIEW OF STATE OF THE ART AND CRITICAL ANALYSIS	64
3.4.1.1 State of the Art in Semantic Web Service-based Systems	65
3.4.1.1.1 Semantic Web Services for Medical Service Flows	66
3.4.1.1.2 Semantic Web Services for Clinical Decision Support	67
3.4.1.1.3 Web Service Composition on the Semantic Web in e-government	68
3.4.1.1.4 Web-Based Decision Support for Personalized Risk Assessment	68
3.4.1.2 State of the Art in Clinical Ontologies	69
3.5 VALUE OF DEVELOPING A SEMANTIC WEB SERVICES-BASED FRAMEWORK FOR HEALTHCARE	71

CHAPTER 4. METHODOLOGY FOR DEVELOPING THE HIGH-RISK PRETERM BIRTH INTEGRATED HYBRID DATA MINING SYSTEM	74
4.1 INTRODUCTION	74
4.2 MEDICAL DATABASES	74
4.2.1 PERINATAL PARTNERSHIP PROGRAM OF EASTERN AND SOUTHEASTERN ONTARIO (PPESO) DATABASE	75
4.2.2 PREGNANCY RISK ASSESSMENT MONITORING SYSTEM (PRAMS) DATABASE	76
4.2.3 RISK FACTORS AVAILABLE IN PPESO AND PRAMS DATABASES	77
4.3 ARCHITECTURAL OVERVIEW OF THE INTEGRATED HYBRID DATA MINING SYSTEM	79
4.4 OVERVIEW OF PERFORMANCE MEASURES USED IN THIS WORK	81
4.4.1 CONTINGENCY TABLE	82
4.4.2 CORRECT CLASSIFICATION RATE	83
4.4.3 CONSTANT PREDICTOR	84
4.4.4 SENSITIVITY AND SPECIFICITY	84
4.4.5 LOGARITHMIC SENSITIVITY INDEX	84
4.4.5.1 Receiver Operating Characteristic (ROC) Curves	85
4.5 ANN DETAILS	86
4.5.1 NETWORK PARAMETERS AND SOFTWARE EXECUTION	88
4.5.2 WEIGHT ELIMINATION COST FUNCTION	89
4.5.3 FEATURE SELECTION	90
4.5.3.1 Variable Reduction Using the Modified Relative Weights Method	91
4.6 DEVELOPING THE PPESO-BASED HIGH-RISK PTB MODEL	95
4.6.1 PPESO PERINATAL DATABASE	95
4.6.2 CREATING THE ARTIFICIAL DATASETS	97
4.6.3 DEFINING THE HIGH-RISK PRETERM BIRTH MODEL: REFINING THE OUTPUT	99
4.6.4 ADJUSTABLE NETWORK PARAMETERS	104
4.6.5 VERIFICATION	105
4.7 DEVELOPING THE PRAMS-BASED HIGH-RISK PTB MODEL	106
4.7.1 PRAMS PERINATAL DATABASE	107
4.7.2 PROCESSING STEPS	112
4.7.3 VARIABLE SIGNIFICANCE IN THE HIGH-RISK PRETERM BIRTH DATASET	113
4.7.4 ADJUSTABLE NETWORK PARAMETERS	114
4.8 DEVELOPING THE HYBRID SCREENING SYSTEMS FOR PERINATAL RISK ASSESSMENT (PTB-SPECIFIC)	115
4.8.1 PERFORMANCE MEASURES: ENSURING CONFIDENCE IN THE MODEL	116
4.8.2 DISTINGUISHING BETWEEN TRUE AND FALSE POSITIVES	119
4.8.3 INCORPORATING MLP ANN-BASED RISK STRATIFICATION ESTIMATIONS INTO THE INTEGRATED HYBRID SCREENING SYSTEM	120
4.8.4 LINKING THE CLASSIFICATION-BASED MLP ANN WITH THE RISK STRATIFICATION ESTIMATION-BASED MLP ANN	121
4.8.4.1 Inconsistent Predictions	126
4.8.5 INCORPORATING A KNN CASE-BASED REASONING TOOL TO ASSESS INCONSISTENT PREDICTED OUTPUTS	127
4.8.5.1 Using ANN Weights for the kNN	127
4.8.5.2 Outcome Prediction Using the k-Nearest Neighbors (kNN) Case-Based Reasoning Tool and the Decision Tree Voting Algorithm	129
4.8.6 VERIFICATION: USING INTEGRATED SCREENING TO PREDICT OUTCOMES FOR NEW PATIENT CASES	134

CHAPTER 5. RESULTS**139**

5.1	INTRODUCTION	139
5.2	PPESO CLASSIFICATION-BASED MLP ANN RESULTS	139
5.2.1	PREDICTING PRETERM BIRTH	139
5.2.2	PREDICTING HIGH-RISK PRETERM BIRTH	141
5.2.3	PPESO VERIFICATION SET RESULTS	143
5.3	PRAMS CLASSIFICATION-BASED MLP ANN RESULTS	144
5.3.1	PREDICTING PTB USING THE 7-VARIABLE PRAMS MODEL	145
5.3.2	PREDICTING PTB USING THE 48-VARIABLE PRAMS MODEL	146
5.3.3	PREDICTING HIGH-RISK PRETERM BIRTH USING THE 48-VARIABLE PRAMS MODEL	147
5.3.4	ASSIGNING RELATIVE WEIGHTINGS (SIGNIFICANCE) TO THE PRAMS 48 VARIABLE HIGH-RISK PTB MODEL	149
5.3.5	PRAMS VERIFICATION SET RESULTS	155
5.4	HYBRID SYSTEM RESULTS	156
5.4.1	PASS 1: RESULTS	156
5.4.1.1	Validation of Risk Stratification Estimation Model – Maximizing for Sensitivity	156
5.4.1.2	Setting the cutoff point: Patient-specific Risk Stratification Outcomes	158
5.4.2	PASS 2: RESULTS	161
5.4.2.1	Validation of Risk Stratification Estimation Model – Maximizing for Specificity	161
5.4.2.2	Setting the cutoff point: Patient-specific Risk Stratification Outcomes	163
5.4.3	PASS 3: RESULTS	166
5.4.3.1	Re-assigning conflicting cases: implementing the decision-tree voting algorithm	167
5.4.3.2	Analysis of Results	174
5.4.4	VERIFICATION ON NEW PATIENT CASES	176
5.4.5	VALIDITY OF RESULTS	178

**CHAPTER 6. DESIGN OF A SEMANTIC WEB SERVICES-BASED
FRAMEWORK FOR PERINATAL CLINICAL DECISION SUPPORT**

180

6.1	INTRODUCTION	180
6.2	DESIGN STEP 1: IDENTIFYING CANDIDATE SERVICES	181
6.2.1	COMPOSITE SERVICES FOR PERINATAL CLINICAL DECISION SUPPORT	181
6.2.2	CORE SERVICES FOR PERINATAL CLINICAL DECISION SUPPORT	184
6.2.2.1	Core Data Mining Services	184
6.2.2.2	Interfacing Services	187
6.3	DESIGN STEP 2: DEFINING SERVICE COMPOSITIONS	187
6.3.1	FUNCTIONAL SYSTEM VIEW	188
6.3.1.1	Diagnostic Clinical Decision Support Interactions	189
6.3.1.2	Researcher Interactions	191
6.3.1.3	Semantic Web Services for Healthcare Interactions	193
6.3.2	SERVICE INTEGRATION ARCHITECTURE	195
6.3.2.1	Overview	195
6.3.2.2	Logical System Model	197
6.4	DESIGN STEP 3: CREATING THE PERINATAL CLINICAL DECISION SUPPORT ONTOLOGY	200
6.4.1	KNOWLEDGE (ONTOLOGY) MODELING WITH PROTÉGÉ	200
6.4.2	METHODOLOGY	202
6.4.3	DIAGNOSTIC DECISION SUPPORT ONTOLOGY: TOP-LEVEL VIEW	203

	ix
6.4.4 DECISION SUPPORT COMPONENT	205
6.4.5 CLINICAL DIAGNOSTIC DOMAIN COMPONENT	209
6.4.6 BUSINESS PROCESS WORKFLOWS	212
6.5 ONGOING CHALLENGES IN DEVELOPING SEMANTIC WEB SERVICES-BASED APPLICATIONS	216
6.5.1 SECURITY	216
6.5.2 PRIVACY	217
<u>CHAPTER 7. CONCLUDING REMARKS</u>	<u>218</u>
7.1 CONCLUSIONS	218
7.2 ACADEMIC CONTRIBUTIONS	221
7.3 FUTURE WORK & RECOMMENDATIONS	223
<u>APPENDIX A – LIST OF PUBLICATIONS FOR CHRISTINA CATLEY</u>	<u>228</u>
<u>APPENDIX B - MODELING COST FUNCTIONS AND DERIVING A TRAINING ALGORITHM</u>	<u>231</u>
<u>APPENDIX C – WEIGHTS & BIASES USED FOR NEW PATIENT CASES</u>	<u>232</u>
<u>APPENDIX D – BIVARIATE CORRELATIONS (PEARSON)</u>	<u>234</u>
<u>REFERENCES</u>	<u>240</u>

LIST OF TABLES

Table 3-1 Waegemann's Standards Areas for Health Informatics	71
Table 4-1 Perinatal Risk Factors Compiled from the Literature.....	77
Table 4-2 Contingency Table.....	82
Table 4-3 Evaluating the Effectiveness of ROC Curves	86
Table 4-4 PPESO Preterm Birth Statistics by Year.....	97
Table 4-5 High-risk PTB Statistics by Year	104
Table 4-6 Network Settings of the Best Performing Double Layer Networks for Predicting PTB, High-Risk PTB and High-Risk PTB Excluding Free Flow Oxygen Cases (PPESO Dataset).....	105
Table 4-7 PRAMS Variables – Prevalence and Statistics	108
Table 4-8 PRAMS PTB Statistics For PPESO 8-Variable Model	112
Table 4-9 Network Settings of the Best Performing Double Layer Networks for Predicting PTB and High-risk PTB Using Prams Dataset.....	115
Table 4-10 Clinical Implications of Prediction Results for Preterm Birth	117
Table 4-11 Logic for Determining Conflicting Cases	127
Table 5-1 Performance Measures of the Best-Performing PPESO-Based Double Layer Networks For Predicting PTB.....	141
Table 5-2 Performance Measures of the Three Best-Performing PPESO-Based Double Layer Networks For Predicting High-risk PTB and High-risk PTB Excluding Free-Flow Oxygen Cases.....	142
Table 5-3 Verification Set Results for all Risk Models (PPESO).....	144
Table 5-4 Test Set Performance Results: Predicting PTB Using the.....	146
Table 5-5 Test Set Performance Results: Predicting Preterm Birth Using PRAMS 48- Variable Model	147
Table 5-6 Performance Results: Predicting High-Risk PTB Using the Prams 48-Variable Model.....	149
Table 5-7 Modified Relative Weightings Using Averaging	150
Table 5-8 American-Specific PRAMS Variables and Possible Canadian Equivalents..	154
Table 5-9 PRAMS 48-variable High-Risk PTB Verification Set results	156
Table 5-10 PRAMS High-Risk PTB Performance Criteria for 'Pass 1'	157
Table 5-11 Risk stratification Estimation ANN Performance	157

Table 5-12 Percentage of Test Cases Assigned to Each Risk Stratification Level for Both Low Risk and High-risk PTBs Based on a Cutoff of 0.4.....	161
Table 5-13 PRAMS High-Risk PTB Performance Criteria for 'Pass 2'	162
Table 5-14 Risk Stratification Estimation ANN Performance Results.....	162
Table 5-15 Percentage of Test Cases Assigned to Each Risk Stratification Level for Both Low Risk and High-risk PTBs Based on a Cutoff of 0.6.....	164
Table 5-16 PRAMS High-Risk PTB Performance Criteria for Pass 3	167
Table 5-17 Final Sensitivity and Specificity Calculations ($k=9$).....	173
Table 5-18 Number of Conflicting Verification Cases After Completion of 'Pass 2' ...	177
Table 5-19 Re-Assigning Conflicting Cases Following Completion of 'Pass 3'	177

LIST OF FIGURES

Figure 1-1 Cohesive Design Overview (Linking Problem #1 and Problem #2).....	8
Figure 2-1 Diagrammatic Representation of a Single Node in A Multilayer Percetron (Ennett, 2003)	20
Figure 2-2 Architecture of the Multilayer Perceptron (Ennett, 2003)	21
Figure 3-1 Perinatal Clinical Outcomes of Interest	64
Figure 4-1 Architectural Overview of Integrated Hybrid Screening System	81
Figure 4-2 Flow of Control for Defining the High-Risk Preterm Birth Prediction Model 103	
Figure 4-3 Flowchart (UML activity Diagram) Overview of the Hybrid Screening Model (‘Pass 1’ and ‘Pass 2’).....	126
Figure 4-4 Flow of Control for Training Hybrid ANN, kNN and Voting Algorithm System.....	133
Figure 4-5 Screenshot of kNN-Based CBR System using ‘AutoMatch’ Function	134
Figure 4-6 Flow of Control for Verifying Integrated Hybrid Screening System	137
Figure 5-1 Risk Stratification Estimation Model for	159
Figure 5-2 Risk Stratification Estimation Model for	163
Figure 5-3 Example ‘Pass 1’ and ‘Pass 2’ classification and Input to ‘Pass 3’	166
Figure 5-4 Summary of Results	175
Figure 5-5 Results for k Versus Performance (Cutoff 0.11)	176
Figure 6-1 Diagnostic Clinical Decision Support Interactions Use Case Diagram	191
Figure 6-2 Researcher Interactions Use Case Diagram	193
Figure 6-3 Semantic Web Services for Healthcare Interactions Use Case Diagram	195
Figure 6-4 System Architecture	197
Figure 6-5 DiagnosticDecisionSupport subclass tree - top level view	204
Figure 6-6 Protégé-OWL Subclass view of DecisionEnvelope class	205
Figure 6-7 Major components of DecisionSupport subclass tree	206
Figure 6-8 Applicability expanded subclass tree	207
Figure 6-9 Purpose expanded subclass tree	208
Figure 6-10 Classification expanded Subclass tree	208
Figure 6-11 Collaboration->Classification expanded Subclass Tree.....	208
Figure 6-12 Major Components of the Clinical Diagnostic Domain Subclass Tree	209

	xiii
Figure 6-13 Decision Problem Expanded Subclass Tree.....	211
Figure 6-14 Decision Output Expanded Subclass Tree	212
Figure 6-15 Semantic web Services Compositions for ‘Training Risk Assessment’	214
Figure 6-16 Semantic web Services Compositions for ‘Classification-based Outcome Prediction’	215

GLOSSARY OF TERMS AND ACRONYMS

Actor: In UML terminology an actor is an outside user who interacts with the system.

AI: Artificial Intelligence

ANN: Artificial Neural Network

APGAR: Appearance; Pulse; Grimace; Activity; and Respiration

Architecture: The set of significant decisions about the organisation of a software system, the selection of the structural elements and their interfaces by which the system is composed together with their behaviour [Jacobson *et al.* 1999].

ASE: Average Squared Error

ASTM: A standards development organization that serves as an open forum for the development of international standards.

CBR: Case-Based Reasoner

CCR: Correct Classification Rate

CDSS: Clinical Decision Support System

Class: A UML term to describe objects that share the same operations, attributes and relationships. Classes can be thought of as the blueprint for the object.

CNN: Canadian Neonatal Network

DCOM: Distributed Component Object Model

DICOM: Digital Imaging and Communications in Medicine

DL: Description Logics

DT: Decision Tree

DTD: Document Type Definition

EPR: Electronic Patient Record

FN: False Negative

FP: False Positive

Framework: Defines concepts, organisational structures, architectural components and ideas that are considered most relevant in achieving a vision.

GEM: Guidelines Element Model

HIS: Hospital Information System

HL7: Health Level 7

Integration: Implies both semantic and technical interoperability, as well as a logical integration flow between interoperable modules.

Interoperability: The ability of two or more systems or components to exchange information and to use the information that has been exchanged.

ISO (The International Organization for Standardization): A body dedicated to the production of standards through international agreement.

Medical Informatics: Medical Informatics combines the disciplines of medicine and information technology to support the healthcare delivery process and improve the quality of care delivered, through initiatives such as CDSSs.

MIRG: Medical Information technologies Research Group

MLP: MultiLayer Perceptron

NICU: Neonatal Intensive Care Unit

Object: A concrete realisation of a class; an object is often referred to as an instantiation of a class.

ODBC: Open Database Connectivity

Ontology: Defines vocabularies of representation terms, this can refer to the ontological commitment to using shared vocabularies.

OWL: Web Ontology Language

PPESO: Perinatal Partnership Program of Eastern and Southeastern Ontario

PRAMS: Pregnancy Risk Assessment Monitoring System

PTB: Preterm Birth

RDF: Resource Description Framework

Registry: A facility that stores relevant descriptive information about registered objects, where a registered object is something important that an author or producer wants to have visible to the world so that it can be discovered and used by a customer or client. A registry can operate independently or it can be paired with a repository that stores the registered objects [Gallagher & Carnahan 2000].

Repository: A repository is simply a storage facility for registered objects, with an access method allowing one to retrieve individual objects by an object reference [Gallagher & Carnahan 2000].

RMI: Remote Method Invocation

ROC: Receiver Operating Characteristic

RPC (Remote Procedure Call): A procedure call in which the calling procedure and the called procedure do not share an address space and may be in physically different locations.

Schema: A schema defines the elements that can appear within a document and the attributes that can be associated with an element, as well as defining the document's structure.

Semantic Web: The next generation Internet, where semantic meaning will be associated with all Web resources, both data and services, such that they can be discovered and consumed autonomously by applications.

Semantic Web service: Using Semantic Web technology to describe Web service's capabilities and content in an unambiguous computer-interpretable language (McIlraith & Martin, 2003).

Server: A role filled by a processing element when it provides services to another processor.

SOAP: Simple Object Access Protocol

TN: True Negative

TP: True Positive

UDDI: Universal Description Discovery and Integration

UML (Unified Modeling Language): A language for visualising, specifying, constructing and documenting the elements of a system.

UMLS: Unified Medical Language System

Use Case: A UML term describing interactions between actors and the system; the complete set of use cases describes the system functionality.

W3C: World Wide Web Consortium

Web Services: Software applications that can be discovered, described, and accessed based on XML and standard Web protocols over intranets, extranets, and the Internet (Daconta et al., 2003).

WIGIT: Wherever I Go, It's There

World Wide Web Consortium (W3C): A forum for information, commerce, communication, and collective understanding, which develops open Web standards and guidelines with the goal of achieving Web interoperability.

WSDL: Web Services Description Language

XML (eXtensible Markup Language): A standard from the World Wide Web Consortium (W3C) that provides tagging of information content within documents. XML offers a means for representing content in a format that is both human and machine readable.

XSLT (extended Stylesheet Language Transformations): Describes a vocabulary recognized by an XSLT processor to transform information from the format specified in the source file into a different format. Simply put, XSLT takes XML content and transforms it into any other format for which an applicable transform has been created.

“Healthcare IT advances are pulling together to manage an expanding universe.”

February 2005 - *Healthcare Informatics Online*

CHAPTER 1. INTRODUCTION

1.1 MOTIVATION

Clinical data is growing at an exponential rate; real-time data repositories acquire data around the clock from monitoring devices, and numerous medical database initiatives exist with the purpose of gathering and storing clinical data related to specialized healthcare domains. Unfortunately, the tools available to clinicians to make effective use of this data, by integrating and sharing, drawing inferences, identifying patterns, extracting similarities, and classifying data, have not kept pace with the ability to acquire data. For this reason, the health informatics environment is considered ‘data-rich’ but ‘knowledge-poor’. While computer applications have been very effective in enabling clinical data acquisition, they have been less successful in helping physicians translate this data into knowledge useful in clinical decision-making activities, such as estimating clinical outcomes and performing patient risk assessment. This is reasonable considering that decision-making is a complex and specialized process requiring that medical practitioners integrate diverse knowledge modalities (Abidi, 2005) before arriving at a final decision. Examples of medical knowledge modalities include a practitioner’s clinical experiences, medical literature, clinical practice guidelines, and clinical data mining applications.

Integration is a serious issue for the healthcare domain, which is defined by a “plethora of distributed data and knowledge, from which complex and timely high-value decisions must be made in a low-tolerance environment” (Turner et al., 2004). Wreder

and Ding elaborate “How to migrate from stovepipe systems to the next generation open healthcare information systems that are interoperable, extensible and maintainable is increasingly a pressing problem for the healthcare industry” (Wreder & Deng, 1999). The Semantic Web and Web services are two emerging and evolving technologies that information technology experts predict will change the face of healthcare delivery (Hoffman, 2003; Lea & Vinoski, 2003), ultimately providing integrated healthcare data, knowledge, and computer-based applications. To date, deployed Semantic Web services in the healthcare arena are rare and those that do exist are limited in complexity, however the potential is apparent. Nevertheless, many technical problems remain, particularly in complex multi-disciplinary areas such as autonomous service composition and semantic ontology-based knowledge representation.

From a medical perspective, this work explores the data-rich domain of perinatal care, focusing on the early prediction of Preterm Birth (PTB), defined as birth before 37 completed weeks gestation. Perinatal care is described as the care of a fetus or newborn before, during, and after delivery from the 28th week of gestation through the 7th day after delivery (MedicalGlossary.org, 2004). Information collected during the perinatal period spans obstetrical, postpartum, and potentially neonatal intensive care data; in all, a large quantity of clinical data relating to both the infant and mother is accumulated and stored in distributed medical databases.

1.2 PROBLEM STATEMENT

1.2.1 PROBLEM #1: DEVELOPING AN INTEGRATED HYBRID DATA MINING SCREENING SYSTEM FOR PTB RISK ASSESSMENT

The first contribution of this thesis is to develop a hybrid data mining system, based on Artificial Intelligence (AI) techniques, and demonstrate the applicability of this system towards estimating the multifactorial phenomenon of preterm birth (PTB). The goal is to develop, with physician support, a re-engineered approach to the early prediction of *high-risk* preterm birth, where high-risk PTB is defined as either: (1) birth at less than 33 weeks gestation; or (2) births between 33-36 weeks gestation that include an additional risk factor (risk factors will be discussed in detail in *Chapter 4*). The AI-based integrated hybrid system screens mothers for risk of PTB; screening is based on heterogeneous maternal populations.

To be considered clinically useful in predicting outcomes for new patient cases, the hybrid system must be able to predict preterm birth using variables available to physicians prior to 23 weeks gestation (23 weeks being the earliest age at which survival is feasible). To be broadly applicable, the system should be low-cost and easy to use, and be relevant to the entire maternal population, rather than to a subset of pre-selected mothers having an elevated risk of PTB. In terms of performance, the hybrid system developed in this thesis should yield a sensitivity approaching 64.7% (the current best sensitivity achieved using invasive techniques), and a specificity approaching 85%. These threshold values were determined in collaboration with physicians and will be discussed in more detail in *Chapter 4*.

1.2.2 PROBLEM #2: DEVELOPING A SEMANTIC WEB SERVICES-BASED FRAMEWORK FOR PERINATAL CLINICAL DECISION SUPPORT

The second thesis contribution involves the conception and design of a unifying framework for composing clinical decision support services relating to the perinatal life cycle. The objective of the clinical decision support services integrated in this work is to provide improved early prediction of high-risk preterm birth on a heterogeneous maternal population. This work presents the design of a Semantic Web services framework and offers a strong case study in which to deploy and evaluate the Semantic Web services paradigm.

The Semantic Web services for healthcare framework will define the integration of clinical decision support applications for perinatal risk assessment. The framework will formalize the practitioner's clinical decision support workflow for estimating PTB. The workflows will be modeled as composition scenarios and supported by the diagnostic clinical decision support ontology. This will define the integration of clinical decision support services and perinatal data using Semantic Web services, resulting in more timely and accurate predictions of high-risk obstetrical outcomes at the point of care. In this way, the framework's composition scenarios will model physician-oriented clinical decision support workflows and will serve as a generic model for diagnostic clinical decision support. Once developed, the framework's preterm birth clinical decision support ontology will be generalizable to other perinatal applications requiring diagnostic clinical decision support (e.g. predicting infant mortality, APGAR score, and delivery type).

Using the concepts of 'Data as a Service' and 'Software as a Service' (Turner et

al., 2003), the framework defines the integration of clinical data and clinical decision support tools in a distributed healthcare environment. The objective is to compose Semantic Web services based on pre-defined physician service composition templates. The composition templates model workflows for perinatal decision support; by formulating the physician's decision-making process it is possible to alert physicians to possible high-risk obstetrical outcomes.

Current research initiatives within the Medical Information technologies Research Group¹ (MIRG) are focused on developing Clinical Decision Support Systems (CDSSs) to aid physician decision-making in high-risk medical environments. The novel hybrid data mining system developed in this work focuses on the combination of Artificial Neural Networks (ANNs) for outcome estimation; and a Case-Based Reasoning (CBR)-Decision Tree (DT) application for outcome refinement (i.e. further improving the sensitivity). These applications are key components of the 'Software as a Service' element of the framework.

Determining appropriate composition scenarios (i.e. specifying how different Web services are combined to produce a useful outcome. In Web services terminology, this information is provided in composition templates) is essential to the framework's development.

The term *semantic* in 'Semantic Web services' implies meaning has been associated to the Web resources; this was accomplished by creating an ontology covering all relevant perinatal care and clinical decision support terminology, essentially a

¹ MIRG Principal investigators: Dr. Monique Frize, P.Eng (Carleton University and University of Ottawa, Ottawa, ON) and Dr. C.R. Walker, MBBS, FRCPC (IWK Health Centre, Halifax, NS)

predefined taxonomy for perinatal clinical decision support.

1.2.3 COHESIVE DESIGN OVERVIEW: LINKING PROBLEM #1 AND PROBLEM #2

Problem #1 and Problem #2 are combined to offer the integrated hybrid screening system for perinatal risk assessment as part of a Semantic Web services for healthcare framework. This integration is explained by examining three key concepts, described below and depicted as three layers in Figure 1-1: (1) *Knowledge Description*; (2) *Clinical Decision Support*; and (3) *Knowledge Integration: Semantic Web Services for Healthcare*. Figure 1-1 can be thought of as a sandwich, where the two outer layers enable and implement the Semantic Web services framework, and the middle layer offers the core clinical decision support functionality offered via the framework.

Layer 1 (Knowledge Description) encompasses all knowledge sources required by the other layers- specifically: the domain vocabulary of the clinical databases; knowledge of clinical decision support systems and their interactions; and the practitioner's tacit domain knowledge, which adds the pragmatics of invocation surrounding clinical decision support services (i.e. the: how, when, why, where, and for whom of service invocation). Each knowledge source must be formalized before being invoked by a Semantic Web service: the Perinatal Clinical Decision Support (CDS) Ontology describes the domain vocabulary, knowledge, and pragmatics, while the CDS Composition Templates model the manner in which services are composed.

Layer 2 (Clinical Decision Support) involves developing a trained perinatal PTB risk assessment model. The resulting risk model is then integrated with the 'Hybrid Data Mining Workflow Model for Perinatal Outcome Estimation', which defines how the

clinical decision support systems interact within the overall hybrid system. The result is an integrated clinical decision support service for PTB risk assessment. It is important to understand that although this thesis focuses on PTB risk assessment, if the Health Informatician were to develop a different risk model (e.g. to predict delivery type or infant mortality), then a different integrated clinical decision support risk assessment application could be instantiated based on the same hybrid workflow model.

Layer 3 (Knowledge Integration: Semantic Web Services for Healthcare) provides the Semantic Web services framework, which is the ‘glue’ that integrates the other layers and the means by which the practitioner accesses the Integrated CDS tool for PTB Risk Assessment. The practitioner inputs new patient case data via the Semantic Web services for Healthcare framework, performs risk assessment, and receives a result. The result is evaluated and ideally stored in the patient’s Electronic Patient Record (EPR). Ongoing discussion with the patient occurs.

The layers are interrelated in that Layer 2’s objective cannot be met until Layer 1’s knowledge sources are described, and the trained prediction tool developed in Layer 2 cannot be ubiquitously invoked, shared and evaluated, until Layer 3’s Semantic Web services for healthcare framework is designed.

FIGURE 1-1 COHESIVE DESIGN OVERVIEW (LINKING PROBLEM #1 AND PROBLEM #2)

1.3 HYPOTHESIS

The integrated hybrid screening system will offer an improved non-invasive risk assessment model for predicting preterm birth prior to 23 weeks gestation in heterogeneous maternal populations. The hybrid system will offer earlier predictions than current risk assessment models trained on symptomatic patients only, and will be relevant to both high-risk and low-risk mothers. The trained hybrid system will be applicable to new patient cases with sensitivity approaching 65%; the current gold standard obtained using invasive testing being 64.7%. The hybrid system can be developed for future use in mainstream obstetrical care.

In detail: PTB will be estimated with high specificity and sensitivity on non-symptomatic maternal populations outside of clinical trial situations, using socio-demographic and obstetrical data readily available to the practitioner during obstetrical consultation prior to 23 weeks gestation. A preterm birth risk assessment model will be developed, which when processed by the trained hybrid system, will improve the physician's ability to identify high-risk mothers. This implies that the model will demonstrate superior classification and sensitivity than the current 50% classification success rate for predicting preterm birth amongst symptomatic pregnancies using non-invasive testing (e.g. those which have been previously identified as high-risk). When used in this manner, as a risk assessment screening tool, the system can avoid the need for further costly and invasive maternal testing on low-risk patients while ensuring that a larger proportion of high-risk patients are correctly classified. Please see *Chapter 3* for a

detailed discussion of the sensitivity and specificity achieved using current PTB prediction models.

1.4 SIGNIFICANCE OF THE RESEARCH

From a health informatics perspective, this research is significant in providing a unique approach to solving a timely medical problem. The early prediction of preterm birth is of particular significance considering that the number of pregnancies ending in preterm birth is currently increasing², and to date no non-invasive risk assessment technique has been widely accepted or proven superior to clinical judgment. Among women identified by clinicians as having an elevated risk of PTB, approximately 50% give birth prematurely, indicating a clear need for more accurate prediction models³. If a mother at risk of delivering prematurely could be identified in advance, physicians would be able to adjust their pregnancy management strategy in an attempt to prevent the mother from delivering prematurely. Common pregnancy management strategies include: prescribing bed rest; ensuring the mother remains properly hydrated; prescribing tocolytic agents to relax the uterus and stop or delay the onset of contractions, prescribing medications to prevent infection and to help the baby's lungs develop more quickly in preparation for birth (pregnancy-info.net, 2006). Advance knowledge of a mother's risk of PTB could also impact pregnancy management by ensuring that advanced postpartum care is available. Particularly in the case of rural hospitals, transferring the mother to a

² Experts speculate the two main reasons for the rising PTB rate are advanced maternal age and increased use of artificial reproductive technologies (which have an associated higher rate of multiple births).

³ This is not because the other 50% of women were prevented from delivering prematurely due to physician intervention, but rather that 50% of the mothers were incorrectly identified as being at high-risk of preterm birth.

tertiary care hospital prior to birth may be advisable. This knowledge could also impact the birth plan, for instance if a high-risk PTB is possible, choosing a home birth may be an unsafe option. Additionally, by being forewarned of their increased risk, parents would be in a position to educate themselves about what to expect following a premature birth. There is also the possibility that this knowledge could lead to prevention without medical intervention, for instance by making an effort to reduce maternal stress.

From a systems engineering perspective, this work provides a Semantic Web services for healthcare framework for the purpose of modeling physicians' decision-making process for clinical decision support. The framework provides a means of translating clinical data into clinical knowledge needed by the physician at the point of care, and integrating different knowledge modalities (in this case clinical decision support systems). This addresses two key motivations of this work, described in Section 1.1: knowledge translation and integration. The perinatal clinical decision support ontology is a necessary prerequisite to integrating data from distributed databases, providing the standardization needed to rectify differences in terminology that occur when multiple databases represent the same information in different ways. No ontology currently exists to support diagnostic clinical decision support. It is important to note that this work does not aim to replace physician decision-making, but rather to use clinical information to its fullest capacity to alert physicians to possible outcomes before they occur.

1.5 THESIS OUTLINE

This thesis is organized as follows:

Chapter 2 provides an overview of related medical and technological material,

including: (i) a literature review of the medical problem of estimating preterm birth; (ii) an overview of the clinical decision support applications used in this work for data mining purposes; and (iii) the technical background required to understand the challenges encountered in designing a Semantic Web services for healthcare-based framework.

Chapter 3 presents a critical analysis of current predictive models for estimating PTB. The state of the art in Semantic Web services for healthcare is described. The difference between what has been accomplished to date, and the work accomplished in this thesis are discussed.

Chapter 4 describes the methodology for developing the integrated hybrid screening system for preterm birth risk assessment system.

Chapter 5 presents experimental results for both the classification-based multilayer perceptron ANN models and the integrated hybrid screening system.

Chapter 6 encompasses all work related to designing the Semantic Web services for healthcare-based framework, including the description of the composition scenarios and the specification of the diagnostic clinical decision support ontology.

Chapter 7 discusses the conclusions generated from the results presented in Chapter 5, outlines the thesis' academic contributions, and includes suggestions for future research.

CHAPTER 2. OVERVIEW OF RELEVANT MEDICAL AND TECHNOLOGICAL CONCEPTS

2.1 EARLY PREDICTION OF PRETERM BIRTH

Preterm birth is the leading cause of mortality occurring before 28 days of age (Health Canada, 2000), accounting for 85% of all neonatal deaths not due to lethal congenital malformations (Dawood, 1980) and nearly half of all cases of congenital neurologic disabilities, including cerebral palsy (Goldenberg & Rouse, 1998). Disturbingly, despite significant advances in healthcare, the number of preterm infants born in Canada each year is currently increasing: more than 7% of Canadian infants are born prematurely, often resulting in newborn death or disabilities that cost the Canadian healthcare system tens of millions of dollars each year (Canadian Institutes of Health Research, 2005). The Perinatal Partnership Program of Eastern and Southeastern Ontario recently prepared a statistical report assessing the quality of perinatal care in Ontario. Their message was clear: “The perinatal period is one of the most important of all the life stages. Growing well during the prenatal period, being born at term and having a safe passage to the external world provide a foundation to support future learning, physical and mental health” (PPESO, 2005).

Due to its direct correlation with newborn mortality, reducing the burden of PTB has become the number one perinatal health priority. A key goal of outcome estimation in perinatal care is to identify which factors best predict continuation to PTB when preterm

labour occurs. This problem is particularly significant in North America, where statistical evidence indicates that 11% of all births are preterm (Goldenberg & Rouse, 1998). In Ontario the number is lower with approximately 8% of babies born prematurely; during the course of a lifetime it is estimated that each preterm baby will use \$676,800 (1995 Canadian dollars) in healthcare services (Allan et al., 2002).

2.1.1 RISK MODEL FOR PRETERM BIRTH

While many studies have attempted to predict women at risk of PTB (Creasy, Gummer, & Liggins, 1980; Dawood, 1980; Fiore, 2003; Goldenberg et al., 1998; Grzymala-Busse, Goodwin, & Zhang, 2003; Iams, 1997; Iams et al., 2001; Lockwood & Kuczynski, 2001; Meis et al., 1998; Mercer, Goldenberg, Moawad, Meis, & Das, 1999; Thorsen et al., 2001); so far, no scoring system has proven itself superior to clinical judgment. One of the major obstacles is that most women who deliver prematurely have no obvious risk factors and over half of all PTBs occur in low-risk pregnancies (Allan et al., 2002; Iams et al., 2001). A literature review of preterm predictive models yields disappointing results in terms of discovering an accepted risk-scoring model for PTB: “Preterm prediction efforts have been thwarted by the relatively insensitive and non-specific risk-screening protocols developed and implemented to date” (Johnston, Williams, Hogue, & Mattison, 2001).

While current risk models have proven to be insensitive and non-specific (Iams, 1997), three groups of women have been identified as being at high risk for PTB (Aerts & Iams, 1999), these include: (i) women pregnant with multiples, (ii) women who have experienced previous PTBs, and (iii) women with uterine or cervical abnormalities.

Currently, risk assessment of PTB in maternal populations where the mothers are pre-selected as having an elevated risk of PTB is largely based on fetal fibronectin measurements and cervical length (Lockwood & Kuczynski, 2001). However, these tests do not predict risk of PTB early in pregnancy, they only become more sensitive in estimating PTB as the estimated date of birth approaches, and they are based on invasive techniques.

2.1.1.1 Higher Order Births

The prevailing trend is towards an increase in the incidence of multiple births, currently representing 3.2% of all births in the United States (and 14% of infant deaths) as compared to only 2% in 1980 (Fiore, 2003). This increase has produced a significant adverse impact on preterm birth rates: "...infants from multiple births are nearly six times more likely to be delivered preterm than are singletons" (Fiore, 2003). Even with the best possible prenatal and neonatal care, multiple births pose serious risk of complications, as plurality is inversely proportional to the length of gestation. A singleton birth has an average gestation of 39 weeks; with twins this drops to an average of 36 weeks, while triplets average 33 weeks and quadruplets average 31 weeks.

2.1.1.2 Previous Preterm Birth

For singleton births, a history of PTB in a previous pregnancy appears to be the factor most predictive of preterm birth in the current pregnancy (Dawood, 1980), with the risk of recurrence increasing with every preterm delivery. A woman's most recent birth is the most predictive of her risk. However, despite the association between obstetric history

and risk of preterm birth, most women who experience a preterm delivery do not deliver prematurely in subsequent pregnancies (Aerts & Iams, 1999).

2.1.1.3 Uterine Abnormalities

Uterine and cervical abnormalities have also been found to play a role as indicators of possible preterm birth. Transvaginal ultrasound assessment of cervical length can help to identify women at risk for preterm delivery in their current pregnancy. Medical experts believe that cervical length appears to reflect the level of cervical competence and plays a role in the occurrence of preterm birth (Aerts & Iams, 1999). Iams' team used transvaginal ultrasonography to note that a shorter cervix at 22-24 weeks gestation had an increased risk of PTB; however, no cervical length was identified below which all women delivered prematurely (Iams et al., 2001).

2.1.1.4 Other Risk Factors

In addition to these three groups of women having an elevated risk of PTB, researchers have identified numerous other risk factors believed to be associated with an increased risk of PTB (please see Table 4-1 for a compiled listing of risk factors from the literature). However, to date experts have not been able to conclusively link these factors with subsequent preterm delivery. A possible explanation is that many studies focus exclusively on proving whether a particular variable is significant in predicting PTB, and do not examine the non-linear relationships between the variables: "The various clinical presentations and presumed mechanisms that can cause this condition have been placed into different categories and have been considered unrelated. In this traditional model,

premature birth is viewed as an endpoint that may ensue from any of these risk factors” (Aerts & Iams, 1999).

While it is possible that in certain clinical situations a single variable may be the cause of preterm birth, such as when the mother has used cocaine or experienced trauma, an alternative multi-variable model is more applicable to most cases of PTB where factors may interact to produce that outcome.

2.2 ARTIFICIAL INTELLIGENCE AND DATA MINING IN MEDICINE

The goal of employing Artificial Intelligence (AI) in medicine in this work is not to support ‘strong AI’ – which aims at creating computer systems whose behaviour at some level is indistinguishable from humans (Open Clinical (Knowledge Management for Medical Care), 2005), but rather to examine the human cognition involved in perinatal care and investigate how it can be supported in critical and challenging situations. These ‘weak’ AI systems are not intended to exist independently, but are a form of ‘cognitive prosthesis’ that supports a human in a variety of tasks (Open Clinical, 2005). There are two main styles of artificial intelligence applications in medical data mining, these are: *expert driven* and *discovery driven* systems. Expert systems, which use decision models acquired from experts, were the first to be introduced into clinical practice and are the most accepted by clinicians. Expert systems are limited by the constraints of the Eigenbaum bottleneck, which stipulates that as the complexity of the domain grows it becomes more difficult for human experts to formulate their knowledge as practical strategies. The increase in the amount of available clinical data, and the associated increase in complexity of the healthcare domain, has advanced the use of

discovery driven systems, where knowledge is not derived directly from experts but is discovered from past data using data mining methods. With discovery driven systems, it is easier to generate and maintain clinical knowledge. Bayesian reasoning is one of the oldest discovery driven data mining approaches and boasts strong acceptance by the clinical community due to its explanatory features and closeness to clinical intuition. Artificial Neural Networks and Case Based-Reasoning systems are two other means of knowledge discovery that are less commonly used in clinical practice; however ongoing research by MIRC suggests that they have an important role to play in managing the increasing volume of complex clinical data being generated by Hospital Information Systems (HISs) and diagnostic and monitoring devices.

2.2.1 ARTIFICIAL NEURAL NETWORKS (ANN)

Artificial Neural Networks (ANNs) are shown to be of value in many areas of medicine (Dayhoff & DeLeo, 2001; Ennett, Frize, & Charette, 2004; Ennett & Frize, 2003; Frize, Ennett, Stevenson, & Trigg, 2001; Kudyba, Hamar, & Gandy, 2005; Lockwood & Kuczynski, 2001); they are effective in classifying data with complex, non-linear associations between variables (Frize et al., 2001); they are among the most powerful general nonlinear modeling techniques available (Goodwin & Maher, 2000), and they are capable of performing multifactorial analysis (Dayhoff & DeLeo, 2001). Additionally, it is possible to develop neural network models that identify considerably more future high-risk patients than many traditional rules-based or regression-based models (Kudyba et al., 2005). A review of evidence of health benefit from ANNs in medical intervention has recently been published (Lisboa, 2002). This surveyed

randomized and non-randomized clinical trials of ANNs in the domains of oncology, critical care and cardiovascular medicine, including four studies in perinatal care. The review notes the potential for extensive benefit but criticizes poor methodology and exaggerated claims in many studies.

A number of studies have compared ANNs with regression models and in each, the ANN approach outperformed the statistical method, but usually by only a small margin, suggesting that these approaches may be complementary (Ambalavanan & Carlo, 2001; Livingstone, Manallack, & Tetko, 1997; Sargent, 2001). There are however only a few studies using ANNs in prediction of outcomes for preterm newborns (Grzymala-Busse et al., 2003; Lockwood & Kuczynski, 2001).

ANNs are commonly described as discovery driven systems that mimic the information processing of the human brain. This is not to be confused with mimicking the way clinicians actually think, but rather with modeling the complex biological connections between the brain's neurons and synapses. An ANN uses a process analogous to information processing by the human brain, acquiring knowledge through a learning process and storing it using inter-neuron connection strengths (e.g. weights).

2.2.1.1 ANN Architecture

A set of artificial neurons, called nodes, is connected and forms an artificial neural network. Figure 2-1 provides a diagrammatic representation of a single node in a MultiLayer Perceptron (MLP) neural network (MLPs contain an input layer, one or more hidden layers and an output layer). Specifically, the vector variables are introduced at input nodes (p_i); both the input nodes and the bias term are subsequently multiplied by a

corresponding weight (w_i), summed in each processing unit, and fed through a transfer function ($f(s)$) to scale the output to a value within a limited range.

