Neural Network Models for Leukaemia

by

Manimagalay Chetty

In fulfilment of the requirements for the degree of Doctorate of Engineering in the Department of Chemical Engineering, University of Kwa-Zulu Natal

As the candidate's Supervisor I agree/do not agree to the submission of this thesis.

_____

Prof. M. Carsky

15 January 2009

Supervisors: Prof. M. Carsky
Prof. E. Kormuth

Declaration

I, Manimagalay Chetty, hereby declare that

(i) the research reported in this thesis, except where otherwise indicated is my original work.

(ii) this thesis has not been submitted for any degree or examination at any other university.

(iii) this thesis does not contain other person's data, pictures graphs or other information, unless specifically acknowledged as being sourced from other persons.

(iv) this thesis does not contain other person's writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:

(a) their words have been re-written but the general information attributed to them has been referenced;

(b) where their exact words have been used, their writing has been placed inside quotation marks and referenced.

(v) where I have reproduced a publication of which I am the author, co-author or editor, I have indicated in detail which part of the publication was actually written by myself and I have fully referenced such publications.

(vi) this thesis does not contain text, graphics or tables from the Internet, unless specifically acknowledged with the source being detailed in the thesis and in the References section.

_____

Manimagalay Chetty

This thesis is dedicated to my late dad, Rajandram Moonsamy Pillay whose wisdom, support and encouragement is sorely missed.

Acknowledgements

I would like to express my sincere gratitude to:

My parents for always believing in me, for encouraging me to reach my goals and also for always being there when I need anything.

My husband Sagren and son Sijen for always limiting the time I spent with them..

My niece Thaneshree and nephew Yolan for helping with the data acquisition.

All staff of the Department of Haematology and Immunology at Inkosi Albert Luthli Hospital, Durban.

The National Research Foundation (NRF) for start-up funding for this research.

The Women in Research Committee and the Postgraduate Development and Support Program at the Durban University of Technology for funding my study leave.

Abstract

Artificial neural networks (ANN) can detect complex non-linear relationships between independent and dependent variables. Properly trained ANNs have repeatedly demonstrated superior predictive accuracy to other predictive technologies when applied to non-linear systems. Currently there are no studies that have been carried out on predicting survival of leukaemia patients at all. The neural network prediction method adopted in this study aims to provide a robust and accurate method for predicting survival of leukaemia patients for both censored and uncensored patient data. The aim of this research was also to find out the effectiveness of neural networks in modelling leukaemia prognosis and to determine the factors that have the most influence. There is ongoing research into finding ways and means of extending the life span of diseased patients. There is great interest in identifying factors that will yield better predictions of survival for terminally ill leukaemia patients. Prognostic factors generally differ with the treatment of leukaemia. Clinicians face the problem of how to choose the appropriate treatment regime, therefore an analysis of prognostic factors that predict success or failure may identify patients who require an alternative approach of specialist or targeted treatment. Being able to predict an individual patient's prognosis will enable clinicians to categorise them into the relevant high and low risk treatment groups for conventional treatment or allow for the patients to be incorporated into specialised treatment schedules and clinical trials if available. In this study there is believed to be relationship that exists between the results gained on diagnosis and the period of survival. A patient's health status is dependent on various symptoms and the complexity of the medical condition is dependent on an individual's biological system. This complexity allows for the application of artificial neural networks (ANN) in predicting outcomes in medical application, especially in prognosis prediction and survival rate. This thesis contains contributions to the development of neural network models for survival analysis of leukaemia patients. The feed forward back propagation algorithm (BPA) modified to the gradient descent BPA was identified for the training and building of the neural network for predicting survival of leukaemia patients. The prognostic factors that affect survival have also been determined by the neural networks. The comparisons of models were based on using combined groups of leukaemia patients and comparing them with individual groups of the sub-types of leukaemia, i.e. acute lymphoid leukaemia (ALL), acute myeloid leukaemia (AML), chronic myeloid leukaemia (CLL) and chronic myeloid leukaemia (CML). A combination of 38 variables was used in the development of the neural networks. The variables were age, race, sex, gender, and results of full blood counts, differential tests and flow cytometry. The survival period of patients was based on the diagnosis date and the date of treatment. Those patients who status of mortality was known as of October 2008 were considered to be uncensored and were used for the 2-year and 3-year case studies. The

patients with unknown mortality were considered as censored patients and used for the censored case study. The patient data was processed into a coded system and used to build the neural networks for each data set. The choice of patient groups used for the model building was prompted by the availability of uncensored data for analysis. For the group of combined leukaemia patients and the sub-group CML-CLL, it is recommended that the 2-year neural network model be used. The main prognostic factors affecting leukaemia survival were found to be the patient's age, the mean haemoglobin concentration, % neutrophils and the markers CD13, CD20 and CD56. The race group, platelet count, % monocytes and the markers CD3, CD4, CD34 and LC lambda were found to significantly affect the CML-CLL group of patients. For the ALL and AML groups the 3-year neural network models were favoured. Prognostic factors for the survival of ALL patients were their age, the mean corpuscular haemoglobin concentration, % blasts and the markers CD8 and CD22. For the AML group the important prognostic factors were the patient's age, the mean corpuscular haemoglobin concentration, the % neutrophils, % lymphocytes, and the markers CD7 and CD34.

Table of Contents

	Page number
List of Figures	xi
List of Tables	xiii
Abbreviations	xvi
Definitions	xx
Nomenclature	xxxi
Chapter 1: Introduction	1
Chapter 2: Review of Artificial neural networks	4
2.1 Introduction	4
2.2 Development of artificial neural networks	5
2.3 Neural network modelling	7
2.4 Artificial network operations	11
2.5 Properties of artificial neural networks	14
2.6 Artificial neural network architecture	15
2.6.1. Single-layer feed forward network	16
2.6.2 Multilayer feed forward networks	17
2.6.3. Feedback networks	19
2.6.4 Cellular networks	20
2.7. Feed forward multilayer artificial neural networks	22
2.8 Learning process or training	25
2.9 Mathematical derivation of the back propagation algorithm	29
2.10 Generalisation and overtraining	39
2.11 Size of training set	41
2.12 Cross validation	42
2.13. Limitations of the back propagation algorithm	43
2.14 Summary	45
Chapter 3: Review of Applications of Neural Networks	46
3.1 Introduction	46
3.2 Use of ANN in engineering applications	46
3.3 Current predictor methods used in medicine	49

3.4 Neural networks in medicine 52
3.5 General survival studies 53
3.6 Health benefit from ANN in medical intervention 59
3.7 Methodology for neural networks in medical applications 62
3.8 Medical prognosis using artificial neural networks 65
3.9 Previous neural network studies on survival of cancer patients 69
3.10 Model selection 72
3.11 Missing values 74
3.12 Classification of data 75
3.13 Fitting networks 75
3.14 Problems with comparisons of NN models 78
3.15 Summary 79
Chapter 4: Review of Blood and Haematological Malignancies 82
4.1 Introduction 82
4.2 The Eukaryotic Cell Cycle 83
4.3 Haematology 86
4.3.1 Erythrocytes (red blood cells) 87
4.3.2 Platelets (thrombocytes) 87
4.3.3 Leukocytes (white blood cells) 87
4.3.3.1 Segmented neutrophil (filamented neutrophil, polymorphonuclear neutrophil) 88
4.3.3.2 Band neutrophil (non-segmented neutrophil non-filamented neutrophil) 88
4.3.3.3 Eosinophil 88
4.3.3.4 Basophil 89
4.3.3.5 Lymphocytes 89
4.3.3.6 Monocytes 90
4.4. Cell development 90
4.5 Analysis of composition of blood 91
4.6 Haematopoiesis 92
4.7 Haematopoietic cell cycle kinetics 93
4.8 Haematological disorders or malignancies 93
4.9 Blood cells in leukaemias 94

