NOTE TO USERS

This reproduction is the best copy available.

UMI[®]

**Development of a Model for Prediction of Repeat Childhood
Injuries in Injured Patients Using Artificial Neural Networks**

By

Yipu Shi

A thesis submitted to

The Faculty of Graduate Studies and Research

In partial fulfillment of the requirements for the degree of

Master of Science

In Information Science

Department of Systems and Computer Engineering

Faculty of Engineering

Carleton University

Ottawa, Ontario

Canada

©September 2004

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: 0-612-97481-2

Our file Notre référence

ISBN: 0-612-97481-2

The author has granted a non-exclusive license allowing the Library and Archives Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

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

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

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

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

Childhood injuries are a major cause of premature mortality and disability and a major source of health care spending in the developed world. The Canadian Hospitals Injury Reporting and Prevention Program (CHIRPP) collect data on child injuries nationwide, providing valuable information to effectively prevent injuries. This thesis aimed to search for models associated with one outcome of childhood injuries, repeat childhood injuries in the CHEO-CHIRPP database, by using artificial neural networks (ANN). A feedforward backpropagation-trained double-layered neural network showed very good prediction ability to the outcome prediction of repeat childhood injuries. Upon further examination of variables included in the ANN model, we were able to eliminate inconsequential variables from the ANN model, and enhance further the predictive capability of the ANN model. The relative importance of the variables in the ANN model was estimated. The developed ANN model may be used for individual outcome analysis and also provide health professionals with useful information to effectively prevent injuries.

ACKNOWLEDGEMENTS

I would like to thank my supervisor, professor, Dr. Monique Frize, who guided me through my entire graduate study at Carleton University. Dr. Frize has been a wonderful supervisor and always offers constructive advice for my studies and research. She always encourages and supports the ideas and activities, making all her students successful. I am sincerely grateful for the opportunity to be part of the MIRG (Medical IDEAS Research Group), which is led by Dr. Frize, to develop my knowledge and skills in this exciting field.

I would also like to thank Dawn Rybchynski and Doaa Swailum for her experience and great help on the neural network techniques.

I appreciated the useful information about the CHEO-CHIRPP database provided by Maureen Brennan-Barnes at the Trauma Program, valuable discussion and opinions offered by Dr. Morag Mackay and Dr. Robin Walker at CHEO.

CONTENTS

ABSTRACT.....	i
ACKNOWLEDGEMENTS.....	ii
CONTENTS.....	iii
LIST OF FIGURES.....	iv
LIST OF TABLES.....	iv
 Chapter 1: Introduction.....	 1
1.1 Motivation.....	
1.1.1 Status of childhood injury in Canada.....	1
1.1.2 About CHIRPP.....	2
1.1.3 Repeat childhood injury.....	3
1.2 Significance of neural network research.....	4
1.2.1 Artificial neural networks applied to medical databases.....	4
1.2.2 Application of neural networks to the CHIRPP database.....	4
1.3 Objectives.....	6
1.3.1 Development of ANN models to predict repeat childhood injuries.....	6
1.3.2 Identification of the important factors leading to the repeat occurrence of childhood injuries.....	7
 Chapter 2: The CHIRPP database.....	 9
2.1 CHIRPP database information.....	9
2.1.1 Coding scheme.....	9
2.1.2 Variables.....	10
2.2 Database handling.....	12
2.2.1 Input variable selection.....	12
2.2.2 Output variable.....	14
 Chapter 3: Artificial neural networks and performance evaluation.....	 16
3.1 Artificial neural networks.....	16
3.1.1 Architecture.....	16
3.1.2 Backpropagation algorithm.....	19
3.1.3 Weight-elimination cost function.....	22
3.1.4 Generalization.....	24
3.2 Performance evaluation.....	26
3.2.1 Contingency table.....	26
3.2.2 Receiver operating characteristic (ROC) curve.....	28
3.3 Selection of important variables.....	29
 Chapter 4: Methodology.....	 33
4.1 ANN design.....	33
4.2 Generation of training and test sets.....	35
4.3 ANN implementation process.....	37
4.3.1 Network calibration.....	37

4.3.2 Automated parameter optimization.....	38
4.3.3 Network stopping criteria.....	40
4.3.4 ANN variable reduction.....	41
Chapter 5. ANN results and Discussion.....	43
5.1 Best performance networks.....	43
5.1.1 Best performance of double-layered ANNs.....	43
5.1.2 Logarithmic-sensitivity index as ANN stopping criterion.....	47
5.2 Variable reduction from the ANN model.....	50
5.3 Identifying the importance of variables leading to repeat childhood injuries.....	55
5.4 Discussion.....	57
5.4.1 Applying artificial neural networks to predict repeat childhood injuries to the CHEO-CHIRPP database.....	58
5.4.2 Variable selection of the ANN model.....	61
5.4.3 Potential of ANN models to the prevention of childhood injuries.....	62
Chapter 6. Conclusions, Contributions and Future work.....	65
6.1 Conclusions.....	65
6.2 Contributions.....	66
6.3 Future work.....	67
REFERENCES.....	69

LIST OF FIGURES

Figure 3.1.1.1: Illustration of an artificial neural network processing unit.....	17
Figure 3.1.1.2: Logistic transfer function and hyperbolic tangent transfer function.....	18
Figure 3.1.1.3: Architecture of a feedforward neural network.....	19
Figure 3.2.2: Comparison of ROC curves.....	29
Figure 5.1.1.1: Graphs of CCR, ASE and ROC curves for the optimal performance of the double-layered ANN in the prediction of repeat childhood injuries.....	45
Figure 5.1.1.2: Graphs of CCR, ASE and ROC curves for the optimal performance of the double-layered ANN with eight-elimination cost function in the prediction of repeat childhood injuries.....	46
Figure 5.1.2: Graphs of CCR, ASE and ROC curves and log-sensitivity index for the optimal performance of the double-layered ANN in the prediction of repeat childhood injuries, log-sensitivity index as ANN stopping criterion.....	49
Figure 5.2.1: Graphs of CCR, ASE and ROC curves for the best performance of the double-layered ANN after variable reduction in the prediction of repeat childhood injuries.....	55

LIST OF TABLES

Table 2.1.2: Initial variable list for thesis experiments.....	11
Table 2.2.1. List of variables presented to ANN.....	14

Table 3.2.1: Truth table.....	28
Table 4.3.2: Comparison of manual and automated network parameter adjusting Ranges.....	40
Table 5.1.1.1: The optimal parameter settings of a double-layered network with and without the weight-elimination cost function.....	44
Table 5.1.1.2: Best performance of the double-layered networks in prediction of repeat childhood injuries.....	44
Table 5.2.1: Weights of input variables for the optimal double-layered ANN.....	50
Table 5.2.2: Performance of the ANNs with each input variable removed.....	52
Table 5.2.3: Performance of ANNs with combined removal of input variables.....	54
Table 5.3.1: Weights of input variables for an optimal double-layered ANN after variable reduction.....	56

CHAPTER 1. INTRODUCTION

1.1 Motivation

1.1.1 Status of childhood injury in Canada

Over the past decade, childhood injuries have emerged as a serious public health concern in Canada. Injuries are a major cause of premature mortality and disability in Canada and some other developed countries in the world [Peclet MH. *et al.* 1990, Tanaka T. 1993, Shanon A. *et al.* 1992]. Each year injuries kill more Canadian children and youth over the age of one than all other causes combined [Injury Reports, 1997].

Childhood injuries not only pose a threat to children's health, but they are also a major source of health care spending in Canada and the United States for children and youth 0 to 21 years of age [Danceco ER. *et al.* 2000]. Statistical studies found that there are 45 hospital admissions and an estimated 1,300 visits to emergency rooms across the country for every injury-related death [CHIRPP publication]. In a study evaluating the cost of childhood unintentional injuries and the value of prevention showed that unintentional childhood injuries in 1996 resulted in an estimated \$ 81 billion in cost in the United States [Miller TR. *et al.* 2000]. In Canada, a Vital Statistical data also showed a total annual cost of \$8.7 billion of health care spending for the injury of children in 1994 [Injury surveillance on-line, 1997].

According to the study of injury surveillance from Health Canada, more than 90% of injuries are predictable and preventable. Injuries are not accidents, and should be prevented by designing effective interventions. It has been well recognized that investing in injury prevention can save money and lives. Several proven child safety

interventions cost less than the medical and other resource costs [Injury surveillance on-line]. Hence, to improve public health and reduce health care spending caused by childhood injuries, further research into all aspects of childhood injuries are helpful to identify the high-risk group, as well as the most common and costly causes of childhood injury.

1.1.2 About CHIRPP

Injuries are a common cause for medical treatment happening in children. Before 1990, there were no systematic collection of data describing the occurrence of child injuries in a population base in Canada and very little was known about the circumstances surrounding injuries of children, or the major factors leading to such injuries. Without such data, injuries arising from new causes could not be identified quickly and it was difficult to evaluate intervention programs for childhood injuries. There has been a great need for a surveillance program to identify the new causes of injuries and to guide the intervention models. The Canadian Hospitals Injury Reporting and Prevention Program (CHIRPP) was created in 1990 for the purposes of providing both a comprehensive picture of various types of injuries occurring in Canadian children and also a firsthand data resource for research on this topic [Mackenzie SG. *et al.* 1999]. CHIRPP is a computerized database on injuries of children (and some adults) seen in the emergency departments of 10 pediatric hospital and 6 general hospitals across Canada. The general hospitals also provide records on adults, but the primary focus of the program is on children and youth 19 years of age and younger. Because of the participation of nation wide of major children's hospitals, the enormous growth of the CHIRPP database in past decade provides health professionals with the potential to

track large amounts of data regarding to substantial and significant childhood injury patient populations. The CHIRPP database can provide invaluable information to researchers who need to know why and to whom injuries are happening and even factors leading to repeat injuries. It was also expected to provide as much information as required by the development of various prevention tools using injuries as the outcome. Thus, the challenge remains to find ways to bring out the information or patterns embedded in the data to develop models that make injury prevention more effective.

1.1.3 Repeat childhood injuries

In a health surveillance program conducted on Canadian children and youth shows that injury mortality rate due to unintentional injury fell by close to two thirds between 1980 and 1996 [Injury surveillance on-line]. This decline is very likely due to the combined effects of prevention of injury-producing events, reduction in injury severity, and improved emergency response and trauma care, which implies that many deaths caused by injuries are preventable if risk factors are effectively identified. These possible injuries related factors have been characterized in a CHIRPP News Letter published on-line for the public.

In a preliminary study used a part of the CHIRPP data, a statistical descriptive study found that more than forty percent of the cases have repeat injuries, that is, for each patient, more than one visit to the emergency room for various types of injuries. Therefore, the question arises as to why some patients have repeat injuries, whereas others not. This motivated a search for features or patterns exist in a group of injured children who have repeat injuries to distinguish with those cases injured only once. These questions will be addressed in this study.

1.2 Significance of artificial neural network research

1.2.1 Artificial neural networks applied to medical databases

Artificial neural networks (ANNs) were first developed in 1959 by Rosenblatt as well as by Widrow and Stearns, and further expanded and popularized by a few of other outstanding researchers in this field [Werbos PJ. 1974, Rumelhart DE. *et al.* 1986]. By far, neural networks are recognized as established analytical tools for biomedical research. Neural networks are briefly described as computer programs consisting of interconnected equations organized in nodes giving a loose association with the animal nervous system. They have nonlinear capabilities to reveal relationships between input variables and an output variable. They are better suited than logistic regression modeling systems for pattern recognition and to predict outcomes when the relationships between the variables are complex, multidimensional, and nonlinear as commonly exist in complex biological systems [Bishop CM. 1995]. Neural networks have achieved successes for a number of disparate clinical diagnosis tasks, for example, diagnosis of myocardial infarction, cancer survival prediction, and mortality in a neonatal intensive care unit (NICU) [Itchhaporia *et al.* 1996, Burke HB. *et al.* 1997, Frize M. *et al.* 2001]. Because of its proven manipulation power for complex data, neural networks are viewed as enhanced tools for extracting outcome-specific information from existing medical databases to human experts for their decision support.

1.2.2 Application of neural networks to the CHIRPP database

In recent years, outcome analysis has emerged as an essential tool for resource allocation, quality assurance and improvement, research into health care delivery, and comparative evaluation of health care centers. Multiple models exist to assist with prediction of the outcome of injured adult patients [Senkowski CK. *et al.* 1999]. Compared to that, there are limited studies concerning the developed predictive models for the evaluation of injured children and adolescents. Moreover, most adult predictive models are logistic regression statistical models, which do not handle complex medical data well if nonlinear relationships exist. Recently, there appeared a study describing models to predict survival in pediatric trauma patients, derived by approaches of neural networks and logistic regression. The two different models were compared and concluded the neural network model better than that of logistic regression [DiRusso SM. *et al.* 2002].