FIGURE 2-1 DIAGRAMMATIC REPRESENTATION OF A SINGLE NODE IN A MULTILAYER PERCETRON (ENNETT, 2003)

Networks can have a number of layers, and each layer can have multiple hidden nodes. Each node has an output that will serve either as the input to the next layer, or if the final layer has been reached, will result in an output. The ANN thus develops a set of outputs based on a system of input conditions. Once sufficient training runs have been performed to minimize the error (where training is based on presenting cases with known outcomes to the ANN), the ANN can model the system automatically. There are many possible ANN architectures to consider: single layer networks; multi-layer feed forward

networks (MLPs); recurrent or feedback networks; and associative networks which detect similarities between new input and stored patterns (Medsker, 1994). Figure 2-2 depicts the architecture of a generalized MLP ANN with one input layer, one hidden layer, and one output layer.

FIGURE 2-2 ARCHITECTURE OF THE MULTILAYER PERCEPTRON (ENNETT, 2003)

2.2.1.2 ANN Learning

ANN learning can be supervised or unsupervised. Supervised learning implies that for each set of inputs, the outputs are known. Learning then occurs by adjusting the network weights, by a number of possible mechanisms, in an attempt to achieve correct results. Unsupervised learning implies that the neural network self-organizes to produce categories into which the inputs fall. No knowledge is supplied about the correct classification and users vary network parameters to control the number of categories into which the network classifies the inputs (Medsker, 1994).

In the case of a medical database, the ANN can be trained to estimate a set of clinical outcomes from the input variables in the database. In this respect the ANN serves a similar function to common statistical techniques. The main advantage of neural

networks over statistical techniques is that the model does not have to be explicitly defined before beginning the experiments. ANNs can recognize the relevant data and patterns, whereas a statistical model requires prior knowledge of the relationships between the factors under investigation (Blum, 1992). Neural networks can be more accurate when applied to complex clinical databases with non-linear relationships between the variables. This presumes that methods have been used to avoid overfitting (over-training), which is the loss of generalization that occurs when the patterns used to train the networks are too closely linked to the test data. An over-trained network will not generalize well to other databases or new data.

2.2.1.3 Training, Testing and Verification Sets

When developing neural network models, three independent data sets are used. The terminology for naming these three sets is not standardized across neural network research; however, this work uses the approach of Dayhoff & DeLeo in their overview of artificial neural networks, specifically naming the three sets: training, testing (also referred to in some work as validation) and verification (Dayhoff & DeLeo, 2001). The training set is used for learning and adaptation, performed by the adjustment of weights during training. The results from the testing set determine when it is possible to stop training. Finally, the verification set runs independently of neural network training providing a benchmark evaluation metric for the performance of the neural network as a classifier when deployed for actual use (Dayhoff & DeLeo, 2001).

2.2.1.4 Advantages of Neural Networks

In summary, neural networks were chosen for this work, as compared to traditional statistical techniques, for the following reasons: ANNs are flexible non-linear systems that demonstrate robust performance when dealing with noisy data, as well as having the ability to generalize from the input data. Users require less formal statistical training and do not need prior model information or extensive consultation with domain experts. ANNs also make it possible to integrate data of different formats (continuous, binary). Additionally, ANNs are well suited for use when relationships between the variables are complex, multidimensional and non-linear and there are a large number of input variables, as is the case with the databases used in this thesis.

While there is no single algorithm that performs better than all other algorithms for any given biomedical problem, both neural network techniques and statistical approaches have advantages and disadvantages. Logistic regression is the clear choice when the goal is to determine possible causal relationships among variables (Tu, 1996), and neural networks are well suited to developing prediction models for non-linear variables (medical data is often believed to have non-linear relationships between input variables and outcomes), where a three layer neural network can approximate any bounded non-linear mapping between input signals and outcome.

All methods should continue to be explored in a complementary manner, although this is outside the scope of a single thesis

2.2.1.5 Prior Uses of ANN within MIRG

Over the past ten years MIRG has employed ANNs to estimate clinical outcomes, such as post-surgery mortality for coronary artery bypass grafting surgery patients (Ennett & Frize, 2003), artificial ventilation needs for adult intensive care unit patients (Buskard et al., 1993; Frize et al., 1995; Frize, Ennett, Stevenson, & Trigg, 2001, Tong, Frize & Walker, 2002), and in-hospital mortality, length of stay and duration of artificial ventilation for Neonatal Intensive Care Unit (NICU) patients (Frize, Ennett, Stevenson, & Trigg, 2001, M. Frize et al., 2004). Efforts have been made to validate the use of ANNs in clinical settings by comparing results with other techniques, such as radial basis functions with principal component analysis (Li, 2005), and fuzzy logic classifiers (Frize et al., 2004).

Within MIRG, ANNs are primarily used to classify patient cases using the input variables in clinical databases for the purpose of outcome estimation. Once a network has undergone sufficient training runs and one or more network structures have been verified in a verification set, the system can be used to classify *new* patient cases. This is accomplished by embedding the structure (e.g. number of nodes and hidden layers), weights and biases from the best-performing ANN prediction model into a CDSS that accepts clinical values from physicians via a user interface. The system then executes a single epoch to deliver outcome estimation for the requested patient case. As a secondary ANN function, it is possible to use the ANN weights to determine a set of indicators or risk factors for a particular clinical condition; these indicators are referred to as a minimal data set (Rybchynksi, 2005).

2.2.2 ANN RESEARCH FRAMEWORK

MIRG has recently added the classification-based MLP ANN tool, described in the previous section, to an Artificial Neural Network Research Framework (ANN-RFW) (Rybczynski, 2005). The main purpose of the ANN-RFW is to improve the efficiency and effectiveness of developing MLP ANN-based models. The ANN-RFW focuses on decreasing the operational time for running experiments and in providing exhaustive experimental results for outcome prediction using MLP ANNs. From an experimental perspective, the ANN-RFW introduces three new functionalities, all used in this work.

(1) *Automated parameter update*: In the original ANN software, experiments were run for each structure individually (where the structure is defined by the number of layers and number of hidden nodes in the network); following each run the researcher extracted the values of each of the nine ANN learning parameters and chose the best parameter, manually transferring the best parameter to the next run (these parameters are described in detail in Section 4.5.1). The ANN RFW introduces the ‘subsequent choice’ method for choosing how to carry forward the best parameter value from a previous run. For each network structure, the parameter start points are set and a single parameter divide and conquer run⁴ is performed. If the results from that attempt are better than the previous best results, the parameter value is carried forward as the new start point for the next parameter’s attempt. This continues until a stagnant point is reached, where the best result of the current parameter trial is not better than the previous best. After finding nine consecutive stagnant points (one attempt for each of the nine ANN learning

⁴ The divide and conquer algorithm recursively breaks a problem down into two or more sub-problems of the same type, until these become simple enough to be solved directly (Wikipedia.org, 2006).

parameters), the structure is complete (Frize, Ennett, & Hebert, 2001; Ennett, Frize & Charette, 2004; Rybchynski, 2005).

(2) *Network Structure Trials*: The previous ANN automation used a divide and conquer algorithm to find prediction results for one neural network structure only, the structures were selected at the discretion of the researcher and it was very rare for more than three different structures to be attempted for a specific outcome researched by an individual (Rybchynski, 2005). The ANN RFW performs an exhaustive search to find the results for all network structures including single layer and spanning across double layer with one hidden node to double layer with $2n+1$ hidden nodes, where n is the number of input nodes. Following steps (1) and (2) the user must manually examine all parameter value – network structure combinations to discover the best performance results for their outcome of interest and corresponding performance criteria, this is a time-consuming procedure.

(3) *Verification Dataset* - A verification program was created to validate the trained and tested MLP ANN on new patient data. The ANN-RFW also introduced a committee of classifiers approach; where ANNs trained on different structures constitute the committee members.

2.2.3 ADDING RISK STRATIFICATION TO THE MLP ANN

Recent work within MIRG has succeeded in adding risk stratification estimations (i.e. *level of risk* estimations) to the clinical outcomes provided by the ANN software, which in the past were based solely on binary outcome classification. A clinical outcome of +1 indicated that the patient had tested positive for the outcome of interest while -1

indicated the outcome was not present. However, an improved solution would allow the CDSS to provide physicians with clinical risk stratification estimations for each outcome. Zhou and Frize's new work using the Canadian Neonatal Network (CNN) database adds risk stratification estimation to the ANN software used for predicting mortality in the Neonatal Intensive Care Unit (NICU) (Zhou & Frize, 2006). Specifically, the MLP ANN with risk stratification estimation incorporates information to stratify patients into groups according to their mortality risk.

The CNN database contains patient physiological variables, diagnoses, and therapies; traditionally ANNs that include risk stratification estimation include probabilistic information in their input variables. Zhou and Frize explain that frequency analysis can be used to add probabilities to the training set, but that this is laborious and time-consuming (Zhou & Frize, 2006). A preferred method would train the MLP ANN using non-probabilistic variables generally available in clinical databases (e.g. physiological, administrative, and patient history), and produce risk stratification-based results.

Networks trained to minimize the cross-entropy or sum-of-squares cost function can be used to create risk stratification estimation outputs; both closely relate to the maximum likelihood concept (a probability theory method used to estimate the unknown distribution of a population from observed samples) and can be interpreted in a Bayesian learning framework (Zhou & Frize, 2006). The following summarizes the recent work of Zhou and Frize (Zhou & Frize, 2006), explaining the rationale for their approach which combines maximum likelihood estimation and neural network learning to approximate

risk stratifications for clinical outcomes.

2.2.3.1 Summary of Zhou and Frize's Approach Combining Maximum Likelihood Estimation and Neural Network Learning

Maximum likelihood estimation is an approach to parametric estimation where the function form of the probability density function is known in advance. Parametric estimation approximates the true distribution by adjusting a set of parameters on a predefined function form, such as a Gaussian or exponential distribution. However, as Zhou and Frize explain, neural network approximations are generally non-parametric estimations where the probability density function cannot be identified.

Based on the theory that a three-layer neural network can approximate any bounded continuous function, assuming that the number of hidden nodes is sufficient, Zhou and Frize propose applying the maximum likelihood estimation to approximate a neural network probability function by changing the network parameters, specifically: the number of hidden nodes, the weights, and the biases. In Zhou and Frize's work this involved identifying the parameter set that would successfully approximate the NICU mortality probability function. In this work, the goal is to approximate the PTB probability function (called the risk stratification estimation) for the clinical databases in question.

Essentially, this means that the cost functions in neural network training (such as the hyperbolic tangent function used in the classification-based MLP ANN models in this thesis) are modeled as maximum likelihood functions in order to *estimate risk stratification*.

A new weight update algorithm was developed with gradient-ascent search for a neural network with a single binary output unit. The new algorithm performs gradient search to maximize the likelihood cost function of the training data, whereas the traditional back-propagation algorithm seeks to minimize the error in training data. To be consistent with previous MIRG work, the MLP ANN with risk stratification estimations developed in Zhou and Frize's new work was coded in Matlab and used the logarithmic sensitivity index as its stopping criteria.

The final outcome of the MLP ANN with risk stratification estimations does not provide the severity of illness, but rather provides the user with an indication of the risk they have that the outcome will occur. For instance, a risk stratification of 1 provides a high level of certainty that the outcome will occur (that is, the birth will be premature), while a risk stratification of 0 will provide a high level of certainty that the birth will be full term. By assigning appropriate cutoff points within the stratification ranges, outcome screening models can be created. Appendix B provides Zhou and Frize's equations for modeling the cost functions and deriving a training algorithm.

2.2.4 CASE-BASED REASONING (CBR) – K-NEAREST NEIGHBOURS (KNN) ALGORITHM

In Case-Based Reasoning (CBR) systems new problems are solved by recalling experience from previously solved problems, which are stored in a memory of cases (Hullermeier, 2005). Case-based reasoning systems are used to provide physicians with diagnostic support, with the idea being that new problems can be solved based on previous cases and the physician's domain knowledge. MIRG's CBR system follows the

rationale that physicians require knowledge about past cases for comparison purposes, but prefer to reason about current cases themselves (Frize & Walker, 2000).

MIRG's newest CBR tool is based on a weighted k-Nearest Neighbor (kNN) algorithm, a similarity-based matching algorithm that finds the k closest matching cases to the query instance (where k is the number of closest matching cases to be retrieved, as specified by the user). The distance from the target case to all other cases in the case-base is calculated and the cases with the smallest distance to the target case are chosen as the most similar (Rybachynski, 2005). The following provides a step-by-step description of the kNN algorithm (Cotea & Jiwani, 2003):

Step 1: Find minimum and maximum of all input variables

Step 2: Calculate the maximum distance (Max_dist)

$$Max_dist = \sqrt{\sum_{r=1}^{r=n} [(max_r * w_r) - (min_r * w_r)]^2}$$

where $w_r(x_i)$ is the weight associated with the r^{th} attribute of instance x

Step 3: For each record in the database:

- Calculate the distance between that record and the target record ($dist$)

$$dist(t_r, x_r) = \sqrt{\sum_{r=1}^{r=n} [(t_r * w_r) - (x_r * w_r)]^2}$$

where t_r and x_r are the target value and the current record value at the r^{th} attribute

- Calculate the similarity

$$similarity(t_r, x_r) = 1 - (dist(t_r, x_r) / Max_dist)$$

The kNN tool is employed by MIRG to perform the following functions: 1) match closest cases for inspection by a clinician and display the ten closest matching cases; 2)

impute missing variable values into patient cases – this work is described in detail by Ennett (Ennett, 2003); 3) extend datasets to include missing minimum data set variables; and 4) Predict outcomes – this work is currently incomplete, but it has been proposed that the kNN algorithm could be used to predict outcomes (Rybachynski, 2005).

2.2.5 DECISION TREE

Decision Trees (DT) are another data mining technique that are useful in extracting comprehensible rules (Wang, Liang X-X., & Sun, 2005) and are considered an effective techniques for acquiring classification knowledge (Tsujino, 1995). A decision tree is a structure that contains leaf nodes and decision nodes, where a leaf node represents a specific class of example, and a decision node specifies a test to be carried out on a specific attribute (i.e. a decision) (Kijisirikul & Chongkasemwongse, 2001). A child of the root node is then created for each possible value of the attribute. A decision tree classifier builds a DT; association rules can then be extracted from the tree (Aslandogan & Mahajani, 2004). In a univariate decision tree, at each node, only one feature is used (i.e. only one decision is made). For continuous data, this takes the form $x_i > c_k$, where x_i is the i th output feature and c_k is the threshold. For categorical data the decision would be based on a test for equality (e.g. if $x_i = c_k$).

2.2.6 PARENTS DECISION SUPPORT (PADS)

MIRG has recently developed a knowledge translation tool, coined PArnts Decision Support (PADS). This CDSS was developed in close collaboration with physicians, and comprises a set of adaptable user interfaces that perform knowledge

translation for the physician.

Frize *et al.* developed the PADS framework to integrate knowledge discovery-based information on a newborn patient's estimated outcomes with the physician's knowledge interpretation. The CDSS's purpose is to help parents make decisions regarding the management of critically ill babies. The parents' perspectives are transformed into codified knowledge. Context sensitive content adaptation delivers personalized and customized information to a variety of users, from physicians to parents. The system provides structured knowledge translation, interpretation, and exchange between all participants in the decision, facilitating collaborative decision-making that involves parents at every stage on whether to initiate, continue, limit or terminate intensive care for their infant (Frize, Yang, Walker, & O'Connor, 2005).

2.3 COLLABORATIVE DATA MINING METHODS

In collaborative data mining, systems are created to combine data mining methods, either in series or in parallel, with the objective of improving the model's ultimate predictive power.

2.3.1 PARALLEL SYSTEMS

In parallel systems, each data mining method produces a unique result; all results are then combined into a final outcome through voting, averaging, or weighted averaging. Rybchynski has added a committee of classifiers to the ANN research framework (Rybchynski, 2005). The committee of classifiers calculates the outcome result for each of the committee members using previously trained member classifiers (defined by their

weights and biases), and then calculates the committee result from the member's results based on a majority rules voting system. The committee of classifiers used by MIRC is an example of a homogeneous committee, where each parallel element is based on the same data-mining technique, in this case a set of ANNs. A heterogeneous committee uses predictors based on different data mining techniques; they can also be trained with different data subsets.

A committee approach to classification is known to produce generally improved results, provided that error rates are less than 50% for each member of the committee. With efficient classifiers that are moderately independent, the classification error rate can drop dramatically. Choosing disjoint feature sets is one way to produce independent classifiers (Jerebko, Malley, Franaszek, & Summers, 2003).

The ANN committee of classifiers verification tool is used when validating the classification-based MLP ANN models used in this work. Developing the classification-based MLP ANN models is the first step in the hybrid system methodology proposed in this thesis. However, the proposed system is better classified as a series (e.g. hybrid) system, with data flowing sequentially through data mining applications.

2.3.2 HYBRID (SERIES) SYSTEMS

In hybrid systems, the models are arranged consecutively with results and/or parameters from the initial model being passed to the next model, and so on, such that successive iterations refine the system's predictive ability. Hybrid systems allow information to be fed through consecutive data mining iterations, where each iteration may be defined by a different data mining technique. This can be considered a process of

refinement, where each step adds value in terms of compensating for the weakness of a previous step in outcome prediction. Hybrid systems are also referred to as series collaborations.

2.4 SEMANTIC WEB SERVICES FOR HEALTHCARE – OVERVIEW OF KEY CONCEPTS

The Semantic Web and Web services are two complementary and evolving technologies that will change the face of information and knowledge management in healthcare. The ongoing efforts to blend the Semantic Web and Web services to provide seamless flow of information and knowledge (Crawford, 2005), offers opportunities for sharing, automating and enabling access to information which will prove invaluable in solving the current information integration challenges so pervasive in the distributed world of healthcare informatics. In fact, many healthcare Information Technology (IT) experts predict that in the future, healthcare services will be offered as interchangeable Semantic Web services used in distributed but related medical domains.

Based on the acceptance of Web services predicted by many IT experts (Daconta, Obrst, & Smith, 2003; Kreger, 2003; Lea & Vinoski, 2003), healthcare providers are starting to leverage Web services as a solution to data and application integration (Chatterjee, 2003) in clinical environments. The Semantic Web provides a mechanism for adding meaning to data, essential in healthcare where the same clinical information can have many different representations. Combining the Semantic Web and Web services in relation to the healthcare domain, results in *Semantic Web services for healthcare* (Catley, Frize, & Petriu, 2006), which will ultimately enable automated interpretation of

clinical data.

Semantic Web services for healthcare open possibilities in clinical Knowledge Management (KM), Clinical Decision Support (CDS), and clinical data and application integration; they have the potential to support an advanced healthcare environment, offering health networks new applications and services (Frize, Walker & Catley, 2006).

2.4.1 WEB SERVICES

“Experts forecast that entire ecosystems of e-services will be built around specific industries, providing specific processes to solve specific problems for specific types of customers, through specific transaction chains” (Hoffman, 2003), this thesis is based on this premise, so that in the future healthcare services will be offered as interchangeable Web services serving distributed but related medical domains (Catley, Frize, Petriu, Walker, & Yang, 2004). The following sections provide both an overview of Web services and offer a detailed look at the implementation technologies.

2.4.1.1 Web Services – Overview

Web services constitute an important emerging technology for which potential applications are unlimited. In simplest terms: “if a provider can imagine a way of delivering something of value to a customer to provide some usefulness... they have a viable Web service” (Stafford, 2003). In the context of this thesis, MIRG is the Web service provider; physicians and allied healthcare workers involved in obstetrical and perinatal care are the customers (indirectly the infants and their parents as well); and the Web services are any tools that the provider can make available to customers to aid their

clinical decision making. Web services can be used for providing service to human end users, as well as to other applications; many believe that the primary goal of Web services is to address the problem of application-to-application integration (Tosic, 2004). In this thesis, consumers of Web services are both healthcare applications and human end users – the physicians.

Web services are based on a system architecture model referred to as ‘service-oriented’; with providers assessing which applications they can offer as services to different groups of potential users. There are three main roles in a Service Oriented Architecture (SOA): a provider, a consumer, and a directory containing Web Service descriptions. The provider publishes descriptions of its Web services in a directory, which is accessible by the consumer. Once the consumer selects a service, the consumer and service are dynamically bound, e.g. at run-time. The ‘publish-find-bind’ model of interaction enables loose coupling of providers and consumers and, thus, increases agility, flexibility, and adaptability of distributed systems (Tosic, 2004). This leads to a simple definition of Web services: *Loosely coupled software components that use standard Internet technologies to interact dynamically, independent of platform or language. Web services facilitate integration and interoperability because the underlying implementation and deployment platform are not relevant to the application invoking the service.*

Candidate Web services can be categorized as being either core or composite services. A core Web service offers basic functionality that will potentially be required by multiple higher-level applications. Composite Web services represent these higher-level

applications; composite services combine two or more core services to offer a complete system application as seen from the user's perspective. In Web services terminology this is referred to as a composition scenario. Core Web services can be combined into many different composition scenarios and can be thought of as customizable building blocks for compositions.

With the ability to offer sophisticated applications to consumers via Web services, a new dimension emerges: applying the service-oriented concept to data and software. Turner *et al.* have coined the terms: Software as a Service (SaaS) and Data as a Service (DaaS) (Turner et al., 2003). The vision is to eliminate many of the problems that occur with developing, updating and evolving software systems (Turner et al., 2004), and indirectly the data accessed by these systems. By providing a highly available infrastructure for connecting business partners anywhere and anytime (de Bruijn, Fensel, Keller, & Lara, 2005), Web Services offer an alternative to costly customized solutions for integrating business applications.

2.4.1.2 Web Services – Detailed Discussion

Web services are accessed via ubiquitous Web protocols and data formats, such as Hyper Text Transfer Protocol (HTTP) and the eXtended Markup Language (XML), respectively, rather than via object model specific protocols such as Distributed Component Object Model (DCOM) or Remote Method Invocation (RMI) (Graham, 2003) - this changes the view of 'middleware' in the system. Web services are invoked over a network, however they do not have to reside on the World Wide Web; they can be located on an Intranet, or anywhere on the network.

There are three key components of Web service systems with three corresponding major XML-based standards provided by the World Wide Web Consortium (W3C) to support functionality over the Web:

Delivery: Comprises all technologies required to transport a service request from client to server, including XML for message encoding, and SOAP (previously known as the Simple Object Access Protocol – now considered a misnomer) for handling data transmission of XML-formatted data. A Web service provides one or more endpoints, called ports. Ports implement an interface, a collection of one or more operations that can be invoked by consumers.

Description: A Web service interface provides a collection of operations accessible through standardized XML messaging and described using the Web Services Description Language (WSDL), which specifies the operations provided by a Web service, including the kinds of objects that are expected as input and the output of the operations (Graham, 2003), the location (URL) and protocol support of a website.

Discovery: The service requestor discovers the Web service via discovery agencies that allow service descriptions to be published and discovered. From a performance perspective this involves the time to look up the service in the Web services directory using Universal Description Discovery and Integration (UDDI). The Universal Description Discovery and Integration is an XML language for defining and interacting with a Web service repository. It is also possible to assume a more simplistic approach, where the client has prior knowledge of the service address.

2.4.1.2.1 XML

From a healthcare perspective XML has an important role to play in promoting the acceptance of computer applications within the medical community. XML has become an integral component in many clinical systems under development, and this number is expected to increase dramatically in the near future (Stalidis, Prentza, Vlachos, Maglavera, & Koutsouris, 2001). XML makes data portable by separating the presentation of the data from the data content; in addition users are able to define original content by creating new XML schemas. Schemas create self-describing data by specifying how the receiving application will interpret the markup tags and can be used to identify errors in the data itself, essential in healthcare where data integrity is crucial. The schemas are interpreted and potentially validated by XML parsers and subsequently directed to enterprise applications. In this way XML enables application-to-application communication and facilitates the development of smart agents, as applications interpret and respond to information intelligently. The eXtended Stylesheet Language Transforms (XSLT) is a W3C standard that transforms XML content from one form into another for the purpose of promoting interoperability between vocabularies. Using XSLT, XML data can be transformed into any format for which the developer creates a matching XSL transform.

2.4.1.2.2 SOAP

SOAP is a W3C XML standard, defining a lightweight protocol for information exchange. Currently its usage centers largely on the exchange of XML-formatted data over HTTP; however SOAP's main purpose is to implement protocol and language

independent Remote Procedure Calls (RPC). A SOAP message comprises an envelope and a body, which contains the method call and its parameters. The SOAP specification defines: 1) rules for using XML to represent data; 2) conventions for representing RPCs; and 3) bindings to HTTP.

SOAP offers application interoperability, which is advantageous when functions from different systems are being combined and when function calls cross platform, operating systems and/or language boundaries.

2.4.2 THE SEMANTIC WEB

Adding semantics (i.e. meaning) to the Web is a necessary component towards realizing a Web service's goal of application-to-application integration. To this end, the next generation Internet will be the Semantic Web, a ubiquitous computing resource that will offer AI-based services to both humans and machines (Ahmed, 2003). The following sections provide an overview and detailed discussion of the Semantic Web.

2.4.2.1 Semantic Web – Overview

Tim Berners-Lee, Director of the W3C, states that the vision of the Semantic Web is to associate meaning to all Web resources such that they can be discovered and consumed autonomously by applications (Berners-Lee, Hendler, & Lasilla, 2001), making the Semantic Web a meaningful indexed repository of documents and services (Lee, Patel, Chun, & Geller, 2004a). Currently, interpretation of Web-based information requires human knowledge and intuition; both humans and machines could interpret the Semantic Web.

Schweiger claims that because the Web and healthcare systems are both “little organized systems of distributed data”, innovations in Web technology are particularly relevant to the healthcare industry (Schweiger, Brumhard, Hoelzer, & Dudeck, 2005). As such, the development of the Semantic Web will impact healthcare in numerous ways, from retrieving information from multiple disparate databases so that patient mobility will not affect the continuity of individual care and transfer of patient information (Sun, 2004), to implementing electronic patient records that support the ‘Wherever I Go, It’s There’ (WIGIT) paradigm, to enabling machines to capture and provide clinicians with the information stored in clinical guidelines and scientific publications.

2.4.2.2 Semantic Web – Detailed Discussion

For many, the driving goal forwarding the development of the Semantic Web is to enable Semantic e-Business. There are numerous similarities between e-Business and e-Health, with developments in the former percolating through to the later. For instance, Semantic e-Business is defined as “an approach to managing knowledge for the coordination of e-business processes through the systematic application of Semantic Web technologies” (Singh, Iyer, & Salam, 2005a), a definition that can be equally well applied to e-Health which “encompasses much of medical informatics but tends to prioritize the delivery of clinical information, care and services rather than the functions of technologies” (Open Clinical (Knowledge Management for Medical Care), 2005).

There are three key components to realizing Semantic Web technology: knowledge representation, ontologies, and intelligent software agents.

2.4.2.2.1 Knowledge Representation

Knowledge representation provides structured collections of information and inference rules for automated reasoning (Singh, Iyer, & Salam, 2005b). Describing business processes and business rules using semantic knowledge representation languages allows intelligent reasoning by automated software agents. The Web Ontology Language (OWL) is a W3C standard for semantic knowledge representation (Singh, Iyer, & Salam, 2005b) based on the Resource Description Framework (RDF) and Description Logics (DL).

Knowledge Representation is also concerned with business processes, which are modeled by workflow descriptions that describe individual tasks at an atomic transactional level, such that semantic languages can describe the individual services offered by organizations. Ontologies can then be used to describe the links between the various resources (Singh, Iyer, & Salam, 2005b).

2.4.2.2.2 Ontologies

Ontologies provide a standard, shared vocabulary that represents the meaning of entities (Singh, Iyer, & Salam, 2005a) by defining the common words and concepts used to describe and represent an area of knowledge (Daconta et al., 2003). The notion of ontologies plays a central role in the Semantic Web, by providing a domain language that is useful to both humans and machines (Knublauch, Dameron, & Musen, 2003).

Central to the goal of ontologies is to provide the ability to enable formal reasoning. While some large healthcare systems use a common data model to integrate information from multiple facilities, the majority of existing clinical information is stored

in heterogeneous databases using different names and different data models (Sun, 2004). Although the Semantic Web provides a means to attach meaning to data, there is no guarantee that everyone agrees with this meaning. In order to deal with numerous and heterogeneous data formats, ontologies are needed to define healthcare standards and the mapping between them. Standard ontologies, schemas and vocabularies are a prerequisite for the Semantic Web (Schweiger et al., 2005). In describing their experience to create a healthcare information broker, Turner *et al.* state that future work will require an ontology service comprising multiple ontologies, as this is considered the only viable option for providing a global view of an information space (M. Turner et al., 2004). While efforts such as Health Level 7 (HL7) and Digital Imaging and Communications in Medicine (DICOM) are providing a common ground for describing healthcare information, they are not currently at a level to resolve all relevant interoperability issues (Y. Lee, Patel, Chun, & Geller, 2004a).

OWL is a standard language for sharing Semantic Web contents (Knublauch et al., 2003) and is now a W3C recommendation (i.e. a standard like HTML and XML). Concepts are stored in an OWL file that resides on a publicly accessible Web server. Other OWL ontologies, agents and services can link to this file and use the ontology's concepts (Knublauch et al., 2003). OWL can use the linking provided by RDF to allow ontologies to be distributed across systems; in this way OWL is specifically engineered for the Semantic Web (Shadbolt, Berners-Lee, & Hall, 2006).

2.4.2.2.3 Agents

An intelligent agent is a system that perceives its environment and acts upon the

information it receives. Intelligent agents interpret and exchange semantically enriched knowledge for users (J. Hendler, 2001). By extending the existing Web with conceptual data readable by machines, this knowledge can then be used by software agents to perform tasks. In this way, developments in semantic technologies make Semantic Web content computer interpretable and amenable to agent interoperability and automated reasoning techniques (McIlraith, Son, & Zeng, 2001).

2.4.3 SEMANTIC WEB SERVICES

Semantic Web services are created by adding explicit semantics to Web service descriptions, providing a higher level of automation in the location and use of Web services (McIlraith et al., 2001). One of the key ways in which the Semantic Web and Web services interact is to facilitate the automated composition of Web services. This requires descriptions of the service capabilities and input/output parameters; as well as indicating in what ways and in what situations the service should be used (Lee, Patel, Chun, & Geller, 2004b).

A Semantic Web service is described based on its functionality (the capability) and the interface (de Bruijn et al., 2005). The capability describes the service functionality in terms of: assumptions about the state before execution of the service, preconditions for the information space before execution, effects of the service on the state, and post conditions describing the information space after execution. The interface describes the kind of input and output messages processed by the service and defines how core services interact to achieve the functionality of a composite service. The service is advertised or published through a service repository.

CHAPTER 3. CRITICAL ANALYSIS

3.1 INTRODUCTION

Section 3.2 presents a critical analysis of the state of the art in perinatal risk assessment, with particular emphasis placed on the current role of clinical decision support in perinatal care. Section 3.3 discusses the value of developing a new approach for preterm birth risk assessment. Section 3.4 presents a critical analysis of the state of the art in Semantic Web services for healthcare, highlighting the differences between the approaches taken in this thesis and current initiatives. This is followed by Section 3.5, which discusses the merit of creating the Semantic Web services framework for perinatal risk assessment.

3.2 EVIDENCE THAT PROBLEM #1 HAS NOT PREVIOUSLY BEEN SOLVED

3.2.1 LIMITATIONS OF CURRENT PREDICTION APPROACHES

Methods for early prediction of PTB based on a) obstetrical history, demographic factors and premonitory symptoms or b) risk-scoring systems based on selected epidemiological, historical, and clinical markers have proven to be insensitive and non-specific (Lockwood & Kuczynski, 2001), indicating that new strategies are needed to successfully achieve early prediction of preterm birth. The following sections describe the current approaches to PTB prediction, and discuss the limitations of each.

3.2.1.1 Risk-Scoring Models

Using the term 'risk model' is misleading as there is presently no universally accepted model to assess maternal risk of spontaneous PTB – approximately 50% of preterm births occur in women with no apparent identifiable risk factors (Aerts & Iams, 1999) and attempts to predict PTB have been largely unsuccessful (Thorsen et al., 2001). An initial attempt to predict PTB in asymptomatic patients using risk-scoring models was offered by the Creasy risk scoring system. The Creasy system was based on a prospective analysis, and predicted approximately 10% of the population would give birth prematurely, however, only one third of these patients delivered before term (Creasy et al., 1980), also the system performed better in predicting multiparis⁵ patients. In general, "risk-scoring schemes tested in both large unselected and in high-risk obstetric populations have not provided accurate means for classifying pregnant women according to their preterm prediction risk" (Johnston et al., 2001). In a review of previous studies, the sensitivity of risk assessment scoring tools ranged from 17%-38% in their ability to predict PTB (McLean, Walters, & Smith, 1993). The highest sensitivity was achieved using sociodemographic factors to develop a risk scoring system for PTB prediction, with a high specificity (98.6%) for primiparas⁶ and 95.05% for multiparas; however, the sensitivity was much lower at 38% (Tan, Wen, Walker, & Kitaw, 2004). The low sensitivity limits the clinical effectiveness of this model.

Tan's team used logistic regression to create a model to predict preterm birth in multiple pregnancies (Tan et al., 2004), note that the preterm birth rate is much higher in

⁵ A woman who has been pregnant two or more times.

⁶ A woman who is pregnant for the first time.

multiple births, as compared to singleton births. In focusing on multiple births, the model ceases to be applicable to all pregnancy cases and is not a broadly applicable screening tool. The model included the variables: plurality, age, race, marital status, parity, smoking, weight gain per week, medical complications, and paternal age. The sensitivity, specificity and positive predictive values were 43.2%, 81.7% and 24.9% respectively. The area under the ROC curve was 0.680.

The general consensus in the literature is that current statistical tools cannot effectively be applied to determine the complex associations between epidemiological, clinical, biochemical and biophysical variables, because preterm birth is such a multifactorial problem (Lockwood & Kuczynski, 2001).

3.2.1.2 Invasive Diagnostics

In the obstetrical community, the most commonly applied predictive technique for estimating PTB is based on obstetric history examined in conjunction with fetal fibronectin testing⁷ and/or cervical factors (Goldenberg et al., 1998). Fetal fibronectin testing is most commonly used for: (1) symptomatic patients (i.e. women who report contractions between 24 and 36 weeks gestation), who have a cervix that is less than 3 cm dilated, and are being evaluated for risk of preterm delivery; and (2) at-risk women for whom a negative test result could eliminate the need for intervention (Aerts & Iams, 1999).

Using fetal fibronectin testing and obstetrical history, Goldenberg et al. were able

⁷ The presence of fetal fibronectin in the cervico-vaginal secretions of symptomatic women during weeks 22 through 34 of gestation indicates an increased risk of preterm delivery (March of Dimes, 2006).

to predict preterm birth in a population of mothers having an elevated risk of PTB (i.e. those who had a history of previous preterm birth) with a sensitivity of 63% (Goldenberg et al., 1997), this study considered singleton births only. Crane et al. found that a combination of positive preterm birth-risk score and fetal fibronectin testing had a sensitivity of 44.4% in a heterogeneous population, although the authors concede the results are limited by the small sample size (140 patients) (Crane et al., 1999). Similar results were achieved by Peaceman et al., who employed fetal fibronectin testing to achieve a sensitivity of 43.9% in a heterogeneous population (Peaceman et al., 1997).

The National Institute of Child Health and Human Development Maternal-Fetal Medicine Units Network's Preterm Prediction Study found that the strongest predictors of PTB are fetal fibronectin and cervical length (Goldenberg et al., 2001). Using these two indicators (i.e. fetal fibronectin testing positive and cervical length ≤ 25 mm), researchers were able to predict PTB in women with a history of prior preterm delivery (i.e. a maternal population having an elevated risk of PTB) with a sensitivity of 64.7% (Iams, 1997; Greenberg 1998). Cervical length alone, measured by cervical ultrasonography, fails to identify over two thirds of the deliveries that occur before 35 weeks gestation (sensitivity: 31%), and can therefore not be considered a good indicator of PTB (Hirsch, 2006).

Further work by Iams et al. attempted to use the same combination of fetal fibronectin testing and cervical length measurement, employed on the elevated risk mothers, to estimate PTB in a low-risk maternal population prior to 35 weeks gestation (Iams, 2001). The reported sensitivity was low (15.6%), indicating that the model is not

applicable to low-risk mothers. The authors conclude that no screening test for preterm birth in low risk mothers, other than obstetric history, can be recommended.

The American College of Obstetricians and Gynecologists suggests that obstetric tools to identify patients at risk for preterm labor may be most helpful in identifying women at low risk: “ultrasonography to determine the cervical length and/or fetal fibronectin screening may be useful in determining who is at high risk for preterm labor, but the benefit may be mainly to rule out those who are not at high risk for preterm birth” (American College of Obstetricians and Gynecologists, 2001).

Using both fetal fibronectin measurements and cervical length, physicians can predict PTB with a sensitivity approaching 65% in an elevated risk maternal population (Goldenberg et al., 1998). However, current evidence does not support the use of cervical length measurement or fetal fibronectin status in screening for risk of preterm delivery in a low-risk population (Aerts & Iams, 1999). Additionally, these tests do not predict risk of PTB early in pregnancy; they only become more sensitive in estimating PTB as the estimated date of birth approaches. This is because fetal fibronectin testing commonly predicts preterm delivery only 7-10 days before advanced cervical dilation, leaving little time for decision making in terms of prenatal treatments (Tan et al., 2004). Based on physician input from Dr. C.R. Walker (formerly neonatologist at the Children’s Hospital of Eastern Ontario and currently Vice-President Medicine at the IWK Health Centre) and from Dr. S.W. Wen (clinical epidemiologist at the Ottawa Health Research Institute), it was decided that the most effective PTB risk assessment model would be able to predict PTB risk before 23 weeks (23 weeks being the earliest age at which survival is usually

possible) and would be applicable to the entire population, not just a subset of symptomatic patients. Fiore corroborates this idea: “Most interventions designed to prevent preterm birth... are not universally effective and are applicable to only a small percentage of women at risk for preterm birth. A more rational approach to intervention will require a better understanding of the mechanisms leading to preterm birth” (Fiore, 2003). Ideally, women at risk of preterm birth could be accurately identified early in the pregnancy using readily available obstetrical and socio-demographic information, allowing preventative measures to be taken (Wen, 2003).

The low-cost, simple to use, and broadly applicable hybrid approach described in this thesis is proposed as a complimentary technique to the current procedure of using costly and invasive clinical testing on selected elevated risk maternal populations. Due to a lack of complete obstetric data, it is often difficult to perform risk estimation based on sophisticated medical markers and extensive medical histories, as published in large clinical trials (Goldenberg et al., 1998). Preterm prediction models based on detailed clinical and laboratory data from large centers often have difficulties generalizing to the population level (Tan et al., 2004). Tan also concludes that preterm prediction models based on vital statistics and hospital discharge data, although traditionally having a lower sensitivity, are more generalizable to the population at large. This is a very desirable feature, since ideally preterm prediction should be applicable to population data, rather than only pre-identified elevated risk groups.

3.2.1.3 Information Technology-Based Approaches

Three groups have published significant research applying information

technology-based approaches to the prediction of PTB. This section will detail the goal, methodology and results of each research effort, followed by a discussion of the approach, including the limitations.

3.2.1.3.1 *Lockwood and Kuczynski*

Goal: The aim of this research is to create a risk-analysis tool based on combining biophysical and biochemical markers of preterm delivery with optimal clinical and epidemiological predictors (Lockwood & Kuczynski, 2001). The goal is to identify at-risk asymptomatic patients with high sensitivity and specificity.

Overview: The model is based on using logistic regression models combined with artificial neural networks (NevProp artificial neural network program developed by the University of Nevada School of Medicine) to assign appropriate weights to each of the pathogenic pathways, with the aim of pinpointing the relative contribution of each pathogenic process. The research project is ongoing, based on long-term observational studies of numerous clinical factors obtained between 18-32 weeks gestation, as well as detailed informational surveys of epidemiological, sociodemographic, psychosocial, and obstetric factors, based on prior sociodemographic surveys of risk.

Methodology: The methodology proceeds in a step-wise fashion as follows:

- 1) Employ univariate logistic regression to assess the significance of each factor;
- 2) Use multivariate logistic regression and an artificial neural network to assess each risk factor and each multivariate model (then compare the results from each);
- 3) Measure accuracy by the area under the ROC curve.

Discussion: The main goal of this research “to identify at-risk asymptomatic

patients with high sensitivity, specificity” is very similar to that of this thesis, and both use an ANN tool as the basis of their techniques. However, Lockwood *et al*’s work takes a different approach in that it depends largely on complex biophysical and biochemical data, not generally available in low-risk pregnancies. Other differences include: Lockwood *et al*’s work is not a population screening system as data comes from elevated risk women only; the hybrid system proposed is based on combining the ANN with logistic regression, rather than using a MLP ANN with risk stratification estimations, nor is a kNN used to maximize knowledge from similar past cases; when selecting a minimal dataset the system does not present all data to the ANN and then allow the ANN to identify the minimal dataset, but rather pre-selects the variables that will be presented to the ANN based on statistical analysis (an approach that does not make full use of the ANN’s ability to identify significant variables in non-linear systems).

3.2.1.3.2 *Goodwin and Mayer*

Goal: Investigate the success of different data mining techniques when applied to the prediction of PTB

Overview: Using Duke Medical Center’s TMRTM data repository, the authors created a data warehouse to be used for predicting PTB. The PTB rate is high, at approximately 22% for the training and test sets (more than double the PPESO PTB rate), but this is typical of the elevated risk referrals in academic medical centers.

Methodology: Goodwin and Mayer developed custom data mining software (Clinical Miner) to facilitate exploring the perinatal database and created a modeling technique (FactMiner) for assessing the amount of information in a database in relation to

a particular outcome and for performing binary classification. The variables included 32 demographic variables and 393 clinical variables measured at each prenatal visit. The ROC curve was given as the measure of performance.

Results: The neural net with demographic variables only had an area under the ROC curve of 0.64 (which classifies it as an average predictor based on the classification scheme presented in Table 4-3); the neural net with demographic variables plus all clinical variables had an area under the ROC curve of 0.66.

Discussion: An important aspect of this work in relation to this thesis lies in the finding that contrary to earlier studies, most of the predictive power from the database (which contained a total of 1233 variables) came from 32 demographic variables, with only insignificant improvements to clinical accuracy when hundreds of clinical variables were added. This corroborates the approach used in this thesis of supplying the ANN with demographic and general obstetrical information, to create a broadly applicable, minimally invasive first screening technique, rather than relying on sophisticated and costly clinical testing. Due to the proprietary nature of the software being developed, the methodology does not supply parameter details for the ANN. The reported results with an area under the ROC curve of less than 0.7 do not meet the minimum performance criteria for ROC values specified in this work. Other differences with this work include: their focus is not on early prediction of PTB and their ANN model is only applicable to elevated risk mothers and cannot be applied to the general population.