4.10 Symptoms and signs of leukaemia 95
4.11 Diagnosis of leukaemia 95
4.12 Causes of leukaemia 96
4.13 Types of Leukemias 97
4.14 Acute leukaemia 99
4.14.1 Acute lymphoblastic leukaemia 99
4.14.2. Acute myeloid leukaemia or myelocytic or myeloblast 100
4.15 Chronic leukaemia 103
4.15.1 Diagnosis of chronic myelogenous leukaemia 103
4.15.2 Diagnosis chronic lymphocytic leukaemia 104
4.16 Prognosis and survival in acute leukaemia 106
4.17 Prognosis and survival in chronic leukaemia 119
4.18 Diagnosis of haematological disorders 122
4.19 Quantitative measures of haematopoietic elements in the blood 123
4.19.1 Full Blood Count 123
4.19.2 Differential 124
4.19.3 Platelet counts 125
4.19.4 Flow cytometry 125
4.19.5 Cytogenetics 128
4.20 Summary 131
Chapter 5: Neural Network Modelling Methodology 133
5.1 Introduction 133
5.2 Data collection 133
5.3 Data processing 135
5.4 Missing values 138
5.5 Coding data 139
5.6 Development of neural network models 144
5.6.1 Genetic algorithm 144
5.6.2 Network type and architecture 145
5.6.3 Network parameters 146
5.7 Summary 149
Chapter 6: Results and Discussion 150
6.1 Introduction 150
6.2 Fitting neural network models 150

6.2.1 2- year case study 153
6.2.2 Two year survival based on type of leukaemia 161
6.2.2.1 ALL 161
6.2.2.2 AML 165
6.2.2.3 CML-CLL 169
6.2.3 3-year case study 173
6.2.3.1. 3-year ALL 177
6.2.3.2. 3-year AML 180
6.2.3.3. 3-year CML-CLL 183
6.2.4. Censored case study 186
6.2.4.1. ALL 191
6.2.4.2. AML 193
6.2.4.3. CML-CLL 197
6.3 Limitations in this research 200
6.4. Summary 200
Chapter 7: Conclusions and Recommendations 202
References 205
Appendices 216
Appendix A: Patient medical file data 217
Appendix B: Leukaemia patient data for models 227
Appendix C: Coded data for leukaemia patients 260
Appendix D: 2-year case study 279
Appendix E: 3-year case study 296
Appendix F: Censored case study 310
Appendix G: Visual Basic programs 331

List of Figures

Figure 2-1 A single biological neuron

Figure 2-2 Non-linear model of a neuron

Figure 2.3 Sigmoid function (Haykin, 1994)

Figure 2-4 Artificial neural network topology (DACS, 1992)

Figure 2-5 Simple network with feedback and competition (DACS, 1992)

Figure 2-6 Fully connected feed forward network

Figure 2-7 Feed forward network with a single layer of neurons

Figure 2-8 Partially connected feed forward networks (Haykin, 1994)

Figure 2-9 Recurrent network with no feedback loops and no hidden neurons (Haykin, 1994)

Figure 2-10 Recurrent network with neurons (Haykin, 1994)

Figure 2-11 (a) One-dimensional lattice of 3 neurons. (b) Two-dimensional lattice of 3-by-3 neurons (Haykin, 1994)

Figure 2-12 Forward propagation of function signals and back-propagation of error signals

Figure 2-13 Non-linear model of a neuron with bias

Figure 2-14 Signal flow graph highlighting the details of neuron j

Figure 2-15 Architectural graph of a three-layered feed forward network and associated sensitivity network (back propagating error signals)

Figure 2-16 (a) Properly fitted data (good generalisation) (b) Over-fitted data (poor generalisation (Haykin, 1994)

Figure 3.1 Nested cross-validation scheme: ten-fold within five-fold

Figure 6-1 Predicted 2-year survival (21-11-1, 70-30)

Figure 6-2 Predicted 2-year survival (16-2-1, 80-20)

Figure 6-3 Initial predicted 2-year survival (24-18-1, 90-10)

Figure 6-4 Final predicted 2-year survival (19-9-1, 90-10)

Figure 6-5 Initial predicted 2-year survival for ALL (17-20-1, 90-10)

Figure 6-6 Final predicted 2-year survival for ALL (17-20-1, 90-10)

Figure 6-7 Initial predicted 2-year survival for AML (19-11-1, 90-10)

Figure 6-8 Final predicted 2-year survival for AML (19-11-1, 90-10)

Figure 6-9 Initial predicted 2-year survival for CML-CLL (21-3-1, 90-10)

Figure 6-10 Final predicted 2-year survival for CML-CLL (24-12-1, 90-10)

Figure 6-11 Initial predicted 3-year survival (21-6-1, 90-10)

Figure 6-12 Final predicted 3-year survival (21-6-1, 90-10)
Figure 6-13 Initial predicted 3-year survival for ALL (19-9-1, 90-10)
Figure 6-14 Final predicted 3-year survival for ALL (19-9-1, 90-10)
Figure 6-15 Initial predicted survival for 3-year AML (20-12-1, 90-10)
Figure 6-16 Final predicted survival for 3-year AML (20-12-1, 90-10)
Figure 6-17 Initial predicted survival for 3-year CML-CLL (17-7-1, 90-10)
Figure 6-18 Final predicted survival for 3-year CML-CLL (17-7-1, 90-10)
Figure 6-19 Initial predicted survival for censored group (15-9-1, 90-10)
Figure 6-20 Final predicted survival for censored group (18-3-1, 90-10)
Figure 6-21 Initial predicted survival for censored ALL model (19-21-1, 90-10)
Figure 6-22 Final predicted survival for censored ALL model (19-21-1, 90-10)
Figure 6-23 Initial predicted survival for censored AML model (23-15-1, 90-10)
Figure 6-24 Final predicted survival for censored AML model (23-15-1, 90-10)
Figure 6-25 Initial predicted survival for censored CML-CLL (21-12-1, 90-10)
Figure 6-26 Final predicted survival for censored CML-CLL (21-12-1, 90-10)

List of Tables

Table 6-1 Statistics for patient data

Table 6-2 Summary of models for 2-year survival

Table 6-3 Prognostic factors for 2-year survival

Table 6-4 2-year survival models based on accepted variables

Table 6-5 Mean survival for 2-year model

Table 6-6 Cross validation for 2-year survival

Table 6-7 Summary of models for 2 year survival of ALL

Table 6-8 Prognostic factors for ALL 2-year survival

Table 6-9 2-year mean survival for ALL

Table 6-10 Cross validation for 2-year ALL

Table 6-11 Summaries of models for 2-year survival for AML

Table 6-12 Prognostic factors for 2-year survival for AML

Table 6-13 2-year mean survival for AML

Table 6-14 Cross validation for 2-year survival for AML

Table 6-15 Summaries of models for 2 year survival for CML-CLL

Table 6-16 2-year mean survival for CML-CLL

Table 6-17 Prognostic factors for 2-year CML-CLL

Table 6-18 Cross validation for 2-year survival for CML-CLL

Table 6-19 Summary of models for 3-year survival

Table 6-20 Prognostic factors for 3-year survival

Table 6-21 3-year mean survival

Table 6-22 Cross validation for 3-year survival

Table 6-23 Summary of models for 3-year ALL

Table 6-24 Prognostic factors for 3-year ALL

Table 6-25 Mean survival for 3-year ALL

Table 6-26 Cross validation for 3-year ALL

Table 6-27 Summary of 3-year models for AML

Table 6-28 Prognostic factors for 3-year AML

Table 6-29 Mean survival for 3-year AML

Table 6-30 Cross validation for 3-year survival for AML

Table 6-31 3-year survival models for CML-CLL

Table 6-32 Prognostic factors 3-year CML-CLL

Table 6-33 Mean survival for 3-year CML-CLL

Table 6-34 Cross validation for 3-year survival for CML-CLL
Table 6-35 Censored models
Table 6-36 Censored mean survival for all leukaemias
Table 6-37 Prognostic factors for censored group
Table 6-38 Cross validation for censored group
Table 6-39 Censored models for ALL
Table 6-40 Prognostic factors for censored ALL
Table 6-41 Mean survival for censored ALL
Table 6-42 Censored AML model
Table 6-43 Prognostic factors for censored AML
Table 6-44 Mean survival for censored AML
Table 6-45 Censored CML models
Table 6-46 Prognostic factors for censored CML-CLL
Table 6-47 Censored mean survival for CML
Table 6-48 Final recommended neural network models for leukaemia

Table A-1 Patient medical file data
Table B-1 Leukaemia patient data for models
Table C-1 Coded data for leukaemia patients

Table D-1 Predicted survival for 2-year model
Table D-2 Transform table for final 2-year model
Table D-3 PREDICT statistics for 2-year models
Table D-4 Cross validation for 2-year model
Table D-5 PREDICT statistics for 2-year ALL
Table D-6 Transform table for 2-year ALL
Table D-7 Cross validation for 2-year ALL model
Table D-8 Results of model with prognostic factors for 2-year model
Table D-9 Predicted survival for 2-year ALL
Table D-10 Predicted survival for 2-year AML
Table D-11 PREDICT statistics for 2-year AML
Table D-12 Cross validation for 2-year AML
Table D-13 Prognostic factors for 2-year AML
Table D-14 Predicted survival for 2-year CML-CLL