Medical decision-making is based on recognizing complex patterns in the existing information from patients. Neural networks have been shown to be very effective and powerful for this type of pattern recognition. They have repeatedly outperformed traditional methods for pattern recognition and clinical outcome prediction, and have demonstrated an exceptional pattern-recognition ability and great potential in medical outcome prediction. Because of their non-linear modeling capability and very least requirements to the input data, which means input data is not necessarily explicitly defined, neural networks have been widely applied to non-linear problems and are a natural choice for modeling large and complex databases of medical information. In MIRG (The Medical Intelligent Research Group at Carleton University

and University of Ottawa), ANNs have been developed and applied to estimate various clinical outcomes including adult and neonatal ICUs, cardiac surgery risk stratification.

In this study, a part of the CHIRPP data, the CHEO-CHIRPP database was analyzed for patterns involved in the repeat occurrence of childhood injuries using the artificial neural networks. Since the CHIRPP was created in 1990, there have been a number of subsets of the database, each based on the injury data collected in various local regions, have been researched for important patterns that may have particular implications to the development of injury prevention programs. They were mostly statistical descriptive studies and no models have been developed regarding the connections of particular patterns of injuries with a particular injury outcome in the CHIRPP data [Bienefeld M. *et al.* 1997, Pickett W. *et al.* 2000]. In recent studies, there have been some childhood injury related models developed to predict pediatric trauma mortality and injury severity by using a distinct database known as the National Pediatric Trauma Registry in the United States [Tepas JJ. 3rd *et al.* 1997, DiRusso SM. *et al.* 2002]. The studies compared approaches of statistical logistic regression and artificial neural networks and concluded that the ANNs were superior to logistic regression [DiRusso SM. *et al.* 2002].

Due to the outstanding features that neural networks proved in pattern recognition of data, this study will take advantage of this popular technique to extend the studies on the subject of childhood injuries with the CHIRPP data to the development of models capable of identifying those patients to have injuries repeated occurred in future.

1.3 Objective

1.3.1 Development of ANN models to predict repeat childhood injuries

First, a model was developed to distinguish patients who had repeat injuries with those patients who do not. The model development and evaluation was accomplished by the approach of artificial neural networks. Our ultimate goal in this category is to train a neural network and then evaluate it, so that it can be used to classify a new individual injured patient on whether a future injury will happen or not. These patients can then be alerted to this possibility, and even be presented to injury prevention programs.

This study's approach uses a feedforward artificial neural network trained with the backpropagation learning algorithm, in combination with or without the weight-elimination cost function to predict the patients with repeat injuries by using the data from the CHEO (the Children's Hospital of Eastern Ontario)-CHIRPP, a subset of the CHIRPP database. The database contains total 1,25,493 cases of child injury data collected between 1990 and 2001 years at CHEO. Various types of ANN architectures were examined and justified for best performance in terms of the combined evaluation criteria. This goal was explicitly implemented using the single-layered or double-layered networks, each combined with or without the weight-elimination cost function.

1.3.2 Identification of the important factors leading to repeat childhood injuries

As a second objective, the largest network weights at the point of optimal performance of the ANN were extracted to identify the variables that have the most influence on the predictions of the outcome. Alternatively, a systematic variable reduction method was also used to eliminate unimportant variables to further justify the correctness of selected variables for the model. This objective aimed to make the model more effective in the prediction of repeat childhood injuries.

We believe that ANNs are an effective tool in model development and outcome prediction, though, determination of the effectiveness of ANN models are not trivial. It is also equally important to associate the predictive results with the data collection process. Because sometimes limited predictive ability may be the results of ineffective factors collected in the data or the neglect of unrecognized factors, or unknown knowledge. The above objectives were addressed and extensively discussed in the thesis.

CHAPTER 2. CHIRPP DATABASE

2.1 CHIRPP database information

2.1.1 Coding scheme

The first version of the CHIRPP database was created in 1990. The main CHIRPP database was transferred and recoded from this previous software version to a custom designed Oracle system that runs on an HP-755 RISC workstation in 1997. The newly applied coding scheme, taking advantage of the hierarchical structure developed in the new version, was able to reduce great number of codes used in representing wide range of values for some of the variables in the data. This simplifies the data for analyzing without losing information. The benefits of this strategy can be typically represented by variable “Mechanism 2”, which means factors involved in injuries, whose codes reduced from 1186 to 728. Another example of the same strategy, the addition of a second variable to the variable such as “Location”, the place of injury occurred, with the variable “Area”, gives more detailed specification about where injuries taking place. This makes the information to be described more explicit to its maximum extent that is preferred by the neural network studies. This hierarchical coding scheme can be seen in a few of the other variables in the database [Mackenzie SG. *et al.* 1999].

In the original CHEO-CHIRPP database, values of most of the variables were coded in numbers, although free text accompanied with each of the variables for reference when needed. Most of the variables in the database are qualitative variables, also referred to as categorical, or discrete variables. In this category, nominal variables classify data into categories labeled by numbers. In contrast, ordinal variables rank the data in terms of degree. For example, a variable “Mechanism 2” describes factors

involved in injuries can have as many as 728 values that were represented by numbers. Therefore, this is a nominal variable. A special case in this category is when there are only two categories, such as variable “follow-up” with two values “permitted” or “refused” which means a follow-up after injury is accepted or refused by the patient. This nominal variable is simply represented by a binary code as 0 and 1, or –1 and 1. As for the example of ordinal variable, the variable “Disposition” defines the severity of the injury which is coded from 1 to 8 in increasing order for the degree of severity of the injury.

Another main category of data type is continuous variables which are the actual values measured. Only a small number of variables belong to this category in the CHIRPP database. For example, a patient’s age is coded exactly by the actual number of months when the injury occurred.

Three free text variables describe how the injury happened, for example, name of playground or address, and details about products such as brand name or model number. Text searches facilitate the retrieval of records that involve items or circumstances for which no specific code exists.

2.1.2 Variables

The CHEO-CHIRPP is part of the CHIRPP system; its scope is limited to data collected in the children’s hospital of eastern Ontario in Ottawa. The data was collected on all patients presenting to the CHEO emergency room or admitted to the hospital for injury or poisoning, verified and coded by a nurse coordinator from medical charts. In addition, the patients or an accompanying adult were asked to provide a description of the circumstances of the injury. After each interview, a summary of the information

provided is recorded in a specially designed form. There are a total of 125,493 records in the CHEO-CHIRPP database obtained for this study by August 2001.

Table 2.1.2: Initial variable list for thesis experiments

Variable	Description	Type
patient_	patient No.	N
name_	patient's name	T
postal_c	home address	T
age	patient's age	C
gender_c	patient's sex	B
ethnicit	ethnicity	N
followup	followup after injuries	B
injury_e	text - detailed description	T
year_of	year of injury occurred	O
month_o	month of injury occurred	O
day_of_w	day of injury occurred	O
injury_t	time of injury occurred	O
er_visit	ER visit time	T
location	the place when injury occurred	N
area_cod	specific part of the place when injury occurred	N
context_	activity of the person at the time of injury	N
intent_c	intentional injury	N
injury_g	incident type	N
breakdown	how things went wrong and led to the injury	N
mechani	describes the type of energy transfer	N
vehseatp	where seated in vehicle related injury	N
safety_d	safety equipment in use when the injury occurred	N
work_inj	work-related injury	B
industry	type of industry involved	N
occupati	occupation	N
breakd2	an item whose failure led to the injury	N
mechan2	factor directly caused the injury	N
contribu	factor involved in the injury that was not the direct cause	N
noi_code	type of injury suffered by the injured person	N
body_par	body part affected by the injury	N
disposit	how a person was treated, an indicator of the severity of the injury	O

* N = nominal variable, B = binary variable, O = ordinal variable, C = continuous variable, T = text variable

To better understand the CHEO-CHIRPP database, all variables that have been clearly defined and completely coded are listed in Table 2.1.2. It contains total 31

variables that were initially considered for experimentation. The database originally contained the following information: Age and sex distribution of injured persons which listed by patient's hospital number, name, post code and patient's age group, sex, home language; patterns of injury were described by variables that each represent a time factor; year, month, day, time; location (the place where the person was when the injury occurred), area (the specific part of the place where the person was when the injury occurred), injury group (facilitates the identification of certain types of injuries); circumstances of injury that include breakdown events (how things went wrong and led to the injury), mechanism of injury (the type of energy transfer or lack of vital substance that caused injury), context (the activity of the person at the time of injury); physician's injury summary which includes variables describing nature of the injury (the type of injury suffered by the injured person), body part injured and disposition (treatment applied).

2.2 Database handling

The CHEO-CHIRPP data were first received and kept in the form of Microsoft Access. It was converted to an SPSS file for data preprocessing because preprocessing of all input and output variables is the vital step for running ANN experiments. SPSS is a comprehensive statistical software designed to handle all steps in an analysis ranging from planning, data collecting, data access, data management and preparation, statistical analysis, reporting and deployment. Moreover, SPSS has a very nice point-and-click interface to make the preparation of data for ANNs much easier.

2.2.1 Input variables selection

For a successful ANN application, selection of input variables is extremely important. A neural network is only as good as the input data used to train it. If important data inputs are missing, then the effect on the neural network's performance can be significant. Therefore, developing a workable neural network application relies considerably on the solid understanding of the problem domain to ensure any related factors are included.

The first step to determine the input variables to include in the ANN experiments was to do simple statistics to delete variables missing more than forty percent of the patient data. Variables to be selected confirmed with clinical experts to insure they are of interest to health professionals from their understanding of the potential connection of various aspects of childhood injuries. It was also suggested to delete variables that only specify certain types of injury, or for circumstances only related to that type of injury; for example, the variable “seat position” referring to passenger seat position, is only valid when a vehicle accident is the type of injury, though for most of the cases this is not applicable. It was also suggested to exclude two variables in the original database such as the variable “injury group” which is used to group the same type of injuries together to facilitate quick identification of certain types of injuries. Because of the hierarchical coding scheme, a smaller value range for some of the variables at the higher level of hierarchical structure are not likely to give detailed information usually required to train a neural network. Another variable, “intent”, which indicates whether an injury is intentional or not, was also left out from the ANN analysis as suggested by the experts in child injury prevention for the same reason. The selected variables that were deemed important and relevant to childhood injuries and included in the ANN analysis are listed in the Table 2.2.1 Note the selection of input

variables for the ANN at this stage of experiments was focusing on including the important and relevant variables as input data. This stage of variable selection is different from what will be done to eliminate the unimportant variables from the models after the neural networks have been trained.

Table 2.2.1. List of variables presented to the ANN

Inputs	Variables	Description	Type*
1	age	patient's age represent in month	C
2	gender_c	patient's gender	B
3	followup	followup after injuries	B
4	month_o	month of injury happened	O
5	day_of_w	day of injury happened	O
6	injury_t	time of injury happened	O
7	location	the place when injury occurred	N
8	area_cod	specific part of the place when injury occurred	N
9	context_	activity of the person at the time of injury	N
10	breakdown	how things went wrong and led to the injury	N
11	mechan2	one directly caused the injury	N
12	noi_code	type of injury suffered by the injured person	N
13	body_par	body part affected by the injury	N
14	disposit	how a person was treated, an indicator of the severity of the injury	O
Output	repeat_inj	whether injury occurred in a person twice and more	B

* C = continuous variable, O = ordinal variable, B = binary variable, N = nominal variable

2.2.2 Output variable (repeat injury)

Due to lack of variables in the CHEO-CHIRPP database indicating whether a patient suffered repeat injuries or not, such a variable needs to be created discriminating the presence of repeat injury or not. This variable was used as the outcome for a supervised learning process in this study, that is, as a dependent variable to be analyzed by the neural networks. The separation of the database into repeat injury group and non-repeat injury group was accomplished by using the unique hospital patient number. Then injuries that occurred for the same patient can be searched using the unique patient number. Patients where an injury occurred only once were regarded as non-repeat group, and those who had two or more injury records were labeled in the repeat-group.

The separation was simply implemented with a small computer program in Perl. The repeating cases were coded as 1 and non-repeat cases coded as 0. As shown in Table 2.2.1, this variable is dichotomous and this characteristic of the dependent variable resembles a probability estimate problem frequently seen in medical risk prediction models that can be handled by statistical analysis of logistic regression, as well as a nonlinear sigmoid function used with neural networks.