3.2.1.3.3 *Van Dyne, Woolery, Gryzmala-Busse, Tsatsoulis*

Goal: Apply machine learning and expert systems to predict PTB with at least a

50% overall accuracy rate.

Overview: Uses inductive learning, where the classifications of a set of example patients are known; to discriminate between different classifications.

Methodology: The system employs the LERS induction program (Gryzmala-Busse, 1989), based on rough-set theory, which has as input a decision table of entity-attribute-value entries. Rules are extracted from the decision table to form the foundation of an expert system that classifies patient cases as preterm or full term gestation. Rough-set theory handles inconsistencies in creating the expert rules, as is the case where two cases have the same input data but different outputs, by providing certain rules for consistent examples (i.e. with a rule confidence of 100%) and possible rules for all other examples. A percentage confidence is given for each possible rule. The prototype system contained 520 rules and experts confirmed the extracted rules prior to inclusion. Objects were defined to store the patient data structure; a patient object represents all attributes found in the dataset and has 48 possible input parameters. System accuracy was measured by the number of times the predicted outcome matched the actual patient outcome. A prioritization scheme was developed for rules in the prototype expert system, shown in Equation 3-1, where the rule base accuracy rate is the accuracy with which LERS predicted gestational outcomes on a training set.

$$\begin{aligned} \text{Rule priority rating} = & (\text{rule base accuracy rate}) * (\text{rule} \\ & \text{confidence}) + (10 * \text{no. of cases used to generate the rule}) \\ & - (10 * \text{no. of cases misclassified by the rule}) \end{aligned}$$

EQUATION 3-1 RULE PRIORITIZATION

Results: Results were given for three different datasets; the accuracy in predicting

PTB using each dataset is: 1) 90.21% for certain rules and 80.1% for possible rules, 2) 66.17% for certain rules (there were no inconsistent results in this dataset so no possible rules were needed), and 3) 54.36% (all inconsistent examples were removed from this dataset).

Discussion: It is worthwhile to note that the authors found that rules generated from very large datasets have an advantage over rules generated from smaller datasets. This helps to justify the generalization abilities of the model developed in this work, which is based on two large databases of over 20,000 cases each. The authors also claim that accuracy is a poor measure of performance when so many women deliver full term, since this provides a false sense of highly accurate predictions while failing to detect women at risk for PTB.

Limitations of Van Dyne *et al.*'s work include the fact rules are generated for each database separately and each database is tested separately. Because different datasets contained different information, and rules were created for each dataset, the published results are difficult to use as a basis of comparison with other work. The authors state that rules need to be tested further to validate applicability to other datasets.

The published results were based on whether predictions were made from the 'certain' rule set or the 'possible' rule set, making it impossible to reproduce results since the authors do not provide the minimal dataset used for any of these datasets. Also, there is no consistency between the results achieved for the three datasets.

Again, this work only classifies women that had been previously identified as high risk for PTB, so the technique is not applicable to the entire population. It is not

known whether the focus is on early prediction as the minimal dataset variables are not provided, therefore it cannot be determined whether they are available before 23 weeks gestation.

3.2.2 OTHER HYBRID SYSTEMS USED IN CLINICAL OUTCOME PREDICTION

The previous section demonstrated that prior approaches to the early prediction of PTB differ significantly from the hybrid methodology proposed in this thesis. This shows that the problem of predicting PTB is both a significant one, and one that has not been previously solved in the manner outlined in this work (or been previously solved with sufficiently successful results to warrant any further development necessary). However, a further question remains: is the hybrid AI-based system for clinical outcome prediction developed in this work novel outside of the domain of preterm birth prediction? This section examines other clinical decision support systems that can be considered 'hybrid systems' based on their combined use of ANN (or multiple ANNs), kNN and DT technology, either individually or in an integrated manner. The focus is on contrasting similar clinical decision support systems used in the domain of soft computing with that developed in this thesis.

3.2.2.1 Hybrid Systems Based on Combining ANNs and CBRs

Ennett examined ANN-CBR hybrid systems extensively. She explains that in relation to ANN-based series collaborations, traditionally researchers have taken the approach of using the nonlinear modeling ability of ANNs to strengthen the reasoning ability of expert systems, such as case-based reasoning systems (Ennett, 2003).

Ennett developed an ANN and CBR hybrid system for imputing missing values that employed weights from a classification-based MLP ANN with one hidden layer as the CBR connection weights. The modeling strengths of the ANN are used to select the matching weights of the CBR and the CBR's query-matching ability permits the use of cases with missing values in the ANN to model the system (C. Ennett, 2003). Shin et al. proposed a system similar to that of Ennett, where they used the values of the connection weights from a three-layer neural network to calculate associated values for use as weights in a memory-based reasoning system (i.e. a CBR system) (Shin, Yun, Kim, & Park, 2000).

Chen and Burrell describe various ANN-based CBR systems (D. Chen & Burrell, 2001); these can be considered CBRs with a neural network structure. Examples of such systems include: Reategui et al.'s work to create a hybrid system which uses a neural network to make hypotheses that guide the CBR system to find a similar case in the case base (Reategui, Campbell, & Leao, 1997); and Myllymaki and Tirri's system which uses a probabilistic neural network and a CBR to replace missing values in the input vector with the closest-matching case (Myllymaki & Tirri, 1993).

The ANN-CBR integration in this work is most closely related to that described by Ennett, in that the classification-based MLP ANN connection weights are used to define the CBR weightings (although use of the modified relative weights technique for determining the meaning of the weights implies that the ANN in question can have any number of hidden nodes, whereas Ennett's technique was limited to MLP ANNs with one hidden layer and one hidden node). However, the hybrid system used in this work is not

meant to replace missing values, but rather to employ MLP ANN weights in the kNN-based CBR tool to aid in classifying cases as true positives or true negatives.

3.2.2.2 Hybrid Systems Based on Combining (ANNs (and/or) CBRs) with Decision Trees

In the hybrid system designed in this work, multiple ANNs are integrated with the CBR system. The results are then fed into a decision tree to generate the final model, essentially linking the decision tree with both the ANN and the CBR. This section examines combinations of ANNs and/or CBR tools, with decision trees, discussing similarities and differences with the hybrid system proposed in this thesis.

Neural networks and decision trees are two of the most popular approaches to solving classification problems. The majority of the published research takes the approach of developing a decision tree that is then used to create a neural network. There are significantly fewer references in the literature on systems combining CBRs and DTs, either in a hybrid or parallel manner. The general practice seems to be for the CBR and DT to work in parallel and their results judged by a voting system. The literature review found no hybrid system containing the same methodology presented in *Chapter 4*.

3.2.2.2.1 Decision Tree -> Neural Network Combinations

Tsujino *et al.* (Tsujino, 1995) constructed a neural network based on the result of DT induction. The authors claim that the combination of ANNs and DTs is so complementary that it is possible to obtain the advantages of both while eliminating the individual disadvantages. Specifically, the DT is far easier to understand for human beings than an ANN. Also, it is possible to use the ANN to extract hidden knowledge and

then express it explicitly in the form of a DT (Tsujino, 1995). They state that placing the DT first in the hybrid system allows them to achieve a reasonable initial network structure. However, with the classification-based MLP ANN, the goal is not to restrict the initial structure, but rather to make all variables available to the network and then, if possible, to remove insignificant variables via the relative weights method.

Other work that develop neural networks from DTs include: Yoo and Sethi's conversion algorithm for translating decision trees to radial basis function neural networks (Yoo & Sethi, 1995); and Lee *et al.*'s hybrid ANN-DT system for automated cervical smear screening (J. S. Lee, Hwang, Davis, & Nelson, 1991), consisting of a decision tree which clusters the training data into distinct sub-classes which are then fed to the neural network.

3.2.2.2.2 *Decision Tree <-> Neural network Combinations*

Tsirogiannis *et al.* (Tsirogiannis et al., 2004) used a multi-level combination scheme for medical classification. They claim that combinations of neural networks and DTs exhibit better performance than single classifiers for complex problems (Tsirogiannis et al., 2004). Using four different medical classification problems, they show that their multi-level approach is effective in reducing the error compared to each stand-alone method. Their work refers to ensemble systems (where the problem is split based on the divide and conquer principle and a different classifier is used in parallel for each region of the problem), which is a different form of collaborative system than the hybrid system proposed in this work. The authors use bagging (where the classifiers are trained on bootstrap replicas of the training set and then combined using a voting

algorithm) and boosting (where the training sets are formed based on minimizing the error and a weighted voting algorithm is used depending on whether the training set is categorized as difficult or easy). The three levels in the multi-level system are as follows: 1) the bottom layer comprises the single classifiers (such as stand-alone ANNs and DTs); 2) the middle layer holds the combination methods (such as bagging and boosting); 3) and the top layer contains the combinations of the bagging and boosting methods. The third level uses a weighted voting scheme where the weights correspond to a reliability measure assigned to each method; reliability is defined as $1 - \text{test error rate}$. Although this ensemble system is clearly different from the hybrid system described in this thesis, the research corroborates the idea that combinations of different AI techniques can be used to outperform stand-alone approaches to classification. Similarities with this work include the voting algorithm used after passes through ANN structures.

Kijsirikul *et al.* (Kijsirikul & Chongkasemwongse, 2001) used neural networks to prune decision trees to minimize over-fitting. Over-fitting of DTs occurs when the size of the tree is too large in relation to the number of training cases (Kijsirikul & Chongkasemwongse, 2001). Their method assigns weights to nodes in the decision tree based on their significance, which is determined by a back propagation neural network. The DT is first converted to rules, and then the neural network is constructed from the rules, followed by the training of the network. Finally, the original DT is pruned based on the ANN weightings. This hybrid system combines both ANN and DTs. Other similarities with this thesis include the use of ANNs to specify the weights in another AI technique. However, this thesis uses ANN weights to assign significance to kNN weights

(which are then used indirectly by the DT), rather than creating the DT first and then pruning using the ANN.

3.2.2.2.3 Neural Network -> Decision Tree Combinations

Wang *et al.* (Wang et al., 2005) applied a hybrid ANN-DT technique to the problem of breast cancer classification. They first use neural networks to obtain a minimal dataset (they refer to this as filtering the noise data in the training set); a decision tree is then employed to extract rules from the pruned training set. Wang *et al.* explain that when DTs are applied to large databases, the number of rules can become excessively large and errors can accumulate from level to level in a large tree (Wang et al., 2005).

Extracting rules from trained ANNs is a popular mechanism for making the ANN results understandable to humans. Many different algorithms have been developed for this task (Boz, 2002), (Setiono, Leow, & Thong, 2000), (Zorman & Kokol, 2002).

Another sophisticated technique is to embed neural networks directly into the decision tree; examples include: NNTree, a DT where each node is an expert neural network (Zhau, 2001); and CHDT where neural networks are embedded in leaf nodes of the decision tree (Z. Zhou & Xiang, 2001). It was decided that this approach was not applicable to the classification problem of PTB, where the number of available variables and cases would make this approach unrealistic (48 variables and over 30,000 available cases).

3.2.2.2.4 *Decision Tree -> kNN*

Aslandogan and Mahajani combine the ‘beliefs’ of three different classifiers, k-Nearest Neighbor, Naïve Bayesian and Decision Trees, and use Dempster’s rule of combination (Shafer, 1976) to arrive at one output (Aslandogan & Mahajani, 2004). Their technique is applied to the problem of breast tumors and skin lesions. They conclude that outcomes based on combined beliefs show better classification results than any single classifier, and that relying on a single classifier may lead to unacceptable misclassifications, corroborating the claim of Tsirogiannis *et al.* This work is a parallel system, where each classifier performs independently; uncertainty is then evaluated and compared for each classifier.

Hullermeier proposes a framework of experience-based decision making which extends case-based decision making (Hullermeier, 2005). In the framework, an agent faced with a new decision uses a CBR to gain experience from past problems. However, rather than only presenting information from past cases, the framework both 1) predicts the utility of potential actions; and/or 2) establishes a relationship between decision problems and past actions. These two functionalities are realized using decision trees. This work represents one of the few examples of a truly hybrid kNN->DT systems, where the results from the kNN are then fed into the DT.

3.3 VALUE OF DEVELOPING A NEW HYBRID DATA MINING APPROACH FOR PREDICTING PTB

Prospectively identifying high-risk and/or low-risk mothers and subsequently adjusting healthcare management interventions based on the patient’s risk stratification

levels (such as offering antenatal steroids (Tan et al., 2004)) can result in improvements in the outcomes for these patients (Kudyba et al., 2005).

Traditionally, diagnosticians are most interested in increasing preterm birth accuracy (i.e. sensitivity). Full term birth accuracy (i.e. specificity) is considered to be of secondary value (Grzymala-Busse et al., 2003). The argument can be made that because approximately 50% of PTBs occur in mothers with no risk factors and approximately 50% of mothers identified as elevated risk do not give birth prematurely, physicians are currently unable to accurately estimate either PTB or full term birth. Of course, a model that maximized both sensitivity and specificity would be ideal; however, there is value in developing a model favoring improved sensitivity. An effective risk assessment tool would accurately identify the majority of mothers in need of further testing, while reducing the number of low-risk mothers subjected to this testing, adhering to the principle that increased predictive capability ultimately helps reduce inefficient allocation of healthcare resources (Kudyba et al., 2005).

3.3.1 OTHER APPLICATIONS

In addition to predicting preterm birth, there are numerous other perinatal outcomes of interest to clinicians, including: low-birth weight, low APGAR score, requirement for advanced neonatal intensive care, congenital malformations, and infant mortality. Neonatal intensive care outcomes include: requirement for artificial ventilation; length of stay; and possible complications. Of particular interest from a maternal perspective is the risk of cesarean section. Figure 3-1 indicates the chronology of these outcomes during the perinatal period. MIRG has previously studied prediction

of APGAR5 score, infant mortality and delivery mode (cesarean or vaginal) (Frize et al., 2004). The hybrid data mining system developed in this thesis is designed to be generalizable: depending on the required outcome, the appropriate minimum data set is retrieved and the approach is similar regardless of the outcome. In this way, the system can be extended to become a sophisticated perinatal risk assessment tool, providing clinicians with complete evaluation of perinatal risk for many possible outcomes.

FIGURE 3-1 PERINATAL CLINICAL OUTCOMES OF INTEREST

3.4 EVIDENCE THAT PROBLEM #2 HAS NOT PREVIOUSLY BEEN SOLVED

3.4.1 OVERVIEW OF STATE OF THE ART AND CRITICAL ANALYSIS

This section surveys existing research in the emerging area of Semantic Web services-based healthcare systems (including ontology development). While this thesis focuses on clinical decision support within the perinatal care environment, the survey

presents relevant initiatives from healthcare in general (e-health), as well as other relevant domains such as e-government and e-business.

3.4.1.1 State of the Art in Semantic Web Service-based Systems

In March 2006 the IEEE-SA Health IT Standards Group⁸ convened in Ottawa. The focus was on identifying gaps in current health IT standards. One of the major gaps was found to be the need for generalizable *Expert Systems Frameworks* for healthcare, particularly to support AI-based clinical decision support. Currently, Semantic Web services are being advocated as a logical means of achieving open healthcare systems; applications span a range of healthcare domains (Catley et al., 2004; Chatterjee, 2003; Y. Lee, Patel, Chun, & Geller, 2004a; McGregor, Heath, & Wei, 2005; M. Turner et al., 2004), and encompass services with diverse goals, from knowledge management to application integration. However, the literature review found no framework presently existing for supporting Semantic Web service-based diagnostic clinical decision support. The framework proposed in this thesis offers a set of Semantic Web-based applications that place the practitioner at the center of an architecture that enhances patient health services. The framework promotes the ‘WIGIT’ paradigm, that is: ‘Wherever I go, it’s there’; where the user is the practitioner, and the ubiquitous applications are the clinical decision support services. The framework also supports low-cost, flexible, real-time clinical decision support at the point of care.

⁸ <http://grouper.ieee.org/groups/hit/>

3.4.1.1.1 *Semantic Web Services for Medical Service Flows*

A notable composition initiative is offered by Lee *et al.*'s research on Semantic Medical Services (Y. Lee, Patel, Chun, & Geller, 2004a; Y. Lee, Patel, Chun, & Geller, 2004b). Their team is determining how heterogeneous medical Web services interoperate in a medical service flow for the cardiovascular domain. The research has three main goals:

1. Develop a model to represent Web service compositional knowledge based on syntactic, semantic and contextual knowledge.
2. Show how heterogeneous medical services interoperate in a medical service flow composed using these three forms of knowledge
3. Perform ontology integration to enable the automatic composition of services

Traditionally, services have been defined based purely on syntactic and semantic constraints; Lee *et al.* propose adding contextual constraints to the knowledge mix. Contextual (or pragmatic) knowledge is a relatively new concept in the area of Semantic Web services. Contextual knowledge identifies the situations in which a service should be used. For example, if a service for choosing a blood lab returns multiple options, select a lab closest to the patient, with the fastest turn around time and lowest cost. Contextual constraints are sophisticated and involve detailed knowledge of the patient's preferences. Lee *et al.* explain "[contextual knowledge] is not about the service itself, but about the way the service relates to the satisfaction of the goals of the consumer of the service" (Lee, Patel, Chun, & Geller, 2004b). Because each user has potentially different goals, contextual knowledge is hardest to formalize as it depends largely on the context (user's

circumstances) in which the service is executed.

Lee's work focuses on Semantic Web services for clinical knowledge management; the key functionalities involve exchanging large volumes of cardiovascular clinical data including administrative data, procedural data and clinical studies. The challenge focuses on integrating information from distributed organizations. Web services include: medical code lookup, patient record validation, and image retrieval. In creating the complete Semantic Web Services for healthcare framework, this thesis expands on Lee's idea of incorporating contextual knowledge into service composition, to meet the needs of diagnostic clinical decision support service integration.

3.4.1.1.2 Semantic Web Services for Clinical Decision Support

A sophisticated healthcare example involves using Semantic Web services to reduce the waiting times for non-critical surgery (Motta, Domingue, Cabral, & Gaspari, 2003), by providing five interacting Semantic Web Services: (1) a diagnostic service to diagnose conditions based on a set of symptoms, (2) a yellow page service indicating which hospitals in Europe provide specific medical services, (3) a cost query service to provide the cost of medical services on a per hospital basis, (4) an ambulance service to determine the cost of shipping patients between hospitals, and (5) an exchange rate service for converting between European currencies. This research is noteworthy in that the services combine to provide added value to the users. The Semantic Web services for healthcare framework draws on this idea by integrating individual data mining applications to form the hybrid screening system for perinatal risk assessment, where the integration of multiple AI-based services yields better results than each service is able to

achieve individually.

3.4.1.1.3 Web Service Composition on the Semantic Web in e-government

Medjahed's team wrapped welfare applications in modular Web services to allow users efficient access to e-government services (Medjahed, Bouguettaya, & Elmagarmid, 2003; Medjahed, Benatallah, Bouguettaya, Ngu, & Elmagarmid, 2003). The Web services combination was based on a set of rules that check composability of services; in e-government applications, as in healthcare, features such as service composition must be transparent to the user. E-government is an example of another highly distributed domain in which it is beneficial to allow seamless access to all related services through a single access point.

3.4.1.1.4 Web-Based Decision Support for Personalized Risk Assessment

EsPeR is a Web-based decision support system for screening, risk assessment and prevention (Open Clinical, 2005). For a given individual, it provides a personalized assessment of his/her main health risks and offers recommendations for controlling or reducing these risks based on the patient's individual profile and the latest French national evidence-based clinical practice guidelines. EsPeR has four main functions:

1. To provide information on the 10 highest risks of avoidable causes of death according to demographic characteristics (age, sex and region of residence).
2. To provide an assessment of the specific risks affecting each individual based on published clinical practice guidelines and risk prediction models.

3. To provide information on cardiovascular risk management, cancer screening, tobacco smoking cessation, alcohol addiction, depression and suicide.
4. To offer a printable summary of advice for physicians and patients of information designed to reduce health risks.

The guideline model used in EsPeR represents the decision tree as an XML Document Type Definition, which is read and executed by a custom-built inference engine. Decision tables are used to validate the decision algorithm and decision trees specify the algorithm.

EsPeR is one of the few examples in the literature of a Web-based system in actual clinical use where clinical decision support is offered via Web-based applications. Currently, EsPeR does not incorporate Semantic Web technology.

3.4.1.2 State of the Art in Clinical Ontologies

One of the major drivers of shared semantics has been e-science and the need for integrating diverse and heterogeneous datasets. This is being achieved in large part by the adoption of ontologies (Shadbolt et al., 2006). Ontologies are currently being defined for many biomedical domains such as anatomy, genetics, and cancer research (Knublauch et al., 2004). A major cancer initiative, the National Cancer Institute's Thesaurus project, is currently under development, to structure cancer-related resources in a machine readable, meaningful way (Golbeck et al., 2003).

A thorough literature review has shown that no formalized and generalizable ontology for clinical decision support currently exists. Clinical ontologies have been proposed, and in some cases developed, in the areas of: clinical practice guidelines (Yale

Medical School, 2005), clinical trials (Sim, 1995) and clinical pathways (H. Chen, Colaert, & De Roo, 2004).

One of the most sophisticated ontologies is offered by <GEM> (Guidelines Element Model) (Yale Medical School, 2005). <GEM> is an XML document which is used as a template to model clinical practice guidelines in XML by translating natural language guidelines into a format that can be understood by computers. The idea is that <GEM> formalizes knowledge used throughout the guideline lifecycle, to model information relating to developing, disseminating, implementing and maintaining clinical guidelines. The <GEM> Document Type Definition (DTD) was balloted as an international standard for the representation of practice guidelines in XML format and has become ASTM standard E2210-02. However, <GEM> does not provide an ontology for diagnostic clinical decision support. An example of a CDSS using clinical practice guidelines is offered in (Shiffman, Michel, Essaihi, & Marcy, 2004), which describes a knowledge-based system for practitioners counseling patients for smoking cessation. The system translates natural language guidelines using <GEM>, information such as decision variables and action items (what, whom, how, when, why) are translated into a computerized clinical decision support workflow detailing the course of action practitioners should follow for different patients.

During the IEEE-SA Health IT Standards Group Ottawa meeting, Waegemann⁹ proposed a nine-layer health informatics standards 'stack', shown in Table 3-1, with decision support systems occupying the top-most layer. In this stack, decision support can

⁹ C.Peter Waegemann, Immediate Past Chair, Committee E31 on Health Informatics

refer to any system that offers support for: admission systems, eligibility, diagnostic support, order entry and test results, etc (MIRG's clinical outcome prediction tools fit into this layer). Waegemann stated that because decision support is such a broad field where new applications are continually being added, the goal is not to achieve a standard, so much as defining vocabularies (referred to as ontologies in this work). He concluded that there are no known standards in the field of diagnostic decision support.

TABLE 3-1 WAEGEMANN'S STANDARDS AREAS FOR HEALTH INFORMATICS

LAYER	DESCRIPTION	STANDARDS
1	Modeling	HL7, CEN TC 251, OpenEHR, DICOM, others
2	Imaging	DICOM
3	Messaging & Communication	HL7, X12, NCPDP, IP/XML
4	Confidentiality/Security	ASTM E31, ISO TC 215, CEN TC 251, Others
5	Financial/Management	Many by country, system
6	Health records and care functions	ASTM, HL7, many others
7	Identification systems	Governments, ASTM, others
8	Medical Concept representation	SNOMED, LOINC, IDC 9, IDC 10, many others
9	Clinical Decision Support	Various proprietary vocabularies

The meeting concluded by identifying five current gaps in health informatics vocabularies and standards; the most significant gap was that in medical expert systems, particularly AI-based initiatives.

3.5 VALUE OF DEVELOPING A SEMANTIC WEB SERVICES-BASED FRAMEWORK FOR HEALTHCARE

Although there is a need for clinical decision support systems to aid physicians in making optimum diagnosis and reducing medical errors (Committee on Quality of

Healthcare in America, Institute of Medicine, 2000; Wilson, Runciman, & Gibberd, 1995; Rideout, 2005), from a technical perspective the advancement of CDSSs has been severely hampered by two key factors:

1. Clinical data is intended for use by humans only; even in the case of the electronic patient records, the data has no attached semantic meaning and is primarily intended for human viewing
2. Clinical data needed to make decisions are heterogeneous, fragmented and distributed geographically, making it difficult to train and deploy AI systems on large medical databases covering broad population demographics for the simple reason that it is difficult for researchers to obtain access to the necessary information. This problem is compounded by the implicitly private nature of healthcare data and the resulting restrictions from an ethical viewpoint.

Kwon reports that as decision support system environments are rapidly changing from centralized and closed to distributed and open, scalability and interoperability features are becoming more crucial to CDSS development (Kwon, 2003). The application of the Software as a Service (SaaS) and Data as a Service (DaaS) paradigms for healthcare are enormous, offering the potential to easily share distributed heterogeneous medical data between researchers and allied healthcare professionals, and processing this data by distributed medical software.

Employing Semantic Web services for integration has benefits for the field of perinatal care, where prenatal and postpartum care is being offered by an increasingly diverse number of centres, including: family practitioners; obstetricians; high-risk units;

midwifery clinics; and between public and private versions of these institutions. Ideally, diagnosticians should be able to access both clinical data and clinical decision support services from any of these locations without having to support the software locally. Relying on customized solutions for sharing services and data between such heterogeneous units (e.g. customized communication lines, proprietary messaging protocols and proprietary descriptions of how to communicate), would not be financially or practical, which is why most healthcare centres operate largely in isolation. Providing an accessible approach to integration will promote continuity and coordination of care between institutions, and even across different sectors and disciplines.

CHAPTER 4. METHODOLOGY FOR DEVELOPING THE HIGH-RISK PRETERM BIRTH INTEGRATED HYBRID DATA MINING SYSTEM

4.1 INTRODUCTION

This chapter details the methodology behind the development of the AI-based integrated hybrid screening tool for identifying mothers at high-risk of PTB, when screening is based on heterogeneous maternal populations consisting of both high-risk and low-risk mothers.

The chapter begins by providing a description of the two perinatal databases used in this work and their relevant variables, followed by a diagrammatic overview of the integrated hybrid system architecture. The implementation details for developing classification-based MLP ANN models for both perinatal databases are presented, followed by the development and validation of the hybrid-screening model.

4.2 MEDICAL DATABASES

Accessing clinical data is a common hurdle in health informatics research, primarily due to concerns for patient confidentiality and privacy. A further obstacle to training is that it is rare for one database to provide all relevant information in relation to outcome estimation and risk assessment. Particularly, in the distributed field of perinatal care, integration of data from heterogeneous sources is likely to remain the norm until a pervasive electronic patient record spanning all patient care and research facilities is

introduced. The perinatal databases obtained for this research are described in the following sections.

4.2.1 PERINATAL PARTNERSHIP PROGRAM OF EASTERN AND SOUTHEASTERN ONTARIO (PPESO) DATABASE

Under the direction of the Perinatal Partnership Program of Eastern and Southeastern Ontario (PPESO), a regional database was created in 1997 to provide perinatal data to PPESO partners. The database has evolved significantly since its inception, and has become a unique co-operative venture with the 37 partners of PPESO (hospitals, health units, Health Information Partnership (HIP), community health centres, and Ottawa and Queen's Universities), the Ministry of Health and Longterm Care (MoHLTC), and CritiCall (CC). The Niday Perinatal Database and the Southwest Perinatal Program Data Warehouse captures 84% of births in the province (PPESO 2005).

A user guide has been prepared to ensure consistency of definitions for each variable among the hospitals. In addition, each hospital has received training to manage the system's data entry and reporting capabilities. An audit of the database in 2001 found high concordance (less than 5% error) between the database and the case room logbook for all but four variables: electronic monitoring (6.6%), induction (7.0%), postal code (8.5%), and antenatal steroid use (> 10%). As of January 1, 2001, the database was enhanced from a stand-alone computer program installed in each hospital, to the web-based CritiCall Ontario system. Hospitals now enter data directly into the database and can generate reports independently.

All experiments were performed on PPESO databases, covering the years 1999-2001; removing cases with missing values for the variables of interest yields a combined total of 48,000 cases. Results were verified on unseen 2002 data (19,710 cases with no missing values).

4.2.2 PREGNANCY RISK ASSESSMENT MONITORING SYSTEM (PRAMS) DATABASE

Initiated in 1987 as a surveillance project by the Centers for Disease Control (CDC) and Prevention, PRAMS collects state-specific, population-based data on maternal attitudes and experiences before, during, and shortly after pregnancy (Centers for Disease Control and Prevention, 2005). By providing detailed social and demographic information not available from other sources, PRAMS is a unique source of data for researchers interested in studying and improving maternal and infant care.

On a yearly statewide basis, PRAMS selects a sample of all women who gave birth that year, creating a heterogeneous subset of mothers, such that the findings can be extrapolated to an entire state's population. Using a standardized data collection methodology (Centers for Disease Control and Prevention, 2005), PRAMS combines two modes of data collection: a survey conducted by mailed questionnaire with multiple follow-up attempts, and a survey by telephone (Centers for Disease Control and Prevention, 2005).

As well as obstetrical and maternal medical history, PRAMS provides non-clinical information concerning pregnancy not available in other databases, such as: barriers to prenatal care, maternal use of alcohol and cigarettes, physical abuse,

contraception, economic status, sources of maternal stress, social support and services, and mental health during pregnancy.

After removing missing values there was a total of 29,104 available cases, which were divided into a verification set (9,701 cases), a 2/3 training set (12,935 cases) and a test set (6,468 cases).

4.2.3 RISK FACTORS AVAILABLE IN PPESO AND PRAMS DATABASES

Table 4-1 presents a categorized list of potential PTB risk factors compiled from the medical literature, notes whether each factor is available prior to 23 weeks gestation, and indicates which factors are available in the PRAMS and PPESO databases.

TABLE 4-1 PERINATAL RISK FACTORS COMPILED FROM THE LITERATURE

TABLE 4. TERMINAL RISK FACTORS COMPILED FROM THE LITERATURE			
RISK FACTOR [RISK CONDITION]	AVAILABLE BEFORE 23 WEEKS	FACTOR PRESENT IN DATABASE	
		PPESO	PRAMS
SOCIAL AND DEMOGRAPHIC FACTORS			
Age [<17-18 , >30]	YES	YES	YES
Ethnicity [African-American]	YES		YES
Lives Alone (marital status)/ Single Parent	YES		YES
Education [<12 years]	YES		YES
Low income	YES		YES
Poor home condition [Domestic violence, physical, sexual, emotional abuse]	YES		YES
Work during pregnancy [long working hours with long periods of standing]	YES		YES
Stress	YES		YES
MEDICAL HISTORY			
Chronic hypertension [greater than 140/90 mm hg at less than 23 weeks]	YES		YES
Cardiac disease	YES		
Diabetes	YES		YES
Chronic pulmonary disease	YES		
Genital tract abnormalities	YES		YES
History of sexually transmitted disease or bacterial vaginosis	YES		YES
PREVIOUS OBSTETRICAL HISTORY			

Previous spontaneous preterm delivery	YES		
Previous indicated preterm delivery	YES	YES	YES
Previous stillbirth	YES		
Shorter or longer interval between pregnancies	YES		YES
¹⁰			
CURRENT PREGNANCY			
Nulliparity	YES	YES	YES
Is having a first baby	YES	YES	YES
Maternal gravidity ¹¹	YES		
Maternal parity ¹²	YES	YES	YES
Interpregnancy interval	YES		YES
Body mass index [<19.8]	YES		YES
Less than 62" in height	YES		YES
Smoked cigarettes during pregnancy	YES	YES	YES
Drank alcohol during pregnancy	YES		YES
Cocaine use during pregnancy	YES		
Cocaine use before pregnancy	YES		
Drug use during pregnancy	YES		YES
Late or no prenatal care	YES		YES
Vaginal bleeding in first or second trimester	YES		YES
Urinary tract infection	YES		
Acute hypertension [140/90 mm hg during gestation < 20 weeks]	YES		YES
Antibiotic treatment	YES		
Hospitalized for reason other than preterm labour	YES		YES
RESULTS OF SPECIAL TESTING ¹³			
Cervicovaginal fetal fibronectin ¹⁴			
Cervical length <25 mm (by transvaginal ultrasonography)			
PERINATAL OUTCOMES OF INTEREST			
Gestational age at birth		YES	YES
Birthweight		YES	YES
Neonatal care required		YES	YES
APGAR score		YES	

¹⁰ A short interval is considered less than 6-9 months between birth and the beginning of the next pregnancy (March of Dimes, 2003).

¹¹ Gravidity is defined as the number of times that a woman has been pregnant (Patient Plus, 2006)

¹² Parity is defined as the number of times a woman has given birth to a foetus with a gestational age of 24 weeks or more, regardless of whether the child was born alive or was stillborn (Patient Plus, 2006).

¹³ Special testing is only performed on high-risk maternal patients; this information is not available before 23 weeks gestation and is not available in either the PRAMS or PPESO databases.

¹⁴ The presence of fetal fibronectin in vaginal secretions after 22 weeks indicates pathologic changes at the maternal-fetal interface.

Resuscitation required	YES	
Congenital malformations		YES
Cesarean section/ mortality?	YES	YES

4.3 ARCHITECTURAL OVERVIEW OF THE INTEGRATED HYBRID DATA MINING SYSTEM

There are two main systems in the integrated hybrid screening system architecture, depicted in Figure 4-1: a training and verification system, and a clinical system. The training and verification system is used by the researcher to train and validate models on known outcomes. The clinical system, to be used by the clinician, employs the optimal settings obtained during training, to provide outputs for new patient cases with unknown outcomes.

Training the integrated hybrid screening system involves using a pipeline of clinical decision support data mining services through which patient data flows. Broadly speaking there are four main data mining services (or steps): 1) classification-based MLP ANN processing; 2) MLP ANN processing with risk stratification estimation; 3) k-Nearest Neighbours (kNN)-based Case-Based Reasoning (CBR) processing for matching cases; and 4) a Decision-tree voting algorithm for the final classification of conflicting cases.

Successful completion of step 1 is essential, in that it proves that a successful classification-based MLP ANN high-risk preterm birth model can be developed for known patient cases, to be used as a model for future cases. Step 2 uses the best parameter values from step 1 to develop a MLP ANN prediction model with risk stratification estimation. This model is first used to maximize network sensitivity,

followed by further processing to maximize network specificity, and finally to achieve a compromise between sensitivity and specificity. Note that the classification-based MLP ANN, with its laborious and time-consuming training, is used only in this first training step to define the optimal network parameters for the MLP ANN with risk stratification estimations. The results from step 2 flow into step 3 and step 4, which integrate CBR results with a decision tree voting algorithm to classify conflicting cases as high or low risk. Conflicting cases are those that were assessed differently by the two ANN models (i.e. the model that maximized network sensitivity and the model that maximized network specificity). Following completion of these four steps, a model has been developed and tested for known patient cases (i.e. those where the outcomes are known). The last training step involves validating the model on new verification data to demonstrate that the trained model can be used on new patient cases with acceptable sensitivity and specificity. After validation, the final settings from the trained integrated hybrid screening model are used in the clinical system. From the user's perspective, the clinical system is very simple: a new patient case is entered and the system processes the data and assesses the risk. The invocation of the different trained CDSSs required to provide risk assessment is completely transparent to the final user.

FIGURE 4-1 ARCHITECTURAL OVERVIEW OF INTEGRATED HYBRID SCREENING SYSTEM

4.4 OVERVIEW OF PERFORMANCE MEASURES USED IN THIS WORK

Many performance measures exist to calculate the effectiveness of experimental results. While it is possible to choose the best ANN structure (i.e. result) based solely on one performance measure, this approach is overly simplistic and is unlikely to yield the most effective, robust and generalizable network. A better technique is to find a compromise between the key performance indicators. Performance measures used to assess the experimental ANN prediction models for estimating PTB are explained in the

following sections.

4.4.1 CONTINGENCY TABLE

The performance measures used in this work are based on the concept of false and true positives, and false and true negatives. The contingency table (or truth table) shown in Table 4-2 illustrates how these values are determined based on comparing the model output with the true output.

TABLE 4-2 CONTINGENCY TABLE

		TRUE OUTPUT		
		NOT PRESENT	PRESENT	SUM
MODEL OUTPUT	NOT PRESENT	TRUE NEGATIVE (TN)	FALSE NEGATIVE (FN)	TN + FN
	PRESENT	FALSE POSITIVE (FP)	TRUE POSITIVE (TP)	FP + TP
	SUM	TN + FP	FN + TP	TOTAL CASES

Based on the contingency table, we arrive at the following definitions:

True Positive (TP): The model output is positive and the true output is positive.

False Positive (FP): The model output is positive and the true output is negative.

True Negative (TN): The model output is negative and the true output is negative.

False Negative (FN): The model output is negative and the true output is positive.

Sensitivity and specificity, defined below, are key measures of performance required for clinical applications, although the necessary performance criteria will vary between different clinical applications. This work focuses on achieving the highest sensitivity possible while maintaining a clinically acceptable specificity. Additionally, the ANN's performance was reported based on published guidelines for documenting

results of medical applications of artificial neural networks (McGowan, Stevenson, & Frize, 1996). The guidelines state that in addition to sensitivity and specificity, authors should report: Correct Classification Rate (CCR) and the area under the Receiver Operating Characteristic (ROC) curve for both the training and test sets. The performance was also compared with the CCR of the constant predictor. For completeness, the Average Squared Error (ASE) is also included. The Average Squared Error (ASE) calculates the difference between each predicted value and its corresponding actual value squared and averaged over all cases.

4.4.2 CORRECT CLASSIFICATION RATE

The Correct Classification Rate (CCR) is a measure of the total number of cases that were predicted correctly, defined in Equation 4-1 by the number of true positives and true negatives divided by the total number of cases.

EQUATION 4-1

$$CCR = \frac{TN + TP}{TN + FN + TP + FP}$$

In the case of highly skewed data the CCR (i.e. the number of cases correctly classified divided by the total number of cases) can be a misleading performance measure (Frize et al., 1998; Ennett & Frize, 2000; Ennett & Frize, 2003). For example if the occurrence of high-risk PTBs is only 5%, then a network that classifies all cases as low-risk would have a CCR of 95%, however, cases of the rare high-risk outcome would not be identified. From a clinical perspective it is often more significant to predict the low occurrence outcome, which is most often associated with an adverse clinical condition,

than to have a high CCR that correctly classifies all the negative cases but incorrectly classifies all the positive cases.

4.4.3 CONSTANT PREDICTOR

The *a priori* statistics define the constant predictor. It is a statistical benchmark that classifies all test set cases as belonging to the outcome class with the highest *a priori* probability of the training set. For highly skewed data, while the constant predictor may have a higher CCR, from a clinical viewpoint it is not considered as useful as a model with higher sensitivity.

4.4.4 SENSITIVITY AND SPECIFICITY

Sensitivity and specificity are key performance measures cited in clinical literature, the ideal value of each is 1. Sensitivity, defined in Equation 4-2, is a measure of the number of true positives predicted correctly (the true positive rate). In this thesis it is a measure of how well the model classifies high-risk PTB. Specificity, defined in Equation 4-3, is a measure of the true negatives predicted correctly (true negative rate). In this thesis it is a measure of how well the model classifies low-risk births.

$$\text{sensitivity} = TP / (TP + FN) \quad \text{EQUATION 4-2}$$

$$\text{specificity} = TN / (TN + FP) \quad \text{EQUATION 4-3}$$

4.4.5 LOGARITHMIC SENSITIVITY INDEX

The logarithmic sensitivity index proposed by Scales was used as the stopping criteria to balance the influence of the skewed outcomes by optimizing both the

sensitivity and specificity of the neural network (Ennett, Frize, & Scales, 2003). The logarithmic sensitivity index, defined in Equation 4-4, expands towards infinity when both sensitivity and specificity are 1, but is slightly weighted in favor of sensitivity. The degree of weighting (or bias) towards sensitivity is controlled by the exponent n (Ennett, Frize, & Scales, 2003), which is equal to 0.75 for the experiments related to this thesis. Using the logarithmic sensitivity index as the stopping criterion results in ANN runs where training usually stops when there has been no logarithmic sensitivity index increase within the last 500 epochs (Ennett, Frize, & Scales, 2003; Ennett, Frize, & Charette, 2004).

$$Index = -sensitivity^n * \log(1 - sensitivity * specificity) \quad \text{EQUATION 4-4}$$

4.4.5.1 Receiver Operating Characteristic (ROC) Curves

The Receiver Operating Characteristic Curve (ROC) assesses the ability of the model to discriminate between outcomes and achieve a balance between sensitivity and specificity; it is a plot of the model's sensitivity versus one minus its specificity obtained by varying the threshold value of the output node across its range of values (Woods & Bowyer, 1997). The area under the ROC curve, the C-index, is generally cited in the clinical literature when assessing a model's performance.

ROC curves used in this work are determined using specificity and sensitivity values). An ROC curve that resembles a step function (i.e. with a C-index approaching 1) is an indicator that all cases have been correctly classified¹⁵, whereas a 45-degree positive line (C-index equal to 0.5) is a useless classifier, indicating that the false positive

¹⁵ Assuming the step function's true positive rate begins to increase directly from the origin.

and true positive rate are equal (Frize, Ennett, Stevenson, & Trigg, 2001; Ennett, 2003). Rough guidelines for evaluating the effectiveness of a predictive test based on C-index values are presented below in Table 4-3 (compiled from (Simon, 2005) and (Tape, 2006)):

TABLE 4-3 EVALUATING THE EFFECTIVENESS OF ROC CURVES

C-INDEX		EFFECTIVENESS
MINIMUM	MAXIMUM	
0.5 <=	< 0.6	FAIL
0.6 <=	< 0.7	AVERAGE
0.7 <=	< 0.75	GOOD
0.75 <=	< 0.92	VERY GOOD
0.92 <=	<= 1.00	EXCELLENT

ROC curves can be influenced by the prevalence of outcomes (C. Ennett, 2003), which is why they are most often used to describe the performance of outcomes that comprise nearly an equal number of positive and negative cases (Metz, 1978). Due to the severely skewed nature of the clinical data considered in this work (less than 10% preterm birth rate), it would be unrealistic to expect the area under the ROC curve to be in the excellent range. However, it is expected to be in the 'very good' range.

4.5 ANN DETAILS

The tool used in this work for developing classification-based MLP ANN models for the early estimation of PTB is written in MATLAB using the Neural Network Toolkit, with extensive modifications, such as automated processing, use of a logarithmic sensitivity index to optimize for both sensitivity and specificity when mining skewed data (Ennett & Frize, 2000), and automatic calculation of Receiver Operating Characteristic (ROC) curves.