Table D-15 PREDICT statistics for 2-year CML-CLL
Table D-16 Cross validation for 2-year CML-CLL
Table D-17 Prognostic factors for 2-year CML-CLL

Table E-1 Predicted survival for 3-year model
Table E-2 PREDICT statistics for 3-year model
Table E-3 Cross validation for 3-year model
Table E-4 Transform table for 3-year model
Table E-5 Predicted survival for 3-year ALL
Table E-6 PREDICT statistics for 3-year ALL
Table E-7 Cross validation for 3-year ALL
Table E-8 Transform table for 3-year ALL
Table E-9 Predicted survival for 3-year AML
Table E-10 PREDICT statistics for 3-year AML
Table E-11 Cross validation for 3-year AML
Table E-12 Transform table for 3-year AML
Table E-14 PREDICT statistics for 3-year CML-CLL
Table E-15 Cross validation for 3-year CML-CLL
Table E-16 Transform table 3-year CML-CLL

Table F-1 Predicted survival for censored model
Table F-2 PREDICT statistics for censored model
Table F-3 Cross validation for censored model.
Table F-4 Transform table for censored model
Table F-5 Predicted survival for censored ALL
Table F-6 PREDICT statistics for censored ALL
Table F-7 Transform table for censored ALL
Table F-8 Predicted survival for censored AML
Table F-9 PREDICT statistics for censored AML
Table F-10 Transform table for censored AML
Table F-11 Predicted survival censored CML-CLL
Table F-12 PREDICT statistics for censored CML-CLL
Table F-13 Transform table for censored CML-CLL

Abbreviations

A	artificial
ABL	non receptor tyrosine kinase Abelson
ACC	accuracy
AI	artificial intelligence
AIDS	acquired immune deficiency syndrome
ALL	acute lymphocytic leukaemia
AMI	acute myocardial infarction
AML	acute myeloid leukaemia
ANLL	acute non-lymphocytic leukaemia
ANN	artificial neural network
APC	antigen-presenting cells
APL	acute promyelocytic leukaemia
Ara-C	ctyarubine
ATHOS	acquired immune deficiency syndrome time-orientated health outcome study
ATRA	all-trans retinoic acid
BCR	breakpoint cluster region
BCR-ABL	fusion gene BCR-ABL
BP	back propagation
BPA	back propagation algorithm
BPL	back propagation learning
C	carbon
CD	cluster differentiation
CDC	cell division cycle
CDI	cyclin dependent inhibitors
CDK	cyclin dependent kinases
CGL	chronic granulocytic leukaemia
CI	confidence interval
CLL	chronic lymphocytic leukaemia
CML	chronic myeloid leukaemia
CR	complete remission
CT	computerized tomography

CVC	Chalkley vessel count
DFS	disease free survival
DLBCL	disuse large B-cell lymphoma
DNA	diribonucleic acid
ECG	Electro-cardiograph
exp	exponential
FBC	full blood count
FFN	feed forward network
FISH	Fluorescent in situ hybridisation
FL	fluorescent
FSC	forward scattered
G	gap
GCS	Glasgow coma score
hct	haematocrit
HLA-DR	human leukocyte antigen complex on chromosome 6 region 6p21.31.
hmb	haemoglobin
hmr	haematocrit
Ialch	Inkosi Albert Luthuli Hospital
ICU	intensive care unit
INV	inversion
ISS	injury severity score
KM	Kaplan-Meier
LC Kappa	light chain isotype of immunoglobulin
LC Lambda	light chain isotype of immunoglobulin
lc	leucocytes
LDH	lactic dehydrogenase
LFS	leukaemia free survival
LMS	least means squares
LR	learning rate
LR	logistic regression
M	mitosis
mc	monocyte
mch	mean corpuscular haemoglobin
mchc	mean corpuscular haemoglobin concentration

mcv	mean corpuscular volume
MDR	multidrug resistance
M-FISH	multiplex-fluorescent in situ hybridization
MLP	multilayer perceptrons
mpo	myeloperoxidase
mpv	mean platelet volume
MRC	Medical Research Council
MRD	minimum residual disease
MRI	magnetic resonance screening
MSE	mean square error
M-TEL	multiple telomere assay
N	nitrogen
ND	non dividing
NISS	new injury severity score
NN	neural network
np	neutrophil
OS	overall survival
P	probability
pc	platelet count
PCR	polymerase chain reaction
PDP	parallel distributed processing
Ph	Philadelphia chromosome
PLT	platelets
PML-RARα	chimeral gene
PMN	polymorphonuclear
PNS	peripheral nerve stimulator
PORC	postoperative residual cicatrisation
Pr	probability
PRS	post remission survival
PTS	paediatric trauma score
R	correlation of output with target value
RBC	red blood cells
RBF	radial basis functions
rc	reticulocyte

rcc	red cell count
rdw	red cell distribution width
RNA	ribonucleic acid
ROC	receiver operating scores
RRP	radical retro-pubic prostatectomy
RT	reverse transcript
RTS	revised trauma score
S	PREDICT test
S	synthesis phase of cell cycle
SFBC	swirling fluidized bed combustor
SKY	special karyotyping
SSC	side scattered
T	train
t	translocations
tanh	hyperbolic tangent
TNM	tumour, nodes and metastasis
UK	United Kingdom
V	validation
WBC	white blood cells
wcc	white cell count
WD	weight decay

Definitions

Albumin:	The major plasma protein responsible for much of the plasma colloidal osmotic pressure and serving as a transport protein for large organic anions (e.g. fatty acids, bilirubin, some drugs) and for some hormones when their specific binding globulins are saturated.
Allogeneic:	Denoting individuals (or tissues) that are of the same species but antigenically distinct.
Allografting:	Tissue that is taken from one person's body and grafted to another person.
Ameboid:	Having an irregular or asymmetric outline with peripheral projections as the outline of a group of cells growing in a nutrient culture and resembling a one cell protozoan.
Anaemia:	Reduced number of erythrocytes, quantity of haemoglobin, or the volume of packed red cells in the blood.
Aneuploid:	An abnormal number of chromosomes in a cell.
Annexin:	A common name for a family of cellular proteins that bind calcium-dependents to phospho-lipid membranes.
Antibodies:	These are specialised cells or proteins of the immune system which can recognise organisms or antigens that invade the body (such as bacteria, viruses, and fungi) and set off a complex chain of events designed to kill these foreign invaders.
Antigen:	It is any substance that is capable of inducing a specific immune response and reacting with the products of that response.

Apoptosis: A pattern of cell death affecting single cells, marked by shrinkage of the cell, condensation of chromatin and fragmentation of the cell into membrane-bound bodies that are eliminated by phagocytosis.

Assay: Determination of the amount of a particular constituent of a mixture, purity or potency of a drug.

Autoimmune: Any deviation from or interruption of the normal structure or function of any body part, organ, or system that is manifested by a characteristic set of symptoms and signs and whose etiology, pathology, and prognosis may be known or unknown.

Basophils: A granular leukocyte with an irregularly shaped, relatively pale-staining nucleus that is partially constricted into two lobes and with cytoplasm containing coarse bluish-black granules of variable size.

B-cell: Type of white blood cell that produces antibodies.

Carboxy: Nitrogen organic compound containing the carboxy group (-COOH), which is weakly ionized in solution forming a carboxylate ion (-COO-).

Chimerical: An organism with different cell populations derived from different zygotes of the same or different species, occurring spontaneously or produced artificially.

Chromatin: The substance of chromosomes, i.e. the portion of the cell nucleus that stains with basic dyes.

Cyanmethaemoglobin: A compound formed by combination of hydrocyanic acid with methemoglobin.

Cyclin dependent kinase: A group of protein kinases originally discovered as being involved in the regulation of the cell cycle.

Cytochemical: The identification and localisation of the different chemical compounds and their activities within the cell.

Cytogenetic: Originating from the development of the cell.

Cytokines: Chemicals made by the cells that act on other cells to stimulate or inhibit their function.

dehydrogenase: An enzyme that catalyzes the transfer of hydrogen or electrons from a donor by oxidizing it or reducing it.