After the database has been separated into two groups, it was found that 54,786 records were repeat injury cases (43.66%) and 70,707 cases in non-repeat injury group (56.34%), respectively.

CHAPTER 3. ARTIFICIAL NEURAL NETWORKS and PERFORMANCE EVALUATION

3.1 Artificial neural networks

3.1.1 Architecture

In this study, we proposed the application of artificial neural networks (ANNs) to predict repeat childhood injuries so that potential models could be developed for prevention. Neural networks have been the subjects of an active field of research that has matured greatly over the past 40 years. The capability of neural networks was expended from linear to nonlinear in 1974 by Warbos, with the intention of developing a method as general as linear regression, but with nonlinear capabilities. ANNs were then popularized in the mid 1980's by Rumelhart and McClelland, and since then have been applied a broad spectrum in the medical field including medical decision support [Penny W. *et al.* 1996]. Some studies have compared ANNs with human expert opinions where ANNs' diagnostic capabilities appeared to exceed that of the experts [Baxt WG. 1995]. The advantages that ANNs offer over statistical multivariate analysis is that the models do not have to be explicitly defined prior to the running of the ANNs, whereas prior knowledge of relationships between independent factors are required before a statistical model can be developed [Blum A. 1992, Livingstone DJ. *et al.* 1997].

ANNs are computer programs that mimic the nervous system of the human brain, and the structure and function of ANNs simulate the interconnected networks of the brain. ANNs are models used to emulate the human pattern recognition function through a similar parallel processing structure of multiple inputs [Penny W. *et al* 1996]. The model is designed to capture the causal relationships between dependent and

independent variables in a given data set. Thus, ANNs have the ability to scan the data for patterns and in particular can be used to construct nonlinear models.

Artificial neurons are nodes in an artificial neural network, and these nodes are processing units as illustrated in Figure 3.1.1.1 [Fundamentals of neural networks].

Figure 3.1.1.1: Illustration of processing unit of an artificial neural network

The nodes receive information from the inputs and each input is multiplied by a corresponding weight, known as synaptic connection. The multiplied values are then summed in each processing unit, and fed through a transfer function to produce the output value between 0 and 1 or -1 and 1 depending on the transfer function used. In ANNs, these internal weights are adjusted during a “training” process, whereby input data along with corresponding known output values are submitted to the network for processing repetitively and, on each repetition, the weights are adjusted incrementally to bring the network’s output closer to the desired values [Dayhoff JE. *et al.* 2001].

The most popular transfer function is the sigmoid function that produces output values from 0 to 1 and values above a threshold (for example 0.5) represent a “Yes” class and otherwise a “No” class of output, a type of values often seen in medical outcomes. The sigmoid function is thus commonly used in the setting of medical outcome prediction in which outputs are often presented in the range from 0 to 1. The

other transfer function, the hyperbolic tangent, or tan-sigmoid function, differs from the former in its sharper shape of transfer that allows faster learning of ANN. This function is considered to be a good trade-off for ANNs, where speed is an important aspect of computation. Furthermore, as the output of a tan-sigmoid function ranges from -1 to 1 with transfer curve across zero, ANNs can learn even if some features are absent from inputs, which makes learning more accurate. The comparison of the two transfer functions is depicted in Figure 3.1.1.2.

Figure 3.1.1.2: Logistic transfer function and hyperbolic tangent transfer function

By far the most popular architecture of neural networks is the multiplayer perceptron (MLP) type of network, as it is applied successfully to many pattern classification problems, comprising an input, one or more hidden, and an output layer of nodes as shown in Figure 3.1.1.3. The nodes are typically organized into layers, and each layer is comprised of a number of nodes corresponding to multiple levels of inputs, which means the output in one layer can be the input for the next layer. Depending on the complexity of the database, the network may or may not require hidden layers. An ANN with no hidden layers is a single-layered ANN, and a double-layered ANN has one hidden layer. The majority of studies utilize double-layered neural networks in

which all the layers are fully interconnected through a hidden layer. This means that each node in one layer is connected to all nodes in the next layer. For a single-layered ANN, its prediction power is similar to the generalized statistical linear model. Theoretically, for ANNs with hidden layers, increasing the number of layers of the ANN structure increases the capability of the ANN to handle more complex input data [Penny W. *et al.* 1996]. In practice, there is no substantial evidence that more than one hidden layer adds to the predictive capability of a network [Penny W. *et al.* 1996]. Moreover, the training of the network would become much complicated if more hidden layers were used, as more parameters need to be adjusted. Therefore, the majority of medical applications prefer to use double-layered networks.

Figure 3.1.1.3: Architecture of a feedforward neural network

3.1.2 Backpropagation Algorithm

The neural network requires training to reach its optimal performance and attain the appropriate parameter values. The network is trained by exposing it to a training data set, which provides the inputs and desired outputs to the network. Weights of networks are adjusted until appropriate output results are the desired outputs. The most commonly used training algorithm, backpropagation algorithm, also known as gradient

descent learning, and the generalized delta rule, was first introduced by Rumelhart *et al* in 1986, and since then a number of computational enhancements have been developed to improve the algorithms [Mehrotra K. *et al.* 1997, Press WH. *et al.* 1988].

Backpropagation is a supervised learning technique with the objective of training the network to transform inputs to a desired output. First, the data is presented to the input layer, the network begins its first training epoch with a set of arbitrary weights assigned to each input nodes, and these are modified in successive iterations by a process of “back propagation”. The network output is compared with the desired output, and differences are fed backward through the system and adjusted so as to minimize the mean square output error. This training process continues until further increments in accuracy are no longer achieved. Adjustment of weights in training is completed by the backpropagation algorithm. This algorithm is correct for any number of layers and function used to compute the outputs of the nodes from their inputs [Werbos PJ. 1974, Rumelhart DE. *et al.* 1986].

In the algorithm, the error represents the discrepancy between the network’s desired output and predicted output for a training set is treated as a function. The weights of the network need to be adjusted along with training. An error leads to a weight to be modified and allows the network to learn by changing its connection weights. The amount of error in a network at a moment can be expressed by the sum of squared errors cost function, which is described as follows:

$$SSE = \sum (\text{desired}_i - \text{predicted}_i)^2$$

In which desired_i is the desired output at time i of entire training period, and predicted_i is the predicted output value after networks having been trained at the same time point.

For an epoch or an iteration, each cycle of training pattern presented to the network exactly once, the output error is propagated backwards to the input layer to adjust weights in each node of the network. This process goes on until the total errors at a minimum by meeting certain stopping criteria. At the end of the training process, the finally adjusted weights represent the final values of the network parameters for a given dataset. Mathematical details were given by [Mehrotra K. *et al.* 1997] and can be found in their studies.

There are also some strategies that have been added to the gradient descent algorithm to make the training more accurate and effective [Mehrotra K. *et al.* 1997, Press WH. *et al.* 1988]. One of the concerns of the backpropagation algorithm to the network is the possibility of a training ending up on a local minimum of the mean squared errors, as opposed to the global minimum. Some techniques, such as using genetic algorithms for network training are developed to find a global rather than a local minimum [Fahlman SE. 1988]. It was also advised to train networks a number of times from different initial weights. [Penny W. *et al.* 1996] The learning rate parameter determines how much the weight should be adjusted. Although a higher learning rate speeds up the training process, it is recommended to start with relatively small values to be able to converge to a global minimum [Penny W. *et al.* 1996]. Another strategy for improvement of the algorithm is the use of momentum. The idea is to add a proportion of the previous weight-change value to the new value to prevent the network from getting caught in local minimum [Penny W. *et al.* 1996, Press WH. 1988].

The time for training a network is greatly related to the number of input variables and training samples. It increases exponentially with the number of inputs and increases polynomially with the size of the training samples. Some techniques are good

for speeding the network's training computation time, as well as provide advantages in the search for a minimal point in the training of networks [Fahlman SE. 1988].

3.1.3 Weight-elimination cost function

Once a network has been successfully trained, it should be capable of predicting the output for the problems in the same domain. However, in some circumstances, training samples may contain features or information that are not present in the broader population of cases. It is easy to create problems of overfitting which means networks are too specific to the training data to extend to the prediction of general cases. One way to solve the problem of overfitting is to add a complexity term to the cost function. Two cost functions, weight-decay and weight-elimination, can be used to reduce memorization [Weigend AS. *et al.* 1990]. Weight-decay adds a penalty term to the cost function. Weight-decay limits the size of the connection weights, and penalizes large weights. The weight-decay penalty term causes the weights to converge to smaller absolute values than they otherwise would. According to [Bartlett PL. 1997], the size of the weights is more important than the number of weights in determining generalization. In addition, large weights can cause excessive variance of the output [Geman S. *et al.* 1992]. Therefore, the adoption of weight-decay can make the network more stable since the output variance can be reduced by the addition of this small term into the cost function formula. Weight-decay is contained in the weight-elimination algorithm. Weight elimination [Weigend AS. *et al.* 1991] uses the following formula besides the main mean squared error function.

$$\lambda \sum (w_i/w_0)^2 / [(w_i/w_0)^2 + 1]$$

where w_i is the individual weight at the moment of time i and λ is a user-specified weight-decay constant. While weight decay using the sum of squared weights tends to

shrink the large weights more than the small ones, weight elimination tends to shrink the small weights more down to zero. Weight-elimination is more suitable for network pruning by eliminating variables that have little impact on the network's prediction capability as unrelated variables may add unnecessary noisy to the training of the networks and cause the generating models less effective.

The generalization ability of the network can depend crucially on the decay constant λ , which means network is sensitive to the value, especially with a small training set. One approach to choosing the decay constant λ is to train several networks with different amounts of decay and estimate the generalization error for each; then we choose the decay constant that minimizes the estimated generalization error using cross-validation. During the training process, the decay constant λ can be iteratively updated [Weigend AS. *et al.* 1990]. Although the approach is well defined and practically effective, the disadvantage of this approach is time consuming. Rognvaldsson TS. introduced a more simple method for estimating the weight decay parameter by combining with an early stopping method, which is also the most common method to shorten training time to avoid overfitting. Empirical studies show that the weight-decay method, merging with an early stopping, could provide good weight-decay parameters [Rognvaldsson TS. 1998].

A fundamental problem with weight-decay is that different types of weights in the network will usually require different decay constants for good generalization. Therefore, in a popular double-layered network, it is recommended that at least three different decay constants for input-to-hidden, hidden-to-hidden, and hidden-to-output weights. While the best estimated generalization errors might be achieved by adjusting all these decay constants, vast amounts of computation are required. Therefore, there are

still, theoretically and practically, problems with this approach. The weight-elimination technique has been extensively explored in our research group for its potential enhancement of the neural network performance [Frize M. *et al.* 2001]. The ANN experiments carried out in this thesis will alternatively apply the weight-elimination technique to enhance the performance of the network.

3.1.4 Generalization

Generalization is the ability of the trained neural network to perform an appropriate prediction task on unseen data. To assess how well a trained network will perform on new data, the original dataset is usually divided into two sets without overlapping: a training set and a test set.

The performance that a network achieves on its training set may not necessarily appear on the test set. We have known that overfitting prevents a network from good generalization. The overfitting problem mostly occurs when the complexity of the trained network exceeds the patterns actually included in the entire dataset or the function being estimated. The increased complexity of the network can be the result of applying complex network architecture such as more layers, more hidden nodes or more input variables. On the other hand, feeding a training network with sufficient samples that reflect the true characteristics of the population of interest is also crucial. Therefore, care needs to be taken not only on having an optimal sample size but also on unbiased sampling, or a random selection of samples.

There are no well-established guidelines when selecting a particular network structure for specific data except some practical rules from a large amount of empirical studies. The number of hidden layers, hidden nodes in a hidden layer, the pattern of

neuron connections and the types of transfer functions are mainly determined by trial and error [Itchhaporia D. *et al.* 1996].

The majority of the decision-making applications with neural network use three-layer networks. It is said that this type of network is capable of approximating any nonlinear function [Penny W. *et al.* 1996]. Therefore, when we design a network with a hidden layer, we only have to decide how many nodes to choose in the hidden layer. Taking advantage of this powerful computational property, the training of the network dramatically speeds up. A theoretical study shows that the relationship of number of hidden nodes with the number of training samples is likely to influence network generalization performance [Baum EB. *et al.* 1989]. [Livingstone DJ. *et al.* 1997] through an empirical study, suggests the following formula to estimate the training samples required in practice.

$$D = mo/w$$

where m is the number of training samples, o is the number of network outputs, and w is the total number of network weights. If D is greater than 3, the network is ensured to have good generalization and to avoid memorization of the training set.