Over the past decade, researchers within MIRG have experimented with a variety of ANN architectures for predicting clinical outcomes. They have achieved good results with the classification-based MLP ANN based on a feed-forward back propagation neural network-learning algorithm using the hyperbolic tangent transfer function with a weight-elimination option. A MLP network is generalized from a single layer network: it contains one input layer, one or more hidden layers, and an output layer. Feed forward implies that results travel directly from the inputs to the outputs. A back-propagation learning algorithm calculates the error backwards (i.e. from output to input) and then updates the connection weights and biases to minimize the error. The hyperbolic tangent (ranging from -1 to 1) exhibits a sharp transition between output values, facilitating faster learning, it offers a sharp sigmoidal transition at zero which results in very few output values near zero, so a nonzero output is more likely (Fausett, 1994).

The classification-based MLP ANN with one hidden layer is very popular for use in medical data mining. Medical data is believed to be nonlinear in nature with many unknown interactions; the hidden layer allows nonlinear interactions between the variables, with hidden nodes acting as calculators of higher order statistics (Rybachynski, 2005). Additionally, there has been little evidence to suggest that using more than one hidden layer provides a better result than using one hidden layer (Buskard, Stevenson, Frize, & Solven, 1994; Penny & Frost, 1996). The more nodes a network has, the better it is able to model the dataset. However, there is always a danger of over-fitting to the training set, where the network cannot generalize to new data. Also, the computing time increases significantly when the number of nodes is large.

4.5.1 NETWORK PARAMETERS AND SOFTWARE EXECUTION

Nine network parameters were used to drive the learning of the back-propagation MLP ANN: learning rate; adaptive learning rate scale factors (learning rate increment and learning rate decrement); weight-decay constant, which determines how strongly the weights are penalized; adaptive weight-decay factors (weight decay constant increment and weight decay constant decrement); weight-elimination scale factor (w_o - defines the sizes of large and small weights); momentum (prevents the learning algorithm from getting caught in local minima and allows it to continue its search for the global minimum (Frize et al., 2001); and error ratio (a user-defined value that measures the change from the current to previous error to determine how the learning rate, weight decay constant and momentum will be adapted to improve the network performance (Ennett, Frize, & Charette, 2004)). The parameter values define how the neural network weights and biases are changed from epoch to epoch as a function of the sum-squared error calculated over all trained cases (Rybachynski, 2005). One complete cycle of presentation of training cases to the MLP ANN algorithm is called an epoch; a minimum of 25 epochs must be performed to give adequate time for learning. The software automatically progresses from one “parameter run”¹⁶ to another, carrying forward the best parameter value from each run until a stagnant point is reached. This progresses until nine stagnant points are found. The user must then examine the results for each parameter value/network structure (defined by the number of hidden layers and the number of hidden nodes in the hidden layers): combination to manually extract the optimal

¹⁶ Where a “parameter run” implies changing the values for one parameter while holding the other eight parameters constant.

performance results for their problem of interest and pre-determined performance criteria (e.g. in this case although some networks yielded a sensitivity of over 70%, their corresponding specificity and ROC values were too low to generate a clinically useful model). The ANN software automatically runs experiments for all network structures: the default is to run for a single layer network with no hidden nodes (e.g. no hidden layer), to one hidden layer with 2 hidden nodes, to one hidden layer with $2n + 1$ hidden nodes, where n is the number of input variables (Rybachynski, 2005). This upper limit was chosen based on the Kolmogorov superposition theory adapted to an ANN context, which implies that the maximum number of hidden nodes necessary would be no larger than two times the number of input variables plus one (Beiu, 2004). Running the automated software based on this criteria (i.e. up to a maximum of 17 hidden nodes for eight obstetrical input variables) with approximately 5,000 test cases and 10,000 training cases takes on average five days per clinical output of interest (for all network structure-parameter combinations)¹⁷. This is to be expected as training time increases proportionally to the number of nodes in the hidden layer and the data set size. Manually extracting the optimal performance results is also a time consuming procedure.

4.5.2 WEIGHT ELIMINATION COST FUNCTION

All experiments employ a weight-elimination term added to the error function, in an effort to avoid the problem of overfitting, where the patterns are related too closely to the training data. Weight decay and weight elimination are two complexity terms that can be added to the cost function. Weight elimination reduces the smaller weights to 0,

¹⁷ Hardware: IntelCore™ Duo Processor (1.6GHz, 2MB L2 Cache, 533MHz FB)

thereby simplifying the complexity of the ANN and minimizing overfitting. Weight elimination is well suited for network pruning by eliminating variables that offer little or no assistance in estimating the outcomes (Trigg, 1997; Frize, Ennett, Stevenson, & Trigg, 2001). Weight decay limits the size of the connection weights; this penalizes large weights by forcing them not to increase. The weight elimination cost function is shown in Equation 4-5, weight decay is actually contained within the weight elimination formula. The first term is the initial cost function, this is the sum of squared errors cost function where $E_0(W)$ is the difference between the target and network output, and the second term is the weight elimination component. In Equation 4-5: W represents the weight vector; lambda (λ) is the weight decay constant that determines how strongly weights are penalized; w_{ij} represents the ANN weights of connecting nodes i and j ; and w_0 is the weight scale factor that defines the sizes of large and small weights. When w_0 is small the small weights will be forced to zero resulting in fewer large weights (weight elimination), if w_0 is large many small weights remain and the large weights are limited (weight decay) (Trigg, 1997; Rybchynski, 2005).

$$E(W) = \sum_{i=1}^p (E_0(W))^2 + \lambda \sum_{ij} \frac{\frac{w_{ij}^2}{w_0^2}}{1 + \frac{w_{ij}^2}{w_0^2}} \quad \text{EQUATION 4-5}$$

4.5.3 FEATURE SELECTION

The purpose of feature selection is to determine which of the original inputs are most useful for predicting the outcome of interest and to eliminate features with little or no predictive information (Kim, Street, Menczer, & Russell, 2003). When developing

systems for outcome estimation it is useful to find ways to reduce the number of variables used to estimate outcomes, that is, define a minimum dataset for each of the models of outcome estimation to be created. When only the most important variables remain, the minimum dataset has been reached; the difficulty lies in knowing which variables to remove. One approach, termed sensitivity analysis, is to reduce the input variables one by one and observe when the performance deteriorates (i.e. the sensitivity of the network's performance to each change) (M. Frize, Ennett, & Hebert, 2001; Zernikow, Holtmannspotter, Michel, & et al, 1999)). If a variable's removal has little or no impact on the results, then it is removed from the dataset and a new 'minimum dataset' has been defined.

Once the ANN's adjustable parameter values have been tuned to achieve optimal performance, it is possible to perform feature selection by comparing the maximum sensitivity achieved using this model with that obtained using a reduced dataset. The goal is to determine whether reducing the network's complexity improves the ANN classification (Frize et al., 2001). Using sensitivity analysis, variables with low absolute connection weights for each network layer and parameter value were identified and eliminated independently to assess the impact on sensitivity and other performance measures. Note that this is performed in addition to enabling the weight elimination cost function, as previously described.

4.5.3.1 Variable Reduction Using the Modified Relative Weights Method

Traditionally, it has been difficult to discern the relevance of each input variable in MLPs with multiple hidden nodes and one hidden layer, due to the network's

complexity and difficulty in interpreting the meaning of the ANN connection weights. Although sensitivity analysis can be used to identify the most significant variables for a particular outcome, while simple, the process is laborious, and might be unrealistic when running experiments with 30,000 cases and 48 input variables, which can take over a week to run using the ANN-RFW. Also, an ideal minimum dataset includes the relevance, or weighting of each variable, and this is not achieved using the sensitivity analysis approach.

A more viable option for variable reduction is to employ the relative weights method. A simplified version of this method was proposed for medical databases trained on single layer ANNs only (Frize, Ennet, & Hebert, 2001): the absolute value of the connection weights for each input-output pair are scaled between 0 and 100, with 100 representing the strongest relative importance. Inputs with smaller weights are removed (with the choice of cutoff determined through experimentation) and the results are analyzed; if no degradation occurred, the results are scaled again and variables with low relative weights are removed. This cycle continues until the performance begins to degrade. Ennett also employed this approach for computing the relative weights of the ANN, by taking the average of the absolute values of all the weights for each input (Ennett, 2003). These values were then used as the connection weights in a case-based reasoning system. While intuitive and simple to use, this method lacks the ability to handle the more complex networks with hidden layers that often yield better results when training non-linear medical databases with many input variables. Garson (Garson, 1991), later modified by Goh (Goh, 1995), proposed and presented a new method for calculating

relative weights for multilayer perceptron ANNs.

Rybachynski provides a detailed evaluation of the Goh method, noting that while the method appears to include weights from the hidden-output connections, further examination indicates that these weights actually cancel out in the equations (Rybachynski, 2005). Rybachynski's work proposes and validates a simplified Garson approach, which uses the input-hidden connections as well as the output-hidden connections in the calculation of the relative weights. Rybachynski clearly demonstrates that her relative weights method provides a more closely aligned set of relative importance values for the inputs when the number of hidden nodes is varied, than is seen with the Goh equations. The closer alignment makes it possible to define a final relative importance which can be applied to more than one structure (e.g. is applicable even when the number of hidden nodes varies). Rybachynski's relative weights method is outlined below.

i : the number of input variables (which determines the number of input nodes)

j : the number of hidden nodes.

k : the number of output nodes (always one in our case).

w_{ij} : the connection weight between input node i and hidden node j .

w_{jk} : the connection weight between hidden node j and output node k .

rw_{ij} : the relative connection weight between input node i and hidden node j .

rw_{jk} : the relative connection weight between hidden node j and output node k .

RI_i : the Relative Importance of input node i .

Relative Weight Equations:

Step 1: Calculate the relative weighting of the hidden-output connections into the output node. For a single output node set $k=1$.

$$rw_{jk} = \frac{abs(w_{jk})}{\sum_j abs(w_{jk})} \quad \text{Equation 4-6}$$

Step 2: Calculate the relative weighting of the input-hidden connections into each hidden node. For each hidden node j :

$$rw_{ij} = \frac{abs(w_{ij})}{\sum_i abs(w_{ij})} \quad \text{Equation 4-7}$$

Step 3: Multiply each relative weighting input-hidden connection value by the relative weighting value of its connected hidden-output weight. For each input i :

$$P_{ij} = rw_{ij} * rw_{jk} \quad \text{Equation 4-8}$$

Step 4: Sum the final relative weighting products (P_{ij}). For each input i :

$$S_i = \sum_j P_{ij} \quad \text{Equation 4-9}$$

Step 5: Divide the sum of the final relative weightings for each input node by the total of all relative weightings. For each input i :

$$RI_i = \frac{S_i}{\sum_i S_i} \quad \text{Equation 4-10}$$

4.6 DEVELOPING THE PPESO-BASED HIGH-RISK PTB MODEL

4.6.1 PPESO PERINATAL DATABASE

One of the limitations of the PPESO database in relation to this work is that because it focuses on perinatal data; out of 37 medical variables, only eight are related directly to the obstetrical period and are available for predicting PTB. Because we are interested in predicting PTB before 23 weeks gestation, it is not possible to keep variables in the dataset that are only available after this gestational age. For instance, one of the variables in the PPESO database indicates whether a baby had a low birth weight – if this variable was used as an input variable for predicting PTB it would provide a highly biased result in that almost all preterm babies are also low birth weight babies. While it might appear that this information would generate a model with excellent classification ability, the model is in fact useless because it includes variables that are only available after birth and are unknown before 23 weeks gestation. As can be seen from Table 4-1, the PPESO database is rich in perinatal outcomes of interest (as compared to the PRAMS database which contains many more possible preterm birth risk factors and fewer perinatal outcomes). Perinatal outcomes, other than gestational age, are important because, as shall be seen when following through the methodology and results for developing the PTB risk model, PTB alone as an outcome is insufficient for obtaining a clinically acceptable sensitivity. To solve this problem, a refined *high-risk* PTB model is defined. The PPESO database will be used to develop this high-risk model. The PRAMS model will then be developed based on the PRAMS database's 48 input variables and a slightly less complete version of the high-risk criteria.

First, the study population must be defined, including the input variables (possible PTB risk factors) and the output of interest (gestational age at birth). Based on physician input from neonatologist Dr. C.R. Walker, the obstetrical variables selected as being available before 23 weeks for predicting PTB include: maternal age, number of babies for this pregnancy, number of previous term babies, number of previous preterm babies, parity (total number of previous children), baby's gender, whether mother has intention to breastfeed, and maternal smoking after 20 weeks gestation. All variables are known to physicians prior to 23 weeks and are therefore available to be used in the early prediction of PTB. Variables that were excluded from the input variable dataset comprise entirely those that are only available at delivery and after birth, such as: labour type, delivery type, maternal pain relief, birth weight and APGAR score.

Table 4-4 summarizes the PTB statistics between 1999 and 2002; the PTB rate is similar for all years ranging from 8.1% to 9.2%. When examining only the preterm birth cases, the prevalence of variables (for binary variables) and mean values (for continuous variables) varies only slightly over the four-year span, indicating that the four databases are very comparable in terms of the eight input variables used in this work. For example, among women who delivered prematurely, the prevalence of women who had experienced previous PTBs ranged from 10.1% (in 2002) to 12.2% (in 2001). In preparing the data for processing, the non-binary-valued inputs were Gaussian normalized¹⁸, the normal value of each variable was subtracted from each input value and the resulting differences were divided by three standard deviations pertaining to each

¹⁸ Experiments were performed using categorical inputs for non-binary variables with poor results.

input variable value over the entire data set. The same processing approach was used for the PRAMS data.

TABLE 4-4 PPESO PRETERM BIRTH STATISTICS BY YEAR

YEAR	1999	2000	2001	2002
ANNUAL PTB RATE	8.7	9.2	8.4	8.1
ANNUAL VARIABLE PREVALENCE (%)				
SMOKING	17.5	17.3	17.1	15.9
INTENTION TO BREASTFEED	71.5	75.4	79.8	78.8
PREVIOUS PTB	10.3	11.3	12.2	10.1
PREVIOUS TERM BIRTH	42.1	42.7	43.3	42.2
SINGLETON BIRTH	75.8	80.6	77.6	76.1
GENDER = MALE	55.2	52.4	51.2	53.2
AGE*	28.9	29.2	29.4	29.5
PARITY*	0.87	0.87	0.87	0.88

* These are continuous variables – mean values for each year is shown. Over all years: Patient's age – range 12 to 48 years, mean 29.24 years. Parity: range: 0 –13 children, mean 0.875 children.

4.6.2 CREATING THE ARTIFICIAL DATASETS

ANNs often have difficulty classifying skewed data, where the outcome of interest is a relatively rare event, resulting in a network with high specificity and low sensitivity. From a medical perspective this is a serious problem, because physicians are interested in predicting rare outcomes. PTB rates in the PPESO dataset range from 8.1-9.2%; initial classification-based MLP ANN results based on this distribution were poor (please see *Chapter 5* for results), suggesting that a new approach was needed. In an attempt to compensate for low PTB rates and to increase the network's ability to predict PTB with high sensitivity, this work examines the effect of artificially increasing the PTB rate in the training set presented to the network during learning.

Two different approaches were considered for altering the *a priori* distributions of

the training set: a) randomly remove the non-preterm cases until the representation of the preterm cases is sufficiently high, or b) randomly resample preterm cases until the representation of the preterm cases is sufficiently high. Because the PPESO database contains a large heterogeneous population of Ontario births (including term and preterm births; babies requiring additional medical care and babies requiring neonatal intensive care), as opposed to a more homogeneous population (e.g. adult intensive care unit data where only the sickest patients are included), method a) was rejected due to the loss of potentially valuable information about non-preterm cases that may lower the network's correct classification power. Method b) has been successfully applied by Ennett and Frize for predicting post-surgical mortality of coronary artery bypass graft patients, by repeatedly presenting non-surviving outcomes to the ANN using random selective sampling; optimal performance was achieved when the under-represented class was artificially increased to a moderate 20% (Ennett & Frize, 2003). The coronary artery bypass graft database contains 6,325 cases and less than 5% mortality; the ANN was trained using 32 input variables describing the patient's medical history. The similarities between this previous work and the PPESO research include the low distribution of the outcomes of interest (mortality and PTB rates) and the fact that both experiments use variables based on patient history collected at admission, and the ANN software and methodology used for training and testing. However, the PPESO database contains a vast wealth of patient cases in contrast with the more limited coronary dataset, and is a population database, whereas the coronary database includes only the sicker patients requiring surgery.

Given the wealth of available PTB cases and the similar variable prevalence for years 1999-2002, method b) is modified by artificially increasing the distribution of the training set using real PTB cases from all years (except the 2002 verification data), rather than randomly re-sampling PTB cases from only one year. The benefits of this approach are that there is no resulting loss of information on the under-represented class, as there would be if only PTBs from one year were included, and there is less chance of over-training the network. Using this technique, the underrepresented class is artificially increased to raise the percentage of high-risk PTB cases in the training set to 20% (10,864 cases). The 5,432 cases in the test set maintain the real distribution of 10%. The number of cases in the training set far exceeded Lisboa's guidelines, which states that the number of observations (e.g. cases) should be 5-10 times the number of available variables (Lisboa, 2002).

4.6.3 DEFINING THE HIGH-RISK PRETERM BIRTH MODEL: REFINING THE OUTPUT

In a further attempt to improve the classification-based MLP ANN's prediction ability, the outcome is refined by defining the concept of high-risk PTB. Consider that PTB is defined as birth at less than 37 weeks gestation; therefore a birth at 37 weeks is full term, whereas a birth at 36 weeks and 6 days is considered premature. However, both births will likely result in normal outcomes with no complications; effectively neither baby is considered a 'high-risk' infant at birth. It is possible that 'low-risk' preterm infants, such as those born near the 37-week mark with no associated complications, may be detrimental to the network's ability to identify 'high-risk' preterm births. The rationale

is that a large proportion of the births occurring between 33 and 36 weeks will have postpartum outcomes similar to term births (Tucker & McGuire, 2005) and have little in common with outcomes from births at less than 33 weeks.

Births at less than 33 weeks are termed “very preterm birth”, and face the highest risk of serious health problems, disability, and death (Perinatal Partnership Program of Eastern and Southeastern Ontario, 2005). Whereas very preterm birth poses the greatest long-term risks, early identification of any high-risk premature birth (including those infants born between 33 and 36 weeks) would be invaluable in pregnancy management.

In an attempt to improve the network’s ability to discriminate between high-risk and low-risk PTBs, the hypothesis is made that PTB as an output, measured by number of week’s gestation, may be misleading and further means of outcome classification are needed. Indicators of high-risk birth available in the PPESO database include:

APGAR score: Developed in 1952 by the late pediatrician, Dr. Virginia Apgar (Alexian Brother’s Medical Center, 2006), the APGAR score tests five attributes: Appearance; Pulse; Grimace; Activity; and Respiration. The score provides an indication of infant health at 1, 5 and 10 minutes after birth. The Apgar score at 5 minutes is considered more clinically significant than the Apgar score at 1 minute, which can be misleading. A score of less than 7 is associated with a higher risk of health and developmental problems for the baby. In Ontario, approximately 1 in 6 very preterm births experience Apgar scores of less than 7 at 5 minutes after birth – as compared to 1 in 100 term babies (PPESO, 2005).

Birth Weight: Definition: 2500 grams (low birth weight), 1500 grams (very low

birth weight), 1000 grams (extremely low birth weight) (Tucker & McGuire, 2005). Babies born at less than 1500 grams, categorized as very low birth weight, are at particular risk of health problems, disability, and death (PPESO, 2005); Approximately 2/3 of low birth weight infants are preterm, although preterm infants may also be small for gestational age and may have neonatal problems additional to those related to shortened gestation (Tucker & McGuire, 2005).

Neonatal Transfer: A transfer to the Neonatal Intensive Care Unit (NICU) following delivery indicates that an infant required advanced medical care, ranging from routine monitoring to more invasive procedures; for future reference, it would be advantageous to include duration of the NICU stay in the model, as this would provide an indicator of the severity of the baby's illness. This information is not currently available with the PPESO database.

Resuscitation: In Ontario, one in 20 babies receive some form of resuscitation at birth (PPESO, 2005). Infant resuscitation techniques documented in the PPESO database include: (a) the common application of free flow oxygen, (b) partial pressure ventilation, or bag-and-mask ventilation, with which most newborns can be successfully resuscitated, (c) chest compressions, administered if the heart rate is absent or unacceptably low, (d) intubation, and (e) administration of drugs such as epinephrine, sodium bicarbonate, and volume expanders. Because a large number of infants routinely receive resuscitation in the form of free flow oxygen, the impact of removing these resuscitation cases from the high-risk PTB definition model is also investigated.

A high-risk PTB model was proposed, where all births at less than 33 weeks (very

preterm births) are high-risk, but only those births between 33 and 36 weeks that include at least one of the four extended criteria are defined as high-risk. Figure 4-2 depicts a Unified Modeling Language (UML) activity diagram showing the flow of control for defining the high-risk PTB model.

Table 4-5 shows the percentage of cases considered high-risk based on the new criteria. The new high-risk PTB criteria results in skewed data, with the high-risk PTB cases severely underrepresented (ranging from 4.4-5.5% of all births).

FIGURE 4-2 FLOW OF CONTROL FOR DEFINING THE HIGH-RISK PRETERM BIRTH PREDICTION MODEL

TABLE 4-5 HIGH-RISK PTB STATISTICS BY YEAR

		1999	2000	2001	2002
ALL HIGH-RISK CASES	PREVALENCE TOTAL BIRTHS (%)	4.4	5.5	4.9	4.7
	PREVALENCE PTBs (%)	56.9	60.7	57.1	58.3
HIGH-RISK CASES EXCLUDING FREE- FLOW OXYGEN	PREVALENCE TOTAL BIRTHS (%)	2.8	4.0	3.5	3.0
	PREVALENCE PTBs (%)	35.9	43.7	40.8	37.2

As was the case for predicting PTB, the underrepresented class is artificially increased to raise the percentage of high-risk PTB cases in the training set to 20%. The test set maintains the real distribution of 5.5% (which was obtained from the original data before artificially increasing the training set). Due to the severely skewed nature of the data, combining high-risk PTB cases from all years does not yield a 20% distribution, and random re-sampling is required to increase the distribution of high-risk PTB cases in the training set.

4.6.4 ADJUSTABLE NETWORK PARAMETERS

Table 4-6 summarizes the network parameter settings at which the best performance¹⁹ was achieved using a classification-based MLP ANN with one hidden layer for each of the four outputs: preterm birth based on the original distribution; preterm birth based on the artificially increased distribution; high-risk preterm birth based on the artificially increased distribution; and high-risk preterm birth based on the

¹⁹ The best performance results are derived from the PPESO test dataset.

artificially increased distribution excluding cases where the patient received resuscitation as free flow oxygen. Settings include the number of hidden nodes and the values for the nine network parameters. Results are presented in *Chapter 5*.

TABLE 4-6 NETWORK SETTINGS OF THE BEST PERFORMING DOUBLE LAYER NETWORKS FOR PREDICTING PTB, HIGH-RISK PTB AND HIGH-RISK PTB EXCLUDING FREE FLOW OXYGEN CASES (PPESO DATASET)

EXPERIMENT	PTB - ORIGINAL	PTB - ARTIFICIAL	HIGH-RISK PTB	HIGH-RISK (-FFO2) PTB
TRAIN PTB RATE (%)	8.6	23	20	20
TEST PTB RATE (%)	8.7	10	5.5	4.0
Hidden nodes	3	3	8	4
Final learning rate*	0.00016649	0.00017404	0.0011203	0.00068864
Learning rate increment	1.2	1.2865	1.001	1.001
Learning rate decrement	0.95	0.9092	0.999	0.999
Final weight decay constant*	0.0004995	0.0004995	0.0005	0.0004995
Weight decay constant increment	1.01	1.01	1	1.001
Weight decay constant decrement	0.99	0.99	0.999	0.999
Final weight scale (w_o)*	0.2	0.501	0.9986	0.9986
Momentum	0.8	0.85	0.607	0.485
Final error ratio*	1	1	1.25	1.25

-FFO2 = Excluding cases where resuscitation was offered in the form of free flow oxygen

* Indicates network settings taken from the final run parameters

4.6.5 VERIFICATION

The best performing model is derived based on the performance of the test set data; a verification dataset is then used to verify the model's ability to generalize to new data. Newly-received PPESO data from 2002 was reserved to form ten independent verification sets of 1971 cases each, following the guideline that the verification set data not be used at all during training or testing so that the ANN results on the verification set can be considered an unbiased prediction of the ANN performance on new data (Dayhoff & DeLeo, 2001). The verification program allows a previously defined ANN estimation model to be applied to a new database to ensure that the ANN was generalizing, rather than overfitting to the training data. Based on the optimal minimal dataset, the

verification tool runs only one epoch, using: 1) the best performing structure (defined by the number of layers and hidden nodes in the network) obtained from the test run; and 2) the corresponding weights. A predicted output is calculated for each presented case. The predicted outputs are then manually compared with the true outputs on a case-by-case basis to identify the number of true negatives, false negatives, true positives and false positives; these values are used to calculate the sensitivity and specificity using Equation 4-2 and Equation 4-3 respectively.

4.7 DEVELOPING THE PRAMS-BASED HIGH-RISK PTB MODEL

Having demonstrated the potential of the classification-based MLP ANN as a means of improving the sensitivity of preterm birth prediction using the PPESO dataset (please see results in *Chapter 5*), the more detailed PRAMS dataset is then employed to:

1. Develop an alternate model for predicting preterm birth on a new database and compare with the PPESO model;
2. Further validate the applicability of the classification-based MLP ANN as a means of predicting both preterm birth and high-risk preterm birth on an enhanced 48-variable model, using a slightly modified definition of the high-risk preterm birth criteria defined for the PPESO model. The new criterion includes congenital malformations but does not provide information on the infant's resuscitation requirements or APGAR score. It is postulated that the lack of APGAR and resuscitation data should not impact the high-risk criterion as this information is largely provided by the NICU admission variable (i.e. most infants will be admitted to a NICU to meet serious resuscitation needs, or if their APGAR scores

are low). Although this may not be true for minor resuscitation needs met by the common admission of free flow oxygen, the PPESO results from *Chapter 5* demonstrate that the high-risk preterm model that did not include cases with resuscitation administered via free-flow oxygen outperformed the model that included free-flow oxygen cases. This indicates that free-flow oxygen, which is commonly administered after birth, is not a good indicator of high-risk preterm birth. Therefore, the PRAMS high-risk criteria can be considered comparable to the PPESO criteria and will provide a second validation of this novel approach for predicting serious preterm cases by considering high-risk preterm cases only.

4.7.1 PRAMS PERINATAL DATABASE

Experiments were performed on PRAMS data, covering the years 2002-2004. The original PRAMS database included 224 variables, not including those created by the PRAMS committee for statistical analysis. After extensive analysis of the database, 48 key inputs were chosen as being available prior to 23 weeks gestation and therefore relevant in the early identification of preterm birth. Table 4-7 details the PRAMS input variables. The bivariate correlations (Pearson) between each of the 48 input variables and the high-risk PTB output are included in Appendix D. The inclusion of more features allows more precise discrimination between true and false positive detection, with the downside being the increased complexity of the model (Jerebko et al., 2003). Variables that were removed included:

1. Confounders for preterm birth (e.g. morbidity-bedrest, premature rupture of the membranes, high number of prenatal care visits). Note that the choice of

confounders was made in conjunction with Dr. C. R. Walker;

2. Variables missing values for greater than 20% of the cases;
3. Variables where information was not available before 23 weeks gestation;
4. Variables that presented duplicate or overlapping information;
5. Variables that were too detailed, and applied only to a small number of patients

(e.g. EP_TRANS, a variable that tells whether the mother received late prenatal care because no transportation was available). Removing overly specialized

variables was done in an attempt to create a network that would be generalizable to other databases.

TABLE 4-7 PRAMS VARIABLES – PREVALENCE AND STATISTICS

VARIABLE DEFINITION	VARIABLE PREVALENCE (%)	STATISTICS (CONTINUOUS VALUES ONLY)		
		AVERAGE	MIN/MAX	ST.DEV
MATERNAL AGE	-	28.8	14 / 53	5.78
PLURALITY	-	1.04	1 / 3	0.205
PREVIOUS PRETERM BIRTH	14.3	-	-	-
PARITY	-	1.78	0 / 18	1.18
GENDER (= MALE)	50.5	-	-	-
BREASTFEEDING	6.4	-	-	-
MATERNAL SMOKING DURING PREGNANCY	14.4	-	-	-
NUMBER OF CIGARETTES PER DAY DURING PREGNANCY	-	1.48	0 / 60	4.42
MATERNAL SMOKING 2 YEARS PRIOR TO PREGNANCY	26.4	-	-	-
MATERNAL DRINKING DURING PREGNANCY	0.074	-	-	-
NUMBER OF DRINKS PER DAY DURING PREGNANCY	-	0.021	0 / 25	0.398
MARRIED	72.0	-	-	-
MATERNAL EDUCATION	-	3.50	1 / 5	1.10
1: 0-8 YEARS	4.1			
2: 9-11 YEARS	12.9			
3: 12 YEARS	35.1			
4: 13-15 YEARS	24.7			
5: >= 16 YEARS	23.3			
MATERNAL RACE		-	-	-
BLACK	13.5	-	-	-
WHITE	63.9	-	-	-

<i>INDIAN</i>	4.94	-	-	-
<i>ALASKA</i>	3.61	-	-	-
<i>HAWAII</i>	5.49	-	-	-
<i>PACIFIC</i>	4.00	-	-	-
<i>OTHER</i>	9.57	-	-	-
HISTORY OF OTHER	-	0.515	1 / 7	0.934
TERMINATIONS				
OBSTETRICAL: MATERNAL	4.18	-	-	-
DIABETES				
OBSTETRICAL: MATERNAL	0.01	-	-	-
FEVER				
OBSTETRICAL: NO MEDICAL RISK	61.0	-	-	-
FACTORS				
INSURANCE	63.4	-	-	-
MEDICAID	16.2	-	-	-
VITAMIN USE DURING	-	1.98	1 / 4	1.29
PREGNANCY				
MATERNAL WEIGHT BEFORE	-	150.7	80 / 409	37.8
PREGNANCY (LBS)				
WHETHER MOTHER WAS TRYING	47.4	-	-	-
TO CONCEIVE				
MONTH PRENATAL CARE BEGAN	-	2.69	1 / 10	1.59
RECEIVED EARLY PRENATAL	81.9	-	-	-
CARE				
WOMEN, INFANT, CHILDREN	43.6			
PROGRAM ²⁰ (WHETHER MOTHER				
WAS PART OF PROGRAM DURING				
PREGNANCY)				
OBSTETRICAL: MORBIDITY ²¹	5.48	-	-	-
HIGH BLOOD PRESSURE				
OBSTETRICAL: MORBIDITY	3.74	-	-	-
BLEEDING				
OBSTETRICAL: MORBIDITY	26.9	-	-	-
NAUSEA				
OBSTETRICAL: MORBIDITY	17.1	-	-	-
KIDNEY				
OBSTETRICAL: MORBIDITY	27.9	-	-	-
CERVIX				
OBSTETRICAL: MORBIDITY CAR	1.99	-	-	-
ACCIDENT				
STRESS: FAMILY	25.8	-	-	-
STRESS: DIVORCE	10.7	-	-	-
STRESS: MOVING	33.3	-	-	-
STRESS: HOMELESSNESS	4.44	-	-	-
STRESS: JOB	13.5	-	-	-
STRESS: EMPLOYMENT	11.0	-	-	-
STRESS: ARGUMENT	29.4	-	-	-
STRESS: COMBINED (COMBINES	9.52			
DRUG USE, IN FIGHT,				

²⁰ Program provides lower income mothers with food and nutritional education during pregnancy.

²¹ Morbidity in this context is defined as a disease, condition or state.

MOM/PARTNER WENT TO JAIL, PARTNER DID NOT WANT PREGNANCY)				
PHYSICAL ABUSE DURING PREGNANCY	9.50	-	-	-
INCOME ('+1' IMPLIES MOTHER RECEIVES INCOME FROM WAGES OR BUSINESS FEES) ²²	9.51	-	-	-
OUTPUT: GESTATIONAL AGE (WEEKS)		37.95	12/43	3.07

The PPESO experimental results from *Chapter 5* indicate that the 'Previous Preterm Birth' variable is considered significant in the early prediction of high-risk preterm birth; this finding is supported by the clinical literature (Meis et al., 1998). After the initial PRAMS pre-processing, there were 91,800 cases available for the 48 variables of interest; however, of these cases only 32,022 had complete data for the 'Previous Preterm Birth' variable. It was decided to focus on these 32,022 cases because:

1. This would allow replication of the PPESO methodology, otherwise the PRAMS dataset would be missing a critical variable from the PPESO model;
2. This number of cases is more than sufficient for developing the ANN models, as researchers suggest that a minimum of 6 to 10 samples per input variable is generally acceptable (Forsström & Dalton, 1995; Weigend, Rumelhart, & Huberman, 1990). In fact, with almost 100,000 cases, the running time of the ANN would be unrealistically long, as it takes over a week to run a complete experiment for all numbers of hidden nodes with just over 30,000 cases.

Further pre-processing on the 32,022 cases now available, resulted in the removal

²² '-1' implies the mother receives income from other sources (aid, unemployment insurance, child support, social security, family or friends)

of cases missing values for the other 47 input variables (not including previous preterm birth), the final PRAMS dataset contains 29,104 complete cases. Considering these complete cases only, Table 4-8 gives statistics for the seven PRAMS input variables that are also available in the PPESO dataset. The PRAMS PTB rate is 17.8%, significantly higher than the PPESO rate of just under 10%. There are several possible explanations for this disparity:

1. When pre-processing the PRAMS data, all cases missing information for Previous Preterm Birth were removed; in clinical databases it is often the case that a missing value implies that the event did not occur (i.e. there was no previous preterm birth). Therefore, the modified PRAMS database may be predisposed towards cases where previous preterm birth occurred, increasing the likelihood of future preterm birth;
2. Because the PRAMS study includes a questionnaire and telephone survey component, it is possible that parents of preterm babies may have been more motivated to participate in the research;
3. Statistically, the US preterm birth rate is higher than the Canadian rate, at approximately 12%, although important racial differences exist, with preterm births occurring in 17.4% of live births to black American women (Public Health Agency of Canada, 2006). Of the preterm birth cases in the PRAMS database, 22.4% occur in black women who make up 13.1% of the entire population – yielding an exceptionally high 30.4% PTB rate.

The variable prevalence for: smoking, singleton birth, maternal age, gender and

parity, are not significantly different between the PPESO and PRAMS databases. The maternal breastfeeding variable is substantially lower, this may be partly attributable to the shorter maternity leave policy in the United States. It is also possible that because the questionnaire and interview component were conducted after the birth, the responses reflected the reality of their breastfeeding situation (as opposed to the PPESO data which indicated only maternal intent). The PRAMS dataset is missing the Previous Term Birth variable, which is not considered detrimental to the outcome prediction, as the PPESO experiments found that this variable was not a good indicator of future preterm birth.

TABLE 4-8 PRAMS PTB STATISTICS FOR PPESO 8-VARIABLE MODEL

PTB RATE	17.8
VARIABLE PREVALENCE (%) – CONSIDERING PTB CASES ONLY	
SMOKING	19.6
MATERNAL BREASTFEEDING	5.78
PREVIOUS PTB	34.0
SINGLETON BIRTH	83.8
GENDER = MALE	50.2
AGE*	28.8
PARITY*	1.89

* These are continuous variables – mean values for all years combined are

provided.

4.7.2 PROCESSING STEPS

Three different sets of classification-based MLP ANN experiments were performed on the PRAMS database:

1. Predicting preterm birth with a non-artificially distributed seven-variable model (the PPESO eight-variable model, minus the ‘Previous Term Birth’ variable).

Artificial distribution was not required as the PTB rate was 17.8% (test) and 18.2% (train); this is above the 15% minimum required for learning to occur (Ennett & Frize, 2000). *Chapter 5* shows that the experimental results with a training and test rate of approximately 18% generate good learning, further justifying this decision. As is the case for each of the sets of experiments, the 29,104 cases were divided into a verification set (1/3) and a remainder set (2/3). The 9,701 cases in the verification set are then divided into ten separate sets (nine sets with 970 cases each, and one set with 971 cases). The remainder set is further divided into a 2/3 training (12,935 cases) and 1/3 test set (6,468 cases).

2. Predicting preterm birth with a non-artificially distributed 48-variable model.
3. Predicting *high-risk* preterm birth with an artificially distributed 48-variable model. The high-risk preterm birth criteria described for the PPESO data is employed here: an infant is considered high-risk if born before 33 weeks gestation, infants born between 33 and 36 weeks are only labeled high-risk if they were admitted to the NICU, or had a birth weight of less than 1500 grams, or had congenital malformations. For the 48-variable model, 8.3% of the cases are considered high-risk; artificially re-sampling was used to increase the distribution of the training set with respect to high-risk PTB cases to 20%.

4.7.3 VARIABLE SIGNIFICANCE IN THE HIGH-RISK PRETERM BIRTH DATASET

From the 48-variable PRAMS model, not all variables have an equal weight in predicting the outcome. In fact some variables might have no, or even a negative impact and too complex a model may result in a less generalizable network. The modified

Garson model described in section 4.5.3.1 was used to rank the relevance of the 48 input variables. Variables identified via the relative weights method as having little impact are removed and the impact on the outcome is assessed. The minimum number of variables required to meet the performance of the complete test set forms the new perinatal high-risk preterm birth minimum dataset. It is also possible, in a multi-factorial problem such as predicting PTB, that all variables are required to achieve the desired sensitivity and specificity values. This work aims to understand the complex interactions that exist between risk factors and pre-term birth, possibly resulting in a new minimum data set for yielding generalizable preterm birth predictions.

4.7.4 ADJUSTABLE NETWORK PARAMETERS

Table 4-9 summarizes the network parameter settings that achieved the best performance²³ results using a classification-based MLP ANN with one hidden layer for each of the three outputs: preterm birth based on the seven-variable non-artificially distributed model; preterm birth based on the non-artificially distributed 48-variable model; and high-risk preterm birth based on the artificially distributed 48-variable model. Network settings are presented for the networks that produced the best compromise between sensitivity and specificity, while favoring sensitivity (i.e. those that achieved the highest sensitivity while maintaining a specificity greater than 85%). Settings include the number of hidden nodes and the values for the nine network parameters. Corresponding performance results are presented in *Chapter 5*.

²³ The best performance results are derived from the PRAMS test dataset.

TABLE 4-9 NETWORK SETTINGS OF THE BEST PERFORMING DOUBLE LAYER NETWORKS FOR PREDICTING PTB AND HIGH-RISK PTB USING PRAMS DATASET

EXPERIMENT	PTB – 7- VARIABLE MODEL	PTB – 48 VARIABLE MODEL	HIGH-RISK PTB – 48 VARIABLE MODEL WITH ARTIFICIAL DISTRIBUTION
Train PTB rate (%)	18.2	17.6	20.0
Test PTB rate (%)	17.8	17.9	7.68
Hidden nodes	8	11	4
Final learning rate*	0.0009356	0.00052188	0.00017406
Learning rate increment	1.001	1.001	1.001
Learning rate decrement	0.999	0.999	0.999
Final weight decay constant*	0.0005005	0.00041995	4.5E-007
Weight decay constant increment	1.001	0.8399	1.0625
Weight decay constant decrement	0.99	0.999	0.0009
Final weight scale (w_o)*	0.01	0.01	0.0275
Momentum	0.5	0.5	0.5
Final error ratio*	1.25	1.25	1.02

* Indicates network settings taken from the final run parameters

4.8 DEVELOPING THE HYBRID SCREENING SYSTEMS FOR PERINATAL RISK ASSESSMENT (PTB-SPECIFIC)

Many clinical screening systems are based on the premise of a ‘double-pass’ system. The first pass offers broad screening that will identify a large proportion of the test population as positive (having incidence of the disease), in an attempt to correctly identify as many true positives as possible (i.e. maximizing for sensitivity). The downside is the correspondingly high number of false positives, which is to be expected since it is challenging to create models that maximize for both sensitivity and specificity. The second pass then refines the specificity, often using more sophisticated and invasive testing, to reduce the number of false positives. Well known examples include maternal screening for fetal Down syndrome, which is generally performed using an initial blood test that measures the levels of three markers for Down syndrome; this test has a high false positive rate. If these initial tests indicate an increased likelihood of a Down

syndrome fetus, there are several more invasive and specific tests available, such as amniocentesis, chorionic villus sampling, and percutaneous umbilical blood sampling, each having an associated risk of complications.

Using the clinical approach of double pass screening as a basis, this work creates a hybrid artificial intelligence-based screening system for the early prediction of high-risk preterm birth in new patient cases. The idea is to develop a multiple pass system that first feeds all new cases through the trained classification-based MLP ANN model that yielded the highest sensitivity. Further passes would improve the specificity (i.e. eliminate as many false positives as possible). The model would serve as an effective preliminary screening tool, based on non-invasive testing; clinical laboratory tests (i.e. fetal fibronectin testing) could be applied to patients identified as high-risk with improved effectiveness and reduced cost.

4.8.1 PERFORMANCE MEASURES: ENSURING CONFIDENCE IN THE MODEL

Different clinical outcomes must maximize (or minimize as the case may be) different performance measures depending on the clinical impact of misclassifying the outcome. Determining which performance measure to maximize for the case of preterm birth presents a highly complex problem. Consider that there is no ‘silver bullet’ therapy or treatment for preventing preterm birth; at best doctors can mitigate risk. Therefore, identifying a woman as high-risk does not necessarily imply that the negative outcome can be avoided. As previously discussed (see section 1.4), preterm birth prediction is used primarily for pregnancy management. This management and monitoring-centred approach differs significantly from the treatment of a disease such as breast cancer, where

failure to identify the disease in its early stages can potentially lead to death. Additionally, follow-up testing for preterm birth by measuring cervical length and fetal fibronectin testing, while invasive and costly, pose little health risks to the mother or fetus. This is not the case with other obstetrical follow-up testing, such as amniocentesis. An initial non-invasive screening system that offers results before 23 weeks gestation would be exceptionally useful in indicating if further testing is necessary. Currently, in the absence of strong risk factors, this testing is not performed.

Table 4-10 details the clinical implications (benefits and drawbacks) of true and false positive results, and true and false negatives results, when considering the problem of preterm birth prediction.

TABLE 4-10 CLINICAL IMPLICATIONS OF PREDICTION RESULTS FOR PRETERM BIRTH

	BENEFITS	DRAWBACKS
TRUE POSITIVE	Follow up invasive testing performed after 23 weeks gestation to re-evaluate mother's risk. Risk is identified in time to modify the pregnancy management strategy in an effort to prevent the PTB (this may include: bed rest, adequate hydration, prescribing tocolytic agents, administering medications to prevent infant infection and to speed infant lung development).	Some mothers may prefer to not know of their elevated risk and avoid receiving extra testing. It is also possible that knowing of the elevated risk could incur additional maternal stress.
FALSE POSITIVE	It is possible that the additional prenatal care could be beneficial, if other maternal or infant conditions were detected and managed.	Mother will receive additional monitoring through increased prenatal visits, resulting in increased costs for the healthcare system. Mother will receive additional, follow-up testing (cervical measurements and fetal fibronectin testing) – note that while invasive this testing does not pose a risk to the infant.