Ectopic: It is located away from normal position and arises from an abnormal site or tissue.

Eosinophils: A leukocyte with coarse, round granules present.

Erythrocytes: Red blood cells.

Erthroleukaemia: A malignant disorder characterised by the proliferation of erythroblastic and leukoblastic tissues.

Erthyroid: These are red blood cells whose principal function is delivering oxygen from the lungs to body tissues *via* the blood.

Extramedullary: Situated or occurring outside any of the medullas, including the medulla oblongata and the medullary cavities of the bones.

Fibrinogen: A protein in the blood plasma that is essential for the coagulation of blood and is converted to fibrin by thrombin and ionized calcium.

Flourochromes: A fluorescent compound used as a dye to mark or stain proteins for examination by fluorescence microscopy.

Fludarabine:	An adenine analogue and purine antimetabolite used as the phosphate salt and as an anti-neoplastic in the treatment of chronic lymphocytic leukaemia.
Gene:	The biologic unit of heredity, self-reproducing and located at a definite position (locus) on a particular chromosome.
Globulins:	A group of proteins in blood plasma whose levels can be measured by electrophoresis in order to diagnose or monitor a variety of serious illnesses.
Granulocytes:	White blood cells.
Haematocrit:	The hematocrit measures how much space in the blood is occupied by red blood cells which is useful when evaluating a person for anemia.
Haematologic:	The branch of medical science that studies the morphology of the blood and blood-forming tissues.
Haematopoiesis:	The formation and development of blood cells, usually taking place in the bone marrow.
Haematopoietic:	Pertaining to the formation of blood or blood cells.
Haemoglobin:	The red respiratory protein of red blood cells that transports oxygen as oxy-haemoglobin from the lungs to the tissues where the oxygen is readily released and the oxy-haemoglobin becomes haemoglobin.
Haemolysis:	The breakdown of red blood cells and the release of haemoglobin that occurs normally at the end of the life span of a red blood cell.
Haemolytic:	The liberation of haemoglobin, consisting of separation of the haemoglobin from the red cells and its appearance in the plasma.

Haemolytic:	The rupture of erythrocytes liberates haemoglobin, i.e. separation of the haemoglobin from the red cells and its appearance in the plasma.
Haemostasis:	The stoppage of bleeding or blood flow through a blood vessel or body part or haemorrhage.
Histopathological:	The science concerned with the cytologic and histologic structure of abnormal or diseased tissue.
HLA-DR:	It is a major histocompatibility complex, MHC class II, cell surface receptor encoded by the human leukocyte antigen complex on chromosome 6 region 6p21.31.
Homeostasis:	The ability or tendency of an organism or a cell to maintain internal equilibrium by adjusting its physiological processes.
Homozygous:	Identical genes controlling a specified inherited trait.
Humoral:	Relating to body fluids, especially serum.
Hypergranular:	It refers to a type of leukaemia: hypergranular promyelocytic leukemia**.**
Hyperphosphorylation:	This occurs when a biochemical with multiple phosphorylation sites is fully saturated and it is also one of the signalling mechanisms used by the cell to regulate mitosis.
Hyperplasia:	An abnormal increase in the number of normal cells in normal arrangement in an organ or tissue which increases its volume.
Hypoxia:	Reduction of oxygen supply to a tissue below physiological levels despite adequate perfusion of the tissue by blood.
Immunoglobulin:	A protein of animal origin with known antibody activity, synthesised by lymphocytes and plasma cells which is found in serum and in other body fluids and tissues.

Immunophenotype:	A phenotype of cells of haematopoietic neoplasms defined according to their resemblance to normal T-cells and B-cells.
Karyotype:	The characterization of the chromosomal complement of an individual or a species including number, form, and size of the chromosomes.
Lactic:	A compound formed in the body during metabolism of carbohydrate, by fermentation of carbohydrates in the rumen and by bacterial action on milk.
leukocytes:	These are white blood cells that protect the body from disease-causing viruses, bacteria, toxins, parasites, and tumor cells.
Lymphadenopathy:	A chronic, abnormal enlargement of the lymph nodes usually associated with disease.
Lymphoblastic:	Pertaining to a lymphoblast or producing lymphocytes.
Lymphocytes:	These are white blood cells of the agranulocyte type, originally from stem cells that produce antibodies which attack harmful cells.
Lymphocytosis:	A condition in which the number of lymphocytes increases above normal levels.
Lymphoid:	Tissues relating to the lymphatic system which has a thin, yellowish fluid called lymph fluid that travels throughout the body, thus the lymphatic system helps control fluids in the body.
Lymphoma:	This is any neoplastic disorder of lymphoid tissue.
Lysosomes:	The self-contained organelles found inside most cells, which contain hydrolytic enzymes that aid in intracellular digestion.

Macrophages : White blood cells (activated monocytes) that protect the body against infection and foreign substances by breaking them down into antigenic peptides recognised by circulating T-cells.

Megakaryocytic: Characterised by the presence of large numbers of megakaryocytes.

Neth-haemoglobin: A haemoprotein composed of globin and haeme that gives red blood cells their characteristic colour and whose function is to primarily transport oxygen from the lungs to the body tissues.

Microblasts: A small nucleated red blood cell.

Mitosis: A method of indirect cell division in which the two daughter nuclei normally receive identical complements of the number of chromosomes characteristic of the somatic cells of the species.

Monoclonal: Derived from a single cell.

Monocyte: A large, circulating, phagocytic white blood cell that has a single well-defined nucleus and very fine granulation in the cytoplasm and that constitutes from 3 to 8 percent of the white blood cells in humans.

Monomorphic: Having one or the same genotype, form or structure through a series of developmental changes.

Mononucleosis: An abnormally large number of mononuclear white blood cells in the blood, especially forms that are not normal.

Mucosal: Refers to tissues that produce mucus, such as the digestive, genital and urinary tracts.

Myeloblasts: An immature cell found in the bone marrow and not normally in the peripheral blood; it is the most primitive precursor in the granulocytic

series which matures to develop into the promyelocyte and eventually the granular leukocyte.

Myelodysplasia: A neural tube defect causing defective development of any part of the spinal cord.

Myelogenous: Produced by or originating in the bone marrow.

Myeloperoxidase: A haemoprotein having peroxidase activity, occurring in the primary granules of promyelocytes, myelocytes and neutrophils and which exhibits bactericidal, fungicidal and virucidal properties.

Neoplastic: Pertaining to a neoplasm.

Neutrophils: White blood cells with cytoplasmic granules that consume harmful bacteria, fungi and other foreign material.

Nucleoside: One of the compounds into which a nucleotide is split by the action of nucleotidase or by chemical means and which consists of a sugar (a pentose) with a purine or pyrimidine base.

Nucleotide: One of the compounds into which nucleic acid is split by action of nuclease; nucleotides are composed of a base (purine or pyrimidine), a sugar (ribose or deoxyribose) and a phosphate group.

Null cell: A lymphocyte that develops in the bone marrow and lacks the characteristic surface markers of the B and T lymphocytes.

Oncogenes: Genes carried by tumor viruses that are directly and solely responsible for the neoplastic transformation of host cells or any genetic element linked to cancer.

Oncoproteins: A protein encoded by an oncogene.

Oxy-haemoglobin:	The red respiratory protein of red blood cells that transports oxygen as oxy-haemoglobin from the lungs to the tissues, where the oxygen is readily released and the oxy-haemoglobin becomes haemoglobin.
Para-immunoblasts:	A lymphocyte that has been activated by an antigen and which will further undergo clonal expansion to increase the number of lymphocytes capable of binding to that antigen.
Phagocytes:	A phagocyte is a cell that ingests and destroys foreign matter such as micro-organisms or debris by a process known as phagocytosis.
Phosphatidyl serine:	A phospholipid found in mammalian cells.
Phosphatise:	To change into phosphates or a phosphate, or to treat with phosphate or phosphoric acid.
Plasma:	The fluid portion of the blood in which the particulate components are suspended.
Pleomorphic:	Refers to a variable appearance or morphology.
Pluripotent:	Capable of affecting more than one organ or tissue.
PML-RAR:	Progressive multifocal leukoencephalopathy.
Polymorphonuclear:	Having a nucleus so deeply lobed or so divided as to appear to be multiple.
Polymorphs:	A colloquial term for a polymorphonuclear leukocyte.
Progenitor:	A direct ancestor or an originator of a line of descent.
Prolymphocytes:	A cell of the lymphocytic series intermediate between the lymphoblast and lymphocyte.