Early stopping is a commonly used approach for getting the best performance of the network, thus no occurrence of overfitting of training set. One of the difficulties in the construction of ANN models is the choice of an appropriate stopping point for training, since it is always possible to continue to reduce the error in a neural network by repeated adjustment of the network weights. This will result in a relative specialized model which would have difficulties in working with generalized data. Early stopping means that training is stopped when the test set error begins to rise. A vast amount of applications of this technique examined on the real data sets showed that the use of the

early stopping criteria appeared to prevent overfitting, and trained networks performed well in prediction on unseen data in the same domain. This approach seems to avoid the problems of both overtraining and overfitting [Livingstone DJ. *et al.* 1997].

3.2 Performance evaluation

The main emphasis in neural networks is classification and prediction accuracy. Thus, the main objective with neural networks is to provide satisfactory results in prediction tests. The majority of medical applications are classification problems. Therefore a network's performance is measured by how well it discriminates between different classes. The outcomes most often seen in medical decision are set in two distinct categories, that is, 1 represent for "Yes" and 0 for "No". If this scheme is used, the output of a network can be interpreted as probability of the occurrence of one outcome to be predicted. There exist many different performance measurements for neural networks. The network's overall classification rate along with sensitivity and specificity on test sets are commonly accepted measures of network performance, as well as the constant predictor and the area under the ROC curves. They will be used as the major evaluation criteria in this thesis.

3.2.1 Contingency table

Though correct classification rate is suitable for network evaluation, it may cause problems if used on its own. It should be quoted in conjunction with the performance of a priori probability, a constant predictor that classifies all cases belonging to the class with the greatest a priori probability [Frize M. *et al.* 2001]. A constant predictor means to always choose the class having the maximum *a priori* probability. This is known as the no-data rule or the default rule [Penny W. *et al.* 1996].

These precautions are necessary because in many medical applications a *priori* probabilities of different classes are usually quite uneven. In the CHEO-CHIRPP database, we have one output including two distinct classes, in which the non-repeat injury children class has more prevalence than repeat injury children class. A priori probability is 54.1 percent, which means a constant predictor would classify all patients in the test set as non-repeated cases, and the correct classification rate would be 54.1 percent as it correctly classifies 54.1 percent of all cases. In terms of a *priori* probability of 54.1 percent in the output, we would like to think that there is no much prevalence for one outcome in the sample data over the other one.

A truth table, shown in Table 3.2.1, is helpful for understanding the definitions of the correct classification rate (CCR), sensitivity and specificity. The correct classification rate identifies the percentage of correctly classified classes. It is calculated by the sum of the number of cases that were correctly classified into their respective known classes out of all cases. Based on the table, CCR is illustrated by the sum of true positive (TP) and true negative (TN) divided by the total number of all cases (TP+TN) in the dataset. The network's sensitivity identifies the rate at which a subject who had repeat injuries is classified as repeat injury case. From the table, the sensitivity is the rate of true positive divided by all known positive cases in the output data. The formula for sensitivity is $TP/(FN+TP)$ where FN represents the number of false negatives. Specificity is the rate at which a patient who had no repeat injury is classified as no-repeat injury case, and can be calculated using the formula as: $TN/(TN+FP)$ where FP is the number of false positives. All these parameters measuring the performance of networks depend on the prevalence of the outcomes. They are considered effective

means of measuring network performance in the data where no apparent prevalence of one outcome over another.

Table 3.2.1: Truth table

Predict output	Known output			
		Yes	No	
	Yes	True negative(TN)	False negative(FN)	TN+FN
	No	False positive(FP)	True positive(TP)	FP+TP
		TN+FP	FN+TP	Total cases

3.2.2 Receiver operating characteristic (ROC) curve

The area under the receiver operating curve (ROC curve) assesses how well the model discriminates between two outcomes. ROC curve originally was developed for dichotomous outcomes in which an event either was or was not a member of the positive class. It later extended to accept continuous values ranging in the 0-1 interval as well as 0 and 1 values [DeLeo JM. *et al.* 1995] The ROC curve is a visual description of the operating points and potential trade-off between the true and false positive rates by varying the threshold value of the output across its range of values [Woods K. *et al.* 1997]. It is important to note that ROC curves are useful only in the presence of a general underlying monotonic relation between independent variables and dependent variables [Dayhoff JE, 1990].

The ROC curve is a plot of a model's sensitivity versus one minus its specificity. A typical ROC curve is illustrated in Figure 3.2.2. A perfect ROC curve would look like a step function representing how well the model separates two output classes. Its maximum area is 1, and a 45-degree diagonal line indicates the uselessness of the model. Thus, the closer the ROC curve resembles a step function, and the value of area is close to 1, the better the model.

Figure 3.2.2: Comparison of ROC curves

3.3 Selection of important variables

Once a network has been trained and tested at the point of its optimal performance, the model may be further validated as a candidate of an established classifier for predicting this kind of data. We are also interested in examining how much each individual input contributes to the model at its maximal performance. Since not all the variables played an equivalent role in the model, the identification of those important variables can be meaningful to medical experts.

The weight-elimination cost function technique has been demonstrated as an effective approach in extracting the most important input variables by reducing the weights of less influential variables down to zero [Frize M. *et al.* 2001]. The CHIRPP database was originally designed to give an overall picture of child injury status in the country, thus includes as much detailed information as possible for various research

purposes. Since there have been no previous existing studies about model development on outcomes related to the subject of child injuries based on this database, the decisions about which of the variables should be used as inputs, or left out from ANN analysis are determined in consultation with experts of childhood injury prevention.

Before introducing many variables to the ANN experiments, one should hypothetically reduce the input variable set to the most significant factors suspected to affect the outcome. The selection of input variables is important as the performance of a network can be improved by eliminating unnecessary input variables [Bishop CM, 1995]. This benefits the training of the neural networks in two ways. First, the selected variables ensure the neural network has been given useful information at a reasonable level of detail; and second, by smoothing the data, the noise entering the network can be reduced. This is important because noise can obscure the underlying relationships within input data from the network, as it must concentrate on interpreting the noise component. The only disadvantage is that worthwhile information might be lost in an effort to reduce the noise, but this tradeoff always exists when attempting to smooth noisy data. Thus, the extensive expert opinions are necessary for high quality variable selection.

The variables have to be a general representation that applies to the majority of the sample cases. There were 33 variables in the original database in which a few of showed to be too specific and applied to very few of the cases. For example, a variable representing whether an injury is work-related or not, might be useful in another study, but is not applicable to most of the cases in this study. Also, a variable recording the seat position of an injured child when a vehicle accident occurs has nothing to do with most cases where no vehicle is involved. Nevertheless, they could be very important

factors to be considered in a vehicle-related injury study. Therefore, with expert's opinion and the pre-examination of input variables, the training of the network can begin with more outcome-oriented meaningful variables, thereby save much computation time.

Also, the use of double-layered networks with a number of hidden nodes enhances the complexity of the model that is capable of dealing with complex non-linear data. But the interpretation of the ANN results becomes more difficult. There are two levels of weights for a double-layered network and the weights obtained at the input layer are considered for ranking the importance of variables. Small weights on a node may be forced to zero by the weight-elimination cost function, but a variable cannot be seen as unimportant unless the weights of all of the nodes attached to that variable are reduced to zero. A variable can be deemed unimportant if the weights of all of the nodes are near zero. For a single-layered network with the weight-elimination function, the interpretation of input weights is straightforward as there is only one weight per input variable. They directly reflect the strength of role played by each of variables in the model.

Warner emphasized another issue about variable selection, suggesting that if more detailed information is contained in the data, they should be included in the ANN analysis. Compared to other types of variables, binary predictors that have only two values are most common in medical data. The network will not be able to learn the effects of the predictor on the outcome ranging between these two values. The neural network can be restricted to a simpler model than would be possible if the predictor variables take on a range of values. Because of the hierarchical structure of the database, only the lower level of variables at the structure that contain more detailed injury-

related information are chosen for the ANN, whereas summarized variables at the higher level of the structure are discarded from the ANN analysis. For example, we would rather use the “Age” variable representing exactly the age of patients when injury occurred to them than an “Age group” variable that groups a range of ages at predetermined values. If there is room for predictions to improve, it is worth considering the appropriate combination of selected variables with desired models.

Since direct analysis of the weights produced by networks with a hidden layer is difficult, a simpler approach is to systematically remove each of the variables one at a time to see if the performance of the network changed compares with the intact data set [Qian N. *et al.* 1988]. Variables can be ranked for their contribution to the model. [Delaurentis M. *et al.* 1994] explored the interaction between predictive variables and survival rate on censored data. They concluded that the network was not a black box but could lead to useful insights into the roles played by different prognostic variables in determining patient outcome. In practice, one common method referred to as leave-one-out method is to rank inputs according to how much the testing error increases when the individual input is not used in the training or testing of the network. This method has been used successfully to find which factors were predictive of admission from a network trained on data from psychiatric emergency room [Somoza E. *et al.* 1993]. One drawback of this method is the requirement of considerable computation. For a network with a high number of inputs, a large number of experiments are needed for the training and testing.

CHAPTER 4. METHODOLOGY

4.1 Our ANN design

The effectiveness of ANNs in the prediction of many clinical outcomes has been enormously evidenced in the literature and, in practice, have been used to aide physicians for decision making in clinical settings. Besides, ANNs, as effective prediction tools, have been developed in the MIRG to predict various clinical outcomes. MIRG is interested in seeing if our ANNs can be applied to a variety of databases. The CHIRPP database is a large-scale and multivariate data set dominated by various categorical variables, it would be interested to see whether our ANNs are able to reveal potential complex relationships in the data. Also, predicting whether a patient will have a repeat injury or not, is a dichotomous output and a good fit to use our ANN tools. Thus, the analysis of the CHEO-CHIRPP database will be carried out by the approach of neural networks in this study.

There are a number of neural network architectures that can be chosen for the study of the current database. Many of the studies compare various types of network architectures and conclude that the best-performing network is the feedforward ANN trained with the backpropagation algorithm [Ennett CM. *et al.* 2004]. So it will be chosen as our network architecture.

Both single-layered and double-layered backpropagation networks can be used with the weight-elimination cost function. There are a number of benefits of using this strategy as discussed on Chapter 3. First, it can be an alternative method of reducing variable by forcing the already small weights to zero. The model's complexity can be reduced when the number of variables is decreased. This effectively prevents overfitting and increasing the model's generalization to new data. Secondly, it has been previously

documented that the inclusion of irrelevant input data confuses the network and therefore reduces its performance [Frize M. *et al.* 1995, Trigg HCE. *et al.* 1997, Somoza E. *et al.* 1993]. Besides, the backpropagated trained networks in conjunction with or without the weight-eliminated cost function has been developed and validated using various clinical data sets by MIRG [Trigg HCE. *et al.* 1997, Frize M. *et al.*, 1998, Ennet CM. *et al.*, 2004].

The double-layered network architecture was mainly adopted in the model's training, though comparisons on the network's performance to other structures of network were also conducted. A double-layered neural network with a single hidden layer has demonstrated its great power to estimate any nonlinear relationships and is substantially used in this study. Besides, a single-layered network was designed for experiments as it is more appropriate to read simple relationships between inputs and output. With single-layered networks, not only the computation time is greatly decreased, but the interpretation of the network weights is straightforward.

The ANNs developed by MIRG use the MATLAB as platform and were described by [Frize M. *et al.* 1998] and [Ennett CM. *et al.* 2004]. This tool incorporates several useful features that allow users to modify the experiments. The choice for the architectures of the networks can easily be implemented by switching from single-layered to double-layered and verse versa, and the number of hidden nodes can be adjusted easily as well. The weight-elimination cost function can be chosen between "ON" or "OFF". The ANN tools also calculate the *a priori* statistics for the training and test set, and present results visually through a graph construction of correct classification rate, average square errors and ROC curves.

4.2 Generation of training and test sets

The CHEO-CHIRPP database with selected variables ready for ANN analysis was separated into a training set and a test set to develop a predicting model of childhood repeat injury. The entire database was divided into a training set and a test set used to test the performance of the trained networks. The training set contains two-thirds of the data and one-third is left for the test set. After discarding the cases with missing values of the variables selected for model development, there were 98,576 cases remaining in the data set. This data set was further separated into repeat injury group, 53,323 cases (54.1%) and non-repeat injury group, 45,253 cases (45.9%) in terms of the output variable, that is, whether a repeat injury or not.