Additional maternal stress.		
TRUE NEGATIVE	Mother will not require additional testing. Less maternal stress knowing there is a lower risk of PTB.	None Identified.
FALSE NEGATIVE	None identified.	Mother will not receive additional prenatal care. Pregnancy management strategy and birth plan will not be modified to reflect the mother's risk of high-risk PTB.

The conclusion is reached that the greatest benefit with the least associated negative impact is achieved by maximizing the true positive rate (i.e. sensitivity). However, it is also important to obtain a reasonably high true negative rate, avoiding unnecessary follow-up testing with the associated stress and increased costs.

The first pass through the hybrid screening system will focus on identifying as many positives as possible (including both true and false positives), these will be referred to as Predicted True Positives (PTPs), and then proceed through further iterations to reduce the number of false positives. Note that predicted negatives will be referred to as Predicted True Negatives (PTNs).

Ideally, the model's final sensitivity should approach or exceed the current gold standard for PTB prediction of 64.7%, achieved by Goldenberg et al. using invasive testing (based on a combination of fetal fibronectin testing and cervical length measurements in a population of mothers with a history of prior preterm delivery). If this sensitivity is achieved, this implies that the hybrid screening system applied to a heterogeneous maternal population, is able to produce the same results as a clinically

invasive test applied to an elevated risk population only (remember that the same clinical testing combination of fetal fibronectin testing and cervical length applied to a low risk population yielded a sensitivity of 15%). Additionally, the non-invasive clinical results are available before 23 weeks gestation, rather than only becoming more accurate as the delivery date approaches. This would also exceed the sensitivity values achieved by Crane et al. (44.4%) and Peaceman et al. (43.9%) using fetal fibronectin testing in a heterogeneous population.

One of the challenges in developing clinical models is maximizing the sensitivity while maintaining a clinically acceptable specificity. Based on discussions with Dr. C. R. Walker, it was decided that a specificity approaching 85% would indicate a good true negative rate, to be considered acceptable the specificity must be greater than 80%.

4.8.2 DISTINGUISHING BETWEEN TRUE AND FALSE POSITIVES

When developing clinical ANN-based research models the training, test and verification set outcomes are known, while actual clinical data have unknown outcomes. When a new patient case (with an unknown outcome) is processed by a trained MLP ANN model, a classification-based outcome is produced (e.g. '+1' or '-1'). It is not possible to distinguish between the false positives and true positives for the positive outcomes, and the false negatives and true negatives for the negative outcomes. What is needed is a means of exploiting the refined MLP ANN model for predicting high-risk PTB with high sensitivity, in conjunction with a mechanism that would allow the trained system to distinguish between the true and false positives and true and false negatives for each patient case. The method would be trained and verified on known outcomes, but

would be applicable to unknown outcomes. The proposed solution involves incorporating a back propagation feed forward MLP ANN with risk stratification estimations into the integrated screening system.

4.8.3 INCORPORATING MLP ANN-BASED RISK STRATIFICATION ESTIMATIONS INTO THE INTEGRATED HYBRID SCREENING SYSTEM

As described in section 2.2.3, a MLP ANN with risk stratification estimations attached has recently been developed by MIRG for use in predicting mortality in the neonatal intensive care unit (Zhou & Frize, 2006). The methodology proposes using this tool, as part of the integrated hybrid screening system, to aid in distinguishing between true and false positives, and true and false negatives.

To date, the MLP ANN with risk stratification estimations has been tested on the Canadian Neonatal Network (CNN) database for the prediction of mortality. Similarly to high-risk preterm birth, mortality is a highly skewed outcome; with mortality rates in the CNN database of 3.75% for the complete database with missing values replaced using Ennett's weight imputation method (C. Ennett, 2003).

This work applies the MLP ANN with risk stratification estimations to the realm of high-risk PTB prediction, with the goal of distinguishing between predicted true and false positives, and predicted true and false negatives, by a) stratifying high-risk PTB into risk levels²⁴, and b) assigning a cutoff point within the stratification ranges to identify the predicted true and false predictions (i.e. if the value is greater or equal to the cutoff, mark as a predicted true positive, else mark as a predicted false positive, or alternatively: if the

²⁴ The number of risk levels will vary depending on the distribution of the dataset; this will become apparent when examining the results in *Chapter 5*.

value is less than the cutoff, mark as a predicted true negative, else mark as a predicted false negative). Implementing these two steps will lead to the development of a high-risk PTB screening tool by assigning appropriate cutoff points within the risk stratification ranges.

4.8.4 LINKING THE CLASSIFICATION-BASED MLP ANN WITH THE RISK STRATIFICATION ESTIMATION-BASED MLP ANN

The classification-based MLP ANN from the ANN-RFW (section 2.2.2), was used to develop and validate the PTB and high-risk PTB classification models for the PPESO and PRAMS datasets. By performing automatic parameter updates and iterating through all network structure trials the MLP ANN from the ANN-RFW exhaustively provides all possible network structure results, from which the optimal results are then manually extracted. The nine network parameters (section 4.5.1) and network structure (defined by the number of hidden nodes) that yielded the best results obtained using the classification-based MLP ANN, are then used in the MLP ANN with risk stratification estimations. The MLP ANN with risk stratification-based estimations uses only one set of parameter values to determine the final risk stratification levels – the classification-based MLP ANN will choose these parameter values.

There are two programs associated with the MLP ANN with risk stratification estimations. The first provides risk stratification-based results over the entire dataset for research purposes only. Each patient case, with a known outcome, is assigned to a risk stratification group. The risk distributions given by the ANN are then compared with the real distribution, to measure the performance of the network in terms of discrimination

ability. This program is useful for developing a risk stratification model, but gives no risk stratification information about specific patient cases. The second program, allows the user to further apply an existing risk stratification estimation model as an outcome-screening tool; stratification information is given for each specific patient case. This functionality plays a key role in training the integrated hybrid screening system. Figure 4-3 illustrates how the classification-based MLP ANN and MLP ANN with risk stratification estimations are integrated in the hybrid screening system. Simply speaking, the risk stratification values obtained from the MLP ANN with risk stratification estimations are used to classify all cases as high or low-risk based on different cutoff values. The selected cutoff values vary depending on the performance criteria. Figure 4-3 offers a more detailed description of the 'Pass 1' and 'Pass 2' sub-systems, depicted in Figure 4-1's architectural overview of the integrated hybrid screening system. It is important to recognize that the figure depicts the first two steps in training the hybrid model, not the application of the model to new patient cases. The circled numbers in Figure 4-3 correspond to those in the numbered list below, which summarizes the integration methodology.

Pass 1: Maximizing for sensitivity

1. The first step is to develop the classification-based MLP ANN models. Because the PRAMS high-risk PTB dataset yielded the highest sensitivity values (see *Chapter 5* results), the artificial training set and the real test set were chosen from this database. A criterion is tested and defined: specifically, while maximizing the sensitivity at greater or equal to 65%, the specificity must be

greater or equal to 80%, with a c-value greater or equal to 0.75. The criterion can be considered an iterative analysis to optimize sensitivity while retaining thresholds that make clinical sense for all performance metrics (Goodwin & Maher, 2000). Higher sensitivity values were tested (with correspondingly lower specificities), but these did not generate a learning model that would be acceptable for estimating PTB on new patient data.

2. Extract the nine classification-based MLP ANN network parameters and number of hidden nodes required to obtain the optimal performance results from Step 1.
3. Using these parameters and network structure, perform risk stratification using the MLP ANN with risk stratification estimations.
 - a. Perform validation by comparing the predicted classification of the test data with the actual risk stratification of the training dataset.
 - b. If the validation is successful, proceed to Step 4, extracting the weights and biases from the MLP ANN with risk stratification estimations validated in Step 3a.
 - c. Else, return to Step 1, and modify the performance criteria as necessary (this implies that the desired criteria cannot be met by the classification-based MLP ANN parameter values chosen in Step 2, and a less stringent criterion is needed).
4. After successful validation of the MLP ANN with risk stratification estimations, generate patient-specific outcomes.

5. Based on analysis of test data generated by the MLP ANN with risk stratification estimations, 'cutoff' values are defined. E.g. if the value is greater or equal to the cutoff value, the case is marked as Predicted True Positive (PTP), else mark as Predicted Negative (PN).
6. At completion of 'Pass 1', each patient case originally classified as '+1' by the classification-based MLP ANN, will have been further classified as a PTP or a PN. This categorization is based on the classification-based MLP ANN network parameters chosen to maximize *sensitivity*, which identified the greatest number of TPs originally.

Pass 2: Maximizing for specificity

1. 'Pass 2' aims to identify the greatest number of true negatives by maximizing the network specificity. Because the results from 'Pass 1' and 'Pass 2' will be compared, it is necessary that 'Pass 2' results be generated using the same dataset as in 'Pass 1' (i.e. the PRAMS 48-variable HR-PTB training set). A criterion is tested and defined: specifically, while maximizing the specificity at greater or equal to 88%, the sensitivity must be greater or equal to 45%, with a c-value greater or equal to 0.75.
2. Extract the nine classification-based MLP ANN network parameters and number of hidden nodes required to obtain the optimal performance results from Step 1.
3. Using these parameters and network structure, perform risk stratification using the MLP ANN with risk stratification estimations.
 - a. Perform validation by comparing the predicted classification of the

test data with the actual risk stratification of the training dataset.

This is important because in order for the results from the MLP ANN with risk stratification estimations to be considered valid, the stratification of the test set must agree with that of the training data.

- b. If the validation is successful, proceed to Step 4, extracting the weights and biases from the MLP ANN with risk stratification estimations validated in Step 3a.
 - c. Else, return to Step 1, and modify the performance criteria as necessary (this implies that the desired criteria cannot be met by the classification-based MLP ANN parameter values chosen in Step 2, and a less stringent criterion is needed).
4. After successful validation of the MLP ANN with risk stratification estimations, generate patient-specific stratification outcomes.
5. Based on analysis of test data generated by the MLP ANN with risk stratification estimations, stratification 'cutoff' values are defined. E.g. if the value is less than the cutoff value, the case is marked as Predicted True Negative (PTN), else mark as Predicted Positive (PP).
6. At completion of 'Pass 2', each patient case originally classified as '-1' by the classification-based MLP ANN will have been further classified as a PTN or a PP by the integrated screening system. This categorization is based on the classification-based MLP ANN network parameters chosen to maximize the

specificity, which would succeed in identifying the greatest number of TNs.

FIGURE 4-3 FLOWCHART (UML ACTIVITY DIAGRAM) OVERVIEW OF THE HYBRID SCREENING MODEL ('PASS 1' AND 'PASS 2')

4.8.4.1 Inconsistent Predictions

After completing 'Pass 1' and 'Pass 2', all patient cases will have been classified once by each subsystem. All cases are examined to determine if the predicted outputs are consistent. For instance, if a case has been marked as both a predicted true positive and a

predicted true negative, this indicates a logical flaw. Cases that are marked consistently by the two subsystems are considered complete. Cases with inconsistent predicted outputs require further assessment. Table 4-11 provides a matrix describing this logic:

TABLE 4-11 LOGIC FOR DETERMINING CONFLICTING CASES		
PREDICTED OUTCOME 'PASS 1'	PREDICTED OUTCOME 'PASS 2'	INPUT TO 'PASS 3'
PTP	PP	TP
PTP	PTN	CONFLICT
PN	PP	CONFLICT
PN	PTN	TN

4.8.5 INCORPORATING A KNN CASE-BASED REASONING TOOL TO ASSESS INCONSISTENT PREDICTED OUTPUTS

MIRG has developed and tested a k-Nearest Neighbors (kNN) case-based reasoning tool for use in the Neonatal Intensive Care Unit (see section 2.2.3); Ennett and Frize have successfully employed a similar tool to replace missing values in the Canadian Neonatal Network database. Described broadly, their approach was to use the ANN connection weights as the weighting terms for the CBR. Ennett and Frize's loosely coupled hybrid system effectively integrates the ANN and CBR for the purpose of replacing missing values in clinical databases. This thesis extends the integration by developing a methodology for integrating the ANN and kNN (which MIRG is now using to replace the original CBR tool which became obsolete) for the purpose of assessing inconsistent predicted outcomes, effectively broadening the functionality of the hybrid ANN-kNN system to include outcome prediction.

4.8.5.1 Using ANN Weights for the kNN

When transferring the ANN weights to kNN format, Ennett simplified the approach by focusing on single layer networks (Ennett, 2003). This meant that the

weights on the connections departing each input node were summed and the relative importance for each input variable was calculated. This approach is not sufficient for predicting high-risk PTB where the best prediction results were obtained using MLPs ANNs with multiple hidden nodes. Section 4.7.3 described the relative weights approach, developed initially by Garson and modified by Rybchynski, used to rank the variables for predicting high-risk PTB. This approach will be used to add meaning to the ANN connection weights when linked with the kNN.

Ennett astutely pointed out that although her CBR weightings were based significantly on the ANN connection weights, there are certain variables that neonatologists expect to find in a clinical scoring system for estimating mortality (Ennett, 2003). The same applies to estimating high-risk PTB. Physicians would have difficulty establishing confidence in a model that fails to place significance on known risk factors, specifically: previous term birth and multiple gestations. The final HR-PTB weighting model will be examined to ascertain whether it corroborates with the clinical literature on known PTB risk factors. However, in an effort to 'let the data speak', this methodology does not advocate changing the weightings as was proposed by Ennett. The rationale being that there is currently no known risk model for early PTB predictions, suggesting that data interactions exist that have not been identified in the clinical literature.

The kNN accepts weights as whole numbers from 0 to 100, with 0 being an input that is to be ignored, such as an administrative field, and 100 being the most important input field. Transferring weights from ANN format to kNN format involves three steps:

1. Extracting each input from the array of connection weights provided by the MLP ANN and using the modified relative weight method to find a weighting value for each input variable;
2. Comparing ANN-based weights with clinical literature to ascertain clinical relevance of the model from Physician's perspective;
3. Performing linear scaling to transform the relative weights, calculated from the ANN, to kNN scaled weights.

4.8.5.2 Outcome Prediction Using the k-Nearest Neighbors (kNN) Case-Based Reasoning Tool and the Decision Tree Voting Algorithm

The kNN finds the closest matching cases, using a matching algorithm that takes into consideration the weights assigned to each input variable. As explained, this thesis uses the MLP ANN connection weights to define the kNN weightings. Recall that the classification-based MLP ANN from 'Pass 1' maximized network sensitivity, and 'Pass 2' maximized network specificity. However, following the first two passes through the system, cases remain that contain conflicting predicted outcomes. 'Pass 3' will strive for a compromise between sensitivity and specificity in order to classify these inconsistent cases. 'Pass 3' integrates three decision support techniques: MLP ANN, kNN-based CBR, and a DT voting algorithm. Figure 4-4 depicts how the techniques are combined for training purposes, with the circled numbers in the figure corresponding to those in the numbered list below, which summarizes the integration methodology for 'Pass 3':

Pass 3: Finding a compromise between sensitivity and specificity

1. Following 'Pass 2' extract all marked cases with conflicting outcomes based on the

logic described in Table 4-11.

2. The classification-based MLP ANN high-risk PTB model based on the PRAMS 48-variable high-risk PTB dataset achieved the best compromise between sensitivity and specificity. The compromise criterion is tested and defined as: a sensitivity approaching 65% and specificity approaching 85%.
3. Extract the ANN connection weights for each hidden layer and hidden node.
4. Transform the ANN connection weights to kNN weightings as described in section 4.8.5.1, in this way the kNN will be programmed to find a compromise between sensitivity and specificity.
5. Create the four database tables required by the kNN tool:
 - a. Create the *weights* table containing the weightings associated with each attribute.
 - b. Create the *index* table containing the minimum and maximum values for each attribute.
 - c. Create the *mean* table containing the average value of each attribute.
 - d. Create the *match* table, which contains all training data from the PRAMS 48 variable high-risk PTB dataset (including the known output values).
6. Develop two *query* database tables, one containing all conflicting cases that are really TPs (called MatchTP – 194 cases), and the other containing all conflicting cases that are really TNs (called MatchTN – 1376 cases). Conflicting cases are

obtained from the test data following 'Pass 1' and 'Pass 2'. The *query* tables contain the data with missing information, which the kNN tool will fill in on a case by case basis using the running average of the k closest matching cases. In this case the missing value is the outcome field (High-risk PTB) which is left blank in the *query* tables.

7. Run the kNN tool
 - a. Connect to the appropriate database via the ODBC connection.
 - b. Build the index (which sets the minimum, maximum, and mean values, and weights).
 - c. Choose k value (the desired number of matches).
 - d. Use the system's automatic matching functionality (AutoMatch) to find the k closest matching cases and, in conjunction with the DT logic, fill in the outputs in the *query* tables for both MatchTP and MatchTN. A screenshot of the kNN tool matching based on the MatchTP database is shown in figure 4-5. The kNN algorithm is described in detail in 2.2.4.
8. Use the decision tree-voting algorithm's cutoff value to classify conflicting cases as 'PTPs' or 'PTN's for both the conflicting actual TP cases from MatchTP, and the conflicting actual TN cases from MatchTN. Use this information to re-assign the conflicting cases for each group. The voting algorithm is explained in detail in section 5.4.3.1. Determine the new sensitivity and specificity. If these values meet or exceed the clinical performance criteria, and the sensitivity value is higher than

any previously achieved, define the current k and cutoff values as the new optimal values. This entire process continues until the optimal combination of k and cutoff values are found.

FIGURE 4-4 FLOW OF CONTROL FOR TRAINING HYBRID ANN, kNN AND VOTING ALGORITHM SYSTEM

CBRS

Step 1: Set Up Connection

Data Source Name:
MatchTP

Username: Database Specific
[]

Password: Database Specific
[]

Number of Primary Keys:
1

Set Connection

Step 2: Make your Selection

Patient ID:
14991

Number of matches:
10

Weights:

MAT_AGE_N	PLURAL_N	PREV_PRE	PARITY_N	GENDER	BF_EVER	MOMSMOKE
38	100	64	25	30	64	18

Set Weights **Save Weights to DB**

Step 3: Send Request

Build Index **Match**

AutoMatch

FIGURE 4-5 SCREENSHOT OF KNN-BASED CBR SYSTEM USING 'AutoMatch' FUNCTION

4.8.6 VERIFICATION: USING INTEGRATED SCREENING TO PREDICT OUTCOMES FOR NEW PATIENT CASES

The integrated hybrid screening system, developed using the PRAMS 48 variable

high-risk PTB training and test sets, is verified on the PRAMS 48 variable high-risk PTB verification set. The verification set comprises 9701 independent complete cases that were not used during training or testing. The goal is to meet or exceed the sensitivity and specificity results achieved during training of the integrated hybrid screening system. The verification methodology is described below and depicted in Figure 4-6.

- 1) Process the data by the trained risk stratification estimation ANN MLP tool using the parameters, weights and biases from Step 4 of 'Pass 1' (i.e. those that achieved the best sensitivity - weights and biases provided in Appendix C).
- 2) Classify resulting cases as HR-PTB based on a cutoff value of 0.4 (e.g. predicted outputs in the range of 0.4 - 0.5 and above are classified as high-risk). Mark cases as PTP or PN. Note the cutoff value was determined through experimentation, results are presented in 5.4.1.2.
- 3) Process the data by the trained risk stratification ANN MLP tool using the parameters, weights and biases from Step 4 of 'Pass 2' (i.e. those that achieved the best specificity - weights and biases provided in Appendix C).
- 4) Classify resulting cases as HR-PTB based on a cutoff value of 0.6 (e.g. predicted outputs in the range of 0.6 - 0.7 and above are classified as high-risk). Mark cases as PTN or PP. Note the cutoff value was determined through experimentation, results are presented in 5.4.2.2.
- 5) Mark conflicting cases following 'Pass 1' and 'Pass 2' based on logic described in Table 4-11.
- 6) Using the same match, index, mean and weights tables, run the kNN tool for a new

query file. The new query file contains all 9701 verification cases with the outcomes removed. Choose a k value, note the k value was determined through experimentation, results are presented in 5.4.3.1.

7) Based on a cutoff of 0.1, reassign all conflicting cases as PTP or PTN (if predicted output $\geq 0.1 \rightarrow$ PTP, else PTN). Note the cutoff value was determined through experimentation, results are presented in 5.4.3.1.

8) Based on the true values, determine sensitivity and specificity and compare with the clinical performance requirements of this thesis to verify the model.

FIGURE 4-6 FLOW OF CONTROL FOR VERIFYING INTEGRATED HYBRID SCREENING SYSTEM

After verification of the results on new patient data, the integrated hybrid screening system can then be used on new patient cases. It is important to understand that the time-intensive classification-based MLP ANN will never be run when applying the

final system to new data. The flow of control for obtaining new outcome predictions is similar to that described in Figure 4-6; however, the model is not verified against the known outcome and the final sensitivity and specificity are not calculated. The kNN tool will only be invoked if the results from 'Pass 1' and 'Pass 2' are found to be in conflict.

To keep the case-base current, new patient cases should be continually added to the *match* database, used for retrieving the k closest matching cases. Additionally, an ideal system would periodically re-train the entire integrated hybrid screening system based on adding new cases to the datasets used during MLP ANN training. The details of this implementation are left for future work, and would include evaluating how often the datasets and case base should be updated, and the impact on the results.

CHAPTER 5. RESULTS

5.1 INTRODUCTION

This chapter begins by presenting the PPESO and PRAMS classification-based MLP ANN results for predicting PTB and high-risk PTB in sections 5.2 and 5.3 respectively. The integrated hybrid screening system results are presented in section 5.4.

5.2 PPESO CLASSIFICATION-BASED MLP ANN RESULTS

This section assesses if changing the PTB rate distribution of the training datasets increases the classification-based MLP ANN performance over one trained on the non-artificially distributed data. Two datasets are compared: a full eight-variable model and a new minimum dataset obtained using weight elimination. This is followed by an investigation of a new high-risk PTB model, developed by refining the initial PTB outcome using two permutations of the concept of 'high-risk'. The new model exhibited highly skewed data. Artificially increasing the underrepresented class in the training set is again used in an attempt to increase the performance of the outcome estimations.

5.2.1 PREDICTING PRETERM BIRTH

Results achieved using the real PTB distribution of 8.6% were poor, with a maximum sensitivity of 20%. The results presented in Table 5-1 indicate that artificially

increasing the distribution of the under-represented outcome in the training set improves the sensitivity of the classification-based MLP ANN at the cost of slightly lower results for specificity and Correct Classification Rate (CCR). However, it can be argued that the classification-based MLP ANN trained on the artificial datasets is of more value to physicians because a greater number of mothers at risk of PTB are correctly classified. The network model trained to predict PTB using the 2001 artificial training dataset outperformed the model trained on the original dataset: sensitivity increased by 13%, peaking at 33.4%, and the area under the ROC curve also increased.

Table 5-1 also shows the results of removing each variable individually. Eliminating 'Previous term birth' (Term) increased the sensitivity from 33.4% to 36.6%, at the expense of lower specificity and CCR. Eliminating 'Intention to breastfeed' (BF) increased the sensitivity only minimally to 33.8%. Running classification-based MLP ANN experiments with the new six-variable minimal dataset yielded a sensitivity of 33.8%, only marginally higher than the best sensitivity achieved using the eight-variable model. Based on these results it is difficult to justify removing both variables from the preterm prediction model, however the increased sensitivity achieved using the seven-variable model without 'Term' suggests that previous term birth is not a good indicator of future PTB. As the PRAMS dataset does not include previous term birth, this finding is of particular interest in that it justifies ignoring this variable.

The classification-based MLP ANN identified four variables as having the highest connection weights, listed in order of significance they are: number of babies this pregnancy, number of previous preterm babies, total number of previous children, and

maternal smoking after 20 weeks gestation. These results help to validate the applicability of the classification-based MLP ANN to clinical PTB estimation, as each of the four variables is commonly believed to be a risk factor associated with an increased frequency of PTB (Meis et al., 1998).

TABLE 5-1 PERFORMANCE MEASURES OF THE BEST-PERFORMING PPESO-BASED DOUBLE LAYER NETWORKS FOR PREDICTING PTB

EXPERIMENT	ORIGINAL	ARTIFICIAL	WEIGHT ELIMINATION: REMOVING (TERM) ²⁵	WEIGHT ELIMINATION: REMOVING (BF) ²⁶	WEIGHT ELIMINATION: REMOVING (TERM) AND (BF)
Train PTB rate (%)	8.60	23.1	23.1	23.1	23.1
Test PTB rate (%)	8.67	10.6	10.6	10.6	10.6
Hidden nodes	3	3	3	3	3
Hidden layers	2	2	2	2	2
Sensitivity (%)	20.4	33.40	36.6	33.8	33.8
Specificity (%)	98.7	96.3	93.9	95.1	95.0
CCR (%)	91.9	90.0	88.2	89.0	88.9
Average Squared Error	0.286	0.410	0.447	0.365	0.595
Area under ROC (test)	0.6455	0.7066	0.70	0.6777	0.6248
Area under ROC (train)	0.6530	0.6875	0.6596	0.6638	0.6122
Constant predictor (%)	91.4	90.0	90.0	90.0	90.0

5.2.2 PREDICTING HIGH-RISK PRETERM BIRTH

The new PPESO-based high-risk PTB model successfully increased the network sensitivity. Table 5-2 shows the performance measures of the three best performing classification-based MLP ANN networks with one hidden layer for predicting both high-

²⁵ Term: Previous term births

²⁶ BF Intention to breast feed

risk PTB and high-risk PTB excluding free-flow oxygen cases. For each model, three results are displayed in order of increasing sensitivity: (1) the network that achieved the best specificity while maintaining a sensitivity of over 36.6% (the best sensitivity value achieved using the seven-variable PTB model with Term removed via weight elimination); (2) the network that achieved the best compromise between sensitivity and specificity; and (3) the network that achieved the best sensitivity while maintaining a specificity of over 85%.

The results show sensitivity increasing to 45.8% (for the high-risk PTB model with 8 hidden nodes), and peaking at 54.8% (for the high-risk PTB model with free flow oxygen cases removed and 4 hidden nodes); a substantial improvement over the original PTB non-artificially distributed result of 20.3%. This was also an improvement over the best sensitivity results obtained using the PTB artificially distributed model (36.6% for the new seven-variable model with Term removed via weight elimination). The area under the ROC curve for both the training and test sets increased (ROC under test: 0.7319 for the 8-node high-risk PTB model and 0.7232 for the 4-node high-risk PTB model with free flow oxygen cases removed), indicating that the new models were more effective at discriminating between outcomes.

TABLE 5-2 PERFORMANCE MEASURES OF THE THREE BEST-PERFORMING PPESO-BASED DOUBLE LAYER NETWORKS FOR PREDICTING HIGH-RISK PTB AND HIGH-RISK PTB EXCLUDING FREE-FLOW OXYGEN CASES

EXPERIMENT	HIGH-RISK PTB			HIGH-RISK PTB (-FFO2)		
TRAIN RATE (%)	20			20		
TEST RATE (%)	5.5			4.0		
HIDDEN NODES	3	9	8	11	3	4
SENSITIVITY (%)	38.3	40.0	45.8	39.1	40.6	54.8
SPECIFICITY (%)	92.9	91.4	87.9	91.5	90.6	85.8
CCR (%)	89.8	88.3	85.5	89.7	87.6	84.8

AVERAGE SQUARED ERROR	0.32	0.36	0.40	0.31	0.36	0.40
AREA UNDER ROC (TEST)	0.71	0.72	0.73	0.73	0.70	0.72
AREA UNDER ROC (TRAINING)	0.71	0.71	0.73	0.72	0.70	0.71
CONSTANT PREDICTOR (%)	94.5	94.5	94.5	96.0	96.0	96.0

-FFO2 = Excluding cases where resuscitation was offered in the form of free flow oxygen

CCR = Correct Classification Rate

In terms of misclassification, it was encouraging to note that the new high-risk models were able to maintain specificity values over 85%, although all models yielded lower specificity values than the value obtained using the seven-variable PTB model with Term removed via weight elimination. Specificity values ranged from 87.9% to 92.9% for the high-risk PTB model, and 85.8% to 91.5% for the high-risk PTB model with free flow oxygen cases removed. Attempts to increase the distribution of the under-represented class beyond 20% yielded unacceptably low specificity values.

5.2.3 PPESO VERIFICATION SET RESULTS

Table 5-3 compares the sensitivity and specificity of the verification sets used for each of the three risk models: PTB, high-risk PTB, and high-risk PTB with free flow oxygen cases removed. The network parameter settings for these modes are provided in Table 4-6. The verification experiments were run using the parameter settings of the networks that yielded the highest sensitivity values for each of the three risk models. The results are the average (mean) performance of the 10 verification sets (1971 cases each) and the standard deviation of those experiments. Although differences in results for the verification set are to be expected, in general the results indicate that the ANN was effective in creating a generalized model. In particular, sensitivity increased from 33.4%

to 35.6% for predicting PTB (specificity 88.6%), and decreased from 45.8% to 37.3% for predicting high-risk PTB (specificity 92.2%). The high-risk PTB model with free-flow oxygen cases removed yielded the highest sensitivity of 50.1% (as compared to 54.8%) and was accompanied by a decrease in specificity to a low of 85.1% from 85.8%, just within the acceptable limit. Verifying the data against ten independent data sets demonstrates that the model developed using data from 1999-2001 can be generalized to data from future years; validating the model's robustness and applicability as a predictor of new patient cases.

TABLE 5-3 VERIFICATION SET RESULTS FOR ALL RISK MODELS (PPESO)

EXPERIMENT	PTB	HIGH-RISK PTB	HIGH-RISK PTB (-FFO2)
SENSITIVITY (%)	35.6 ± 0.062	37.3 ± 0.057	50.1 ± 0.083
SPECIFICITY (%)	88.6 ± 0.0058	92.2 ± 0.0053	85.1 ± 0.0084

-FFO2 = Excluding cases where resuscitation was offered in the form of free flow oxygen

5.3 PRAMS CLASSIFICATION-BASED MLP ANN RESULTS

The classification-based MLP ANN PRAMS results begin by repeating the PPESO seven-variable model experiments for predicting PTB (this is the PPESO model with Term removed). This serves to validate the PPESO minimal dataset and to demonstrate that the high-risk criterion increases the network's ability to predict PTB. Results are compared with the PPESO model; differences in the results are discussed.

Next, results obtained from the 48-variable PRAMS model are presented, for predicting both PTB and high-risk PTB. The 48-variable classification-based MLP ANN model for predicting high-risk PTB yields the best overall results, and is consequently used to determine the ranking of the variables based on the modified Garson's relative weights method.

5.3.1 PREDICTING PTB USING THE 7-VARIABLE PRAMS MODEL

Results are presented in Table 5-4, categorized based on their sensitivity and specificity values as: 1) highest sensitivity; 2) best compromise between sensitivity and specificity (i.e. favoring sensitivity while maintaining a specificity of over 85%); 3) compromise between sensitivity and specificity (favoring specificity while maintaining a sensitivity of over 30%); 4) compromise between sensitivity and specificity (no bias); and 5) highest specificity. The PRAMS results perform similarly to the PPESO dataset for predicting PTB, particularly when selecting a compromise between sensitivity and specificity, with the sensitivity higher (PRAMS: 45.3%; PPESO: 36.6%) and specificity slightly lower (PRAMS: 90.1%; PPESO: 93.9%). The similarities in the results prove that the PPESO minimal dataset (excluding previous term birth) is a generalizable model, which can be applied to different databases to yield equivalent or improved results. The results for the best compromise between sensitivity and specificity show a higher sensitivity being achieved with the PRAMS model (PRAMS: 45.3%; PPESO: 36.6%). The highest specificity reached with the PRAMS model of 95.8% (sensitivity 27.7%) was slightly lower than that of the PPESO model - 96.3% (sensitivity 33.4%). The best sensitivity achieved with the PRAMS model of 51.5% far exceeded the best achieved with the PPESO model (36.6%), however the corresponding specificity was low (82.6%).

The PRAMS database contains a greater number of cases collected from a more diverse population (i.e. all 52 American states as compared to the province of Ontario only) and the results indicate that it is a more sensitive classifier of PTB.

TABLE 5-4 TEST SET PERFORMANCE RESULTS: PREDICTING PTB USING THE PRAMS SEVEN-VARIABLE MODEL

EXPERIMENT	HIGHEST SENS.	COMPROMISE (FAVORING SENSITIVITY)	COMPROMISE (FAVORING SPECIFICITY)	BEST COMPROMISE BETWEEN SENS. & SPEC.	HIGHEST SPEC.
Train PTB rate (%)	18.2	18.2	18.2	18.2	18.2
Test PTB rate (%)	17.8	17.8	17.8	17.8	17.8
Hidden layers	0	2	2	2	2
Hidden nodes	1	8	6	9	9
Sensitivity (%)	51.5	46.7	38.9	45.3	27.7
Specificity (%)	82.6	89.3	92.1	90.1	96.0
CCR (%)	82.6	81.7	82.6	82.2	83.8
Average Squared Error	0.573	0.528	0.495	0.498	0.495
Area under ROC (test)	0.7041	0.7071	0.7162	0.7014	0.7022
Constant predictor (train %)	81.8	81.8	81.8	81.8	81.8

5.3.2 PREDICTING PTB USING THE 48-VARIABLE PRAMS MODEL

The improved sensitivity values achieved using the PRAMS seven-variable model validates the usefulness of the PRAMS data in predicting PTB using a classification-based MLP ANN. The logical next step is to measure the effectiveness of the PRAMS 48-variable model in predicting PTB, to see if better results can be achieved with more input variables. Table 5-5 presents the performance results for predicting PTB using the 48-variable PRAMS model. The same five categories were used as with the previous PRAMS results. The highest sensitivity increased to 60.7%; however the corresponding specificity is under the 85% threshold (83.0%). Nevertheless, this model offers a clinically acceptable sensitivity, approaching the best sensitivity achieved from the

literature of 65%. The 'compromise in favor of sensitivity' gives a sensitivity of 54.7% (an improvement over the best sensitivity of 51.5% achieved using the 7-variable PRAMS model), with a specificity of 86.1%. The highest specificity of 94.4% is less than that of 95.8% reached with the 7-variable model; however the corresponding sensitivity was much higher. The 'best compromise' result gave a sensitivity of 50.2% with a specificity of under 90%.

TABLE 5-5 TEST SET PERFORMANCE RESULTS: PREDICTING PRETERM BIRTH USING PRAMS
48-VARIABLE MODEL

EXPERIMENT	HIGHEST SENS.	COMPROMISE (FAVORING SENSITIVITY)	COMPROMISE (FAVORING SPECIFICITY)	BEST COMPROMISE BETWEEN SENS. & SPEC.	HIGHEST SPEC.
Train PTB rate (%)	17.6	17.6	17.6	17.6	17.6
Test PTB rate (%)	17.9	17.9	17.9	17.9	17.9
Hidden layers	2	2	2	2	2
Hidden nodes	11	11	6	2	4
Sensitivity (%)	60.7	54.7	40.9	50.2	37.9
Specificity (%)	83.0	86.1	93.8	89.5	94.4
CCR (%)	79.1	80.5	84.3	82.5	84.3
Average Squared Error	0.626	0.590	0.475	0.508	0.478
Area under ROC (test)	0.7717	0.7617	0.765	0.7627	0.7643
Constant predictor (train %)	82.5	82.5	82.5	82.5	82.5

5.3.3 PREDICTING HIGH-RISK PRETERM BIRTH USING THE 48-VARIABLE PRAMS MODEL

Given the success of the PRAMS 48-variable model in predicting PTB with high sensitivity, the PRAMS high-risk criterion (see section 4.5.2) is then applied to the 48-variable model. It is reasonable to hypothesize that because the PPESO high-risk criterion improved the PPESO dataset results, a similar improvement will occur for the

PRAMS dataset; the results are presented in Table 5-6.

The PRAMS 48-variable high-risk preterm birth model generates the best ROC values of any of the models, being the first to exceed 0.80 – classifying the model as a ‘very good’ predictor. Although the highest specificity of 93.2% is slightly lower than that generated by the other models, the ROC value of 0.8028 and the correct classification rate of 89.3% were the best achieved to date; the corresponding sensitivity of 43.3% is also higher than that achieved for the ‘highest specificity’ category of the other models. Notably, the best sensitivity of 66.4% exceeds the gold standard from the clinical literature. This model will be essential in developing the first pass through the integrated hybrid screening system, which optimizes for sensitivity. The best compromise results are similar to the best results achieved using the PPESO model for predicting high-risk PTB (sensitivity: 55.1% as compared to 54.8%; specificity: 86.1% as compared to 85.8%), the ROC value is also higher (0.7735 as compared to 0.705, which is the difference between a ‘good’ and ‘very good’ classifier).

These results demonstrate that the PRAMS 48-variable high-risk model outperforms both the PPESO seven-variable model and the PRAMS seven-variable model when processed by a classification-based MLP ANN, suggesting that in this clinical outcome, better results can be achieved with a greater number of variables describing the patient’s obstetrical and medical history.

The results from the PRAMS 48-variable high-risk model validate the high-risk criterion identified for the PPESO model and surpasses the highest sensitivity of 65% achieved clinically for predicting preterm birth using invasive testing. It also confirms

that the 'transfer to NICU' variable is an appropriate substitute for the variables 'APGAR 5' and 'Resuscitation requirements'.

TABLE 5-6 PERFORMANCE RESULTS: PREDICTING HIGH-RISK PTB USING THE PRAMS 48-VARIABLE MODEL

EXPERIMENT	HIGHEST SENS.	COMPROMISE (FAVORING SENSITIVITY)	COMPROMISE (FAVORING SPECIFICITY)	BEST COMPROMISE BETWEEN SENS. & SPEC.	HIGHEST SPEC.
Train PTB rate (%)	20.0	20.0	20.0	20.0	20.0
Test PTB rate (%)	7.68	7.68	7.68	7.68	7.68
Hidden layers	2	2	2	2	2
Hidden nodes	6	4	3	3	5
Sensitivity (%)	66.4	56.5	45.1	55.1	43.3
Specificity (%)	80.3	85.2	90.7	86.1	93.2
CCR (%)	79.2	83.0	87.2	83.7	89.3
Average Squared Error	0.644	0.385	0.355	0.383	0.347
Area under ROC (test)	0.7795	0.768	0.7819	0.7735	0.8028
Constant predictor	80.0	80.0	80.0	80.0	80.0

5.3.4 ASSIGNING RELATIVE WEIGHTINGS (SIGNIFICANCE) TO THE PRAMS 48 VARIABLE HIGH-RISK PTB MODEL

The following presents the relative weighting of all 48 PRAMS variables. The modified Garson relative weights method is used. The approach taken was to first determine the relative weightings for each of the results presented in Table 5-6. Because each of the different network structures were chosen based on different performance criteria, they have different weights and biases and therefore will generate different ANN weighting terms. Slight variations in the final weightings obtained for each result using the modified relative weights method were expected and observed. In order to reduce the effect of this variability, the second step was to average the weightings of the two results geared towards improved sensitivity ('Highest sensitivity' and 'Compromise favoring

sensitivity') with the results from the 'Best Compromise Between Sensitivity and Specificity' category, in an effort to balance the weighting between sensitivity and specificity, while maintaining a clinically acceptable sensitivity. Table 5-7 presents the results of this averaging process, the variables are ranked from most to least significant.

TABLE 5-7 MODIFIED RELATIVE WEIGHTINGS USING AVERAGING

VARIABLE RANKING	MODIFIED RELATIVE WEIGHTS
PLURAL_N	100.0
MM_BLEED	88.2
PREV_PRE	64.9
BF_EVER	64.9
MM_HBP	64.8
PACIFIC	58.4
OTHER	56.6
STRS_FM3	55.1
MORB_CAR	54.5
MOMDRINK	53.3
STRS_COM	53.1
ALASKA	52.3
BLACK	52.0
MOMDRKS_N	52.0
MOMCIG_N	51.7
INC_JOB	48.5
MAT_ED_N	47.8
SMK_2YRS	47.4
STRS_MOV	46.3
MORB_CRV	43.6
MM_FEVER	43.6
WHITE	41.9
INDIAN	39.4
VITAMIN_N	39.1
PREG_TRY	39.1
MAT_AGE_N	38.2
MM_DIAB	38.0
MM_NOMD	35.0
HAWAII	34.3
MOM_LB_N	33.3
PNC_MTH_N	32.5
OTH_TERM_N	32.5
GENDER	30.5
PAD_HUS	28.0
WIC_PREG	27.0
MORB_NAU	27.0
MARRIED	26.2
PARITY_N	25.6
STRS_ARG	25.0
STRS_HOM	24.8
STRS_WRK	23.8
STRS_DV3	22.5
INSURE	21.8
MORB_KID	21.1
ERLY_PNC	20.5

MEDICAD3	20.3
STRS_JOB	19.7
MOMSMOKE	19.0

The ranking scheme identifies the multiplicity of birth (PLURAL_N) as the number one predictor of PTB; this is corroborated in the PPESO findings where ‘number of babies this pregnancy’ was found to be the most significant of the eight PPESO input variables. Clinically this result is also reasonable, as on average twins are born at 36 weeks, with higher order births born even earlier. Maternal bleeding before 23 weeks (MM_BLEED) and previous preterm birth (PREV_PRE) were identified respectively as the second and third best indicators of PTB. The PPESO eight-variable model also identified previous preterm birth as the second highest ranking variable (early maternal bleeding was not available in the PPESO model).

Interestingly, maternal smoking (MOMSMOKE) was not given a relatively high significance (weighting: 19.0); however the number of cigarettes smoked by the mother during pregnancy appears to be of greater significance (MOMCIG_N; weighting: 51.7). This is not the case with maternal drinking, where it appears that whether or not a mother drinks during pregnancy (the binary variable MOMDRINK; weighting: 53.3) is equally as important as the number of drinks consumed daily (MOMDRKS_N; weighting: 52.0).

Other variables that were found to be significant include: maternal high blood pressure (MM_HBP), black race (BLACK), family income (INC_JOB), maternal education (MAT_ED), whether the mother was trying to conceive (PREG_TRY), stress (STRS_COM and STRS_FM3), and morbidity of the cervix (MORB_CRV). It has also been suggested in the literature that socio-economic status and race can impact PTB and

this is corroborated by these results.

The quality of stress appears to be important. For instance, stress in the family is a stronger predictor than stress on the job. Whether the mother is registered with Medicaid, or has insurance does not appear to be a significant indicator of future PTB (which is perhaps counter-intuitive when considering that family income was identified as an important indicator). Early prenatal care and the month prenatal care began, do not appear to be particularly strong indicators of PTB.