Proto-oncogene c-ABL: A normal gene that with slight alteration by mutation or other mechanism becomes an oncogene which is mostly believed to normally function in cell growth and differentiation.

Pseudopods: A temporary projection of the cytoplasm of certain cells or of certain unicellular organisms, especially amoebas, that aids in locomotion and phagocytosis.

Purine: A colorless crystalline organic base that is the parent compound of various biologically important derivatives, e.g. uric acid, caffeine, adenine and guanine.

Reticuloendothelial: Of or relating to or being the widely diffused bodily system constituting all phagocytic cells except certain white blood cells.

Retinoblast: Development of this tumour is initiated by mutations that inactivate both copies of the gene that codes for the retinoblastoma protein.

Splenomegaly: Is an enlargement of the spleen which usually lies in the left upper quadrant of the human abdomen and can be caused by leukaemia.

T-cell: Type of white blood cell produced in the thymus gland that regulates the immune system's response to diseased or malignant cells.

Thrombocytopenia: Is an abnormal drop in the number of blood cells (or platelets) involved in forming blood clots.

Thymus: A lymphoid organ that is located in the superior mediastinum and lower part of the neck and is necessary in early life for the normal development of immunological function.

Transcriptional: The process by which mRNA is synthesized from a DNA template resulting in the transfer of genetic information from the DNA molecule to mRNA.

Trisomy: An additional chromosome in the normal complement to ensure that in each nucleus a chromosome is represented three times rather than twice.

Ubiquitination: A polypeptide found in all eukaryotic cells including plant cells that participates in a variety of cellular functions including protein degradation.

Vacuoles: A small cavity in the cytoplasm of a cell, bound by a single membrane and containing water, food, or metabolic waste.

Nomenclature

Symbols

d	output vector
E	energy function
e	error
f	femto
f	function
g	gram
g	logistic function
L	log likelihood
L	litre
M	past samples
n	discrete time
o	overall output vector
o	outer scaling vector
p	pico
Pr	probability function
s	inner scaling factor
u	sum function
v	internal activity level of neuron
w	weight
x	input
y	function
z^{-1}	unit delay operator
α	momentum
γ	control step-size parameter
δ	local gradient of neuron
Δ	small change
ε	sum of errors
η	learning rate parameter
θ	threshold

φ	activation function
Є	symbol for “belongs to”

Subscripts

av	average
i	neuron left of *j* neuron
j	neuron
k	index of perceptron
N	number of samples
p	number of inputs to neuron

CHAPTER 1
INTRODUCTION

Cancer is a major cause of disease related to human deaths in many developed countries. It has been observed frequently that the prognosis of cancer patients with the same clinical diagnosis can be different. Survival analysis used to define prognostic indices for survival or recurrence of a disease, and treatment outcome. These methods are commonly used in oncology. Clinicians wish to avoid using these further treatments unless the risk of recurrence is high; as the side effects may be unpleasant or dangerous. Cost is another consideration as some treatments are very expensive. The risk of recurrence must be estimated using information available at the time of diagnosis and initial treatment. There is no consensus amongst clinicians as to the best way of integrating the different data. Since prediction is not always easy new variables may be suggested frequently. It becomes quite an expensive exercise if one has to collect all the possible information for each patient. It has been recognised in the medical literature that neural networks have much to contribute to the model of cancer survival. It is therefore vital that the prognosis of cancer patients be accurately determined to ensure that adequate treatment is proposed. Accuracy in survival prediction would ensure that clinicians would immediately, on diagnosis, be able to direct a patient into the appropriate treatment protocol and group. For those patients where the survival period predicted is short, preparations can be made by both the clinicians and the families for a palliative care program, to ensure that the needs of the patients are met during their remaining days. If the model predicts a long survival period then the patients will be directed to a low risk treatment group or regime. For shorter survival predictions the patients will be placed into a high risk group. Low risk patients will receive standard or conventional therapy, and high risk patients will have their treatment targeted to a specific abnormality. Alternately there may be clinical and drug trials that the high risk patients can be exposed to. Since their prediction of survival is low anyway, the possibility of extended mortality will always be quite encouraging to a terminally ill patient who is willing to be part of new research methods and procedures. Cancer patients are sensitive to radiotherapy and chemotherapy. This is caused by multiple factors because the mechanisms of cancer development (or malignancy) are quite complex (Takahashi *et al*, 2007).

Leukaemia is a cancer of the blood and on diagnosis a series of blood and marrow samples are taken from the patient for analysis which aids in the diagnosis of the sub-type of leukaemia, thus indicating to the clinicians the treatment protocol to be adopted for the patient. Clinical data of both censored and uncensored leukaemia patients were used to build neural network models to predict

their survival. There are currently no studies that have been carried out for the prediction of survival of leukaemia patients. The purpose of this study was to find out if neural network modelling can be a reliable method for prediction of survival and prognostic factors for leukaemia. A neural network model has the ability to find hidden patterns in complex data having multiple variables. This is highlighted when the variables are related to one another in non-linear relationships, as is the case with the data used in this research. Results available from patient‟s medical records were used to develop a database of possible variables that could be used to predict a patient‟s survival and prognosis. The final variables adopted for the neural network modelling in this study comprised a range of 38 variables. Only patients with all known information were initially eligible for the study. In order to obtain sufficient numbers for credibility of the research, only patients who had a maximum of 3 missing variables were incorporated into the study. Average values calculated with all patient data (censored and uncensored) were used as replacement for the missing values. Patient‟s demographics: age, gender and race, type of leukaemia, full blood count, differential, flow cytometry and chromosome analysis were used to develop the neural network models. In a clinical setting, patients may enter and leave a treatment program or institution at any point from diagnosis till remission or death. The data for these types of patients may not be reliable as their current status of life or death would be unknown. In order to maintain large patient numbers and thus retain credibility of an analysis method, this censored (unknown mortality status) data is usually incorporated into the model building process. A 2-year case study and a 3-year case study were carried out on uncensored patients. A case study comprising both the censored and uncensored patients was also investigated. The neural network models were developed for each of the above mentioned groups for the prediction of survival. Each of the above groups was divided into the leukaemia sub-types, i.e. acute lymphoid leukaemia (ALL), acute myeloid leukaemia (AML), chronic lymphocytic leukaemia (CLL) and chronic myeloid leukaemia (CML). A feed forward back propagation algorithm (FFBPA) was used in the development of the neural network models. This robust and complex algorithm has successfully been used in some difficult problems in all fields of research. The patients‟ data from the medical files were processed into a coded system and used as an input for the building of the neural network models. A final combination of 38 variables was used in the development of the neural networks. The variables were age, race, sex, gender, and results of full blood counts, differential tests and flow cytometry. The prognostic factors were determined by the combined groups and the individual sub-types based on the statistics of the model building process. Prognostic factors generally differ with treatment of leukaemia. Clinicians face the problem of how to choose the appropriate treatment regime, therefore an analysis of prognostic factors that predict for success or failure may identify patients who require an alternative

approach of specialised or targeted treatment. Being able to predict an individual patient‟s prognosis will enable clinicians to categorise them into the relevant high and low risk treatment groups for conventional treatment or allow for the patients to be incorporated into specialised treatment schedules and clinical trials if available. Reliability and high accuracy of models will dictate the type of treatment to be administered to the patient and if the prediction does not improve the patient‟s health then other pharmacological or behavioural therapies can be adopted.

CHAPTER 2
REVIEW OF ARTIFICIAL NEURAL NETWORKS

2.1 Introduction

"Neural networks are computational methodologies that perform multi-factorial analyses with emphasis on high parallelism and high interconnectivity. Neural networks are also called artificial neural networks (ANNs) to distinguish them from real biological neural networks. They are also called connectionist systems and parallel distributed processing systems (PDP)" (Venkataraman, 2004). ANNs have emerged as a viable tool for non-linear modelling technique. ANNs are complex electronic models based on the brain"s neural network structure. The brain is trained from birth and basically over time learns from experience. This modelling of the brain is a less technical way of developing machine solutions. These biologically inspired computing methods are currently being revered as the next breakthrough in the computer industry. The "simple" brains of animals can perform functions which modern high speed computers cannot replicate. Advances in biological research have enabled a better understanding of the mechanism of the natural thinking process. Research shows that the brain stores patterns as information. Some are very complicated, e.g. being able to recognise individual faces from different angles. Pattern recognition usage is used widely in the computer field to solve problems. Traditional programming is not favoured but involves the development of complex parallel networks for the training of these networks to solve a particular problem. Non-traditional computing words like behave, react, self-organize, learn, generalize and forget are used in this field.