A preliminary experiment with the entire data set showed that the training of the network was extremely time-consuming even for one parameter. This may be explained by the empirical studies discussed above, as the training time is related exponentially to the number of input variables and has a polynomial increase with the number of cases. Hence, it is important to search for alternative sampling strategy to optimize the training process, and to ensure important features in the input data are not missing by the sampling process. This can be solved either eliminating unrelated input variables or reducing the sample size. A lot of effort and discussion has been made to determine the least amount of variables for the model. Regarding the sample size, a random selection of 20 percent cases for each of the two groups, repeat injury and non-repeat injury, was done and this satisfied the sample size criteria to train a model without causing problems (Refer to the formula in Paragraph 3.1.4). The resulting non-repeat injury group and repeat injury group contained 10,639 cases and 9010 cases, respectively. The training set was developed with two-thirds of these cases from each of the two groups

and then mixed together with a MATLAB program. The resulting training set contained 13,100 cases with *a priori* statistics being 54.1 percent. The rest of the one-third from each of two groups formed test set containing 6,549 cases.

Once the most appropriate raw input data and sufficient sample size had been selected, it had to be preprocessed, as ANNs require data to be prepared in a certain format; otherwise, the neural network will not produce accurate results. As our ANNs used the hyperbolic tangent function as the transfer function, it requires the input values to be normalized between -1 and 1 . The decisions made in this phase of development are critical to the performance of a network. Most of the CHEO-CHIRPP database was preprocessed using the statistical software SPSS. The SPSS program randomly creates both the training set and test set according to the above criteria. The existing functionalities of the program provide comprehensive tools to prepare data for ANNs. This includes data formatting, transformation and detailed statistical analysis such as correlation between each pair of variables.

Data normalization is the final preprocessing step for ANNs. Normalization is a transformation performed on data inputs to distribute the data evenly and scale it into an acceptable range for the network. The goal is to ensure that the statistical distribution of values for each input and output is roughly uniform. This can speed up the training of the network and also increase the network's ability to learn the association between inputs and outputs. As the transfer function used in these experiments is the hyperbolic tangent function and the resulting output is ranges between -1 and 1 , all input binary variables were reduced to -1 and 1 to represent an absence of a feature or a presence, respectively. The continuous and nominal variables were normalized by subtracting the mean and dividing by a number of standard deviations. This is a relatively simple

normalization method that utilizes a statistical measure of central tendency and variance to help remove outliers, and spread out the distribution of the data, which tends to increase uniformity, though.

4.3 ANN implementation process

ANN experiments were conducted using fourteen variables preprocessed as inputs and one variable as output from the CHEO-CHIRPP database. The training of neural networks was performed on different types of network structures including single-layer and double-layer networks with and without the weight-elimination cost function. The calibration of the ANN parameters is focused on the achievement of the network's optimal performance, which is defined as the maximum classification rate, and maximum sensitivity and specificity levels. The evaluation of the results is based upon the model's classification rate, sensitivity versus specificity and the ROC curves.

4.3.1 Network calibration

The backpropagation algorithm is a popular and useful learning technique for tuning the network parameters to fit a training set of input-output pairs. This algorithm has been incorporated into the training programs routinely used in all the prediction tasks of ANNs in the MIRG. The following network parameters need to be adjusted to gradually reach the optimal performance of the networks: learning rate, learning rate increment, learning rate decrement, momentum, error ratio, and weight-elimination constant, weight-elimination scale factor if the weight-elimination cost function switches "ON", as well as the number of hidden nodes in the hidden layer when a double-layered network is applied. Two approaches, manual and automated parameter adjustment methods are compared in terms of the effect on ANN performance. The manual approach requires many trial simulations to explore the reasonable starting

points of values for each of the parameters. An effective way to avoid the network to be caught in local minima is to start network training with different initial weights. The random number generator function in MATLAB could be seeded differently to select different initial network weights. The initial weights for the input layer were randomly chosen to be between -1 and 1 , whereas the hidden layer weights were between -0.1 and 0.1 as recommended by [Weigend AS. *et al.* 1990, Trigg HCE. *et al.* 1997, Ennett CM. *et al.* 2004]. From preliminary experiments, reasonable ranges can be identified to achieve acceptable performance of the network. Then these parameters are fine-tuned to find the optimal settings for that network [Trigg HCE. *et al.* 1997, Ennett CM. *et al.* 2004].

4.3.2 Automated parameter optimization

The automated network is developed to optimize each network parameter without user supervision. The network is implemented using the minimum, maximum and midpoint values of a parameter and compared the test set performance evaluated by CCR. The parameter values generating the two best set measures and their midpoint were used again in a recursive fashion until two equal classification measures were found or until no further parameter values were possible. Thus the search space for that parameter was recursively reduced to half of the entire value range, which considerably saved running time.

The automated process can start from one of the parameters optimized using the manual approach, while the remaining eight parameters were set to constant default values, which were individually found to be optimal values by manual approach. The process was repeated for each of the nine parameters and compared the performance for

each experiment until the optimal performance was achieved [Frize M. *et al.* 2000, Ennett CM. *et al.* 2004].

Previous experiments have confirmed that automated networks performed equally well as the manual approach. The benefits of the automated ANN over manual ANN are a minimum supervision and significant time saving in the parameter tuning process. In a previous manual ANN, the training can run for 3000 epochs and then the user must identify from the graphs the CCR and average square errors for the training and test sets and at what epoch the point of optimal classification performance occurred. If a parameter has a large range of values to adjust, it will require a long time and supervision at every step to optimize the network. For the automated experiments, a new stopping method was used to shorten the running time. Depending on which of the stopping criteria is chosen, the ANN monitors the correct classification rate or log-sensitivity index for each run. If the network performance does not improve within 500 epochs, the training is stopped and the test set performance is recorded. The choice of 500 epochs is reasonable as the performance over this number of epochs is stable and reliable. This approach gives the flexibility to stop training in a time effective manner if the performance did not improve after a sufficient number of epochs [Frize M. *et al.* 2000, Ennett CM. *et al.* 2004] In manual approach, the ANN is set to execute 3000 epochs and after that, the performance must be examined to determine at what epoch the best performance is achieved with the graphs of ANN results. It is obvious that the automated parameter adjustment method shortens the computer running time considerably when taking into consideration the number of experiments required to optimize the total nine network parameters [Frize M. *et al.* 2000, Ennett CM. *et al.* 2004]. The comparison of the adjusting ranges for nine parameters: learning rate,

adaptive learning rate scale factors, weight-elimination constant, adaptive weight-elimination factors, weight-elimination scale, momentum, and error ratio between manual and automated networks is shown in Table 4.3.2 [Ennett CM. *et al.* 2000].

Table 4.3.2: Comparison of manual and automated network parameter adjusting ranges [Ennett CM, 2000]

Parameter	Adjusting range	
	Manual	Automated
Learning rate (lr)	[5e-6, 5e-3]	[5e-6, 5e-3]
Learning rate increment (lr_inc)	[1, 1.005]	[1, 1.305]
Learning rate decrement (lr_dec)	[1, 0.997]	[1, 0.695]
Weight-elimination constant (λ)	[0, 1e-6] to start [0, 0.0008] after 500 epochs	[0, 1e-6] to start [1e-6, 8e-4] after 500 epochs
Weight-elimination increment (λ_{inc})	[1, 1.005]	[1, 1.305]
Weight-elimination decrement (λ_{dec})	[1, 0.997]	[1, 0.695]
Weight-elimination scale (w)	[0.001, 1]	[0.001, 1]
Momentum	[0, 0.99]	[0, 0.99]
Error ratio (err_ratio)	[1.001, 1.04]	[1.001, 1.04]

4.3.3 Network stopping criteria

A stopping criterion is necessary to determine when the automated network has reached a satisfactory level of performance. The network can be set to stop iterating after reaching the highest correct classification rate, the lowest mean square errors and the highest value of the log-sensitivity index for the test set [Scales N. *et al.* 2001]. In addition, the automated network can also be programmed to stop training if each of above stopping criteria did not improve in the subsequent 500 epochs.

The measure of log-sensitivity index as stopping criterion is introduced here. Recall that we also have two other performance measures, sensitivity and specificity. The sensitivity means the repeated injury cases correctly classified as repeated cases against the actual repeated injury cases. Since the class of repeated injury group is not

the one with a priori class and it seems only meaningful for the network to be capable of identifying more potential injury developing cases. Thus, a performance index that reaches a maximum when both sensitivity and specificity approach one, but is slightly non-symmetric in favor of sensitivity is substantially used in attempts to improve the sensitivity of the network [Scales N. *et al.* 2001, Ennett CM. *et al.* 2004].

4.3.4 ANN variable reduction

As a very important aspect after the optimal performance obtained with the entire input data, decisions will be made to eliminate inconsequential inputs from the model. The extraction of the weights can be easily achieved with a single-layered network because there is only one weight per input variable. So this method is well suited to the single-layered network. While for double-layered networks, the method is still applicable but the interpretation of multiple weights associated to one variable is more complex unless the weights of all of the hidden nodes attached to that specific input variable are reduced to zero, then that variable can be eliminated from the model. Experience shows it is not always easy to see that the weights of the nodes from an input variable all become zero or near zero, even in the presence of the weight-elimination cost function, therefore, that variable cannot be eliminated because the nodes with non-zero input weights still use the information from the variable to aid in the output classification.

Other strategies may be adopted. The variable selection is based on the altered performance of neural networks in response to the changes of the input variables. One of this methods will be used in reducing variables for the prediction of repeat childhood injury model using the CHEO-CHIRPP database. This method, however, may require considerable computation since the network must be trained and tested each time one

input is removed from the model. The implementation is simply to remove each of the input variables one by one, train the network, and compare its performance on the test set with the performance of the network trained with all the input variables. This gives us an understanding about the variables that are most important to the network, since the removal of the most important variables caused the most deterioration to the network's performance. We can see which input has the most deteriorating effect on the network when removed individually. The method that a single variable reduction at a time can be explored to the combining removal of a few of the variables that have less effect on the performance of the network, or even conversely enhance the performance of the network. This can be achieved by a number of manual trials based on the network performance from previous step of single variable reduction.

CHAPTER 5. ANN RESULTS and DISCUSSION

This chapter outlines the experimental results of the ANN approach to predict repeat childhood injury in the CHEO-CHIRPP database.

5.1 Best performance networks

As described in the previous section, the ANN was trained and tested for the best performance based on those previous clearly defined criteria, correct classification rate, sensitivity and specificity, as well as ROC curves. Preliminary experiments showed the optimal performance for the network using single-layered architecture was not in a reasonable range. The unacceptable low values of classification rate and sensitivity meant the model created through the single-layer network was not useful. This showed that the single-layered network was too simple to effectively cope with more complicated data. The complexity of information in input data exceeds the prediction capability of this type of neural networks, in other words, they do not match with each other. The network stopped learning and the performance was not improved at all (data not shown). Preliminary experiments proved double-layered network performed better than single-layered network and was used as the primary networks in all the subsequent experiments.

5.1.1 Best performance of double-layered ANNs

The double-layered network architecture with one hidden layer adds necessary complexity to the ANN so that it is capable of approximating more complex relationships in the data [Penny and Frost, 1996]. Table 5.1.1.1 summarizes the parameter settings of double-layered neural networks with and without the weight-elimination cost function reaching at their optimal performance. Table 5.1.1.2 provides

a summary of the performance measures based on the test set performance of double-layered neural networks with and without the weight-elimination cost function: CCR, sensitivity and specificity, and ROC curves. Note that the CCR of the constant predictor is also included in the table 5.1.1.2.

Table 5.1.1.1: The optimal parameter settings of a double-layered network with and without the weight-elimination cost function

Parameter	Optimal values	
	nwt	wt
lr	0.0037512	0.0045319
lr_inc	1.004	1.004
lr_dec	0.7906	0.7189
λ	0	3.2e -6
λ_{inc}	0	1.001
λ_{dec}	0	0.999
w	0	0.0025
Momentum	0.974	0.964
err_ratio	1.0297	1.0297
Hidden nodes	5	5
epoch	2734	2340

* nwt: no weight-elimination cost function, wt: with weight-eliminated cost function

Table 5.1.1.2: Best performance of the double-layered networks predicting repeat childhood injuries

Type of experiment	Stop criteria	Sens (%)	Spec (%)	Epoch	Test CCR	CP (%)	ROC
nwt	Training	CCR	63.3	84.5		74.8	0.8427
	Test		62.2	84.2	2734	74.1	0.8366
wt	Training	CCR	65.5	85.6		76.4	0.8425
	Test		65.2	84.9	2340	75.8	0.8346
nwt	Training	Log_Sens_Index	70.5	71.8		71.2	0.8164
	Test		69.3	71.6	3462	70.5	0.8096

* nwt: no weight-elimination cost function, wt: weight-elimination cost function, ROC: area under receiver operating characteristic curve

The best performance of the network is also available and visualized with graphs of the CCR curves, ASE and the ROC curves, which is shown in Figures 5.1.1.1. and 5.1.1.2.