The weightings shown in Table 5-7 emphasize the multi-factorial nature of PTB: no variable has a weighting of less than 10. This appears to indicate that all variables are contributing, to some extent, to the prediction of PTB (i.e. a mother's risk of giving birth prematurely stems from a number of different influences). The weightings can also be deemed clinically reasonable in that the variables selected as the strongest indicators of PTB (previous PTB and multiple gestation) are given significant weightings.

Using the weightings from Table 5-7, experiments were performed in an effort to achieve a new PRAMS-based minimal dataset for high-risk PTB prediction. Variables with low absolute connection weights were removed individually to assess the impact on the classification-based MLP ANN performance. The results indicate that all 48 variables are needed to maintain a sensitivity of over 65%. Removing the lowest ranking variable (MOMSMOKE) decreased the sensitivity to 56.5% (specificity increased to 85.2%). Results generated by individually removing each of the ten lowest ranking variables failed to realize sensitivity values matching that achieved using the PRAMS 48-variable high-risk PTB model. In developing the hybrid system, it is essential that the sensitivity

be maintained above 65%, leading to the conclusion that all 48 PRAMS variables are needed in the PRAMS minimal dataset.

Comparing this finding with previous MIRC work, it was discovered that when Ibrahim-Swailum attempted to determine a new minimal dataset for predicting delivery type (Ibrahim-Swailum, 2006), the lowest ranking variables had very low relative importance (e.g. PRETERM 0%; Intention to Breastfeed 0%) as compared to the rankings for high-risk PTB prediction. Removing just one of these very low ranking variables (PRETERM) reduced the sensitivity to 51.8% from 53.8%. Ibrahim-Swailum removed all variables individually, but was never able to match the sensitivity achieved with her full 24-variable model.

As previously discussed, Ennett used ANN weightings as the CBR connection weights. In developing her new model for neonatal mortality prediction, Ennett used the approach of keeping all variables that had an absolute weighting of greater or equal to 19 (Ennett, 2003). The lowest connection weight in the PRAMS 48-variable dataset is 19.0, which would qualify it for inclusion in the final weighting set based on Ennett's criteria. The weightings presented in Table 5-7 and the subsequent experiments with weight elimination, indicate that all 48 PRAMS variables are needed to maintain a clinically significant sensitivity.

The PPESO high-risk model was based on seven commonly available obstetrical variables, making it an ideal minimal dataset. However, the experimental results have demonstrated that the model trained on the PRAMS 48-variable high-risk PTB dataset produced much higher sensitivity values. Ideally, a reduced PRAMS minimal dataset

would have been created to successfully achieve improved results; however, it appears that all variables have some positive impact on the final performance. From a clinical viewpoint, all variables in the PRAMS 48-variable model are available prior to 23 weeks gestation and are obtainable during maternal obstetrical visits, as the variable set comprises general information that is discussed during these consultations, such as: previous and current obstetrical history, medical history, and demographic factors.

The PRAMS 48-variable model does contain variables that could potentially limit the model's applicability outside of the United States. In attempting to create a generalizable North-American model, that would be equally applicable to Canadian mothers, it would be beneficial if these American-centric variables could be either removed or, in the future, replaced by a Canadian equivalent. Table 5-8 lists the three variables that are specific to the United States and discusses whether there exists a Canadian equivalent.

TABLE 5-8 AMERICAN-SPECIFIC PRAMS VARIABLES AND POSSIBLE CANADIAN EQUIVALENTS

PRAMS AMERICAN- SPECIFIC VARIABLE	DEFINITION	SUGGESTIONS FOR CREATING A CANADIAN MODEL
Medicaid	An American healthcare program, funded by the states and the federal government, that reimburses hospitals and physicians for care provided to those identified as being unable to cover their medical expenses.	There is currently no equivalent Canadian program. It is probable that this variable is related to family income. The PRAMS dataset collects information on the source of a patient's income, and gives information as to whether the mother is receiving social aid, but does not provide a family's net income. If net income were available for Canadian families it is hypothesized that this

Women Infants Children (WIC) program	The Special Supplemental Nutrition Program for Women, Infants, and Children, serves to safeguard the health of low-income women, infants, and children who are at nutritional risk by providing nutritious foods to supplement diets, information on healthy eating, and referrals to healthcare (USDA 2006).	would have similar impact to the Medicaid variable ²⁷ . Similar initiatives exist in Canada: Canada Prenatal Nutrition Program funded by Health Canada (Canada-wide) Montreal Diet Dispensary (Quebec) Best Babies Possible (Ontario)
Alaska	Whether mother was born in Alaska.	Would be interesting to create a Canadian database that contained the mother's province or territory of birth, to investigate whether the 'Alaska' variable has similar impact as a possible 'Northern Canada' variable.

5.3.5 PRAMS VERIFICATION SET RESULTS

The verification experiments were run using the parameter settings of the networks that yielded the best results for each of the three risk models: highest sensitivity, best compromise between sensitivity and specificity, and highest specificity (original results shown in Table 5-6). The results are the average (mean) performance of the 10 verification sets (9701 total cases/10) and the standard deviation of those experiments. Although differences in results for the verification set are to be expected, in general the results indicate that the classification-based MLP ANN was effective in creating a generalized model using the PRAMS 48-variable model and the high-risk PTB criteria. While the highest sensitivity decreased slightly (65.0% as compared to 66.4% from the training model), it is still within clinically acceptable limits, and the specificity

²⁷ This data may be difficult to obtain, as income is a sensitive issue and not something that is discussed during obstetrical visits.

was virtually unchanged (80.2% as compared to 80.3%). The highest sensitivity achieved with both the ‘best compromise’ and ‘highest specificity’ model outperformed that achieved during training (62.5% as compared to 55.1%; 46.1% as compared to 43.3%). In both cases the specificities decreased slightly.

TABLE 5-9 PRAMS 48-VARIABLE HIGH-RISK PTB VERIFICATION SET RESULTS

EXPERIMENT	HIGHEST SENSITIVITY	BEST COMPROMISE BETWEEN SENSITIVITY & SPECIFICITY	HIGHEST SPECIFICITY
Sensitivity (%)	65.0 ± 0.04	62.5 ± 0.0273	46.1 ± 0.907
Specificity (%)	80.2 ± 0.0113	82.3 ± 0.0099	90.7 ± 0.0120

5.4 HYBRID SYSTEM RESULTS

5.4.1 PASS 1: RESULTS

Recall that at the input to Pass 1, the PRAMS HR_PTБ model had succeeded in achieving a sensitivity of 66% and specificity of 80%; however, this model provides performance results over the entire dataset only, rather than providing outcome estimations for each patient case. The hybrid system switches the focus to predicting outcomes for individual patient cases, as would be done in actual clinical practice.

5.4.1.1 Validation of Risk Stratification Estimation Model – Maximizing for Sensitivity

Before using the risk stratification estimation model as an outcome screening tool, the risk stratification model must be developed and compared with the performance of the classification-based MLP ANN. The optimal parameter values obtained from the classification-based MLP ANN are used by the risk stratification estimation model. In ‘Pass 1’, the goal is to optimize sensitivity while meeting the criteria outlined in Table

5-10. Specifically, while maintaining the sensitivity at a clinically acceptable 65%, the specificity should be greater or equal to 80% and the c-value should be greater or equal to 0.75.

TABLE 5-10 PRAMS HIGH-RISK PTB PERFORMANCE CRITERIA FOR 'PASS 1'

PERFORMANCE CRITERIA	GREATER OR EQUAL TO
Sensitivity	65%
Specificity	80%
C-value	0.75

The performance results from the risk stratification estimation model for maximum sensitivity are presented in Table 5-11 (based on the results of the test set) and compared with the best sensitivity results achieved using the classification-based MLP ANN model.

TABLE 5-11 RISK STRATIFICATION ESTIMATION ANN PERFORMANCE
RESULTS FOR MAXIMIZING SENSITIVITY

EXPERIMENT	BEST SENSITIVITY – CLASSIFICATION-BASED MLP ANN RESULTS	BEST SENSITIVITY – MLP ANN WITH RISK STRATIFICATION ESTIMATION RESULTS
Train PTB rate (%)	20.0	20.0
Test PTB rate (%)	7.68	7.68
Sensitivity (%)	66.4	65.2
Specificity (%)	80.3	80.4
CCR (%)	79.2	79.2
Area under ROC (test)	0.7795	0.7942
Constant predictor	80.0	80.0

Comparing the two results from a perspective of discrimination ability shows that the risk stratification estimation model performance results meet or exceed the performance criterion for 'Pass 1', with the area under the ROC curve for the test set increasing to just under 0.8 (indicating that the model is a very good discriminator).

While the sensitivity value fell minimally, it still exceeds 65%, and both the CCR and specificity increased marginally; essentially the differences in the two models are insignificant and it can be concluded that using the classification-based MLP ANN parameters in the MLP ANN risk stratification estimation model, quickly yielded a model with similar performance results.

5.4.1.2 Setting the cutoff point: Patient-specific Risk Stratification Outcomes

The weights and biases from the new risk stratification estimation model can now be used in the output screening program, which assigns a specific risk stratification estimation to each patient case. The outputs of the risk stratification estimation model range from 0 to 1. Generally, a lower cutoff point results in higher sensitivity of the positive outcome. The sensitivity and specificity values are used to determine thresholds for identifying high-risk PTB cases; in this way the risk stratification estimation model can be employed as a classification model by setting the cutoff point.

Figure 5-1 shows the risk stratification for both low risk cases (defined by a risk classification of 0 and depicted by light grey bars in Figure 5-1) and high-risk PTB cases (defined by a risk classification of 1 and depicted by dark grey bars in Figure 5-1). The risk stratification shows the number of cases classified as both low risk and high risk for ranges between 0.0 and 1.0. Note that the results generated from the MLP ANN with risk stratification estimations have been assigned to these ranges. All results are based on the PRAMS 48-variable high-risk PTB test set data.

FIGURE 5-1 RISK STRATIFICATION ESTIMATION MODEL FOR
LOW RISK BIRTHS (0) AND HIGH-RISK PTBs (1)

In maximizing for sensitivity, different cutoff values were tested between 0.1 and 0.5 (cutoff values above 0.5 would favor specificity because fewer high-risk cases would be correctly classified). Through experimentation, cutoff values in 'Pass 1' were evaluated based on their ability to achieve the highest possible sensitivity while still enabling 'Pass 3' to meet its final performance criteria. This was a time consuming iterative process, where different cutoff values for both 'Pass 1' and 'Pass 2' were tested concurrently, and then the final results in 'Pass 3' were evaluated. If the final performance results from 'Pass 3' were not met, the cutoff values from 'Pass 1' and 'Pass 2' were modified, and the process continued accordingly. When maximizing for sensitivity it was found that the best results were achieved based on a cutoff of 0.4. This

means that an ANN risk stratification estimation output of greater or equal to 0.4 would classify a baby as a high-risk PTB, whereas an output of less than 0.4 is classified as a low-risk case.

The percentage of test cases assigned to each risk stratification group, for both high risk and low risk cases, is shown in Table 5-12. For both low risk cases (HR-PTB = 0) and high risk cases (HR-PTB = 1), the shaded area in Table 5-12 depicts the cases that will be correctly classified based on a threshold of 0.4. Specifically, of the 497 cases that are high-risk (HR-PTB = 1), a cutoff of 0.4 will correctly classify 377 of these cases (sensitivity: 75.9%). Of the 5971 cases that are low-risk (HR-PTB = 0), a cutoff of 0.4 will correctly classify 4234 of these cases (specificity: 70.9%).

Table 5-12 also compares the risk stratification of the test and training set, by showing the percent of low and high-risk PTB cases for both the training and test sets. This validates that the stratification of the test set agrees with that of the training data.

The goal in training the hybrid system is to develop a model that can be applied to new patient cases with unknown outcomes. Having demonstrated that the classification-based MLP ANN model can be applied to a risk stratification estimation model while maintaining its classification ability, the next step is to identify all Predicted True Positives (PTPs) based on the prescribed cutoff value. The approach taken is to mark cases as 'PTP' if their probabilistic estimation is greater or equal to 0.4. The remaining cases are marked 'PN' for Predicted Negative. 'Pass 2' will identify which of these negative cases are actual 'PTNs'. In performing this step the only information that can be used is the risk stratification estimation output, although the outcomes are available at

this stage they must be ignored.

TABLE 5-12 PERCENTAGE OF TEST CASES ASSIGNED TO EACH RISK STRATIFICATION LEVEL FOR BOTH LOW RISK AND HIGH-RISK PTBS BASED ON A CUTOFF OF 0.4

HR-PTB (0 OR 1)*	RANGE (RANGE1 ≤ VALUE < RANGE2)	TOTAL CASES (TEST)	
% LOW RISK (TEST / TRAIN)			
0	0 – 0.1	683	11.4 / 11.6
	0.1 – 0.2	1672	28.0 / 28.1
	0.2 – 0.3	1196	20.0 / 20.0
	0.3 – 0.4	683	11.4 / 11.6
	0.4 – 0.5	566	9.5 / 9.3
	0.5 – 0.6	754	12.6 / 12.5
	0.6 – 0.7	360	6.0 / 6.1
	0.7 – 0.8	48	0.8 / 0.8
	0.8 – 1.0	9	0.2 / 0.1
TOTAL (0)		5971	
% HIGH-RISK PTB (TEST / TRAIN)			
1	0 – 0.1	15	3.0 / 1.9
	0.1 – 0.2	31	6.2 / 5.4
	0.2 – 0.3	36	7.2 / 7.6
	0.3 – 0.4	38	7.6 / 7.2
	0.4 – 0.5	53	10.7 / 11.7
	0.5 – 0.6	121	24.3 / 26.0
	0.6 – 0.7	160	32.2 / 29.2
	0.7 – 0.8	34	6.8 / 9.3
	0.8 – 1.0	9	1.8 / 1.6
TOTAL (1)		497	

*0 represents low-risk birth and 1 represents high-risk PTB

5.4.2 PASS 2: RESULTS

5.4.2.1 Validation of Risk Stratification Estimation Model – Maximizing for Specificity

As was the case in 'Pass 1', the optimal parameter values obtained from the classification-based MLP ANN are used by the risk stratification estimation model to determine the risk stratification levels. In 'Pass 2', the goal is to optimize specificity

while meeting the criteria outlined in Table 5-13. Specifically, while maintaining the specificity at a clinically acceptable 88%, the sensitivity must be greater or equal to 45% and the c-value must be greater or equal to 0.75. Note that these values constitute intermediary criteria and do not reflect the final performance criteria referred to in the hypothesis.

TABLE 5-13 PRAMS HIGH-RISK PTB PERFORMANCE CRITERIA FOR 'PASS 2'

PERFORMANCE CRITERIA	GREATER OR EQUAL TO
Sensitivity	45%
Specificity	88%
C-value	0.75

The performance results from the risk stratification estimation model are presented in Table 5-14 (based on the results of the test set) and compared with the specificity results achieved using the classification-based MLP ANN model.

TABLE 5-14 RISK STRATIFICATION ESTIMATION ANN PERFORMANCE RESULTS
FOR MAXIMIZING SPECIFICITY

EXPERIMENT	SPECIFICITY – CLASSIFICATION-BASED MLP ANN RESULTS	SPECIFICITY – MLP ANN WITH RISK STRATIFICATION ESTIMATION RESULTS
Train PTB rate (%)	19.99	19.99
Test PTB rate (%)	7.68	7.68
Sensitivity (%)	45.1	50.9
Specificity (%)	90.7	88.8
CCR (%)	87.2	85.9
ASE	0.355	0.0992
Area under ROC (test)	0.782	0.789

Comparing the two results from a perspective of classification and discrimination ability shows that the risk stratification estimation model performance exceeds the performance criterion for sensitivity from 'Pass 2', with the sensitivity rising to just over

50%. Although the specificity value fell slightly; it still exceeds the criteria of 88%.

5.4.2.2 Setting the cutoff point: Patient-specific Risk Stratification Outcomes

Figure 5-2 shows the risk stratification estimation for both low risk cases (defined by a risk categorization of 0 and depicted by light grey bars in Figure 5-2) and high-risk PTB cases (defined by a risk categorization of 1 and depicted by dark grey bars in Figure 5-2). The risk stratification shows the number of cases classified as both low risk and high risk for ranges between 0.0 and 1.0. All results are based on the PRAMS 48-variable high-risk PTB test set data.

FIGURE 5-2 RISK STRATIFICATION ESTIMATION MODEL FOR LOW RISK BIRTHS (0) AND HIGH-RISK PTBs (1) BASED ON A CUTOFF OF 0.6

A higher cutoff point results in higher specificity of the positive outcome. In maximizing for specificity different cutoff values were tested between 0.5 and 1.0 (cutoff

values below 0.5 would favor sensitivity because fewer low risk cases would be correctly classified).

Through experimentation, cutoff values in 'Pass 2' were evaluated based on their ability to achieve the highest possible specificity while still enabling 'Pass 3' to meet its final performance criteria. When maximizing for specificity it was found that the best results were achieved based on a cutoff of 0.6; this means that an ANN risk stratification estimation output of greater or equal to 0.6 would classify a baby as high-risk PTB, whereas an output of less than 0.6 is classified as a low-risk case.

The percentage of test cases assigned to each risk stratification group, for both high risk and low risk cases, is shown in Table 5-15. For both low risk cases (HR-PTB = 0) and high risk cases (HR-PTB = 1), the shaded area in Table 5-15 depicts the cases that will be correctly classified based on a cutoff of 0.6. Specifically, of the 5971 cases that are low risk (HR-PTB = 0), a cutoff of 0.6 will correctly classify 5301 of these cases (specificity: 94.8%). Of the 497 cases that are high-risk (HR-PTB = 1), a cutoff of 0.6 will correctly classify 183 of these cases (sensitivity: 36.8%).

TABLE 5-15 PERCENTAGE OF TEST CASES ASSIGNED TO EACH RISK STRATIFICATION LEVEL FOR BOTH LOW RISK AND HIGH-RISK PTBS BASED ON A CUTOFF OF 0.6

HRPTB (0 OR 1)*	(RANGE1 ≤ VALUE < RANGE2)	TOTAL CASES (TEST)	% LOW RISK (TEST)
0	0 – 0.1	1678	28.1
	0.1 – 0.2	2025	33.9
	0.2 – 0.3	808	13.5
	0.3 – 0.4	460	7.7
	0.4 – 0.5	330	5.5
	0.5 – 0.6	309	5.2
	0.6 – 0.7	361	6.0
	0.7 – 0.8	0	0
	0.8 – 1.0	0	0

TOTAL (0)		5971	
			% HIGH-RISK PTB (TEST)
1	0 – 0.1	36	7.2
	0.1 – 0.2	55	11.1
	0.2 – 0.3	53	10.7
	0.3 – 0.4	52	10.5
	0.4 – 0.5	48	9.7
	0.5 – 0.6	70	14.1
	0.6 – 0.7	183	36.8
	0.7 – 0.8	0	0
	0.8 – 1.0	0	0
TOTAL (1)		497	

* 0 represents low-risk birth and 1 represents high-risk PTB

After completion of 'Pass 2', cases are marked as Predicted True Negative ('PTN') if their risk stratification estimation is less than 0.6. The remaining cases are marked 'PP' for Predicted Positive. The next step is to compare the results from 'Pass 1' and 'Pass 2' and discover any discrepancies in the predicted outcome from each pass. If 'Pass 1' classified a case as 'PN' and 'Pass 2' classified the same case as 'PTN', then the outcomes are logically consistent and the final outcome is 'PTN'. The same situation occurs if 'Pass 1' classified a case as 'PTP' and 'Pass 2' classified the same case as 'PP'. However, if 'Pass 1' and 'Pass 2' give inconsistent outcomes, the case is marked as 'Conflict' and must be further assessed in 'Pass 3'. Figure 5-3 provides a screenshot of the marking process highlighting these representative situations. The first column indicates the input to 'Pass 1', which is the value obtained from the MLP ANN with risk stratification estimation. This value is assigned a range, and then classified based on 'Pass 1's' cutoff of 0.4, as a PTP or PN. The third column indicates the input to 'Pass 2', obtained from the MLP ANN with risk stratification estimation. This value is also

assigned a range, and then classified based on 'Pass 2's' cutoff of 0.6, as a PTN or PP. Finally, the 'Pass 2' classification is compared with the true value, to determine whether there is a conflict in the classifications between the first two passes. This information becomes the input to 'Pass 3'.

Value Pass 1	Range: Pass 1	Pass 1 Classification	Value Pass 2	Range: Pass 2	Pass 2 Classification	TRUE VALUE	Input to Pass 3
0.35	0.3 to 0.4	PN	0.19	0.1 to 0.2	PTN	PTN	PTN
0.55	0.5 to 0.6	PTP	0.30	0.3 to 0.4	PTN	PTN	Conflict
0.43	0.4 to 0.5	PTP	0.30	0.3 to 0.4	PTN	PTN	Conflict
0.69	0.6 to 0.7	PTP	0.66	0.6 to 0.7	PP	PTP	PTB

FIGURE 5-3 EXAMPLE 'PASS 1' AND 'PASS 2' CLASSIFICATION AND INPUT TO 'PASS 3'

5.4.3 PASS 3: RESULTS

Following the PRAMS analysis, the maximum sensitivity achieved on test data for known patient cases was 66% with a corresponding specificity of 80%, recall that this represented test results over the entire dataset and did not provide results for individual patient cases. In an attempt to use the parameter values and variable rankings from the trained PRAMS HR-PTB model as a basis for a system that classifies individual new patient cases as PTP or PTN, 'Pass 1' in the hybrid system is based on maximizing the sensitivity, achieving a sensitivity of 75.8% and a specificity of 70.9%. 'Pass 2' focused on maximizing specificity, resulting in the sensitivity dropping to 36.8% and the specificity increasing to 94.8%. The goal in 'Pass 3' is to improve upon the performance results achieved with the classification-based MLP ANN HR-PTB model; however these results will be applicable to new patient cases through an outcome screening tool.

In 'Pass 3', the goal is to find the final compromise between sensitivity and

specificity while meeting the criteria outlined in Table 5-16, this corresponds to the final performance criteria outlined in the hypothesis. Using the weightings from Table 5-7, both the MatchTP and MatchTN query databases have had their outputs (i.e. HR-PTB outcomes) filled in using the automatic matching functionality provided by the kNN. Through experimentation, the optimal value for k and the corresponding cutoff values were determined. Ideally, the final results should demonstrate stability over a range of k values, as a limitation of the kNN algorithm is that the results can be very sensitive to the choice of k .

TABLE 5-16 PRAMS HIGH-RISK PTB PERFORMANCE CRITERIA FOR PASS 3

CLINICAL PERFORMANCE CRITERIA	APPROACHES
Sensitivity	65%
Specificity	85%

5.4.3.1 Re-assigning conflicting cases: implementing the decision-tree voting algorithm

The decision tree voting algorithm is implemented in conjunction with the kNN algorithm to assign all conflicting cases following 'Pass 2' as either PTP or PTN. First, the kNN algorithm finds the k closest matching cases, and obtains a similarity ranking for each match. The DT logic then uses the matching values and their similarity rankings, as well as the determined cutoff value, to classify each outcome as a PTP or PTN. Part of the decision tree logic is based on the observation that, in general, the k matches found for actual TP results have higher similarity values than the matches found for actual TN results. Specifically, the majority of matches found for the actual TP cases have similarity values greater or equal to 90%. This information, in addition to the determined cutoff

values, provides a useful mechanism for distinguishing between TPs and TNs and allows the system to more accurately classify the cases as PTP or PTN.

The component of the kNN/DT algorithm that focuses on producing a predicted output for each conflicting case is described in the following pseudo code. The logic focuses on first determining, for each matching case, whether the similarity value is greater than 90%, then a vote is incremented if the output of the matching case is equal to 1 (indicating a positive outcome). Note that at this point the k closest matching cases have been found (using the kNN algorithm as described in 2.2.4) and the similarity value for each case has been calculated (stored in the array: similarity[k]). Next, the logic iterates through the number of columns in the database, and for each match, based on the similarity count and the vote, the missing value is determined.

For every matching case:

```

for (int i=0; i<k; i++) {
    Determine whether similarity value for each match is greater than 90%
    if (similarity[i] >= 0.9)
        Increment the similarity count (simCount)
    if more results exist {
        for j = 0 : number of columns in the database {
            If (similarity[i] >= 0.9)
                Calculate running total (aSum[j])

            If (similarity[i] < 0.9) {
                Calculate running total (bSum[j])
                If predicted outcome for match = 1
                Then increment the vote (vote++)
            }
        }
    }
}
} // end outer for loop

```

Next, iterate through all columns of the database to fill in the missing values
for (int j=0; j< nbColumns; j++) {

```

If (simCount == k) {
    query[j] = aSum[j] / k }

if (simCount == 0) {
    If (vote >= 2) {
        query[j] = bSum[j] / k }
    Else
        query[j] = 0
    }

if ((simCount < k) AND (simCount >= 1)) {
    If (vote >= 2){
        query[j] = (aSum[j]/simCount) + (bSum[j]/(k - simCount)) }
    Else
        query[j] = aSum[j] / simCount;
    }
} // end outer for loop

```

It was determined that a k value of 9 with a cutoff of 0.11, a similarity ranking of 90% and a vote of 2 yielded optimal performance. A cutoff of 0.11 favors sensitivity, as a greater number of high-risk cases will be correctly classified. Specifically, when considering the conflicting PTP cases from MatchTP, all cases with risk stratification values greater or equal to 0.11 will be classified as a PTP. Table 5-17 presents the calculations and results that led to this conclusion. The results presented in Table 5-17 are explained in detail by the pseudo code and calculations discussed in this section.

Based on the new outcomes determined for the conflicting cases using the kNN/DT algorithm, steps (1) and (2) involve determining the number of PTPs and PTNs using different cutoff values. This calculation is performed for both the conflicting TP cases from MatchTP (step 1), and for the conflicting TN cases from MatchTN (step 2). Pseudo code for this part of the decision tree logic is provided below:

```

For all conflicting TP cases from MatchTP {
    If conflicting TP case  $\geq cutoff$ 
        Add case to Sum_PTP_1
    Else
        Add case to Sum_PTN_1
} // end summation for all conflicting TP cases from MatchTP

For all conflicting TN cases from MatchTN {
    If conflicting TN case  $< cutoff$ 
        Add case to Sum_PTN_2
    Else
        Add case to Sum_PTP_2
} // end summation for all TN conflicting cases from MatchTN
}}
```

Step (3) involves re-assigning these conflicting cases as either PTN or PTP based on the chosen cutoff value. These values will be used to determine the final sensitivity and specificity. Pseudo code for this part of the decision tree logic is provided below. Note that the k value referred to in the code, is that used in the kNN's automatch functionality, incrementing the k value implies that automatch is being performed for all values of k up to MAX.

Define: Best_Sensitivity = 0, Best_Specificity = 0, Best_k, Best_cutoff, FN, FP, MAX

```

For (int k = 1; k < MAX; k++) {
  For (int cutoff = 0.05; cutoff < 0.95; cutoff++) {
    For all conflicting TP cases from MatchTP {
      PTP = Total_TP + Sum_PTP_1
      PTN = Total_TN + Sum_PTN_1
      PFN = PTN
      Sensitivity = PTP / (PTP + PFN)
    }
    For all conflicting TN cases from MatchTN {
      PTN = Total_TN + Sum_PTN_2
      PTP = Total_TP + Sum_PTP_2
      PFP = PTP
      Specificity = PTN / (PTN + PFP)
    }
    If Sensitivity -> 0.65 and Specificity -> 0.85{
      If Sensitivity > Best_Sensitivity
        Best_Sensitivity = Sensitivity
        Best_Specificity = Specificity
        Best_cutoff = cutoff
        Best_k = k
    }
  }
}

```

Considering a concrete example, of the 497 cases that are actual TPs, 194 cases are classified as conflicting before 'Pass 3'. Based on the cutoff value, these cases are re-assigned as either PTP (meaning that they will be correctly classified by the system) or PTN (meaning that they will be incorrectly classified by the system and are actually Predicted False Positives (PFPs)). With a cutoff value of 0.11, the new values after reassigning the conflicting cases become:

Re-assigning cases for TRUE VALUE = TP:

$$\begin{aligned} \text{PTP} &= \text{Total_TP} + \text{Sum_PTP_1} \\ &= 183 + 138 = 321 \\ \text{PTN} &= \text{Total_TN} + \text{Sum_PTN_1} \\ &= 120 + 56 = 176 \Rightarrow \text{Really FN} \end{aligned}$$

Because these cases are all actual TPs, it is now apparent that the values that Pass 3 marked as PTN were classified incorrectly and are actually PFN. These values will be used to determine the sensitivity:

$$\begin{aligned} \% \text{ Sensitivity} &= (\text{PTP} / (\text{PTP} + \text{PFN})) * 100 \\ &= (321 / (321 + 176)) * 100 = 64.58\% \sim 65\% \end{aligned}$$

Next, of the 5971 cases that are actual TNs, 1376 cases are classified as conflicting after 'Pass 3'. Based on the cutoff value, these cases are re-assigned as either PTN (meaning that they will be correctly classified by the system) or PTP (meaning that they will be incorrectly classified by the system and are actually Predicted False Positive (PFP)). Based on a cutoff value of 0.11, the new values after reassigning the conflicting cases become:

Re-assigning cases for TRUE VALUE = TN:

$$PTN = \text{Total_TN} + \text{Sum_PTN_2}$$

$$= 4234 + 802 = 5036$$

$$PTP = \text{Total_TP} + \text{Sum_PTP_2}$$

$$= 361 + 574 = 935 \Rightarrow \text{Really FP}$$

Because these cases are all actual TNs, it is now apparent that the values that 'Pass 3' marked as PTP were classified incorrectly and are actually PFPs. These values will be used to determine the specificity:

$$\% \text{ Specificity} = (\text{PTN} / (\text{PTN} + \text{PFP})) * 100$$

$$= (5036 / (5036 + 935)) * 100 = 84.3\% \sim 84\%$$

TABLE 5-17 FINAL SENSITIVITY AND SPECIFICITY CALCULATIONS (K=9)

HR-PTB		TOTAL CASES	Cutoff Values
(1) FOR CONFLICTING TP CASES FROM MATCHTP			
TOTAL		194	
SUM_PTP_1			138
SUM_PTN_1			56
(2) FOR CONFLICTING TN CASES FROM MATCHTN			
TOTAL		1376	
SUM_PTP_2			574
SUM_PTN_2			802
(3) RE-ASSIGNING CONFLICTING CASES			
TRUE VALUE	INPUT TO PASS 3	TOTAL	NEW COUNT AFTER RE-ASSIGNING CONFLICTING CASES
TN	CONFLICT	1376	
	PTN	4234	5036
	PTP	361	935
TN TOTAL		5971	
TP	CONFLICT	194	-
	PTN	120	176
	PTP	183	321
TP TOTAL		497	
SENSITIVITY			65%
SPECIFICITY			84%

5.4.3.2 Analysis of Results

The results in Table 5-17, based on a k value of 9 and a cutoff of 0.11, show that the integrated hybrid screening system has successfully met the clinical performance criteria outlined in this thesis, that is a sensitivity of 65% and specificity of 84%. The best results from the classification-based MLP ANN presented in Table 5-6 show that the previous highest sensitivity value of 66% had a corresponding specificity of 80% and attempting to achieve a specificity approaching 85% led to a dramatic decrease in sensitivity, dropping to 56%. The integrated hybrid screening system has successfully maintained the best sensitivity from the PRAMS 48-variable model while increasing the specificity to a clinically acceptable 84%. Figure 5-4 summarizes the performance results achieved over the course of this work, beginning with training the non-artificially distributed PPESO 8-variable model, and working through to the final hybrid results. It is important to understand that only the hybrid system is clinically viable, in that it provides individual results for unknown outcomes. The other results are for training purposes only, in that they provide generalized results and are based on known outcomes.

For training purposes only (i.e. generalized results and known outcomes)

For clinical use (i.e. individual results for unknown outcomes)

FIGURE 5-4 SUMMARY OF RESULTS

To illustrate the stability of the results k values of 8, 10 and 11 were also attempted. A k value of 8 yields a sensitivity of 63.3% and a specificity of 84.4%, a k value of 10 yields a sensitivity of 65% and a specificity of 83%, and a k value of 11 yields a sensitivity of 65.5% and a specificity of 82%. In general, as the value of k increases, the sensitivity increases and the specificity decreases, until a stagnant point is reached for k values of greater or equal to 30 (sensitivity: 74%; specificity: 73%). These results are illustrated in Figure 5-5.

FIGURE 5-5 RESULTS FOR K VERSUS PERFORMANCE (CUTOFF 0.11)

5.4.4 VERIFICATION ON NEW PATIENT CASES

The following presents the results based on the verification methodology outlined in section 4.8.6, using the PRAMS 48-variable high-risk PTB verification set with 9701 cases. Although the sensitivity and specificity are presented at completion of each pass, this information is for research purposes only. A clinical implementation of the results would not have known outcomes, making it impossible to measure these performance measures (the idea being that once the model has been sufficiently trained and tested, it is believed that the required performance parameters will be met).

After completion of 'Pass 1', based on a cutoff of 0.4:

'Pass 1' Sensitivity = 67.8%

'Pass 1' Specificity = 80.5%

After completion of 'Pass 2', based on a cutoff of 0.6:

'Pass 2' Sensitivity = 37.2%

'Pass 2' Specificity = 94.4%

After completion of 'Pass 2', the integrated hybrid system found 2403 conflicting cases, representing 24.8% of the total cases. Of these conflicting cases 87.8% are actual TNs, and 12.2% are actual TPs. Table 5-18 presents the distribution of these cases, showing the number of conflicting cases that are actual TNs and actual TPs.

TABLE 5-18 NUMBER OF CONFLICTING VERIFICATION CASES AFTER COMPLETION OF 'PASS 2'

ACTUAL VALUE	PREDICTED INPUT TO PASS #3	TOTAL
TN	CONFLICT	2110
	PTN	6341
	PTP (REALLY FP)	505
		8956
TN TOTAL		
TP	CONFLICT	293
	PTN (REALLY FN)	175
	PTP	277
		745
TP TOTAL		
GRAND TOTAL		9701

After completion of 'Pass 3', based on a cutoff value of 0.11 and a k value of 9, the conflicting cases are re-assigned, as shown in Table 5-19. The final sensitivity and specificity values are calculated from this table.

TABLE 5-19 RE-ASSIGNING CONFLICTING CASES FOLLOWING COMPLETION OF 'PASS 3'

ACTUAL VALUE	PREDICTED OUTPUT FROM PASS #3	TOTAL
TN	PTN	7469
	PTP (REALLY FP)	1487
		8956
TN TOTAL		
	PTN (REALLY FN)	268
	PTP	477
		745
TP TOTAL		
GRAND TOTAL		9701

Final Sensitivity = 64.0%

Final Specificity = 83.4%

The final sensitivity and specificity values, of 64% and 83.4% respectively, meet the clinical performance requirements determined for this thesis and verify that the integrated hybrid screening model is applicable to new patient data.

These results verify the thesis hypothesis. An integrated hybrid screening system has been developed, based on the high-risk PTB criteria and the PRAMS 48 variable dataset, which achieves a sensitivity of 65% and a specificity of 84% during training. Additionally, these results are reproducible on new patient cases, obtaining a sensitivity of 64% and a specificity of 83.4%, using the PRAMS verification dataset of 9701 independent complete cases.

5.4.5 VALIDITY OF RESULTS

The results obtained while developing this clinical decision support system have been validated at three different levels, as stipulated by Hudson and Cohen (Hudson & Cohen, 2000). Firstly, the integrity of the data was validated. This was accomplished by demonstrating that for both the PRAMS and PPESO the data was representative of the population under study and ensuring that missing values had been removed. The accuracy of both databases was verified, the PRAMS database had undergone extensive statistical analysis by the Center for Disease Control and the PPESO database had undergone audits to compare the database with the case-room logbook. For both databases the data was normalized and there were sufficient cases for the number of input nodes. There was also a reliable gold standard for determining the correct classification as the infant's true

gestational age at birth was known. Secondly, the technical integrity of the algorithm was verified by demonstrating that the algorithm behaved as expected on the test and verification data. As suggested by Hudson and Cohen, the primary measure of evaluation is the performance on the test set, to be supplemented by measuring the performance on the verification set. The contributing parameters were analyzed to show that they made sense from a clinical point of view and also contained parameters that have been historically considered important. Lastly, the choice of algorithm was appropriate for the application when considering practical considerations such as training time. In an ideal world it would be possible to compare all approaches, however this is not realistic in practice, and it could be argued is not needed in cases where the proposed system meets the clinical benchmarks.

CHAPTER 6. DESIGN OF A SEMANTIC WEB SERVICES-BASED FRAMEWORK FOR PERINATAL CLINICAL DECISION SUPPORT

6.1 INTRODUCTION

From the perspective of a clinical system user, the integrated hybrid screening system presented and validated in the previous two chapters, functions as a simple clinical decision support service. New patient data is entered and the mother is screened for high-risk PTB based on the CDSS's pre-determined optimal settings, as shown by the clinical system in Figure 4-1's architectural overview diagram. The interactions between data mining services needed to obtain this final outcome are transparent to the user. Chapter 6 presents the design of a Semantic Web services for healthcare framework in which the trained integrated hybrid screening system would be deployed. The framework presents the clinical decision support application, developed in Chapter 4 and implemented in Chapter 5, as a clinical decision support semantic Web Service (clinical decision support *services*). Within the framework, these services are: transparently integrated based on pre-defined composition scenarios, invoked over the Internet from a clinician's office (possibly accessing required patient data from other medical databases); used to generate real-time results at the point of care; and employed to automatically add new results to the electronic patient record.

This chapter presents the design of the Semantic Web services-based framework for perinatal risk assessment. The framework formalizes the practitioner's clinical

decision support workflow for perinatal outcome estimation. The workflows are defined by service compositions and supported by a perinatal CDSS ontology, enabling integration of CDSSs and perinatal data, and resulting in more timely and accurate predictions of high-risk preterm birth (and ultimately other perinatal outcomes).

6.2 DESIGN STEP 1: IDENTIFYING CANDIDATE SERVICES

The first step in the framework's design involves identifying candidate applications to be offered as Semantic Web services for perinatal risk assessment. The core functionalities for perinatal risk assessment are derived from the methodology described in *Chapter 4*. There are two potential user groups: healthcare professionals (and indirectly the patients they are advising) and MIRG's medical researchers. Before exposing an application as a Semantic Web service, it is important to consider how different system functionalities can be combined to provide added value to these two user groups. The assumption is made that while physicians' interest lies in the result from a composite service, researchers will be interested in accessing each component that leads to a final decision (i.e. individual core services).

6.2.1 COMPOSITE SERVICES FOR PERINATAL CLINICAL DECISION SUPPORT

The following composite Web services can be invoked and applied to the perinatal care area; these services include:

Risk Assessment: After the physician specifies the patient for which he/she desires risk assessment, and provides the required patient input variables, the system invokes the hybrid data mining system described in *Chapter 4*. The patient's risk assessment report is

provided to the physician, warnings are generated as applicable, and the EPR is updated. Note that this service currently focuses on risk assessment for high-risk PTB. In the future minimal datasets could be offered for other perinatal outcomes of interest.

Training Risk Models: The risk assessment models invoked by the physician are trained by the MIRG researchers as per the methodology described in section 4.8. Different risk models can be developed for different clinical databases, minimal datasets and outcomes of interest.

Classification-based Outcome Prediction: This service provides a binary output (+1 or -1) and is developed from the classification-based MLP ANN models trained in section 4.5 and 4.7 using the PPESO and PRAMS datasets. The chosen model will depend on the available input variables and required performance criteria (i.e. the PRAMS model provides the highest sensitivity and the best compromise between sensitivity and specificity, while the PPESO model offers the highest specificity at the expense of sensitivity and other performance measures).

Risk Stratification-based Outcome Prediction: This service provides a risk stratification-based output and is developed from the MLP ANN models with risk stratification estimation, trained in section 4.7 using the PRAMS dataset. The chosen model will depend on the available input variables and required performance criteria.

Matching Cases: The kNN-based CBR tool matches an individual patient's condition to the most similar past cases (user can specify the number of cases to match). The kNN-based CBR tool displays outcomes, complications, and medical information of the matched patients. There is also an option to store the results.

Replace Missing Values (kNN): The kNN-based CBR tool offers a service for replacing missing values in patient data using the running average of the k closest matching cases. This service could be used by the physician to replace missing values in past cases, and by MIRG researchers to replace values prior to data mining pre-processing.

Replace Missing Values (kNN-ANN): It is also possible for researchers to impute missing clinical values using a more complex hybrid approach, as described in *Chapter 2*, where weights from the MLP ANN are extracted and used as match weights for the CBR.

Develop Patient Risk Profile and Pregnancy Management Model: Future work will involve extending the framework functionality to provide perinatal decision-making, and the ability to link with the hybrid system. For instance, if a woman has an elevated risk of delivering prematurely, the CDSS could be used in conjunction with physician consultation to help the patient fully understand the procedure for further screening, possible benefits, and to develop a personalized treatment plan for minimizing the risk of preterm birth. This service allows the Physician to integrate the results from the *Risk Assessment* service with a patient-centered CDSS, and possibly to overlay the Physician's own comments and recommendations.

Patient Decision Support: The information on the patient's pregnancy risk profile and pregnancy management model, developed by integrating the *Risk Assessment* results with the Physician's recommendations, is presented to the Patient through this service. The goal is to offer user-centered content-adaptable decision-making support, similar to the PADS system described in section 2.2.6.

6.2.2 CORE SERVICES FOR PERINATAL CLINICAL DECISION SUPPORT

Each composite service requires combining multiple core services. Candidate core services for compositions are divided into two categories: the individual data mining services that make up the hybrid system; and the services that allow users to interface, and data to be integrated, with the system.

6.2.2.1 Core Data Mining Services

Classification-based MLP ANN processing: Uses trained classification-based MLP ANN to predict patient outcomes, provides a binary outcome (i.e. Preterm birth = false (-1)). Supports composite services: *Risk Assessment*, *Classification-based Outcome Prediction* and *Replace Missing Values (kNN-ANN)*.

Classification-based MLP ANN training: Uses MIRG's classification-based MLP ANN to develop trained prediction models. Supports composite services: *Training Risk Models*.

Develop Minimal Dataset: Using either weight elimination, sensitivity analysis, and/or the modified relative weights method, the minimal number of variables required to meet the acceptable performance criteria are determined. This includes ranking the variables such that the ANN connection weights can later be used as the match weights for the kNN. Supports core services: *Classification-based MLP ANN training* and *kNN AutoMatch*.

MLP ANN processing with risk stratification estimation: Uses trained MLP ANN with risk stratification estimations to predict patient outcomes, provides risk stratification estimation for each new patient case (i.e. this patient has a 70% chance of delivering

before 37 weeks gestation). Supports composite services: *Risk Assessment* and *Risk Stratification-based Outcome Prediction*.