ANNs are self learning mechanisms that do not require the traditional skills of programmers. Researchers claim that these neuron-inspired processors have limitless capabilities, but potential users have tried and failed to solve their problems for neural networks. But this confusion has come from the industry itself. Numerous articles have been published with unique claims and specific examples all promising a large assortment of neural networks. Currently only a few are being used commercially, with the feed forward back propagation network being widely used. Neural network structures represent models for "thinking" that are continuously being developed. These networks are just tools to be used by the network architecture for the learning of patterns. One of the prominent features of ANNs is their ability to approximate any non-linear mapping. Hornick *et al* (1989) showed that multi-layer propagation networks with continuously differential activation functions are capable of modelling any continuous non-linear function to an arbitrary degree of

accuracy. This attribute of ANNs is most attractive from an engineering viewpoint. Additionally, the nature of this non-linearity differs greatly from one system to another (Narotram, 1999). A traditional method for solving problems of non-linear systems was to use some linearisation technique to solve the model.

2.2 Development of artificial neural networks

Artificial neural networks replicate the neural activity of the brain which consists of densely connected networks of neurons that are the building blocks of the central nervous system. The human brain consists of 10^{14} neurons of which there are over 100 types in the human nervous system. Each of these neurons can connect with up to 200 000 other neurons, although 1000 to 10000 connections is typical of a network. The power of the mind is based on these multiple connections. The neurons and their connections form a process that is not binary, not stable and not synchronous. This basic working structure of the brain has been used to develop artificial neural networks that can help solve complex problems like pattern recognition, information processing and adaptation. People are always looking to create mechanical devices that can mimic human behaviour. Practically the human brain behaves like a computer. The human brain is capable of generalizing from abstract ideas, recognising patterns, recalling memories, understanding, interpreting, acting on possible events (such as "maybe it will rain tomorrow"), making inferences and judgements, and relating them to situations that have never been encountered before. Even when a person has a brain injury they can still function. In numerical computations the computer is faster than the human brain, but the brain‟s capabilities outperforms the fastest of computers. This ability has researchers motivated to constantly model the human brain" (Warner *et al*, 1996). Therefore, it is nothing like the current available computers or even artificial neural networks. ANN can only replicate the most fundamental elements of this complicated but versatile and powerful organism. For the software engineer who is trying to solve complex problems, neural computing is never about replicating the brain, but about using machines and a new way of solving problems.

The neuron is similar to a chemical processing plant. Chemicals are transported through the dendritic tree to the synaptic bulb until it is fully charged, henceforth releasing the chemicals across the synaptic gap to the body of the neuron. The neuron receives inputs from various sources, combines them in a particular way, performs nonlinear operations on the result and then outputs the final result. The relationship between the four parts is illustrated in Figure 2-1.

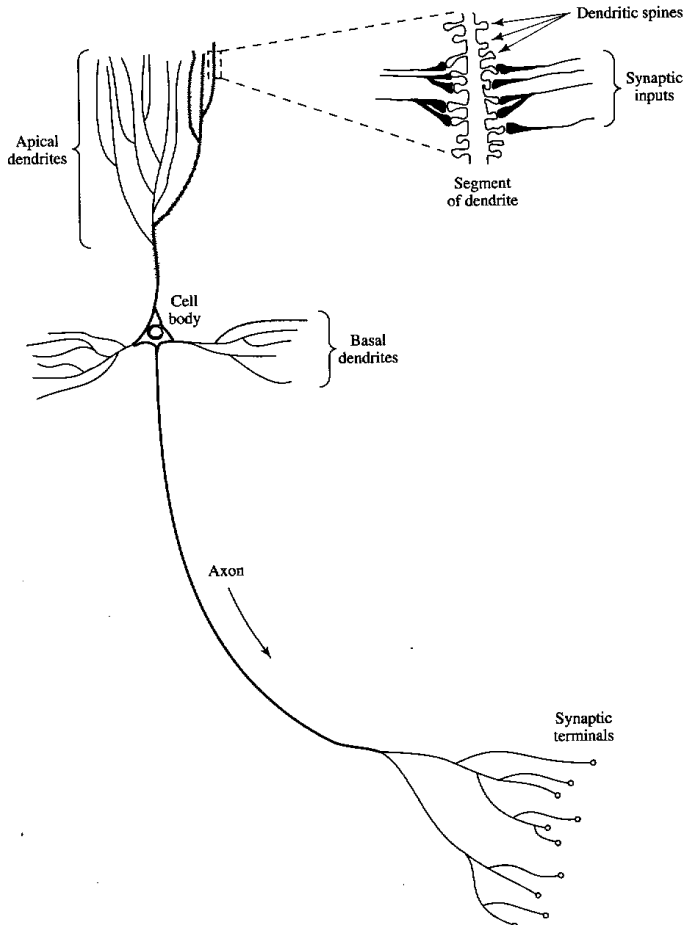

Figure 2-1 A single biological neuron

Neurons are comprised of four basic components, dendrites, soma (cell), axon and synapses. Dendrites have extensions of the soma which are used as input channels. Inputs received through the synapses of other neurons are processed in the soma over time. The soma then outputs the processed value to other neurons through the axon and the synapses. The nature of the synaptic gap increases or decreases the activity of the neuron and the size determines the magnitude of the influence. When the neuron is sufficiently charged it releases the chemicals through the axon into the dendritic tree. The mixture of chemicals governs the strength of the signal passed from the axon to the dendrite of the new cell. Local interaction of the synapses occurs non-linearly. The frequency and arrival time of pulses affects neuron activity. Learning is achieved by chemically adjusting the strength of the synaptic connections between neurons. This can be explained mathematically as "weighting" factors which are used as inputs to the processing unit. The final state of the brain is a

set of chemical "weights". Similarly, fully trained networks store "knowledge" in the form of weight matrices.

Individual biological neurons have minimal computational ability. Remarkable computational properties arise when neurons are interconnected. Individually neurons are slow but when connected in parallel it gives the brain remarkable speed. Current research into the biological nervous system focuses on understanding how networks of biological neurons collectively learn and compute thereby resulting in certain phenomena and behaviours.

2.3 Neural network modelling

Artificial neural networks are classified as "black box" mathematical non-linear regression tools. They learn and identify correlative patterns between sets of input data and corresponding outputs. As in the human brain, learning starts with known data and eventually the brain is trained to adapt to changing environments but the basis is always what was learnt during the training period. As the saying goes "good judgment comes from experience, experience comes from bad judgment," so the neural network needs "experience" to learn to make "good judgments". This experience comes from training it with rich characteristic data of the system under consideration. Data from new input sets can then be used to predict corresponding outputs. The training of these neural networks is a form of non-linear regression and has developed into a tried and tested technique. For the neural network to be successfully implemented in any form, the data used has to be an accurate representation of the system under consideration. The model used for the system must be trained on data that reveals the full range of the expected parameters. The model will then be better equipped to deal with variations in the system outputs (Dunwoodie, 2001).

Artificial neural networks exploit the concept of densely connected networks of simple processing units that are used as powerful tools for practical computations. Engineers and analysts use neural networks as compact ways of discovering complex formulae which have very little to do with simulating intelligence. The following factors have allowed neural network technology to be used as a useful tool for data analysis: technological advances and understandings in neural network algorithms, recognition of commercial potential, advances in desk-top computer speed and price reductions.

Mathematical modelling of the neuron is based on neurophysiology of biological neurons. An artificial neural net is a network of artificial neurons (also referred to as "processing elements (PEs)", "neurons" or "nodes") having several input paths and one output path. The basic unit of neural network, the artificial neuron, simulates the four basic functions of natural neurons. The mathematics of a typical artificial neuron may be represented by Figure 2-2 (Haykin, 1994). Comparing this with Figure 2-1, pulses are converted to pulse rates or frequencies (x_p). The effects of the synaptic gap on activation of the neuron are modelled by weights (w_{kp}) which are multiplied by frequencies (x_p). The "weighted" neurons are then summed up as shown in equation (2.1) to form the internal activity (u_k) of the neuron. The result is processed by a transfer function to produce the outputs.