Figure 5.1.1.1: Graphs of CCR, ASE and ROC curves for the optimal performance of the double-layered ANN in the prediction of repeat childhood injuries

Figure 5.1.1.2: Graphs of CCR, ASE and ROC curves for the optimal performance of the double-layered ANN with the weight-elimination cost function in the prediction of repeat childhood injuries

The best performance of double-layered networks with and without the weight-elimination cost function is compared and can be seen in Table 5.1.1.2. The test set CCR was 74.1 percent for no weight-elimination double-layered network and 75.8 percent for the same network with weight-elimination function. For both types of the networks, the sensitivity was 62.2 percent for the former, and 65.2 percent for the latter. The specificity was 84.2 percent and 84.9 percent, respectively. The double-layered network with weight-elimination performed only slightly better than the no weight-elimination network in terms of the performance measures on the CCR and sensitivity values. Taking a closer look at one of the parameters for the weight-elimination term, the weight-elimination constant, λ , the value is very small at the point where the

network is optimal, implying the addition of this function had little effect on the network. Another parameter of weight-elimination cost function, the scale factor, w , was also very small when the network was optimal, suggesting that the small weights should be forced to zero, resulting in fewer large weights. Although none of the variables had all of their input weights reduced to zero to be able for them eliminated from the model, the weight-elimination cost function did play a positive role in improving the performance of the network. The area under the ROC curve was around 0.83, approximately the same for the two types of networks. This result confirmed a close classification power between the two types of networks.

5.1.2 Logarithmic-sensitivity index as ANN stopping criterion

A higher sensitivity value is always desirable because it represents the ratio at which the network correctly predicts cases where a future injury is likely to occur. Therefore, the next focus was set on how to improve the ANN in finding the highest sensitivity. Recall the previously mentioned log-sensitivity index for the automated ANN stopping criterion, maximizing both the sensitivity and specificity, but slightly favoring sensitivity. The sensitivity was at 62.2 percent for the double-layered network and 65.2 percent for the same network with weight-elimination. In order to train the network with a higher sensitivity, log-sensitivity was chosen as the stopping criterion in the training of the automated ANN. The resulting optimal performance of double-layered ANN is shown in the last two rows in Table 5.1.1.2. From the table, for the network trained with log-sensitivity as a stopping criterion, sensitivity on test set reached 69.3 percent, 7 percent higher compared with previous results using the CCR and ASE as stopping criteria. Specificity, however, dropped about 12 percent simultaneously, and therefore, substantially caused a CCR to be 4 percent less than

previous networks. The figures show the graphs of log sensitivity index, CCR, ASE and ROC curves. The log sensitivity index graphs visually provide a good understanding of the range of the best possible values for each network parameter since the differences in sensitivity can be observed with changes to the network parameters. This approach showed that the use of the log-sensitivity index as a stopping criterion for automated neural networks did improve the network's sensitivity by 7 percent; however, this lowered the specificity by 12 percent, and as a result, the CCR dropped 4 percent. Thus, there's always a trade-off between the levels of the sensitivity, specificity and the CCR, The selection of stopping criteria for the network depends on which of the measure is the focus of the network. The ROC curves showed the area under the curve was 0.8164 for the training data set, and 0.8096 for the data of test set, which is a good result for this problem.

Figure 5.1.2: Graphs of CCR, ASE and ROC curves and log-sensitivity index for the optimal performance of the double-layered ANN in the prediction of repeat childhood injuries, log-sensitivity index as ANN stopping criterion

5.2 Variable reduction from the ANN model

Once the optimal neural network trained upon all input data was found, the second objective was to narrow down the input variables to the most important ones for the model. One way of doing this was by adding the weight-elimination cost function to the training of the network, where small weights of the input variables are forced to zero and eliminated, so that variables with large weights remain as the most important variables in the model. However, the major drawback of this approach is the difficulty of interpretation of the weights of input variables with a hidden layer because the double-layered networks were used throughout the study. Table 5.2.1 gives the detailed weights for the fourteen variables on five nodes of the hidden layer, when a double-layered ANN without weight elimination at its optimal. The variables represented by numbers can be referred to Table 2.2.1 for details. From the table, none of the variables have their connection weights on each of nodes reduced to zero, so no variables can be eliminated from the model. The relative importance of the variables can be evaluated by extracting their input weights and ranked in a decreasing order, which seems very difficult.

Table 5.2.1: Weights of input variables for the optimal double-layered ANN

Nodes	Variables						
	v1	v2	v3	v4	v5	v6	v7
1	-0.0132	0.0344	-0.0639	-0.2969	0.1108	0.2983	0.124
2	-0.2148	0.0189	-0.126	0.1019	0.0213	-0.1292	-0.0209
3	-3.23	0.0962	0.1166	0.2721	0.1353	0.6537	-1.2792
4	-0.0977	-0.0183	0.0191	0.0802	-0.0258	-0.0808	-0.03
5	0.2476	0.0472	-0.0426	0.0494	0.01	-0.161	0.0622

(Next page continued)

Nodes	Variables						
	v8	v9	v10	v11	v12	v13	v14
1	0.0536	0.1223	0.3358	-0.3057	0.1659	-10.9269	-0.0174
2	0.051	-0.2333	-0.3078	0.2664	-2.6174	-2.27	0.5422
3	0.4221	-0.0363	0.3178	-0.1549	3.7749	-0.1161	0.0759
4	-0.02	0.0234	-0.0811	0.0312	-0.4746	14.4526	0.028
5	0.0067	-0.242	-0.2538	0.2404	-6.3606	-0.382	0.5179

* V1: age, V2: gender, V3: follow-up, V4: month, V5: day, V6: time, V7: location, V8: area, V9: context, V10: breakdown, V11: mechanism, V12: nature of injury, V13: body part, V14: disposition

Alternatively, the “leave one out” method was applied to remove the unimportant variables from the model. The idea is to remove each of the fourteen variables from the model first and compare the performance of the networks when using all input variables. In terms of the least deterioration of network performance, each resulting network is ranked in decreasing order by their performance on the CCR and sensitivity values. The optimal double-layered ANN without the weight-elimination function was used as the optimal ANN model with all input variables and then we proceed to variable reduction. The optimal ANN model with the weight-elimination algorithm was not chosen for this set of experiments. The results are shown in Table 5.2.2. The value of the constant predictor (CP) was also noted. As can be seen in the table, the variables “Injury month” and “Injury time” that indicate the details about when a repeat injury occurred, had the least influence on the performance of the networks. The variables “Area”, about where the injury occurred, “Injury day”, which day in a week, and “Body part”, the injured body parts, had to some degree deteriorated the network’s performance. Other remaining variables in the table demonstrated severe deterioration of the performance upon their removal from input data set. This suggests that these variables are highly predictive for the classification of repeat injuries or not.

Table 5.2.2: Performance of the ANNs with each input variable removed

Delete variables	Ranks	Data set	CCR(%)	Sensitivity(%)	Specificity(%)	CP(%)
Optimal ANN(dl-nwt) Non		Training	74.8	63.3	84.5	54.1
		Test	74.1	62.2	84.2	54.1
Injury_month	1	Training	75.7	65.5	84.3	54.1
		Test	75.4	64.9	84.3	54.1
Injury_time	2	Training	74.5	61.6	85.4	54.1
		Test	73.7	60.1	85.2	54.1
Area	3	Training	67.2	51.9	80.5	54.1
		Test	66.5	49.9	80.5	54.1
Injury_day	4	Training	65.8	56	74	54.1
		Test	65	54.2	74.2	54.1
Body_part	5	Training	63.5	52.8	72.6	54.1
		Test	62.6	51.5	72	54.1
Disposition	6	Training	59.5	44.3	72.5	54.1
		Test	58.5	42.1	72.3	54.1
Location	7	Training	54.6	4.4	97	54.1
		Test	54.7	4.3	97.4	54.1
Context	8	Training	54.5	4.2	97.2	54.1
		Test	54.5	4.3	97.1	54.1
Follow_up	9	Training	54.2	3.4	97.3	54.1
		Test	54.4	3	97.9	54.1
Age	10	Training	54.1	9	92.3	54.1
		Test	54.4	9.7	92.3	54.1
Breakdown	11	Training	54.4	1.6	99.1	54.1
		Test	54.4	1.6	99	54.1
Mechanism	12	Training	54.4	1.9	98.8	54.1
		Test	54.2	1.9	98.6	54.1
Nature_of_injury	13	Training	54.2	2	98.5	54.1
		Test	54.2	2	98.4	54.1
Gender	14	Training	54.2	0	1	54.1
		Test	54.2	0.1	1	54.1

The magnitude of the deterioration from the removal of each the input variable may also depend on the combined effect of the removal of multiple variables, as relationships may exist between them. In the subsequent round of ANN experiments, variables from the highest ranking in the Table 5.2.2 continued to be removed from the model with different combinations. The neural networks trained upon different input data sets and their performance on CCR, sensitivity and specificity were listed and

ranked. The representative results are summarized in Table 5.2.3. The removal of both variables “Injury month” and “Injury time” caused only a little decrease on the CCR value of three percent, but the sensitivity value remained unchanged and even increased a little, whereas a relatively higher decrease on specificity was seen. According to the ranking, the above two variables were then combined with a third variable “Area” removed from the inputs. Surprisingly, the neural network outperformed all of the networks that had been obtaining so far. The CCR was 82.2 percent, seven percent higher than the optimal ANN achieved with all input variables. There was a significant increase in the sensitivity value by up to fifteen percent compared to the network with all input variables, and with no deteriorations of the specificity. The specificity was also two percent higher when compared to the network trained with all input variables. In other words, excluding the above three variables from inputs resulted in a significant overall improvement of the model. This showed that the inclusion of unnecessary inputs not only complicates the network, but also confuses it and reduces its performance. The identification of these variables and their elimination from the model definitely improved the predictive power of the resulting model. As experiments continued, another variable “Injury day” was further removed combining with the above three variables from the inputs; the performance of the network started to decline on both the CCR and sensitivity, which means that the variable being removed played a positive role in the model. Notice that, when the variable “disposition”, which represents the severity of the injury, was also removed, the network performance on the CCR was higher again; however, the sensitivity was the lowest and specificity was very high. This resulted network is not acceptable because the performance measures have to be based on both the CCR and sensitivity, specificity levels. Thus, the determination of variable

selection on the neural network trained in the absence of all five variables was not appropriate. Many ANN experiments were conducted while removing various input variables in different combinations and all of the resulting models caused severe deterioration of the performance of the network. A part of result on network performance with various combinations of input variables removal is summarized below in Table 5.2.3. As can be seen from the table, a more effective ANN model could be extracted from the second row by a combined removal of the three least important variables. This model shows a strong, increased predictive power in the prediction of childhood repeat injuries. The best performance of the ANN after removal of least important variables is shown in Figure 5.2.1.

Table 5.2.3: Performance of ANNs with combined removal of input variables (part)

Delete variables		Data set	CCR(%)	Sensitivity(%)	Specificity(%)	CP(%)
Injury_month, Injury_time		Trainig	72.6	65.8	78.4	54.1
		Test	71.5	64.4	77.4	54.1
Injury_month, Injury_time Area		Training	82.2	77.9	85.8	54.1
		Test	82.2	76.8	86.7	54.1
Injury_month, Injury_time Area, Injury_day		Training	67.5	57.8	75.7	54.1
		Test	66.1	55.9	74.8	54.1
Injury_month, Injury_time Area, Disposition		Training	54.1	0.7	99.5	54.1
		Test	54.3	0.8	99.6	54.1
Injury_month, Injury_time Area, Injury_day, Disposition		Training	74.7	59.4	87.6	54.1
		Test	74.7	59	88	54.1

Figure 5.2.1: Graphs of CCR, ASE and ROC curves for the best performance of the double-layered ANN after variable reduction in the prediction of repeat childhood injuries

5.3 Identifying the importance of variables leading to repeat childhood injuries

Through variable reduction, we found the most effective ANN model and least number of variables required to maintain its prediction ability. The next question was to find the most important variables of the model. This is particularly meaningful from the aspect of injury prevention, as health professionals want to know the important factors responsible for repeat childhood injuries. We have discussed the two methods of

evaluating the relative importance of input variables. Table 5.3.1 gives the weights of the eleven variables for the optimal ANN model obtained after variable reduction.