MLP ANN with risk stratification estimation training: Uses MIRG's MLP ANN with risk stratification estimation to develop trained prediction models. Supports composite services: *Training Risk Models*.

Model Verification: Allows newly developed outcome prediction models to be trained on new patient data. Supports core services: *MLP ANN with risk stratification estimation training* and *Classification-based MLP ANN training*.

Network Selection: In both of the above cases the relevant networks are chosen based on the physician's input parameters and the desired outcome, allowing dynamic invocation of different models based on available patient information. Supports composite services: *Risk Assessment*, *Classification-based Outcome Prediction* and *Risk Stratification-based Outcome Prediction*.

kNN Processing: Core functionality that supports the CBR tool, using a k-Nearest Neighbours matching algorithm, the k closest matching cases are retrieved. Supports composite services: *Risk Assessment*, *Replace Missing Values (kNN)*, *Match Cases*, and *Replace Missing Values (kNN-ANN)*.

kNN AutoMatch: Core functionality that supports the CBR tool, using a k-Nearest Neighbours matching algorithm, the k closest matching cases are retrieved for all files missing values in a *query* database. Supports composite services: *Training Risk Models*.

Decision-Tree Voting Algorithm Processing: Uses a trained decision tree voting algorithm to classify conflicting patient cases as true positive or true negative. Supports

composite service: *Risk Assessment*.

Create Warnings: Evaluates clinical outputs from the risk assessment system to determine if a warning is required (for outcomes other than PTB an alert might be required, for PTB a warning is considered more suitable than an alert, since this is not an acute situation). Supports composite services: *Risk Assessment*, and in the future, *Develop Patient Risk Profile and Pregnancy Management Model*.

Data validation, a core service to ensure that the physician specified patient input data conforms to the acceptable boundaries for neural network processing (for each trained network).

Data Integration: For perinatal outcomes other than PTB, it may be necessary to retrieve patient data distributed across multiple databases (this service would not be needed if a universal electronic patient record were available). For example, in the case of perinatal infant care, it may be advantageous to retrieve earlier maternal obstetrical information. Access to each individual database is hidden in a lower-level Web service, and XML is employed to transfer the data to a higher-level core service that understands each database schema and integrates the data for CDSS processing (based on the perinatal clinical decision support ontology described in this work, or other possible ontologies), this integration is transparent to the user and is considered part of the underlying framework middleware. Supports composite service: *Risk Assessment*, *Classification-based Outcome Prediction* and *Risk Stratification-based Outcome Prediction*.

6.2.2.2 Interfacing Services

A *Web services interface* accessed directly by the user to specify the composite service to be invoked.

An *Application interface* for the user to interact with the requested clinical decision support services, specifying the requested outcome, possible threshold limits on this outcome, the required input variables and their boundaries. The user interface can be developed using the Protégé OWL forms functionality, meaning that it is created directly from the ontology.

A *Warning display interface* for formatting and displaying warnings to physicians.

A *Patient Risk Profile and Pregnancy Management Model Interface* for displaying a patient's risk profile, possible future screening, pregnancy management strategies and therapeutic interventions.

6.3 DESIGN STEP 2: DEFINING SERVICE COMPOSITIONS

Web service composition has recently taken center stage as an emerging research area (B. Medjahed et al., 2003). As more Web services become available, it is possible to offer consumers more complex services (composite services) by combining simpler ones (core services) using predefined composition scenarios (also termed service flows). While Web service delivery, description, and discovery are largely technical problems under the domain of IT experts, service composition represents a complex multi-disciplinary problem. It is essential to understand the human systems and interactions that combine to form a knowledge driven activity; healthcare experts are needed to identify

value-added medical services, and to determine appropriate service composition scenarios, including contextual knowledge that adds pragmatics to service compositions (this is often referred to as tacit knowledge (Abidi, 2005)).

The framework allows physicians and researchers to access composite Web services, while keeping the underlying integration of services transparent to the users. In this way the Web-based system is user-friendly and simple to use; from the user's perspective a detailed understanding of the Web service technologies is not required.

Compositions are based on a set of pre-determined rules that specify the composability of core Web services relevant to a particular patient's problem domain. This section lays the design basis for defining service composability in the domain of perinatal diagnostic clinical decision support. The approach is as described in the following section.

6.3.1 FUNCTIONAL SYSTEM VIEW

The functional system view describes all perinatal clinical decision support use cases in system use case diagrams depicting all actors, their interactions with the Semantic Web services, and dependencies between the services themselves. Functional system objectives are identified by three generic UML use cases: 1) **Diagnostic Clinical Decision Support Interactions**; 2) **Researcher Interactions**; and 3) **Semantic Web Services for Healthcare Interactions**. The full set of use cases collectively account for all current and future medical behavior that this framework is designed to support and form a business-level model for describing MIRC's clinical information system initiatives. In the following text, use case names are shown in bold.

6.3.1.1 Diagnostic Clinical Decision Support Interactions

All use cases required to support diagnostic clinical decision support interactions are depicted in Figure 6-1. The **Risk Assessment** use case, which offers the integrated hybrid screening system for preterm birth risk assessment, is deployed as a Web service and proceeds as follows. The Healthcare Professional begins by interacting with the user interfaces (both the Web Services interface and the application interface), specifying the relevant patient data and physician parameters, such as the outcome of interest. An XML document is created containing all available information. **Risk Assessment** parses the XML document to extract the necessary parameter values and invokes **Data Validation** to determine if the data corresponds to its standardized XML schema and ensure data integrity prior to processing. As described in *Chapter 4's* methodology for employing the trained screening system, **Risk Assessment** includes **Risk Stratification-based Outcome Prediction**, **Replace Missing Values (kNN)**, and **Decision Tree Processing**. Based on the results generated by the **Risk Assessment** use case, **Create Warnings** may be invoked (i.e. if a high-risk patient was identified). **Risk Assessment** then interacts with **Develop Patient Risk Profile & Pregnancy Management Model** to generate the patient-specific pregnancy management and risk mitigation plan. This may include retrieving closest matching cases via **Matching Cases** (the Healthcare Professional can also interact directly with this use case to view closest matching cases). The patient view of the information generated by **Develop Patient Risk Profile & Pregnancy Management Model** is presented to the Patient in **Patient Decision Support**; the Healthcare Professional may also review the 'next steps' with the Patient via this use

case. **Standardize Results** ensures that the outcome(s) generated by **Risk Assessment** are integrated in a standardized way into the Electronic Patient Record (EPR). This is based on available XML schemas describing the Perinatal Clinical Decision Support Ontology (presented in 6.4).

The Healthcare Professional can also interact directly with **Classification-based Outcome Prediction**, **Risk Stratification-based Outcome Prediction**, **Replace Missing Values (kNN)**, and **Matching Cases**. Based on the application, the required eXtended Stylesheet Language Transform (XSLT) converts the data into the appropriate format.

FIGURE 6-1 DIAGNOSTIC CLINICAL DECISION SUPPORT INTERACTIONS USE CASE DIAGRAM

6.3.1.2 Researcher Interactions

Before the Healthcare Professional can invoke the **Risk Assessment** composite service, the underlying core services must be trained. Training the prediction model can

encompass any or all of the steps detailed in *Chapter 4's* methodology. Completing all steps implies that the hybrid system described in *Chapter 4* has been fully developed. All use cases required to support researcher's interactions are shown in Figure 6-2. **Training Risk Models** is a use case that coordinates other use cases. It includes four use cases that implement Pass 1, Pass 2, and Pass 3, as described in section 4.7: **Classification-based MLP ANN Training**, **MLP ANN with risk stratification estimation training**, **kNN AutoMatch**, and **Decision Tree Voting Algorithm Processing**. Both **Classification-based MLP ANN Training** and **MLP ANN with risk stratification estimation training** interact with the abstract use case **Model Verification** to verify the trained models on new patient data.

Classification-based MLP ANN Training includes **Develop Minimal Dataset**, which finds the minimum number of clinical variables needed to obtain good results for each outcome. This includes developing the relative weightings for each variable, to be used when performing the **kNN AutoMatch** functionality.

It is also probable that the data will include missing data values, which can negatively impact ANN training. **Replace Missing Values (kNN-ANN)** uses the connection weights from a MLP ANN as the match weights in the kNN to replace missing data with meaningful values.

FIGURE 6-2 RESEARCHER INTERACTIONS USE CASE DIAGRAM

6.3.1.3 Semantic Web Services for Healthcare Interactions

This use case package contains system functionality required to support Semantic Web services-based deployment of clinical decision support services. First, MIRG Researchers **Describe core Web services**, then in collaboration with the Healthcare Professional, **Determine Web service compositions** prescribes how these services will interact to provide user-value. Web service descriptions are provided by the MIRG Researcher via **Describe Web service Compositions**.

As the last step in this process, the MIRG Administrator invokes **Publish Web**

services to make the services available to the Healthcare Professionals. **Define CDS Transforms, Define Schemas, and Publish Schemas**, are three XML-based use cases required to support Web service processing, respectively providing required XSL transforms, and defining and publishing XML schemas (which are ontology-based).

All medical data must be standardized for possible integration, a Perinatal Clinical Decision Support ontology is defined by the MIRG Researcher and refined in conjunction with the Healthcare Professional (**Develop Ontology**). When possible, existing ontologies are reused (**Ontology re-use**). Lastly, the MIRG Administrator publishes the ontology (**Publish Ontology**).

FIGURE 6-3 SEMANTIC WEB SERVICES FOR HEALTHCARE INTERACTIONS USE CASE DIAGRAM

6.3.2 SERVICE INTEGRATION ARCHITECTURE

6.3.2.1 Overview

Figure 6-4 presents an overview of the framework architecture. Arrows in the diagram depict dependencies within and between layers, with requests flowing in the direction of the arrows, in general, downwards from higher to lower layers.

Initially, the Web Service Coordinator package receives service requests which

are managed by the *Request Coordinator* module. The first step after receiving the service request is to determine appropriate composition scenarios for the request based on pre-defined composition templates created by physicians and researchers; this step is performed by the *Service Composition* module. Different composition templates are required based on whether this is a researcher or physician-initiated request.

Having identified the composite Web service(s) required to complete the composition scenario, the *Service Locator* module locates the services using the *Service Discovery Registry* subsystem. Once the results are returned to the *Request Coordinator* the appropriate composite Web service is invoked, resulting in the invocation of one or more core Web services. It is worthwhile to note that certain composition scenarios require the invocation of more than one composite Web service and that both composite and core Web services can access medical domain databases (via the core Web service *Data Integration*).

The Offered Web Services subsystem divides Web services into four categories: 1) composite Web services that provide added value to the Researcher; 2) composite Web services that provides added value to a Healthcare Professional; 3) composite Web services that provides added value to the Patient; and 4) core Web services that can be invoked by any of the composite services (or combinations thereof), or by another core Web service.

FIGURE 6-4 SYSTEM ARCHITECTURE

6.3.2.2 Logical System Model

Figure 6-5 presents the logical model of the system, depicting dependency relationships between the packages and classes of the higher layers with the lower level middleware and system software. Initially, the Web Service Coordinator package

receives service requests, which are managed by the Request Coordinator class. The first step after receiving the service request is to determine appropriate composition scenarios for the request based on pre-defined composition templates created by physicians and researchers; this step is performed by Service Composition which accesses Composition Templates, containing pre-defined compositions of core and composite Web services, described in WSDL. Having identified the composite Web service(s) required to complete the composition scenario, the Service Locator class locates the specified Web services using Service Discovery's Service Discovery Registry class, based on UDDI. Once the results are returned to the Request Coordinator the appropriate Web services are invoked via the Web Server. The Web Server interacts with MIRG's Application Coordinator to interface with the CDSSs via the CDSSInterfaces class; the fact that the CDSSs are accessed as Web services is transparent to the system users. The CDSS Control class coordinates classes required to perform CDSS processing. This class is responsible for initiating all XML-related processing and parsing required for clinical decision support, storing medical data for auditing purposes, and transforming medical databases from a relational model to XML. CDSS Control stores all session parameters for a particular request based on input received from parsing the received XML document and uses these parameters to determine the appropriate schemas and transforms required for CDSS processing. For example, based on the desired clinical decision support output and the requested CDSS identifier, a specific XML schema will be retrieved by the 'XML Resource Manager'; similar logic applies when retrieving pre and post-processing XML transforms. The 'XML Resource Manager' controls all access to XML resources

such as schemas and transforms.

FIGURE 6-5 OVERVIEW PACKAGE DIAGRAM

6.4 DESIGN STEP 3: CREATING THE PERINATAL CLINICAL DECISION SUPPORT ONTOLOGY

The third phase of the design involves creating and defining an ontology for supporting perinatal diagnostic clinical decision support. The premise is that the knowledge acquired about the application domain is the most important cornerstone of a successful software product (Knublauch, 2003). Ontologies allow developers to acquire knowledge from domain experts and capture it in a domain model. The critical analysis from *Chapter 3* showed that no similar ontology currently exists, however such an ontology is required to add semantics to the core and composite Web services described in the previous two sections, that is: mapping the concepts in the Web resources (i.e. the services and their related descriptions) to the ontological concepts.

By instantiating the Perinatal Clinical Decision Support Ontology (i.e. creating instances of the OWL: class for each relevant class in the ontology), it is possible to attach the semantics of clinical decision making to Web services. The resulting Semantic Web services integrate: 1) syntactic knowledge describing required inputs, their preconditions and post conditions; 2) semantic knowledge describing rules specifying the order in which services are composed; and 3) pragmatic knowledge defining the explicit situation in which services should be used.

6.4.1 KNOWLEDGE (ONTOLOGY) MODELING WITH PROTÉGÉ

Protégé-OWL is a widely used open source ontology modeling environment with support for the Web Ontology Language (OWL) (Knublauch et al., 2004). Protégé's

resulting models can be used as the basis for decision support systems, as well as for capturing software requirements, generating Java classes and UML documents, and accessing the Semantic Web. Protégé's conceptual models and knowledge bases can be accessed using open-source Java APIs. While UML 2.0 was used in this chapter to model system behavior and architecture, in general, UML is largely a language for object-oriented programming, that few medical domain experts understand, which is why it is not used for ontology development in this work. In comparison, Protégé is optimized for building domain models.

Protégé projects are called ontologies; they represent a collection of domain concepts and their relationships. Protégé is used for: modeling classes, their attributes and relationships (called slots); generating interactive forms that allow users to enter valid instances (i.e. instantiations of the classes – objects in UML); model processing, by relying on a library of plug-ins to define semantics, perform queries, and define logical behavior. The ontology can be saved in XML, UML, or RDF (Resource Description Framework) format. The OWL semantics specifies how to derive a class' logical instances from its classes. This might be based on facts not present in the ontology, but entailed by the semantics (see the OWL Web Ontology Language Guide) (Stanford University, 2006).

Once an ontology has been developed, Protégé enables integration of the ontology with two generic applications (Stanford University, 2006), including: 1) domain-specific knowledge-acquisition tools, which allow users to build a new knowledge base by acquiring instances; and 2) applications used in solving end-user problems such as

expert-systems and decision-support systems. In particular, Protégé can automatically generate customizable user interfaces from class definitions, enabling rapid knowledge acquisition (Knublauch et al., 2004). There are three different versions of OWL. This work is based on the most complete OWL full, which is a union of OWL syntax and RDF (Resource Description Framework – a graphical formalism for representing metadata).

6.4.2 METHODOLOGY

The ontology being developed in this work can be classified as an application ontology in that it combines a domain ontology (describing the vocabulary related to CDS and perinatal care) as well as tasks related to these domains. The following methodology has been compiled and adapted from the literature for developing Protégé ontologies:

1. Determine the domain and scope of the ontology:
 - a. Who will use the ontology? The ontology should be designed so that it can be used to generate and customize an appropriate knowledge acquisition tool for a specific set of users.
 - b. Who will maintain the ontology?
 - c. What applications will the ontology support? More specifically, what sets of problems do the domain experts need to solve? What problem-solving methods will be used?
 - d. What activities must the knowledge base support?
2. Build an initial ontology of classes and slots.
 - a. Define the classes and class hierarchy

- Class hierarchy should represent an 'is-a' relationship
 - Subclasses should have additional properties that the superclass does not have, or have different restrictions, or be in different relations than the superclass.
 - Deciding between creating a new class or a property value: look at the scope that has been defined for the ontology and how important the concept is in the domain.
- b. Define class properties (slots)
- Slot cardinality
 - Slot value type
3. Modify the Protégé-2000 knowledge acquisition forms to constrain the input data to the desired values. Use the forms to acquire slot values for test instances.
 4. Discuss the ontology and the instance forms with the expected users.
 5. Perform revisions as necessary to both ontology and the forms.
 6. Build a larger knowledge base that can be tested with the domain application.
 7. Test the full application with end-users. This step can lead to further revisions to the ontology and the knowledge acquisition forms.

6.4.3 DIAGNOSTIC DECISION SUPPORT ONTOLOGY: TOP-LEVEL VIEW

The top-level class `DiagnosticDecisionSupport` has two disjoint subclasses: `DecisionEnvelope` and `DecisionBody`. The `DecisionBody` class contains all classes that

implement the ontology. Figure 6-5 provides an overview of the ontology.

FIGURE 6-5 DIAGNOSTICDECISIONSUPPORT SUBCLASS TREE - TOP LEVEL VIEW

The Diagnostic Decision Support ontology is geared towards supporting decision making in the perinatal domain; however, it can be equally applied to other clinical application domains. This is possible because the decision making component and the clinical domain are modeled as two disjoint subclasses of the DecisionBody super class: ClinicalDiagnosticDomain and DecisionSupport. The ClinicalDiagnosticDomain is generalizable to other clinical domain for which diagnostic decision-making is required, while the DecisionSupport class supports diagnostic decision making regardless of the domain on which the decisions will be executed.

The DecisionEnvelope class contains all identification, status and versioning

information required to distinguish and access the ontology, a screen shot of the Protégé OWL editor showing the class hierarchy is shown in Figure 6-6.

<u>Decision Support Ontology</u>	
☐	owl: Thing
☒	Diagnostic Decision Support
☒	Decision Body
☐	Decision Envelope
☐	MIRG Identifier
☐	Developer
☒	Expertise
☒	Name
☒	University
☐	Documentation
☒	Author
☐	Availability
☒	Contact
☐	Electronic
☒	URL
☐	Print
☒	Location
☒	Grant
☐	MultiDisciplinary Team
☒	Biomedical
☒	Clinical
☒	Technical
☒	Status
☐	Versioning
☒	Release Date
☒	Version Number

FIGURE 6-6 PROTÉGÉ-OWL SUBCLASS VIEW OF DECISIONENVELOPE CLASS

6.4.4 DECISION SUPPORT COMPONENT

There is a well-established field of knowledge in the area of clinical decision

support systems that must be contained within the Decision Support component of the ontology. In keeping with the methodology, when developing the ontological classes, emphasis was placed on the applications that the ontology must support. Given that the main purpose of this ontology is to support the pragmatics of clinical decision-making, in essence, the ontology should allow the clinical decision support systems to be self-explanatory. Figure 6-7 depicts the eight disjoint subclasses of the DecisionSupport super class.

FIGURE 6-7 MAJOR COMPONENTS OF DECISIONSUPPORT SUBCLASS TREE

These classes have four main functions:

1. To support the **choice** behind clinical decision support systems - choice is described by the subclasses: Applicability (expanded subtree shown in

2. Figure 6-8), Purpose (expanded subtree shown in Figure 6-9) and Clinical Goal. Choice answers questions such as: Why and when would physician use this service? For which sets of patients? In what circumstances?

FIGURE 6-8 APPLICABILITY EXPANDED SUBCLASS TREE

FIGURE 6-9 PURPOSE EXPANDED SUBCLASS TREE

3. To support **domain knowledge**. Relevant classes include: Classification (expanded subtree shown in Figure 6-10), Collaboration->Classification (expanded subtree shown in Figure 6-11) and InputParameters.

FIGURE 6-10 CLASSIFICATION EXPANDED SUBCLASS TREE

FIGURE 6-11 COLLABORATION->CLASSIFICATION EXPANDED SUBCLASS TREE

4. To support clinical decision support tasks. Task-based classes include: Training and Testing.
5. To support clinical decision support interactions: relationship-based classes include the Collaboration->Workflow class.

6.4.5 CLINICAL DIAGNOSTIC DOMAIN COMPONENT

The Decision Support component of the ontology is designed to be generalizable to any clinical diagnostic domain. This ontology focuses on perinatal care; however, the Clinical Diagnostic Domain component of the ontology must be able to specify any clinical domain for which the practitioner desires decision support. Figure 6-12 depicts the five disjoint subclasses of the ClinicalDiagnosticDomain super class.

FIGURE 6-12 MAJOR COMPONENTS OF THE CLINICAL DIAGNOSTIC DOMAIN SUBCLASS TREE

The Database class contains all information required to describe the database(s) used in developing the model, specifically the: name, year, number of cases, percentage

of missing values, and population details (e.g. heterogeneous or homogeneous).

The `InputVariables` class contains the minimum data set required and constraints on each of the input variables in this dataset (as required).

The `PatientCriteria` class specifies the patient's sex and age, including, if applicable, the minimum and maximum ages for which the model will produce results. Both an inclusion and exclusion criteria are provided for successful model performance.

Of primary interest to the Practitioner is the `DecisionProblem`. Figure 6-13 shows how the `DecisionProblem` subclass models perinatal decision-making explored by MIRG. These problems include all those for which MIRG is currently or planning to provide clinical decision support, covering the prediction of PTB, delivery type, APGAR score, infant mortality, length of stay in a NICU, duration of ventilation, and possible complications. Other decisions can be easily added to the ontology.

FIGURE 6-13 DECISION PROBLEM EXPANDED SUBCLASS TREE

DecisionOutput, shown in Figure 6-14, is a superclass with four subclasses used by the Practitioner to evaluate and critique the clinical decision support's output: 1) TestParameters provides the model performance related information, allowing the practitioners to assess for themselves the clinical usefulness of the model; 2) CertaintyCriteria allows the Practitioner to compare the TestParameter values with their own defined minimum certainty criteria for each parameter (i.e. the model's sensitivity must be greater than 50%); 3) Recommendation details the proposed course of action (if any) offered by the CDS, as well as the related harm and benefits; and 4) ClinicalView allows the Practitioner to compare the CDS output with their own clinical judgment

(FinalDecision) and to insert any comments (PractitionerComment) .

FIGURE 6-14 DECISION OUTPUT EXPANDED SUBCLASS TREE

6.4.6 BUSINESS PROCESS WORKFLOWS

This section applies the ontological concepts described above to the business process workflows for selected composite Web service defined in section 6.2.1, specifically *Risk Assessment* (shown in Figure 6-15) and *Classification-based Outcome Prediction* (shown in Figure 6-16). This illustrates the approach for defining a concrete composition model by combining the chosen Web services with the ontological concepts.

UML models are considered sufficiently expressive for defining Web service compositions and they improve the readability of the compositions, as compared to low-level XML descriptions (Skogan, Gronmo, & Solheim, 2004). However, there is currently a lack of established standards for modeling Web services involving semantic knowledge. The approach taken in this work was to model the Web service compositions using UML activity diagram notation, and use interfaces to the stereotyped Web service states to model the semantic knowledge inputs from the external user to the system, and between Web services themselves. This approach is based on the suggestion of Groenmo and Jaeger in their work modeling Semantic Web service compositions (Groenmo & Jaeger, 2005). The interfaces in Figure 6-16 and Figure 6-16 are shown as rectangles, where all information contained in the interfaces comes from the imported OWL ontology Perinatal Clinical Decision support package.

Both concrete composition models have many similarities, illustrating the generalizable nature of the design. The main difference is the healthcare professional's choice of diagnostic service(s); the other supporting core Web services vary only slightly from application to application. Note that the 'Create Warnings' Web service is only invoked in situations where warnings are required based on the output from the integrated hybrid screening system.

FIGURE 6-15 SEMANTIC WEB SERVICES COMPOSITIONS FOR ‘TRAINING RISK ASSESSMENT’

FIGURE 6-16 SEMANTIC WEB SERVICES COMPOSITIONS FOR 'CLASSIFICATION-BASED OUTCOME PREDICTION'

6.5 ONGOING CHALLENGES IN DEVELOPING SEMANTIC WEB SERVICES-BASED APPLICATIONS

6.5.1 SECURITY

While Web services offer benefits and opportunities, they also present many challenges in healthcare environments, such as the need to comply with health insurance regulations, and increased requirements for reliability, security, and monitoring (Chatterjee, 2003). Currently, one of the biggest obstacles in deploying Semantic Web services is security (Daconta et al., 2003). This is of particular concern in healthcare where maintaining privacy of sensitive patient data is paramount. Authentication, authorization, confidentiality, data integrity, and non-repudiation are all security concerns that relate to Web services and must be addressed before services can be exposed externally. The Web Services industry is working to overcome security concerns with initiatives such as XML Signature for validating message integrity and non-repudiation, XML Encryption for data confidentiality, the Security Assertion Markup Language (SAML), an Organization for the Advancement of Structured Information Standards (OASIS) standard, that provides assertions of trust between parties, and WS-Security, which combines XML Signature and XML Encryption with standard SOAP-messaging. Given that Web service security is still evolving, the majority of applications described in the literature are based on deployment of Web services on intranets, which nevertheless provides great potential for integration of data and services across large healthcare organizations.

6.5.2 PRIVACY

While Web services offer benefits and opportunities, they also present many challenges in healthcare environments. One hurdle is the need to comply with health insurance regulations, such as the Health Insurance Portability and Accountability Act of 1996 (HIPAA) in the United States, applicable to any Canadian organization doing business in the US. In Canada the Personal Information Protection and Electronic Documents Act (PIPEDA) has similar personal information protection requirements.

CHAPTER 7. CONCLUDING REMARKS

This chapter summarizes the conclusions generated from the results presented in the previous two chapters, specifies the academic contributions provided by this work, and discusses ideas for future research, both in the domain of hybrid data mining systems for perinatal care and in the field of Semantic Web services for healthcare.

7.1 CONCLUSIONS

The following discusses the important findings of this thesis work:

1. Succeeded in developing a low cost, flexible, and simple to use integrated hybrid screening system for the prediction of high-risk preterm birth that is applicable to all pregnancy cases, rather than a subset of elevated risk mothers only. The success of this model suggests positive benefits could be obtained from the inclusion of an additional early screening system to identify mothers who would benefit from further invasive testing later in their pregnancy.
2. The best performance achieved by the integrated hybrid screening system shows the model to be a viable clinical predictor for the early prediction of high-risk PTB when measured against clinical benchmarks determined in collaboration with physicians (i.e. sensitivity approaching 65% and specificity approaching 85%). The integrated hybrid screening system was successfully applied to 9701 new patient cases, demonstrating that the model is generalizable to new cases and was not over-trained to

the test and training data. Moreover, once the model was developed new patient cases were verified quickly and easily. Training is the laborious component, using the trained model to predict outcomes for new patient cases was very effective and efficient.

3. The new high-risk PTB criteria developed in this work was applicable to two different perinatal databases, and succeeded in obtaining better results than the traditional PTB prediction models. This illustrates that the low-risk PTBs were in fact detrimental to the network's ability to classify the more critical PTB (or high-risk PTB) cases.
4. The hybrid system shows the complementary predictive abilities of the different data mining techniques used in this work, where the sensitivity and specificity were iteratively refined with each successive step through the hybrid system, where each step had a predefined performance criteria that must be met. The technique of using multiple data mining techniques to distinguish between true and false positives (for outcomes classified as positive by the classification-based MLP ANN) and between true and false negatives (for outcomes classified as negative by the classification-based MLP ANN), for known outcomes, was successful in developing a trained model that could then be used to classify conflicting outcomes as: FP/TP or FN/TN.
5. The parameters obtained from the classification-based MLP ANN developed in this work were applicable to the MLP ANN with risk stratification estimations, yielding similar results and accurate risk stratification models (when compared with the training set).

6. Specification of new cutoff points for the MLP ANN with risk stratification estimations, resulted in a new screening tool for high-risk PTB prediction. The MLP ANN with risk stratification estimations was applied to a perinatal population, further validating the approach used previously by Zhou and Frize (Zhou & Frize, 2006) in the neonatal intensive care unit.
7. Demonstrated that the modified relative weights method modified by Rybchynski could be used to employ the weights of MLP ANNs with one hidden layer, as the weighting terms of a kNN-based CBR tool.
8. Showed that all 48 PRAMS variables identified by clinicians were required to maintain sensitivity values greater than 65%. Although no reduced dataset was developed during this work, the 48 variables can be considered a minimal dataset for high-risk preterm birth prediction, as it was shown that all variables are useful in predicting the multi-factorial problem of PTB. The new PRAMS minimal dataset includes rankings, determined using the modified relative weights method, assigning weighting to all variables. The highest ranking variables are also those deemed by physicians as being most significant, that is: previous preterm birth and multiple births. This finding will hopefully serve to validate the model from a clinical viewpoint. Provided suggestions for transferring three variables to Canadian terms.
9. Presented the design of a Semantic Web services for healthcare-based framework that generalizes the training and deployment of the hybrid system to any perinatal outcome of interest (i.e. the integration of the hybrid system is not preterm birth-specific). The

use cases, system architecture, and business process workflows are applicable to future perinatal outcomes being investigated by MIRG.

10. Created a generalizable Perinatal Clinical Decision Support Ontology, containing a Decision Support component and a Clinical Diagnostic Domain component. In this way, the clinical decision support terminology is kept separate from the clinical application. The ontology supports the pragmatics of clinical decision support (that is, the: who, why, when, where, how, why?) and provides a mechanism for attaching practitioners' tacit domain knowledge into future implementations of clinical decision support systems.

7.2 ACADEMIC CONTRIBUTIONS

The following lists the academic contributions of this work:

1. Successfully developed a re-engineered and improved risk assessment model for the prediction of high-risk preterm birth in a heterogeneous population, to be used as an initial screening tool prior to 23 weeks gestation. The model sensitivity reaches 65% with a specificity of 84%, outperforming the current gold standard achieved using invasive techniques and providing earlier outcome predictions.
2. Demonstrates that high-PTB can be estimated, in non-symptomatic populations, equally well (and earlier) using readily available obstetrical, medical history and socio-demographic variables, as compared to using techniques based on costly and invasive clinical testing, as has been widely publicized in clinical fetal-fibronectin trials. This is contrary to the findings from the clinical literature, which suggest that

invasive diagnostics are the only means of achieving sensitivity values greater than 50%.

3. Creation of a hybrid system for integrating a classification-based MLP ANN, a MLP ANN with risk stratification estimations, a kNN-based case-based reasoner, and a decision tree voting algorithm, to develop a screening tool for preterm birth risk assessment to be used for predicting outcomes on new patient cases. The hybrid system methodology offers a new approach to screening, by integrating data mining modalities for the purpose of distinguishing between true and false positives and true and false negatives. The screening system is trained on known data and tested and verified on new data with equally successful performance.
4. Specification of a new high-risk preterm birth criterion used in successfully increasing the sensitivity of preterm birth risk assessment in a heterogeneous population. The model is adapted to two different perinatal databases.
5. Design of a Semantic Web services healthcare framework for enabling diagnostic clinical decision support. The framework is based on a novel and generalizable Diagnostic Clinical Decision Support Ontology for Perinatal Care, which supports the pragmatics of clinical decision-making using data mining modalities. Addresses issues surrounding the complex multi-disciplinary problem of service composition, such as adding the pragmatics of when and how to combine services. Creates composition templates to describe the service compositions of the hybrid system. Models workflows for perinatal clinical decision support, enabling autonomous invocation of

clinical decision support services in specific physician-initiated situations. These workflows can be extended easily to estimate other perinatal outcomes of interest.

7.3 FUTURE WORK & RECOMMENDATIONS

1. Use the integrated hybrid screening system developed in this thesis for high-risk PTB prediction to predict other perinatal outcomes of interest, such as those described in the Clinical Diagnostic Domain's DecisionProblem class (e.g. APGAR, Delivery type, and complications). Update the Perinatal Clinical Decision Support ontology as required.
2. In conjunction with physicians, implement a clinical trial of the integrated hybrid screening system and compare the model's findings with clinical judgment and actual maternal outcomes.
3. Based on the finding that all 48 PRAMS variables were required to maintain adequate sensitivity, it appears that the addition of demographic, obstetrical and medical variables can improve screening effectiveness. Based on Table 4-1, obtain new data that includes a greater number of possible risk factors (not just those available from the PRAMS and PPESO models), and determine if the addition of more variables can further improve screening results without overtraining.
4. Develop a new minimal dataset that would not include any nationalistic variables (e.g. relative to the United States or Canada only) in an attempt to develop a generalizable North American model. While this model might not be able to maintain the same

sensitivity values, it is likely the specificity would increase and the model would be more easily tested on new data.

5. Determine if cutoff ranges obtained from the MLP ANN with risk stratification estimations can be used to predict the severity of PTB when integrated into the hybrid screening system; that is, can higher cutoff values be used to predict Extremely PTB (EPTB), with slightly less stringent cutoffs used to predict Very PTB (VPTB), and the least stringent used to select PTB cases only. This would mean that separate models could be developed for EPTB and VPTB.
6. The knowledge obtained from the hybrid screening system must be translated into information useful for the clinicians and ultimately the patient. In the neonatal care research previously described, a patient decision support tool was created which embedded an ANN prototype in user-friendly graphical interfaces (M. Frize et al., 2005); physician feedback indicates that a similar approach would be of interest in the prediction of preterm birth. This would involve implementing the Develop Patient Risk Profile & Pregnancy Management Model and Patient Decision Support use cases from the Diagnostic Clinical Decision Support Interactions use case diagram.
7. Develop an approach for updating the hybrid screening system when major changes occur in the original training data. For example, changing trends in breastfeeding or increased occurrences of multiple births. This could include developing statistical models for ongoing learning.
8. Recommendation: In the future, when training new models, it may advantageous to include duration of the neonatal intensive care unit stay in the high-risk PTB criteria,

as this would provide a further indicator of the severity of the baby's illness. This information is not currently available with the PPESO and PRAMS databases. The recommendation will be made to the PPESO committee that including duration of NICU stay may aid in predicting high-risk PTB.

9. The PRAMS dataset originally had over 100,000 cases. Over 70,000 cases were removed due to missing values: specifically missing data for 'previous preterm birth', which is a known indicator of future preterm birth. The PRAMS project is ongoing and it is possible that future data may not have this missing value problem. Obtain new PRAMS data and re-develop the model as appropriate.
10. Assess the impact on system classification when missing values are replaced with imputed values using Ennett and Frize's technique on the PRAMS database, which contains a large number of missing values for the previous PTB variable.
11. The PPESO 1999 and 2000 datasets have many missing values, this would be a good application to test the imputation system developed by Ennett using the modified-relative weights method on new data and see if can emulate the results in this thesis.
12. Test the classification performance of the integrated hybrid model on specific maternal populations: <37 weeks gestation; <35 weeks gestation; and <32 weeks gestation.
13. Develop multiple ANNs minimal datasets, with one pertaining to information related to the fetus and the other pertaining to information related to the mother. In this instance the majority of the variables relate to the mother, however, if predictions

were made at greater than 23 weeks, more fetus-related data would be available and this could be an interesting approach.

14. Consult a panel of medical experts to determine what weights they would assign to the PRAMS 48-variable model. Modify the kNN match weights to reflect the changes and compare the final screening results with those obtained using MIRG's modified relative weights method.
15. Implement a system that would periodically add new patient cases to the *match* database, used for retrieving the k closest matching cases, and to the datasets used during MLP ANN training. Such a system would include evaluating how often the datasets and case base should be updated and the impact on the results.
16. Develop a distance-weighted kNN algorithm to neutralize the effect of inappropriate selection of k values. The goal is to make the near neighbors have greater impact on the final decision than far neighbors; this becomes particularly important for larger values of k .
17. Conduct a usability study to evaluate physician support for the use of Semantic Web services in healthcare. Determine physician views in areas of concern such as: security, privacy, ethical dilemmas, other?
18. Using the Semantic Web services for healthcare framework, and based on the clinical trial suggested in future work (2), implement the Semantic Web-based version of the integrated hybrid screening system for the early prediction of high-risk PTB. Update the framework and ontology as required.

19. Using the Semantic Web services for healthcare framework, implement the training portion of the integrated hybrid screening system, generalized for any perinatal domain, to be offered as Semantic Web services. This involves considering the myriad of security and privacy issues surrounding accessing large proprietary clinical databases, which often have stringent requirements on who can access the data and how it is stored, over the Internet.
20. Apply the Perinatal Clinical Decision Support Ontology to other perinatal outcomes of interest, adapting the ontology as needed.

APPENDIX A – LIST OF PUBLICATIONS FOR CHRISTINA CATLEY

Articles Published in Refereed Journals and Books

Catley, C., Frize, M., Walker, C.R., and Petriu, D.C. (2006) Predicting high-risk preterm birth using artificial neural networks. *IEEE Transactions of Information Technology in Biomedicine.*;10(3):540-9.

Catley, C., Frize, M., and Petriu, D.C. (2006) Semantic Web services for healthcare. Chapter in: *Handbook of Research on Informatics in Healthcare and Biomedicine*. Ideas Group Reference. Editor: Lazakidou, A. 262-269.

Frize, M., Walker, R.C., and **Catley, C.** (2006). Healthcare Knowledge Management: Knowledge Management in the Perinatal Care Environment. In Bali, R., Dwivedi, A. (Eds.), *Healthcare Knowledge Management: Issues, Advances and Successes*. Springer. 286 p.

Papers Published in refereed Conference Proceedings

Frize, M., Ibrahim, D., **Catley C.,** and Walker, C.R. (2006) Using Artificial Intelligence to Estimate Outcomes in Perinatal Medicine. Proceedings of the Canadian Conference on Electrical and Computer Engineering.

Catley, C., Petriu, D.C., and Frize, M. (2005) A UML framework for Web services-based clinical decision support. Proceedings of the 14th International Conference on Intelligent and Adaptive Systems, 174-179.

Catley, C., Frize, M., Walker, C. R., and Petriu, D.C. (2005) Predicting preterm birth using artificial neural networks. Proceedings of the 18th IEEE Symposium on

Computer-Based Medical Systems, 103-108.

Catley, C., Frize, M., Petriu, D.C., Walker, C.R., and Yang, L. (2004) Towards a Web services infrastructure for perinatal, obstetrical and neonatal clinical decision support. Proceedings of the 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2: 3334-3337.

Yang, L., Frize, M., Walker, C.R., and **Catley, C.** (2004) Towards a framework for ethical decision support and knowledge management in neonatal care. Proceedings of the 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2: 3420-3423.

Catley, C., Petriu, D.C., and Frize, M. (2004) Software performance evaluation of a Web service-based clinical decision support system architecture for intensive care units. Workshop on Software and Performance (WOSP), 130-138.

Frize, M., Herry, C., and **Catley, C.** (2003) Examples of biomedical engineering applications to modern medicine. First International Women's Conference on BIEN (Bio, Information, Environment and Nano) Technology.

Catley, C., Frize, M., Walker, C.R., and St-Germain, L. (2003) Integrating clinical alerts into an XML-based healthcare framework for the neonatal intensive care unit. Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2: 1276-1279.

Herry, C., **Catley, C.,** and Ennett, C.M. (2003) Student leadership: biomedical engineering initiatives at Carleton University. Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 4:

3490-3493.

Catley, C. and Frize, M. (2003) A prototype XML-based implementation of an integrated 'intelligent' neonatal intensive care unit. Proceedings of the 4th International IEEE EMBS Special Topic Conference on Information Technology Applications in Biomedicine, 322-325. (Student paper competition finalist)

Catley, C. and Frize, M. (2002) Design of a healthcare framework for medical data interoperability and application integration. Proceedings of the 24th Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society and the Engineering in Medicine and Biology Society, 3: 1952-1953. (Student paper competition finalist)

Catley, C. and Frize, M. (2002) Overcoming medical standardization barriers with XML. Proceedings of the 12th International Conference of Women Engineers and Scientists, 364-365.

APPENDIX B - MODELING COST FUNCTIONS AND DERIVING A TRAINING ALGORITHM

The following is a summary of Zhou and Frize's work in modeling cost functions and deriving a training algorithm to add risk stratification estimation to the MLP ANN, this work is described in detail in Zhou's Master's thesis (Zhou, 2006).

Let D be the training data D composed of patient records:

$$\{ \langle x_1, y_1 \rangle, \langle x_2, y_2 \rangle, \dots, \langle x_m, y_m \rangle \}$$

Where x_m represents the values of a set of input variables for the m^{th} patient and y_m corresponds to an outcome, which takes values of 1 or 0

Rewrite the likelihood function in the following form:

$$L(\Theta) = f(\langle x_1, y_1 \rangle, \langle x_2, y_2 \rangle, \dots, \langle x_m, y_m \rangle; \Theta) = \prod f(x_i, y_i | \Theta)$$

Where $f(x_i, y_i | \Theta)$ denotes the probability of y_i , given the input as x_i in the i^{th} patient record

The conditional probability that $y_i=1$ given x_i will be denoted as $P(y_i=1|x_i)$. This value can be given by the neural network output $O(x_i)$. It follows that $P(y_i=0|x_i)=1-O(x_i)$.