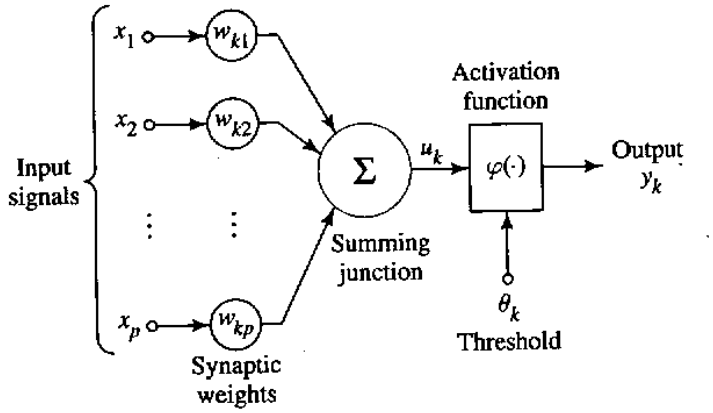

Figure 2-2 Non-linear model of a neuron

The output of the neuron may be calculated using a combination of equation (2.1) and (2.2)

$$u_k = \sum_{j=1}^{p} x_j w_j \tag{2.1}$$

where p is the number of inputs to the neuron. The output of the neuron is given by

$$y_k = f(u_k) \tag{2.2}$$

u_k is the total level of voltage excitation to a neuron, and y_k is the intensity of the resulting output from the neuron (Werbos, 1990b) also referred to as the activation level of the neuron. A constant input called a "bias" is used to simulate thresholds and to simplify the mathematics. Weight zero is associated with the "bias" and the input zero is always equal to 1.

The pre-requisites for a transfer function are that they should have bounded derivatives. Easily differentiated non-linear functions are usually used. The commonly used sigmoidal function is shown below in equation (2.3). The sigmoid function was used for the input and output layers in this study, and is illustrated in Figure 2-3. A similar sigmoidal function has been observed in the human nervous system (Morris *et al*, 1994).

$$f(z) = \frac{1}{1+e^{-z}} \tag{2.3}$$

The differentiated form of (2.3) is (2.4)

$$\begin{aligned} f'(z) &= \frac{e^{-z}}{(1+e^{-z})} \\ &= f(z)(1-f(z)) \end{aligned} \tag{2.4}$$

The first derivative of equation (2.3) is convenient when using the back propagation training algorithm. Outputs of the summing function are sent to a transfer function. This function turns the number into a real output *via* some algorithm. This algorithm takes the input and turns it into a zero or one, a minus one or a one, or some other number. In addition to the sigmoid function (Figure 2-3), the transfer functions that are commonly supported in neural network modelling are: the hyperbolic tangent (2.5) and the bipolar sigmoidal (2.6) functions (Zhu *et al*, 1994).

$$f(x) = \tanh\left(\frac{x}{2}\right) = \frac{1-\exp(-x)}{1+\exp(-x)} \tag{2.5}$$

$$f(x) = \frac{2}{(1+e^{-2x})} - 1 \tag{2.6}$$

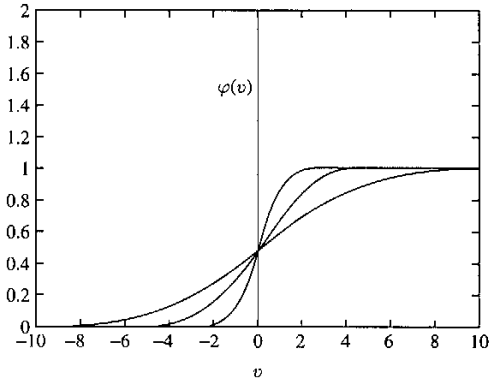

Figure 2-3 Sigmoid function (Haykin, 1994)

Transfer functions can also scale the output or control its value *via* threshold functions. When the slope parameter approaches infinity, the sigmoid function is reduced to a threshold function. On the other hand when a threshold function takes on the value 0 or 1, the sigmoid transfer function assumes a continuous range of values between zero and one. The result of the transfer function output from the processing element is used as an input into other processing elements or to an external connection which is dependent on the network architecture. ANNs are then constructed from these building blocks, i.e. the processing element or the artificial neuron. In a linear transfer function, a single neuron represents a linear equation. The weights in the neuron are equivalent to the parameters in the linear equation thus allowing standard linear regression techniques to be used to solve for the weights. In neural network literature the weights refer to coefficients and for observations they use terms. The functional relationship is dependent on the data being used to find the values for the weights. The advantage of this method is that the network is able to approximate any continuous function and the functional term is a reflection of real data (Warner *et al*, 1996). Conventional modelling involves hypothesising using some algebraic expression that describes the system thereafter using data to fit the model. In linear regression functions, the coefficients can be interpreted in relation to the problem. With neural networks it is difficult to interpret the network. Various functions that are used in conventional modelling and many others can be built into the summation and transfer functions of a neural network. Some networks need to work on problems with multiple responses. These applications are used widely in the robotic industry. The "intelligence" processes are used as inputs to a device which in turn results in the performance of an action or output. These inputs to the network which may come in bursts of 30 seconds, due to limitations of sensors would have to be smoothed. To achieve this, inputs are accepted, data summed and an output produced, e.g. applying a hyperbolic tangent as a transfer function. Output

values from this type of network are continuous, while other applications use summation of data with comparisons to a threshold that produces one of two possible outputs, a 0 and a 1. Some functions can integrate the input data over time to create time-independent networks.

2.4 Artificial network operations

The "art" of using neural networks is based on the interconnected nature of the neurons. In the human mind the information is processed in a dynamic, interactive and self-organizing way. Biological neural networks are made up of complex three-dimensional microscopic components that form infinite interconnections. For existing man-made networks there are always restrictions. With current technology, integrated circuits developed are made up of two-dimensional devices that have multiple layers which are interconnected. Silicon is used for circuit building and the physical constraints impacts on the software capabilities of neural networks (DACS, 1992), i.e. there is a limit to the computing time and the amount of data that can be processed by the software and hardware. Neural networks consist of simple clusters of primitive artificial neurons which form sections connecting them. The connection of these layers is based on learning the "art" of engineering networks find solutions. All artificial neural networks have a similar structure or topology as shown in Figure 2-4.

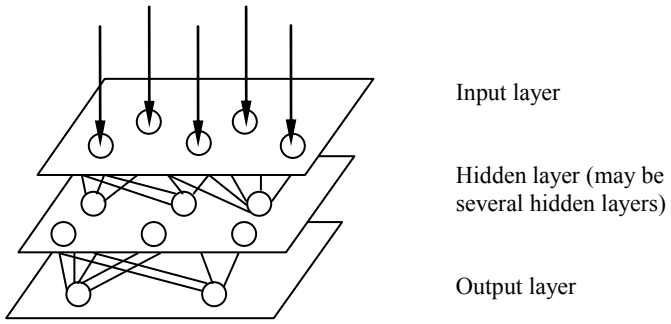

Figure 2-4 Artificial neural network topology (DACS, 1992)

A neural network consists of neurons that are connected in a structured manner like the brain. The functioning of a neural network is dependent on how the neurons are grouped into layers of elements, the connections between these layers, the summation and the transfer functions. Generally three layers form the basis of the network, i.e. the input, hidden and output layer. The input layer receives data from files or from electronic sensors in real-time applications. The information from the output layer is relayed to a secondary computer process or to other process such as mechanical

control systems. The hidden layer which is between the input and output layer can contain many hidden layers which may have various interconnected structures. The signals from the input layer are transmitted to the neurons in the hidden layer. When the neuron performs its function it transmits its output to all the neurons in the layer below it. This process creates a feed forward path to the output. The communication links between the neurons are essential for neural networks as they are the "glue" that keeps the neural network structure in place. They also influence the changing strength to an input. There is a summing mechanism addition by neurons, while the other is used to subtract. In some networks the neuron is used for inhibition within the same layer. This phenomenon is used in the output layer and is referred to as "lateral inhibition". For example, in identifying a character "P" the probability is 0.85, but as "F" is 0.65, the choice of highest probability could be adopted while the others are inhibited. This is also referred to as competition. The architecture is based on the connections between neurons and determines the operation of the network. Some professional software development packages allow the user to prescribe the type of architecture. The range of parameters can be manipulated to excite or inhibit the connections between neurons (DACS, 1992). Another type is a feedback network, where the output layer routes one layer back to a previous layer. An example is shown in Figure 2-5.