Table 5.3.1: Weights of input variables for an optimal double-layered ANN after variable reduction

	Variables										
Nodes	v1	v2	v3	v4	v5	v6	v7	v8	v9	v10	v11
1	0.0279	0.012	0.0114	-0.032	-0.019	-0.0423	-0.0849	-0.0066	-0.522	13.4562	0.0995
2	0.2128	-0.267	-0.3604	0.1409	-0.153	0.025	-0.6238	0.0513	-1.565	-3.0943	-1.5309
3	0.0004	-0.023	-0.037	0.004	0.0486	-0.0255	-0.0545	-0.059	6.414	-0.5914	-6.9324
4	0.2636	0.0191	0.0083	-0.0656	-0.067	-0.1864	-0.1923	0.0873	-0.802	10.3588	0.2253
5	0.0165	-0.031	-0.0487	0.0041	0.0533	-0.042	-0.0865	-0.0689	7.55	-0.5832	-15.934

* V1: age, V2: gender, V3: follow-up, V4: day, V5: location, V6: context, V7: breakdown, V8: mechanism, V9: nature of injury, V10: body part, V11: disposition

From this table, the weights of the eleven input variables are estimated and ranked in a decreasing order: body parts, disposition, nature of injury, breakdown, age, gender, follow-up, context, location, day, and mechanism.

The estimation of the relative importance of the input variables was also confirmed by reducing each variable one at a time on the best ANN obtained with the eleven variables and the resultant networks were compared for their performance. The contribution each variable made to the model was estimated by how much the network deteriorated or improved with their removal (data not shown). The importance of the variables was ranked in a decreasing order: body parts, disposition, nature of injury, breakdown, age, gender, context, location, day, follow-up, mechanism. The two methods showed consistency on the estimation of the most important variables in the most predictive ANN model.

In Table 5.3.1, the weights for the last three variables: body part, disposition and nature of injury, were prominent, suggesting they are most important to the model as

their weights dominates over all other eight variables. To further explore the combining contribution of these three most important variables would have made to the model, the model only with these three variables was examined. The result showed that the CCR was slightly improved 83.89% compared to 82.2 (+1.6%), resulted by significantly increased sensitivity 95.9% compared to 76.8% (+19%) but compromised by reduced specificity 86.7 compared to 73.6% (-13%). This result confirms further the above three variables the most predictive variables in the model and on the other hand, also implies the less important variables should not be simply eliminated from the model as they keep model's specificity.

The identification of the most important variables responsible for repeat childhood injuries will provide health policy makers useful information when making injury prevention program.

5.4 Discussion

In this study, a portion of CHIRPP data, the CHEO-CHIRPP database was analyzed to assess the feasibility of predicting repeat childhood injuries using artificial neural networks. This is the first study that applies artificial neural networks to this problem with the CHIRPP data. As can be seen from the results, the neural network technique used with the CHEO-CHIRPP data showed an improvement of classification capability of approximately 30 percent over the constant predictor, which suggests that the model may have some clinical usefulness and potential to be used in childhood injury prevention.

5.4.1 Applying artificial neural networks to predict repeat childhood injuries to the CHEO-CHIRPP database

The CHIRPP attempts to provide good data on childhood injuries nationwide for health professionals and community health organizations. It is always a challenge to bring out the information concealed in the set of data. Model development is an approach that keeps identifying patterns from a population with a known outcome with a capability of distinguishing between the presence and the absence of this outcome. A statistical approach is generally accepted in the development of outcome prediction models. However, this has limitations. Several statistical techniques such as discriminate analysis and logistic regression can be used to address a new classification problem. Statistical models, however, are sensitive to issues such as assumption of normality, model specification, correlated variables, and noisy data sets. Violations of these assumptions can adversely affect the performance of statistical models. A well-known fact for model development based on an underlying linear regression model is that the form of the model has to be specified in advance. If the data can be fitted well by a linear model, then the choice may be appropriate. However, for most of the medical databases, relationships in data are complex and better fitted by nonlinear models. However, even in such cases, problems also arise as to how to construct a correct nonlinear model. This is especially meaningful for databases where quantitative variables are rarely available. In the CHIRPP data, the majority of the variables are of the nominal type and few are continuous, which adds difficulties in data analysis. Our early studies of the CHIRPP database were unsuccessful, indicating the database is relatively complex and difficult. Neural networks allow models of any complexity to be fitted between the output and input variables, and the form of the model need not be

specified in advance, as it is for a linear regression model. In MIRG, the neural network techniques have been extensively researching and successfully predicting, for example, infant mortality and length of stay in a neonatal intensive care unit (NICU) at CHEO [Frize M. *et al.* 2001]. Moreover, similar topics in recent studies have developed some childhood injury related models to predict pediatric trauma mortality and injury severity by using, though distinct database, known as the National Pediatric Trauma Registry in the United States. The studies compared the approaches of artificial neural networks and statistical logistic regression as classifiers and concluded that the ANNs were superior to logistic regression [DiRusso SM, *et al.* 2002]. Therefore, the ANN techniques were selected as an analyzing tool for the CHIRPP database. Our experiment showed that a single-layered network was not sufficient to handle complex relationships in the CHIRPP data, suggesting further a statistical linear model cannot be an appropriate choice. In addition, noisy and incomplete data sets are unavoidable problems, especially for large-scale databases collected over a large time period. Artificial neural networks overcome many of the shortcomings of statistical approaches, and therefore have been chosen as an analysis tool in this study.

Another advantage of models developed using neural networks is the flexibility in response to updates. The independence of the backpropagation algorithm makes neural networks flexible and allows newly recognized variables to be easily incorporated into the training set and thus can update the network constantly with the latest information. The flexibility is particularly important in applications to medical decision making, as the dramatic changes could have been the results of ongoing research, altered disease epidemiology, improvements in the diagnosis and treatment, and even changes in social psychology. All these changes need to be incorporated into

the models and the ANNs trained again with the new data. In addition, the normalization of input data benefits the neural networks by eliminating the outliers in the data. Moreover, data representation should be designed to use variables taking on more values to provide the networks with sufficient information.

Although neural networks do offer a number of advantages over some traditional statistical data modeling methods and can be complementary of statistical techniques, the shortcomings of this promising technique require more vigorous research to address and improve them [Livingstone DJ, *et al.* 1997, Ennett CM, *et al.* 2004]. For example, neural networks can suffer from overfitting, but may be avoided by selecting suitable network structure and monitoring by a test dataset as have shown in this thesis. The neural networks are also susceptible to chance effects and do not give confidence intervals. Finally, the interpretation of network weights is not straightforward. These problems need to be further explored and addressed.

A neural network using backpropagation algorithm is subject to learning bias meaning that the behavior of the neural network learning process is unpredictable. Insufficient training data can be a factor of learning bias to influence on the classification of networks. Other factors including network structure, parameters of learning algorithm, sequencing of training data and the stop criteria (a strong bias factor) of the learning are referred to as learning bias. Neural network researchers are aware of this problem and have been trying to improve available algorithms to deal with learning bias. One of the studies showed that the effect of learning bias can be controlled by using multiarchitecture monotonic function neural networks [Archer NP, *et al.* 1992]. In this thesis, the efforts we made on sufficient data sample size and

problem domain knowledge were very useful to control the bias which evidenced by a good generalization of our networks.

In this study, we proposed the application of ANNs as outcome prediction tools for the evaluation of patients with repeat injuries. Our ANN model showed very good classification ability to differentiate patients with the above distinct outcomes, which confirms the suitability of applying the neural network technique to this problem.

5.4.2 Variable selection for the ANN model

One of the most common problems encountered in data modeling is the choice of input variables for inclusion in the model. The neural networks can be used to isolate factors responsible for a given outcome, and can be used to study the relative importance of the various factors. The relative importance of each variable as a predictive property of the model is difficult to calculate, but the relative contribution of each variable to the model can be assessed in a number of ways. In the study on repeat childhood injuries with the CHEO-CHIRPP data, many attempts have been made to understand which variables are important for the prediction. Two strategies have been adopted in this study. First, the input weights of variables can be used to estimate the proportion of contribution each variable makes to the model, and it is a simple process for neural networks with no hidden layers when the output and input relationship is linear. However, it is not usually the case because most of the large multi-variable medical databases often have complex, nonlinear relationships that must be handled with more complex neural network structures. In the analysis on the CHIRPP database, the estimation of output weights was difficult because the neural network with the best performance was trained with one hidden layer. The weight-elimination cost function

can be added to the neural networks as an effective way for further reducing inconsequential variables and accordingly, the performance of the neural networks can be improved. Although the addition of this function slightly improved the performance of the network both on the CCR and sensitivity, no variables can be excluded from the model with this approach. Another variable selection strategy is based on the measures of altered performance of the neural networks in responses to the changes of the number and the composition of the input variables. The results were encouraging; three variables “Injury month”, “Injury time” and “Area” were removed, and the performance of the neural network was greatly improved, namely, the prediction power of the model was enhanced. This demonstrated that the existence of irrelevant data could adversely affect the prediction ability of the neural networks. Hence, neural networks can be seen as a variable selection tool to provide different insights into a data set, as shown here for the CHEO-CHIRPP data. Moreover, the benefit of eliminating unnecessary variables is to reduce the complexity of the network.

5.4.3 Potential of ANN models to the prevention of childhood injuries

The CHEO-CHIRPP database reflects the child injury status in the region of Eastern Ontario over the past 10 years. The ANN model we have generated demonstrates a strong power to predict repeat injuries in injured patients in the database. Also, the relative importance of variables in the predicting model has been identified. Our findings in this study may provide health workers, schools and even parents useful information to prevent repeat childhood injuries. For example, the model can be used to predict the possible occurrence of repeat injuries for a specific patient in the future and parents may be alerted to keep a close eye on this child. The sensitivity value for the

model is approximately 80 percent, which means 80 percent of patients will be correctly predicted for repeat injuries. When a parent wants to know whether an injured patient will or will not have a repeated injury, the specificity value of the model would give you the correct prediction of no repeat injuries at 86 percent. Moreover, the identification of the important variables in the ANN model could be important for injury prevention because it allows further studies focusing on the details of the information provided by these important variables to the problem.

Our ANN model showed a very good prediction for the outcome of repeat childhood injuries. It is very likely this model to be distributed to emergency rooms for individual outcome analysis. In addition, our successful ANN model adds more evidence to show that our powerful ANN tools are capable of outcome analysis for a variety of medical databases. We hope to apply this power tool to injury outcome analysis on other aspects of injuries for the CHIRPP database.

There is no doubt that our ANN tools developed in the MIRG be a powerful analytical option in the analysis of the CHEO-CHIRPP database. The ANN model we developed correctly predicts over 80 percent of patients with repeat injuries or not. So if any room remains for considering further improvement of models prediction ability to this problem, we would speculate that might come from some unrecognized factors containing information associated with the current problem. No doubt that the CHIRPP is a well-defined database and has included abundant injury-related information. It is also possible that some patients who tend to have repeat injuries are related to some factors, such as social-economic factors that seems not directly related to child injuries, therefore are not the focus and not included in the current database. Therefore, we also

hypothesize the possibility of the current model further improved if such information carrying variables were included in the model.

CHAPTER 6. CONCLUSIONS, CONTRIBUTIONS and FUTURE WORK

6.1 Conclusions

The first objective was to develop a model to predict repeat childhood injuries using the CHEO-CHIRPP database and the artificial neural networks as the analytical tool. The developed neural network model showed very good prediction ability to the prediction of repeat childhood injuries. The successful removal of three inconsequential variables from the prediction model significantly improved the model's overall classification ability, which is approximate 30 percent higher over the constant predictor. The sensitivity and specificity level of the model is promising, which is particularly meaningful for the parents when the first time injury occurred to their children. Therefore, it is highly suggested that we should develop a clinical prototype based on our ANN model for individual outcome analysis in hospitals.

The second objective of the thesis was to identify the most important factors leading to the repeat childhood injuries with the CHEO-CHIRPP database. The relative importance of the variables in the model was estimated by input weights of variables. It was also confirmed by the approach of variable reduction by performance assessment for each variable removed. The identification of the important variables for repeat childhood injuries is especially useful for health care professionals and researchers to develop effective childhood injury prevention programs.

6.2 Contributions

The following is a summary of the contributions to new knowledge that resulted from the completion of this thesis.