Next, rewrite the $f(\langle x_i, y_i \rangle | \Theta)$ in the likelihood function as:

$$f(x_i, y_i | \Theta) = O(x_i)^{y_i} [1-O(x_i)]^{(1-y_i)}$$

Assuming the observations of patient records are independent, the likelihood function on the training data with m records is obtained as:

$$\prod O(x_i)^{y_i} [1-O(x_i)]^{(1-y_i)}$$

Which can be rewritten in a log-likelihood form:

$$L(\Theta) = \ln[L(\Theta)] = \sum y_i * \ln O(x_i) + (1-y_i) * \ln(1-O(x_i))$$

APPENDIX C – WEIGHTS & BIASES USED FOR NEW PATIENT CASES

Optimizing for sensitivity:

```

GW1=[ 0.8723   -0.5249   0.2080   -0.1346   -0.7749   0.9128
      -0.3933   -0.5251   0.1949   -0.5215
           0.4241   -0.3156   -0.8705   0.1109   -1.0868   0.6466
           0.0614   0.5524   0.8534   -0.0972
      -0.8005   0.8444   0.8044   0.4601   -0.9628   -0.5277
           0.6936   0.4681   0.8638   0.5103
      -0.3789   -0.4878   0.9793   0.5937   -1.0285   0.4884
           0.0642   0.2534   0.9090   0.9618
      -0.7750   0.7300   0.0329   0.5048   -0.0908   0.7017
           0.7607   0.2902;
      -0.7103   -0.8101   0.2591   0.4156   0.1027   -0.5747
           0.8980   -0.2900   0.3347   -0.0670
           0.8464   -0.7835   -0.3378   -0.4630   0.6266   0.4521
      -0.1754   0.4109   0.3614   0.5604
      -0.8779   -0.0020   0.5799   -0.5713   -0.3248   -0.5949
           0.6186   -0.4634   0.5132   0.2171
      -0.1895   -1.0303   -0.0035   0.8510   -0.9121   -0.0846
      -0.4277   0.7219   0.4285   0.4897
      -0.3019   -0.2366   0.6755   0.6404   0.8865   1.0702
      -0.2434   -0.1805;
      -0.4359   -2.1202   -0.9803   -0.0387   -0.0323   0.6745
           0.2231   -0.4244   -0.3724   -0.0619
           0.2209   0.0997   0.5061   -0.7578   -0.0468   0.2917
           0.5673   1.1123   0.8394   0.5847
      -0.3785   -0.0195   -0.1623   0.2620   -0.0415   0.0537
      -0.2022   0.3985   0.0351   0.2550
      -0.0160   0.1434   -0.9365   -1.5124   -0.2027   -0.1551
      -0.4958   0.4339   0.2986   0.0757
      -0.0224   0.0844   0.0533   0.1105   0.1933   -0.2538
      -0.9496   0.1393;
           0.7391   -0.0491   0.2615   -0.8854   -0.3998   -0.6566
           0.5391   0.0195   -0.2261   -0.4935
           0.6675   1.2450   -0.2920   -0.3266   -0.2213   -0.1389
           0.2028   -0.4592   -0.1020   -0.4991
           0.3013   -0.2337   0.6086   0.8317   0.7063   0.2676
      -0.5709   0.4266   -0.0643   -1.0447
           0.7263   -0.5514   0.8634   0.2871   0.0009   0.2955
      -0.2770   0.7841   -0.1497   0.5229
      -0.3316   -0.3138   -0.2011   0.5183   -0.3810   -0.8414
      -0.5124   -0.8955;];

```

```

GB1=[
      -0.7215;
      -0.8575;

```

```
-1.2250;
-0.2827;];
```

```
GW2= [ 0.6467    0.7851   -3.3599   -0.4395];
GB2= [ 0.8584];
```

Optimizing for specificity:

```
GW1= [ -0.3510   -2.7633   -1.8420    0.2622   -0.0489    0.5964
        0.4403   -0.8246   -0.5873    0.7850    0.0574    0.0137
        0.4291
       -1.3702    0.0546   -0.0390    0.8669    1.5925    0.9585
        0.4992   -0.0724    0.1589   -0.0331    0.8002    0.0795
        0.8019   -0.2551   -0.5741    0.0483    0.0762    0.3110
        0.4893
       -1.2537   -2.1956   -0.3767   -0.1081   -1.2046    0.6985
        0.4115    0.0545    0.2033    0.2356   -0.2796    0.2655
        0.2059   -0.8299   -0.1869   -0.6730;
        0.5503   -0.7288   -1.2335   -0.4176    0.1102   -0.6216
0.9327
       -0.4530   -0.4559    0.1690   -0.6673    0.3551   -0.1153
        0.2831   -0.8226    1.0805    0.5153    0.2096    0.6754
0.6454
       -0.5471   -0.4206    0.3866   -0.0296    0.4236   -0.2273
        0.5844    0.8870    0.8970   -0.0039    0.2748    0.5963  -
1.0936
       -0.0809    0.6186    0.5730   -0.3899   -0.3877   -0.4550
0.1195
       -0.8030    0.5311    1.1553   -0.1804   -0.5688   -0.1109
0.6454
       -1.0871;
        0.5708   -1.1577   -0.0373   -0.3175   -0.0416    0.2307
       -0.4539   -0.0762   -0.5451   -0.1202   -0.1079    0.8152
        0.9749   -0.5916    0.5734    1.3786   -0.3079    1.2485
        0.9335    1.0919   -0.3847    0.7499    0.0690    0.8496
       -0.2550   -0.3092   -0.7803    0.5488   -0.2813   -0.4314
       -0.3521   -0.6492    0.0957   -1.0600   -0.4599   -0.5799
       -0.4056   -0.5572    0.0319    0.2109   -0.1809    0.5174
        0.1123   -0.4224    0.3135   -0.9133   -0.3374   -0.8231;  ];
GB1= [ -0.2507;
       -0.4579;
       -0.9112;
];
GW2= [ -2.0175   -0.5464   -1.0487 ];
GB2= [ 0.7761 ];
```

APPENDIX D – BIVARIATE CORRELATIONS (PEARSON)

Correlations		HR-PTB
MATAGE	Pearson Correlation	-.014(*)
	Sig. (2-tailed)	0.015
	Sum of Squares and Cross-products	-73.706
	Covariance	-0.003
	N	29,104
PLURAL_N	Pearson Correlation	.181(**)
	Sig. (2-tailed)	0.000
	Sum of Squares and Cross-products	928.134
	Covariance	0.032
	N	29,104
PREV_PRE	Pearson Correlation	.166(**)
	Sig. (2-tailed)	0.000
	Sum of Squares and Cross-products	1,807.432
	Covariance	0.062
	N	29,104
PARITY_N	Pearson Correlation	.033(**)
	Sig. (2-tailed)	0.000
	Sum of Squares and Cross-products	173.079
	Covariance	0.006
	N	29,104
GENDER	Pearson Correlation	0.000
	Sig. (2-tailed)	0.979
	Sum of Squares and Cross-products	-2.404
	Covariance	0.000
	N	29,104
BF_EVER	Pearson Correlation	.077(**)
	Sig. (2-tailed)	0.000
	Sum of Squares and Cross-products	588.405
	Covariance	0.020
	N	29,104
MOMSMOKE	Pearson Correlation	.053(**)
	Sig. (2-tailed)	0.000
	Sum of Squares and Cross-products	576.979
	Covariance	0.020
	N	29,104
MOMCIG_N	Pearson Correlation	.053(**)
	Sig. (2-tailed)	0.000
	Sum of Squares and Cross-products	274.482
	Covariance	0.009
	N	29,104

SMK_2YRS	Pearson Correlation	.058(**)
	Sig. (2-tailed)	0.000
	Sum of Squares and Cross-products	787.231
	Covariance	0.027
	N	29,104
MOMDRINK	Pearson Correlation	0.005
	Sig. (2-tailed)	0.371
	Sum of Squares and Cross-products	13.928
	Covariance	0.000
	N	29,104
MOMDRKS_N	Pearson Correlation	.015(*)
	Sig. (2-tailed)	0.012
	Sum of Squares and Cross-products	76.776
	Covariance	0.003
	N	29,104
MARRIED	Pearson Correlation	-.086(**)
	Sig. (2-tailed)	0.000
	Sum of Squares and Cross-products	-
	Covariance	1,203.410
	N	-0.041
MAT_ED_N	Pearson Correlation	29,104
	Sig. (2-tailed)	-.056(**)
	Sum of Squares and Cross-products	0.000
	Covariance	-286.723
	N	-0.010
BLACK	Pearson Correlation	29,104
	Sig. (2-tailed)	.129(**)
	Sum of Squares and Cross-products	0.000
	Covariance	1,360.567
	N	0.047
WHITE	Pearson Correlation	29,104
	Sig. (2-tailed)	-.034(**)
	Sum of Squares and Cross-products	0.000
	Covariance	-513.227
	N	-0.018
INDIAN	Pearson Correlation	29,104
	Sig. (2-tailed)	-0.006
	Sum of Squares and Cross-products	0.284
	Covariance	-42.222
	N	-0.001
ALASKA	Pearson Correlation	29,104
	Sig. (2-tailed)	-0.011
	Sum of Squares and Cross-products	0.062
	Covariance	-63.292
	N	-0.002
HAWAII	Pearson Correlation	29,104
	Sig. (2-tailed)	-.050(**)
		0.000

PACIFIC	Sum of Squares and Cross-products	-351.084
	Covariance	-0.012
	N	29,104
	Pearson Correlation	-.021(**)
	Sig. (2-tailed)	0.000
OTHER	Sum of Squares and Cross-products	-130.633
	Covariance	-0.004
	N	29,104
	Pearson Correlation	-.064(**)
	Sig. (2-tailed)	0.000
OTHER_TERM_N	Sum of Squares and Cross-products	-587.556
	Covariance	-0.020
	N	29,104
	Pearson Correlation	.058(**)
	Sig. (2-tailed)	0.000
MM_DIAB	Sum of Squares and Cross-products	298.847
	Covariance	0.010
	N	29,104
	Pearson Correlation	.015(*)
	Sig. (2-tailed)	0.013
MM_FEVER	Sum of Squares and Cross-products	90.002
	Covariance	0.003
	N	29,104
	Pearson Correlation	.044(**)
	Sig. (2-tailed)	0.000
MM_NOMD	Sum of Squares and Cross-products	134.880
	Covariance	0.005
	N	29,104
	Pearson Correlation	-.114(**)
	Sig. (2-tailed)	0.000
INSURE	Sum of Squares and Cross-products	-
	Covariance	1,716.934
	N	-0.059
	Pearson Correlation	29,104
	Sig. (2-tailed)	-.055(**)
MEDICAD3	Sum of Squares and Cross-products	0.000
	Covariance	-814.828
	N	-0.028
	Pearson Correlation	29,104
	Sig. (2-tailed)	.030(**)
VITAMIN_N	Sum of Squares and Cross-products	0.000
	Covariance	347.640
	N	0.012
	Pearson Correlation	29,104
	Sig. (2-tailed)	-.022(**)
	Sum of Squares and Cross-products	0.000
	Covariance	-113.439
		-0.004

MOM_LB_N	N	29,104
	Pearson Correlation	0.002
	Sig. (2-tailed)	0.754
	Sum of Squares and Cross-products	9.500
	Covariance	0.000
PREG_TRY	N	29,104
	Pearson Correlation	-.055(**)
	Sig. (2-tailed)	0.000
	Sum of Squares and Cross-products	-849.824
	Covariance	-0.029
PNC_MTH_N	N	29,104
	Pearson Correlation	.026(**)
	Sig. (2-tailed)	0.000
	Sum of Squares and Cross-products	133.209
	Covariance	0.005
EARLY_PNC	N	29,104
	Pearson Correlation	-.039(**)
	Sig. (2-tailed)	0.000
	Sum of Squares and Cross-products	-468.770
	Covariance	-0.016
WIC_PREG	N	29,104
	Pearson Correlation	.035(**)
	Sig. (2-tailed)	0.000
	Sum of Squares and Cross-products	545.444
	Covariance	0.019
MM_HBP	N	29,104
	Pearson Correlation	.109(**)
	Sig. (2-tailed)	0.000
	Sum of Squares and Cross-products	766.145
	Covariance	0.026
MM_BLEED	N	29,104
	Pearson Correlation	.144(**)
	Sig. (2-tailed)	0.000
	Sum of Squares and Cross-products	849.338
	Covariance	0.029
MORB_NAU	N	29,104
	Pearson Correlation	.059(**)
	Sig. (2-tailed)	0.000
	Sum of Squares and Cross-products	814.364
	Covariance	0.028
MORB_KID	N	29,104
	Pearson Correlation	.060(**)
	Sig. (2-tailed)	0.000
	Sum of Squares and Cross-products	701.275
	Covariance	0.024
MORB_CRV	N	29,104
	Pearson Correlation	.098(**)

	Sig. (2-tailed)	0.000
	Sum of Squares and Cross-products	502.463
	Covariance	0.017
	N	29,104
MORB_CAR	Pearson Correlation	.017(**)
	Sig. (2-tailed)	0.004
	Sum of Squares and Cross-products	73.759
	Covariance	0.003
	N	29,104
STRS_FM3	Pearson Correlation	-0.005
	Sig. (2-tailed)	0.406
	Sum of Squares and Cross-products	-66.062
	Covariance	-0.002
	N	29,104
STRS_DV	Pearson Correlation	.064(**)
	Sig. (2-tailed)	0.000
	Sum of Squares and Cross-products	615.344
	Covariance	0.021
	N	29,104
STRS_MOV	Pearson Correlation	.031(**)
	Sig. (2-tailed)	0.000
	Sum of Squares and Cross-products	452.915
	Covariance	0.016
	N	29,104
STRS_HOM	Pearson Correlation	.025(**)
	Sig. (2-tailed)	0.000
	Sum of Squares and Cross-products	158.646
	Covariance	0.005
	N	29,104
STRS_JOB	Pearson Correlation	.025(**)
	Sig. (2-tailed)	0.000
	Sum of Squares and Cross-products	264.875
	Covariance	0.009
	N	29,104
STRS_WRK	Pearson Correlation	.052(**)
	Sig. (2-tailed)	0.000
	Sum of Squares and Cross-products	506.446
	Covariance	0.017
	N	29,104
STRS_ARG	Pearson Correlation	.061(**)
	Sig. (2-tailed)	0.000
	Sum of Squares and Cross-products	866.026
	Covariance	0.030
	N	29,104
STRS_COM	Pearson Correlation	.160(**)
	Sig. (2-tailed)	0.000
	Sum of Squares and Cross-products	1,456.439

PAD_HUS	Covariance	0.050
	N	29,104
	Pearson Correlation	.160(**)
	Sig. (2-tailed)	0.000
	Sum of Squares and Cross-products	1,456.439
INC_JOB	Covariance	0.050
	N	29,104
	Pearson Correlation	.160(**)
	Sig. (2-tailed)	0.000
	Sum of Squares and Cross-products	1,456.439
HR-PTB	Covariance	0.050
	N	29,104
	Pearson Correlation	1
	Sig. (2-tailed)	
	Sum of Squares and Cross-products	8,256.851
	Covariance	0.284
	N	29,104

** . Correlation is significant at the 0.01 level (2-tailed).

* . Correlation is significant at the 0.05 level (2-tailed).

REFERENCES

Abidi, S. S. R. (2005). Medical knowledge morphing: Towards case-specific integration of heterogeneous medical knowledge resources. *Proceedings of the 18th IEEE Symposium on Computer-Based Medical Systems*, 208-213.

Aerts, M., and Iams, J. (1999). Prevention of spontaneous preterm birth. *Contemporary OB/GYN*, 128-136.

Ahmed, S. (2003). *Ontologies of electronic devices on DAML+OIL for automated product design services in the semantic web*. Master of Engineering in Telecommunication Technology Management, Carleton University, Ottawa, ON.

Alexian Brother's Medical Center (2006). Life with baby: Pregnancy and beyond. Last accessed: August 24, 2006 from <http://www.alexian.org/progserv/babies/thirdtri/apgarscore.html>.

Allan, N., Aylward, D., Berry, E., Bowie, J., Brooks, F., and Burgoyne, W. (2002). *Preterm birth: Making a difference*. Best Start Resource Centre, Toronto, ON.

Alpaydin, E. (1993). Multiple networks for function learning. *IEEE International Conference on Neural Networks*, 1, 9-14.

Ambalavanan, N., and Carlo, W. A. (2001). Comparison of the prediction of extremely low birth weight neonatal mortality by regression analysis and by neural networks. *Early Hum Dev*, 65, 123-137.

American College of Obstetricians and Gynecologists. (2001). Assessment of risk factors for preterm birth. *Obstetrics and Gynecology*, 98, 709-719.

Aslandogan, Y. A., and Mahajani, G. A. (2004). Evidence combination in medical data mining. *Proceedings. International Conference on Information Technology: Coding and Computing*, 465-469.

Beiu, V. (2004). A novel highly reliable low-power nano architecture when von neumann augments kologorov. *Proceedings of the 15th IEEE International Conference on Application-Specific Systems, Architectures and Processors*, 167-177.

Berners-Lee, T., Hendler, J., and Lasilla, O. (2001). The semantic web. *Scientific American*, 284, 34-43.

Blum, A. (1992). *Neural networks in C++*. John Wiley & Sons, Inc., New York, NY.

Boz, O. (2002). Extracting decision trees from trained neural networks. *Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 456-461.

Buskard T, Stevenson M, Frize M, and Solven F. (1994). Estimation of ventilation, length of stay, and mortality using artificial neural networks. *Proceedings of the Canadian Conference on Electrical & Computer Engineering*, Location, 726-729.

Canadian Institutes of Health Research. (2005). *Child health*. Last accessed: October 19, 2005 from <http://www.cihr-irsc.gc.ca/e/28894.html>.

Catley, C., Frize, M., and Petriu, D. (2006). Semantic web services for healthcare. In A. Lazakidou (Ed.), *Handbook of research on informatics in healthcare and biomedicine*. Idea Group Reference, 262-269.

Catley, C., Frize, M., Petriu, D. C., Walker, C. R., and Yang, L. (2004). Towards a web services infrastructure for perinatal, obstetrical, and neonatal clinical decision support. *Proceedings of the 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 3334-3337.

Centers for Disease Control and Prevention. (2005). *Pregnancy risk assessment monitoring system (PRAMS): Home*. Last accessed: October 19, 2005 from <http://www.cdc.gov/reproductivehealth/PRAMS/index.html>.

Chatterjee, S. (2003). Developing enterprise web services and applications: Opportunities and best practices for the healthcare industry. *Proceedings of the 5th International Workshop on Enterprise Networking and Computing in Healthcare Industry*, Location, 159-159.

Chen, D., and Burrell, P. (2001). Case-based reasoning system and artificial neural networks: A review. *Neural Computing and Applications*, 10, 264-276.

Chen, H., Colaert, D., and De Roo, J. (2004). Towards adaptable clinical pathways using semantic web technology: Position paper for W3C Workshop Semantic Web for Life Science. Last accessed: August 24, 2006 from <http://www.w3.org/2005/04/swls/>.

Committee on Quality of Healthcare in America, Institute of Medicine. (2000). In Kohn L. T., Corrigan J. M. and Donaldson M. S. (Eds.). *To err is human: Building a safer health system*. National Academy Press, Washington, D.C.

Cotea, C., and Jiwani, S. (2003) *A Case-based reasoning system for the neonatal intensive care unit*. 4th Year Project Report, University of Ottawa, Ottawa, ON.

Crane, J.G., Armson, A., Dodds, L., Feinberg, R.F., Kennedy, and W. Kirkland, S. (1999). Risk Scoring, Fetal Fibronectin, and Bacterial Vaginosis to Predict Preterm Delivery. *Obstetrics and Gynecology*, 93, 517-522. Last accessed October 31, 2006. Available at: <http://www.greenjournal.org/cgi/content/full/93/4/517>

Crawford, C. (2005). Editorial pointers. *Communications of the ACM*, 48, 5-6.

Creasy, R., Gummer, B., and Liggins, G. (1980). System for predicting spontaneous preterm birth. *Obstetrics and Gynecology*, 55, 692-695.

Daconta, M. C., Obrst, L. J., and Smith, K. T. (2003). *The semantic web: A guide to the future of XML, web services, and knowledge management*. Wiley Publishing Inc., Indianapolis, IN.

Dawood, Y. (1980). The obstetric view of premature labor. In G. F. Smith, & D. Vidyasagar (Eds.), *Historical review and recent advances in neonatal and perinatal medicine*. Mead Johnson. Last accessed: August 24, 2006 from <http://www.neonatology.org/classics/mj1980/ch10.html>.

Dayhoff, J. E., and DeLeo, J. M. (2001). Artificial neural networks – opening the black box. *American Cancer Society (Cancer Supplement)*, 91, 1615-1635.

De Bruijn, J., Fensel, D., Keller, U., and Lara, R. (2005). Using the web service modeling ontology to enable semantic e-business. *Communications of the ACM*, 48, 43-47.

Ennett, C. M., Frize, M., and Scales, N. (2003). Evaluation of the logarithmic-sensitivity index as a neural network stopping criterion for rare outcomes. *Proceedings of the 4th International IEEE EMBS Special Topic Conference on Information Technology Applications in Biomedicine*, 207-210.

Ennett, C. M., Frize, M., and Charette, E. (2004). Improvement and automation of artificial neural networks to estimate medical outcomes. *Med Eng Phys*, 26, 321-328.

Ennett, C. (2003). Imputation of missing values by integrating artificial neural networks and case-based reasoning. PhD Electrical Engineering, Carleton University, Ottawa, ON.

Ennett, C. M., and Frize, M. (2003). Weight-elimination neural networks applied to coronary surgery mortality prediction. *IEEE Transactions of Information Technologies in Biomedicine*, 7, 86-92.

Ennett, C. M., and Frize, M. (2000). Selective sampling to overcome skewed a priori probabilities with neural networks. *Proceedings of the AMIA Symposium*, 225-229.

Fausett, L. (1994). *Fundamentals of neural networks*. Prentice-Hall, Englewood Cliffs, NJ.

Fiore, E. (2003). Multiple births and the rising rate of preterm delivery. *Contemporary Obstetrics and Gynecology*, 48, 67-77.

Forsström, J. J., and Dalton, K. J. (1995). Artificial neural networks for decision support in clinical medicine. *Annals of Internal Medicine*, 27, 509-517.

Frize, M., Solven, F.G., Stevenson, M., Nickerson, B.G., Buskard, T., and Taylor, K. (1995). Computer-assisted decision-support systems for patient management in an intensive care unit. *Proceedings of Medinfo*, 1009-1012.

Frize, M., Wang, L., Ennett, C.M., Nickerson, B., Solven, F.G., and Stevenson, M.H. (1998). New advances and validation of knowledge management tools for critical care using classifier techniques. *Proceeding of the AMIA Symposium*, 553-558.

Frize, M., and Walker, R., (2000). Clinical decision-support systems for intensive care units using case-based reasoning. *Medical Engineering and Physics*, 22, 671-677.

Frize, M., Ennet, C. M., and Hebert, P. (2001). Improving the efficiency and effectiveness of artificial neural networks as clinical decision-support systems. *Proceedings of Medinfo*, 574.

Frize, M., Ennett, C. M., Stevenson, M., & Trigg, H. C. E. (2001). Clinical decision-support systems for intensive care units using artificial neural networks. *Medical Engineering & Physics*, 23, 217-225.

Frize, M., Ibrahim, D., Seker, H., Walker, R. C., Odetayo, M. O., and Petrovic, D. (2004). Predicting clinical outcomes for newborns using two artificial intelligence approaches. *Proceedings of the 26th Annual International Conference of the Engineering in Medicine and Biology Society*, 3202-3205.

Frize, M., Yang, L., Walker, R. C., and O'Connor, A. (2005). Conceptual framework of knowledge management for ethical decision-making support in neonatal intensive care. *Transactions of Information Technology in Biomedicine*, 9, 205-215.

Frize, M., Walker, R.C., and Catley, C. (2006). Healthcare Knowledge Management: Knowledge Management in the Perinatal Care Environment. In Bali, R., Dwivedi, A. (Eds.), *Healthcare Knowledge Management: Issues, Advances and Successes*. Springer. 286 p.

Garson, G. D. (1991). Interpreting neural-network connection weights. *AI Expert*, 6, 46-51.

Goh, T. C. (1995). Back-propagation neural networks for modeling complex systems. *Artificial Intelligence in Engineering*, 9, 143-151.

Golbeck, J., Fragoso, G., Hartel, F., Hendler, J., Parsia, B., & Oberthaler, J. (2003). The national cancer institute's thesaurus and ontology. *Journal of Web Semantics*, 1(1), 89-123.

Goldenberg, R. L., Mercer, B. M., Iams, J. D., Moawad, A. H., Meis, P. J., and Das, A. (1997). The preterm prediction study: Patterns of cervicovaginal fetal fibronectin as predictors of spontaneous preterm delivery. *American Journal of Obstetrics and Gynecology*, 177, 8-12.

Goldenberg, R. L., Iams, J. D., Mercer, B. M., Meis, P. J., Moawad, A. H., and Copper, R. L. (1998). The preterm prediction study: The value of new vs. standard risk factors in predicting early and all spontaneous births. *American Journal of Public Health*, 88, 233-238.

Goldenberg, R. L., and Rouse, D. J. (1998). Prevention of premature birth. *New England Journal of Medicine*, 339, 313-320.

Goodwin, L., and Maher, S. (2000). Data mining for preterm birth prediction. *Proceedings of the 2000 ACM Symposium on Applied Computing*, 46-51.

Graham, S. (2003). Building web services with java: Making sense of XML, SOAP, WSDL, and UDDI. SAMS Publishing, Indianapolis, IN.

Gryzmala-Busse, J. (1989). An overview of the LERS1 learning system. *Proceedings of the 2nd International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems*, 838-844.

Grzymala-Busse, J. W., Goodwin, L. K., and Zhang, X. (2003). Increasing sensitivity of preterm birth by changing rule strengths. *Pattern Recognition Letters*, 24, 903-910.

Guler, N., Yildiz, O. T., Gurgun, F., Varol, F., and Alpaydin, E. (2001). Discriminant functions and decision tree induction techniques for antenatal fetal risk assessment. *Proceedings of the International Joint Conference on Neural Networks*, 2712-2717.

Health Canada. (2000). *Perinatal health indicators for Canada: A resource manual*. Minister of Public Works and Government Services Canada. Last accessed: August 24, 2006 from <http://www.phac-aspc.gc.ca/rhs-ssg/phic-ispc/index.html>.

Hendler, J. (2001). Agents and the semantic web. *IEEE Intelligent Systems*, 16, 30-37.

Hirsch, E. (2006). Born too soon cervical ultrasonography to predict preterm delivery is unproven. *OBGYN.net*, Last accessed: July 3, 2006 from <http://www.obgyn.net/us/us.asp?page=/us/cotm/0002/hirsch>.

Hoffman, D. (2003). Marketing + MIS = e-service. *Comm. ACM*, 46, 29-34.

Horrocks, I. (2006). An introduction to description logics. Last accessed: August 24, 2006 from <http://www.cs.man.ac.uk/%7Ehorrocks/Teaching/cs646/Slides/pt2-dlintro.pdf>.

Hudson, D.L. & Cohen, M.E (2000). *Neural Networks and Artificial Intelligence for Biomedical Engineering*. IEEE Press Series in Biomedical Engineering, New York, NY.

Hullermeier, E. (2005) Experience-based decision making: A satisficing decision tree approach. *IEEE Transactions on Systems, Man and Cybernetics*, Part A, 35, 641-653.

Iams, J. D. (1997). The preterm prediction study: A model for estimation of risk of spontaneous preterm birth in parous women. *American Journal of Obstetrics and Gynecology*, 176:, 652-655.

Iams, J. D., Goldenberg, R. L., Merber, B. M., Moawad, A. H., Meis, P. J., and Das, A. F. (2001). The preterm prediction study: Can low-risk women destined for spontaneous preterm birth be identified. *General Obstetrics and Gynecology*, 184, 652-655.

Ibrahim-Swailum, D. (2006). *Identifying Risk Factors for Newborn Outcomes using Artificial Neural Networks*. Master of Applied Science, University of Ottawa, Ottawa, ON.

Jerebko, A. K., Malley, J. D., Franaszek, M., and Summers, R. M. (2003). Multiple neural network classification scheme for detection of colonic polyps in CT colonography data sets. *Academic Radiology*, 10, 154-160.

Johnston, R. B. J., Williams, M. A., Hogue, C. J. R., & Mattison, D. R. (2001). Overview: New perspectives on the stubborn challenge of preterm birth. *Paediatric and Perinatal Epidemiology*, 15(Suppl. 2), 3-6.

Kijsirikul, B., and Chongkasemwongse, K. (2001). Decision tree pruning using backpropagation neural networks. *Proceedings of the International Joint Conference on Neural Networks*, 1876-1880.

Kim, Y., Street, W. N., Menczer, F., and Russell, G. J. (2003). Feature selection in data mining. In J. Wang (Ed.), *Data mining: Opportunities and challenges*, Idea Group Publishing, 80-105.

Knublauch, H. (2003). *An AI tool for the real world: Knowledge modeling with protege*. Last accessed: August 1, 2006 from <http://www.javaworld.com/javaworld/jw-06-2003/jw-0620-protege.html>.

Knublauch, H., Dameron, O., and Musen, M. A. (2004). *Weaving the biomedical semantic web with the protege OWL plugin*. Stanford Medical Informatics, Stanford University, Stanford, CA. Last accessed: August 24, 2006 from <http://protege.stanford.edu/plugins/owl/publications/KRMed2004-protege-owl.pdf>.

Kreger, H. (2003). Fulfilling the web services promise. *Comm. ACM*, 46, 29-34.

Kudyba, S., Hamar, G. B., and Gandy, W. M. (2005). Enhancing efficiency in the healthcare industry. *Communications of the ACM*, 48, 107-110.

Kwon, O. B. (2003). Meta web service: Building web-based open decision support system based on web services. *Expert Systems with Applications*, 24, 375-389.

Lea, D., and Vinoski, S. (2003). Middleware for web services. *IEEE Internet Computing*, 7, 28-29.

Lee, J. S., Hwang, J., Davis, D. T., and Nelson, A. C. (1991). Integration of neural networks and decision tree classifiers for automated cytology screening. *Proceedings of the International Joint Conference on Neural Networks*, 257-262.

Lee, Y., Patel, C., Chun, S. A., and Geller, J. (2004a). Compositional knowledge management for medical services on semantic web. *Proceedings of the 13th International World Wide Web Conference on Alternate Track Papers and Posters*, 498-499.

Lee, Y., Patel, C., Chun, S. A., and Geller, J. (2004b). Towards intelligent web services for automating medical service composition. *Proceedings of the IEEE International Conference on Web Services*, 384-391.

Li, H. (2005). Assessment of two AI approaches to predict mortality in adult intensive care units. Master of Applied Science, Carleton University, Ottawa ON.

Lisboa, P. J. G. (2002). A review of evidence of health benefit from artificial neural networks in medical intervention. *Neural Networks*, 15, 11-39.

Livingstone, D. H., Manallack, D. T., and Tetko, I. V. (1997). Data modeling with neural networks: Advantages and limitations. *Journal of Computer Aided Molecular Design*, 11, 135-142.

Lockwood, C. J., and Kuczynski, E. (2001). Risk stratification and pathological mechanisms in perterm delivery. *Paediatric and Perinatal Epidemiology*, 15, 78-89.

March of Dimes. (2003). *The growing problem of prematurity*. Last accessed: August 24, 2006 from <http://www.smfm.org/attachedfiles/PrematurityFact.pdf#search=%22%22march%20of%20Dimes%22%20%2B%20perinatal%20statistics%202003%22>.

March of Dimes. (2006). *Fetal Fibronectin (fFN): A Test for Preterm Delivery*. Last accessed: October 31, 2006 from http://www.marchofdimes.com/professionals/681_1149.asp.

McGowan, H. C. E., Stevenson, M., and Frize, M. (1996). A reporting guideline for medical applications of artificial neural networks. *Proceedings of the CMBEC*. Last accessed: October 3, 2006 from <http://www.sce.carleton.ca/faculty/frize/charpap.htm>

McGregor, C., Heath, J., and Wei, M. (2005). A web services based framework for the transmission of physiological data for local and remote neonatal intensive care. *Proceedings of the IEEE International Conference on e-Technology, e-Commerce, and e-Service*, Location, 496-501.

McIlraith, S., Son, T. C., and Zeng, H. (2001). Semantic web services. *IEEE Intelligent Systems*, 16, 46-53.

McLean, M., Walters, W. A., and Smith, R. (1993). Prediction and early diagnosis of preterm labor: A critical review. *Obstetrical and Gynecological Survey*, 48, 209-225.

MedicalGlossary.org. (2004). *Perinatal care*. Last accessed: October 3, 2005 from http://www.medicalglossary.org/patient_care_perinatal_care_definitions.html.

Medjahed, B., Bouguettaya, A., and Elmagarmid, A. K. (2003). Composing web services on the semantic web. *The International Journal on very Large Data Bases*, 12, 333-351.

Medjahed, B., Benatallah, B., Bouguettaya, A., Ngu, A. H. H., and Elmagarmid, A. K. (2003). Business-to-business interactions: Issues and enabling technologies. *The International Journal on very Large Data Bases*, 12, 59-85.

Medsker, L. R. (1994). *Hybrid neural network and expert systems*. Kluwer Academic Publishers, Boston, MA.

Meis, P. J., Goldenberg, R. L., Mercer, B. M., Iams, J. D., Moawad, A. H., and Miodovnik, M. (1998). The preterm prediction study: Risk factors for indicated preterm births. *American Journal of Obstetrics and Gynecology*, 178, 562-567.

Mercer, B. M., Goldenberg, R. L., Moawad, A. H., Meis, P. J., and Das, A. F. (1999). The preterm prediction study: Effect of gestational age and cause of preterm birth on subsequent obstetric outcomes. *American Journal of Obstetrics and Gynecology*, 181, 1216.

Metz, C. E. (1978). Basic principles of ROC analysis. *Seminars in Nuclear Medicine*, 8, 283-298.

Motta, E., Domingue, J., Cabral, L., and Gaspari, M. (2003). IRS-II: A framework and infrastructure for semantic web services. *Proceedings of the International Semantic Web Conference*, 306-318.

Myllymaki, P., and Tirri, H. (1993). Massively parallel case-based reasoning with probabilistic similarity metrics. *Proceedings of the 1st European Workshop on Case-Based Reasoning*, 144-154.

Open Clinical (Knowledge Management for Medical Care). (2005). *e-Health*. Last accessed: August 24, 2006 from <http://www.openclinical.org/e-Health.html>.

Patient Plus. (2006). *Gravidity and parity in pregnancy*. Last accessed: August 24, 2006 from <http://www.patient.co.uk/showdoc/40000158/>.

Peaceman, A. M., Andres, W. W., Thorp, J. M., Cliver, S. P., Lukes, A., and Iams, J. D. (1997). Fetal fibronectin as a predictor of preterm birth in patients with symptoms: a multicenter trial. *American Journal of Obstetrics & Gynecology*, 177, 13-18.

Penny, W., and Frost, D. (1996). Neural networks in clinical medicine. *Medical Decision Making*, 16, 386-398.

Perinatal Partnership Program of Eastern and Southeastern Ontario (PPESO). (2005). *Perinatal services in Ontario: How are we doing?* Last accessed: August 24, 2006 from www.pppeso.on.ca/english/Ontario%20Perinatal%20Report%20PDF%202005.pdf.

Pregnancy-info.net. (2006) *Pregnancy complications*. Last accessed: August 24, 2006 from http://www.pregnancy-info.net/pregnancycomp_8.html

Public Health Agency of Canada. (2006) *Perinatal surveillance system*. Last accessed: June 24, 2006 from http://www.phac-aspc.gc.ca/rhs-ssg/factshts/pterm_e.html.

Reategui, E. B., Campbell, J. A., and Leao, B. F. (1997). Combining a neural network with case-based reasoning in a diagnostic system. *Artificial Intelligence in Medicine*, 9, 5-27.

Rideout, K. (2005). Adverse medical event reduction technologies: identification of primary risk factors of adverse medical events using artificial neural networks. Master of Applied Science, Carleton University, Ottawa, ON.

Roman, D., (2005). Web service modeling ontology. *Journal of Applied Ontology*, 1, IOS Press.

Rybachynski, D. (2005). Design of an artificial neural network research framework: Efficient development of more effective prediction models for medical outcomes. Masters of Applied Science, University of Ottawa, Ottawa, ON.

Sargent, D. J. (2001). Comparison of artificial neural networks with other statistical approaches – results from medical data sets. *Cancer*, 91, 1636-1642.

Schweiger, R., Brumhard, M., Hoelzer, S., and Dudeck, J. (2005). Implementing healthcare systems using XML standards. *International Journal of Medical Informatics*, 74, 267-277.

Setiono, R., Leow, W. K., and Thong, J. Y. L. (2000). Opening the neural network black box: An algorithm for extracting rules from function approximating artificial neural networks. *Proceedings of the Twenty First International Conference on Information Systems*, 176-186.

Shadbolt, N., Berners-Lee, T., and Hall, W. (2006). The semantic web revisited. *IEEE Intelligent Systems*, 21, 96-101.

Shafer, G. (1976). *A mathematical theory of evidence*. Princeton University Press.

Shiffman, R. N., Michel, G., Essaihi, A., and Marcy, T. W. (2004). Using a guideline-centered approach for the design of a clinical decision support system to promote smoking cessation. *Proceedings of the Symposium on Computerized Guidelines and Protocols*, 152-156.

Shin, C. K., Yun, U. T., Kim, H. K., and Park, S. C. (2000). A hybrid approach of neural network and memory-based learning to data mining. *IEEE Transactions on Neural Networks*, 11, 637-646.

Sim, I. (1995). *Developing a clinical trials ontology: Comments on domain modeling and ontology reuse*. Section on Medical Informatics, Stanford University School of medicine, Stanford, CA. Last accessed: August 24, 2006 from http://smi-web.stanford.edu/pubs/SMI_Reports/SMI-95-0584.pdf.

Simon, S. (2005). *Explanation of ROC curves*. Last accessed: August 24, 2006 from <http://www.cmh.edu/stats/ask/roc.asp>.

Singh, R., Iyer, L. S., and Salam, A. F. (2005a). Semantic e-business. *International Journal of Semantic Web and Information Systems*, 1, 19-35.

Singh, R., Iyer, L. S., and Salam, A. F. (2005b). The semantic e-business vision. *Communications of the ACM*, 48, 38-41.

Skogan, D., Gronmo, R., and Solheim, I. (2004). Modeling Web Service

Composition in UML. *Proceedings of the 8th International IEEE Enterprise Distributed Object Computing Conference*. Available at: <http://www.cis.uab.edu/info/edoc2004/presentations/gronmo.pdf>

Stafford, T. (2003). E-services. *Communications of the ACM*, 46, pages.

Stalidis, G., Prentza, A., Vlachos, I., Maglavera, S., and Koutsouris, D. (2001). Medical support system for continuation of care based on web technology. *International Journal of Medical Informatics*, 64, 385-400.

Stanford University (2006). *Protégé 2000 documentation home page*. Last accessed: August 24, 2006 from <http://protege.stanford.edu/overview/protege-owl.html>.

Sun, Y. (2004). Methods for automated concept mapping between medical databases. *Journal of Biomedical Informatics*, 37, 162-178.

Tan, H., Wen, S. W., Walker, M., and Kitaw, D. (2004). *Early prediction of preterm birth by logistic regression in multiple pregnancy*. Report obtained October 2004, from Dr. Shi Wu Wen, OMNI research group, Dept. of Obstetrics and Gynecology, Faculty of Medicine, University of Ottawa, Ottawa, ON.

Tape, T. (2006). *The area under an ROC curve*. Last accessed: August 24, 2006 from <http://gim.unmc.edu/dxtests/roc3.htm>.

Thorsen, P., Schendel, D. E., Desphande, A. D., Vodel, I., Dudley, D. J., and Olsen, J. (2001). Identification of biological/biochemical marker(s) for preterm delivery. *Paediatric and Perinatal Epidemiology*, 15, 90-103.

Tong, Y., Frize, M., and Walker, R. C. (2002). Extending successful predictions of ventilation needs using artificial neural networks from adult to neonatal ICU patients. *IEEE Transaction on Information Technology in Biomedicine*, 6, 188-191.

Tosic, V. (2004). *Service offerings for XML web services and their management applications*. PhD Electrical Engineering, Carleton University, Ottawa, ON.

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

Tsirogiannis, G. L., Frossyniotis, D., Stoitsis, J., Golemati, S., Stafylopatis, A., and Nikita, K. S. (2004). Classification of medical data with a robust multi-level combination scheme. *Proceedings of the International Joint Conference on Neural Networks*, Location, 2483-2487.

Tsujino, K. (1995). Hybrid knowledge acquisition by integrating decision trees and

neural networks. *Proceedings of the IEEE International Conference on Neural Networks*, 1379-1383.

Tu, J.V (1996). Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *J Clin Epidemiol.* 49:1225-31.

Tucker, J., and McGuire, W. (2005). ABC of preterm birth: Epidemiology of preterm birth. *StudentBMJ*. Last accessed: August 24, 2006 from <http://www.studentbmj.com/issues/05/04/education/146.php>.

Turner, M., Budgen, D., and Brereton, P. (2003). Turning software into a service. *IEEE Computer*, 36, 38-44.

Turner, M., Rigby, M., Zhu, F., Kotsiopoulos, I., Russell, M., and Budgen, D. (2004). Using web service technologies to create an information broker: An experience report. *Proceedings of the 26th International Conference on Software Engineering*, 552-561.

USDA Food and Nutrition Service (2006). *Women, Infants and Children*. Last accessed: October 23, 2006 from <http://www.fns.usda.gov/wic/>.

Wang, X., Liang X-X., and Sun, J. (2005). An integrated approach of neural network and decision tree to classification. *Proceedings of the International Conference on Machine Learning and Cybernetics*, 2055-2058.

Weigend, A. S., Rumelhart, D. E., and Huberman, B. A. (1990). Back-propagation, weight-elimination and time series prediction. *Proceedings of the Connectionist Models Summer School*, 875-882.

Wen, S. W. Personal Communication, November 27, 2003. University of Ottawa, Ottawa, ON.

Wikipedia.org. (2006) Divide and conquer algorithm. Last accessed October 30, 2006 from: http://en.wikipedia.org/wiki/Divide_and_conquer_algorithm

Wilson, R. M., Runciman, W. B., and Gibberd, R. W. (1995). The quality in Australian healthcare study. *Medical Journal Australia*, 163, 458-471.

Woods K, Bowyer KW. Generating ROC curves for artificial neural networks. *IEEE Transactions on Medical Imaging*, 1997;16(3):329-337.

Wreder, K., and Deng, Y. (1999). Architecture-centered enterprise system development and integration based on distributed object technology standard. *Proceedings of the 23rd Annual International Computer Software and Applications*

Conference, 250-258.

Yale Medical School (2005). *Guideline elements model*. Last accessed: August 24, 2006 from <http://gem.med.yale.edu/default.htm>.

Yoo, J. H., and Sethi, I. K. (1995). Design of radial basis function networks using decision trees. *Proceedings of the IEEE International Conference on Neural Networks*, 1269-1272.

Zernikow, B., Holtmannspotter, K., and Michel, E. (1999). Predicting length-of-stay in preterm infants. *European Journal Pediatrics*, 158, 59-62.

Zhau, Q. (2001). Evolutionary design of neural network tree - integration of decision tree, neural network and GA. *IEEE Congress on Evolutionary Computation*, 240-244.

Zhou, D., and Frize, M. (2006). Predicting probability of mortality in the neonatal intensive care unit. *Proceedings of the 27th International Conference of the Engineering in Medicine and Biology Society*.

Zhou, D. (2006). Adding probability to neural network prediction of mortality. Master of Applied Science, University of Ottawa, Ottawa, ON.

Zhou, Z., and Xiang, G. (2001). Constructive hybrid decision tree. *Chinese Journal of Computers*, 24, 1057-1063.

Zhu, B. P., Rolfs, R., Nangle, B. E., and Horan, J. M. (1999). Effect of the interval between pregnancies on perinatal outcomes. *New England Journal of Medicine*, 340, 589-594.

Zorman, M., and Kokol, P. (2002). Hybrid NN-DT cascade method for generating decision trees from backpropagation neural networks. *Proceedings of the 9th International Conference on Neural Information Processing*, 2003-2007.