- 1) This is the first study using the CHIRPP data that developed a model for predicting repeat childhood injuries.
- 2) This is also the first time that artificial neural networks were applied to this problem. The study showed that the application of a feedforward backpropagation ANN to predict repeat childhood injuries in the CHIRPP database produced promising results. The effectiveness of this approach was demonstrated by the performance of the ANN model developed. It can correctly predict 80 percent of patients who are likely to suffer repeat injuries, which is about 30 percent higher than the constant predictor.
- 3) The developed ANN model has the potential to be used in a hospital's emergency rooms for individual outcome analysis. The model's sensitivity and specificity values reached approximately 80 percent in prediction accuracy; this could provide parents with useful information about whether or not their children will suffer from repeat injuries and an incentive to develop prevention measures.
- 4) The relative importance of the variables identified in the ANN model is expected to provide injury prevention experts insight into how to more effectively prevent injuries.

- 5) This study demonstrates that the ANN tools developed in MIRG are effective and powerful in the analysis of the CHIRPP database, which is among a variety of databases successfully analyzed by MIRG.

6.3 Future work

During the completion of this thesis, the following issues were considered for continuous studies in this subject with the CHIRPP database.

The neural network development:

- 1) The neural network tool needs to be further improved to reduce the training time to meet the needs of an increasing number of variables and a large sample size.
- 2) Search for an automated approach to identify important factors in the ANN model.
- 3) Develop guidelines for the interpretation of input weights of the variables in the neural network with a hidden layer.

Statistical method development

- 1) Use appropriate statistical methodology to explore the relationships between variables.
- 2) Develop statistical logistic regression models for the same problem and compare with the neural network models.
- 3) Use statistical approach to find important variables for the statistical models and compare with the findings in this study.

CHIRPP database improvement

- 1) Since the CHEO-CHIRPP data is only a part of the entire CHIRPP database, the findings from this study may be region-specific. Extending the studies to the data from other regions of the CHIRPP database would be helpful.
- 2) Other possible predicting factors need to be investigated and may be added to the CHIRPP database in an ongoing data collection. Those factors may include, for example, an injured child's social economic status, which may make our current model more predictive.
- 3) We believe that there are still many injury-related questions whose answers may be found by further studies of this database using effective approaches. The ultimate goal for the studies of this database is to utilize well-established models and incorporate them into existing or developing local programs to effectively prevent childhood injuries.

REFERENCES

- [Archer NP. *et al.* 1993]
Archer NP, Wang S. Learning bias in neural networks and an approach to controlling its effects in monotonic classification. *IEEE Trans, Pattern analysis and machine intelligence* 15;9:962-966, 1993.
- [Bartlett PL. 1997]
Bartlett PL. For valid generalization, the size of the weights is more important than the size of the network," in Mozer, M.C., Jordan, M.I., and Petsche, T., (eds.) *Advances in Neural Information Processing Systems 9*, Cambridge, MA: The MIT Press, 1997: pp. 134-140.
- [Baum EB. *et al.* 1989]
Baum EB, Haussler D. What size net gives valid generalization? *Neural Comput.* 1989;1:151-160.
- [Baxt WG. 1995]
Baxt WG. Application of artificial neural networks to clinical medicine. *Lancet* 346:1135-1138, 1995.
- [Bienefeld M. *et al.* 1997]
Bienefeld M, Pickett W *et al.* A descriptive study of childhood injuries in Kingston, Ontario, using data from a computerized injury surveillance system. *Chronic diseases in Canada*. Vol. 17: No.1, 1997.
- [Bishop CM. 1995]
Bishop CM. *Neural Networks for Pattern Recognition*, Oxford: Oxford University Press. 1995.
- [Blum A. 1992]
Blum A. *Neural networks in C++*. John Wiley & Sons, Inc. New York, 1992.
- [Burke HB. *et al.* 1997]
Burke HB, Goodman PH, Rosen DB. *et al.* Artificial neural networks improve the accuracy of cancer survival prediction. *Cancer* 1997;79:857-862.
- [CHIRPP publication]
CHIRPP. A serious child health problem..
http://www.hc-sc.gc.ca/pphb-dgspsp/publications_e.html
- [Danceco ER. *et al.* 2000]
Danceco ER. *et al.* Incidence and costs of 1987-1994 childhood injuries: demographic breakdowns. *Pediatrics* 2000 Feb;105(2):E27
- [Dayhoff JE. 1990]
Dayhoff JE. *Neural network architectures: an introduction*. New York: Van Nostrand Reinhold, 1990.
- [De Laurentiis M. *et al.* 1994]
De Laurentiis M, Ravdin PM. A technique for using neural network analysis to perform survival analysis of censored data. *Cancer Lett.* 1994;77:127-38.
- [DeLeo JM. *et al.* 1995]
DeLeo JM, Campbell G. Receiver operating characteristic (ROC) methodology in artificial neural networks with biomedical applications. In: *Proceedings of the World Congress on Neural Networks*. Washington, DC, July 17-21, 1995.
- [DiRusso SM. *et al.* 2002]

- DiRusso SM. *et al.* Development of a model for prediction of survival in pediatric trauma patients: comparison of artificial neural networks and logistic regression. *J Pediatr Surg* 2002;37:1098-1104.
- [Ennett CM. *et al.* 2002]
Ennett CM, Frize M, Scales N. Logarithmic sensitivity index as a stopping criterion for neural networks. *Proc IEEE-EMBS/BMES*, 2002.
- [Ennett CM. *et al.* 2004]
Ennett CM, Frize M, Charette E. Improvement and automation of artificial neural networks to estimate medical outcomes. *Med Eng Phys.* 2004 May;26(4):321-8.
- [Fahlman SE. 1988]
Fahlman SE. An empirical study of learning speed in backpropagation networks. Technical Rep. CMU-CS-88-162. Pittsburgh: Carnegie-Mellon University, 1988.
- [Frize M. *et al.* 1995]
Frize M, Solven FG, Stevenson M, Nickerson B, Buskard T, Taylor K. Computer-assisted decision support systems for patient management in an intensive care unit. *Medinfo.* 1995;8 Pt 2:1009-12.
- [Frize M. *et al.* 1998]
Frize M, Wang L, Ennett CM, Nickerson BG, Solven FG, Stevenson M. New advances and validation of knowledge management tools for critical care using classifier techniques. *Proc AMIA Symp.* 553-7.1998.
- [Frize M. *et al.* 2000]
Frize M, Ennett CM, Charette E. Automated optimization of the performance of artificial neural networks to estimate medical outcomes. *Proceeding of IEEE Information Technology Applied to Biomedicine & International Telemedical Information Society.* 2000;168-73.
- [Frize M. *et al.* 2001]
Frize M, Ennett CM, Stevenson MH, Trigg HCE. Clinical decision support systems for intensive care units: using artificial neural networks. *Med Eng Phys* 2001;23:217-225.
- [Frize M. *et al.* 2001]
Frize M, Ennett CM, Hebert P. Improving the efficiency and effectiveness of clinical decision-support systems using artificial neural networks. *Medinfo.* 2001.
- [Fundamentals of neural networks]
<http://beyonddream.com/NeuralNet/fundamental.htm>
- [Geman S. *et al.* 1992]
Geman S, Bienenstock E, and Doursa R. Neural Networks and the Bias/Variance Dilemma. *Neural Computation*, 1992: 4, 1-58.
- [Guyer B. *et al.* 1990]
Guyer B, Ellers B. Childhood injuries in the United States: mortality, morbidity and cost. *Am J Dis Child* 1990;144:649-52.
- [Injury Reports, 1997]
Injury Reports, Health Canada. Canadian Injury Data Mortality – 1997 and Hospitalizations – 1996-97. <http://www.hc-sc.gc.ca/pphb-dgspsp/injury-bles/index.html>
- [Injury surveillance on-line]

- Injury surveillance on-line. The economic burden of unintentional injury in Canada. <http://www.hc-sc.gc.ca/pphb-dgspsp/injury-bles/>
- [Itchhaporia D. *et al.* 1996]
Itchhaporia D. *et al.* Artificial neural networks: current status in cardiovascular medicine. *J Am Coll Cardiol* 1996;28(2):515-521.
- [Izenberg SD. *et al.* 1997]
Izenberg SD. *et al.* Prediction of trauma mortality using a neural network. *Am Surg* 1997;63: 275-81.
- [Lang EW. *et al.* 1997]
Lang EW. *et al.* Outcome after severe head injury: An analysis of prediction based upon comparison of neural network versus logistic regression analysis. *Neurol Res* 1997;19:274-80.
- [Livingstone DJ, *et al.* 1997]
Livingstone DJ, *et al.* Data modeling with neural networks: Advantages and limitations. *J Computer-Aided Molecular Design*, 1997;11:135-142.
- [Mackenzie SG. *et al.* 1999]
Mackenzie SG, Pless IB. CHIRPP: Canada's principal injury surveillance program. *Injury Prevention* 1999;5:208-213.
- [Mehrotra K *et al.* 1997]
Mehrotra K *et al.* Elements of artificial neural networks. Cambridge, MA: A Bradford Book, MIT Press, 1997.
- [Miller TR. *et al.* 2000]
Miller TR. *et al.* The cost of childhood unintentional injuries and the value of prevention. *Future Child*. Spring-Summer; 2000;10(1):137-63.
- [Peclet MH. *et al.* 1990]
Peclet MH, Newman KD, Eichelberger MR, Gotschall CS, Guzzetta PC, Anderson KD, *et al.* Patterns of injury in children. *J Pediatr Surg* 1990;25(1):85-91.
- [Penny W. *et al.* 1996]
Penny W, Frost D. Neural networks in clinical medicine. *Med Decis Making* 1996;16:386-398.
- [Pickett W, *et al.* 2000]
Pickett W, *et al.* Youth injury data in the Canadian hospitals injury reporting and prevention program: do they represent the Canadian experience? *Injury prevention* 6:9-15, 2000.
- [Press WH. *et al.* 1988]
Press WH. *et al.* Numerical recipes in C. New York: Cambridge University Press, 1988.
- [Qian N. *et al.* 1988]
Qian N. *et al.* Predicting the secondary structure of globular proteins using neural network models. *J Molec Biol.* 1988;202:865-84.
- [Rosenblatt F. 1962]
Rosenblatt F. Principles in neurodynamics. Washington, DC: Spartan Books, 1962
- [Rögnvaldsson T. 1998]
Rögnvaldsson T. A Simple Method for Estimating the Weight Decay Parameter. *Tricks of the Trade*, Springer-Verlag, 1998

- [Rumelhart DE. *et al.* 1986]
Rumelhart DE, McClelland JL. Parallel distributed processing. Vols.1 and 2. Cambridge, MA: MIT Press, 1986.
- [Scales N. 2001]
Scales N. Mortality prediction in an ICU and coronary surgery database using neural networks. Term project, Carleton University, Ottawa, ON, Dec 2001.
- [Senkowski CK. 1999]
Senkowski CK, McKenney MG. Trauma scoring systems: A review. *J Am Coll Surg* 1999;189:491-503.
- [Shanon A, *et al.* 1992]
Shanon A, *et al.* Nonfatal children injuries: a survey at the Children's Hospital of Eastern Ontario. *Can Med Assoc J* 1992;146(3):361-5.
- [Somoza E. *et al.* 1993]
Somoza E. *et al.* A neural network approach to predicting admission decisions in a psychiatric emergency room. *Med Decis making* 1993;13:273-80.
- [Tanaka T. 1993]
Tanaka T. Children injuries in Japan. *Acta Paediatr Jpn* 1993;35(3):179-85.
- [Tepas JJ 3rd, *et al.* 1997]
Tepas JJ 3rd, Veldenz HC, Discala C, *et al.* Pediatric risk indicator: An objective measurement of childhood injury severity. *J Trauma* 1997;43:258-261.
- [Trigg HCE. 1997]
Trigg HCE. An investigation of methods to enhance the performance of artificial neural networks used to estimate ICU outcomes. MscEE Thesis, University of New Brunswick, 1997.
- [Weigend AS. *et al.* 1990]
Weigend AS. *et al.* Backpropagation, weight-elimination and time series prediction. In DS Tourestzky, JL Elman, GE Hinton, eds. *Proc of the 1990 Connectionist Models Summer School*. Morgan Kaufmann: San Mateo. 1990:105-116.
- [Werbos PJ. 1974]
Werbos PJ. Beyond regression: new tools for prediction and analysis in the behavioral sciences. [Ph.D. thesis]. Cambridge, (MA): Harvard Univ., 1974.
- [Widrow B. *et al.* 1985]
Widrow B, Stearns SD. Adaptive signal processing. Englewood Cliffs, NJ: Prentice-Hall, 1985.
- [Woods K. *et al.* 1997]
Woods K, Bowyer KW. Generating ROC curves for artificial neural networks. *IEEE Trans on Medical Imaging* 1997;16(3):329-337.