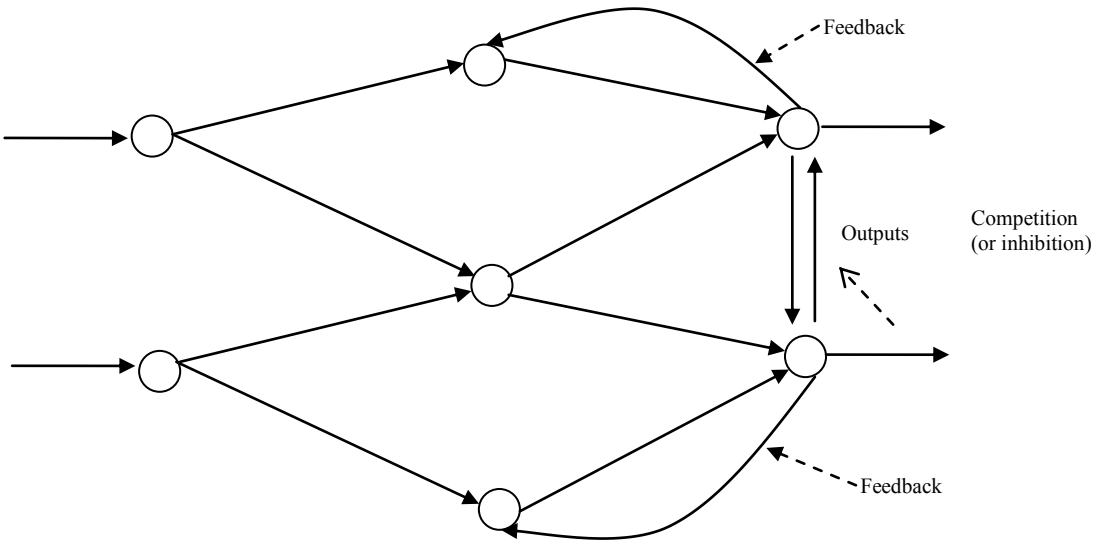

Figure 2-5 Simple network with feedback and competition (DACS, 1992)

Feed forward neural networks are widely applied in the domain of classification, pattern recognition, approximation, forecasting and control. The most prominent models with this

architecture are multilayer perceptrons (MLP) shown in Figure 2-4, radial basis functions (RBFN) and competitive learning models employing "winner-take-all". Extensions and variations of these models exist as: shortcut connections between non-adjacent layers in MLPs, variations in back propagation learning (usually seen as the learning method of choice for MLPs), variations in the basis function for RBFNs (usually seen as the weight setting method of choice for those networks), etc. (DACS, 1992). MLPs are the common name for layered feed forward neural networks. The neurons provide the outputs and there are "hidden units" which form one or more layers of processing units, which link the inputs to the outputs. Classic architectures have one or more layers of "hidden units" and total connectivity between the layers. A standard learning algorithm uses continuously varying transfer functions of which sigmoid transfer functions are most commonly adopted. MLPs are one of the most generally used applied learning models, e.g. used as general classifiers and universal function approximators. For a given problem, the network architecture (number of neurons and number of layers) proposed cannot guarantee that it can perform the task for which it was trained for. Also the objective which is to obtain an accurate generalisation of the new data based on the model is derived from the training data. The network applied depends either on rough guesses as to which might be more appropriate, or an experimental comparison of several types used side by side. Even if the architecture adopted is able to solve the problem, there is no guarantee that it learnt the correct generalisation of the particular data set. Appropriate and optimal application of neural networks for given data sets can only come from adopting a more generalised view which opens up a larger number of variations (*viz*. larger than two or three) and permits a more dedicated use of network solutions (Venkataraman, 2004).

2.5 Properties of artificial neural networks

In order to justify the usefulness of ANNs in the medical field it is necessary to list the general characteristics of ANNs. The role of ANNs may be determined by comparing the characteristics to the needs of predicting survival of diseased patients. ANNs can be used to approximate any non-linear mapping. Hornik *et al* (1989) showed that multi-layer back propagation networks with continuously differentiable activation functions are capable of modelling any continuous non-linear function to an arbitrary degree of accuracy. The degree of non-linearity differs greatly from one system to another; consequently there exists no generally applicable theory or methodology to design non-linear systems. Even though processing elements are connected in many ways, the most popular architecture is the MLP as shown in Figure 2-4. Most ANNs are conceptualised in the layered form (Figure 2-6.) which consists of 3 layers: the input, hidden layer and the output layer.

This graphical representation of a neural net could also be written as a formula involving many summations and transfer functions. This formula maps vectors received at the input layer (must be buffered) and transforms them into vectors in the output values. A MLP can synthesise any function to a desired level of accuracy if given a sufficient number of hidden layers and processing elements.

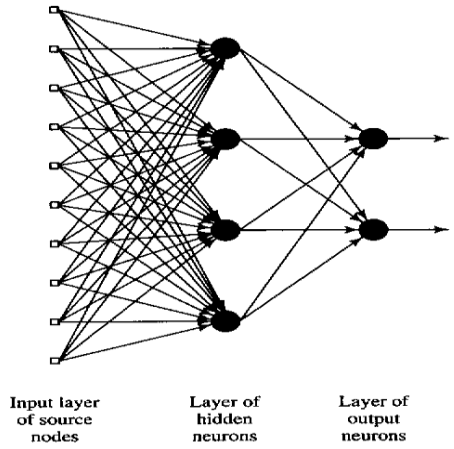

Figure 2-6 Fully connected feed forward network

Normally a single hidden layer is used but the use of more than one layer has been reported by various authors, e.g. Haykin (1994) and Hassoun (1995). By using an appropriate connectionist structure an arbitrary number of hidden layers can be represented by a single layer. The number of neurons in the input and output layers is fixed and based on the nature of the problem under investigation. There is no fixed "rule" for specifying the number of hidden neurons. Most authors will increase the number of hidden neurons until the error of the fully trained network remains constant even when more hidden neurons are added. As the number of hidden neurons increases the computational load is increased since there are more "weights" to be determined. One has to be conservative when specifying the number of hidden neurons, but too few hidden neurons may result in an incomplete mapping by the FFNN of the system under investigation.

There are practical limitations to achieving a high level of accuracy since real world data is typically incomplete and noisy. There would have to be a trade-off between accuracy and the generalization ability of the formula. Like regression, solving for the parameters or weights requires data. This data must consist of a set of input records that have corresponding "target" output records. This relationship between the input and output data provides historical examples which the neural net training algorithm uses to determine and learn the mapping. This combination of PEs,

connections, weights and transfer functions form the network architecture. This architecture then represents a complex mathematical formula that has been derived from historical data.

2.6 Artificial neural network architecture

A vast array of ANNs has been developed for a variety of purposes. Although they differ in structure, implementation and principle of operation, they all share common features. ANNs are computing systems that consist of a number of interconnected signal or information processing units (artificial neurons) which have the following similar features (Venayagamoorthy, 1998). Hardware cannot be separated from the software in the structure because the processing of information and memory is distributed throughout the whole structure. ANNs are trained, rather than programmed to perform particular tasks. Complex interconnections of neurons imply that the state of one neuron affects the potential of the large number of neurons to which it is connected according to the weights (or strengths) of connection. Connection weights (synaptic strengths) are usually adaptive and can take place anywhere in the structure thus allowing for a distribution of memory in ANNs. Processing units (neurons) typically contain non-linear activation functions, i.e. the new state of a neuron is a non-linear function of the signals produced by the firing activity of the other neurons. Networks often use imprecise and unreliable elements but they are characterised by a high degree of robustness (insensitivity) to noisy input data and element failure by the use of a highly redundant distributed structure (DACS, 1992). ANNs demonstrate remarkable robustness since their functionality is not affected by parameter variations over a wide range. The manner in which the neurons of a neural network are interconnected is closely linked with the learning algorithm used to train the network. Learning rules used in the design of neural networks are therefore referred to as being structured (Haykin, 1994). The different architectures of ANNs can be divided into four large categories: single-layer feed forward networks, feed forward (multilayer) networks, feedback (recurrent) networks, and cellular (lattice) networks.

2.6.1. Single-layer feed forward network

A layered neural network is a network of neurons organized in layers. The simplest form is an input layer of source nodes that project onto an output layer of neurons (computation nodes) but not *vice versa*. This is purely a feed forward type. The case of four nodes in both the input and output layer is illustrated in Figure 2-7 (Haykin, 1994).

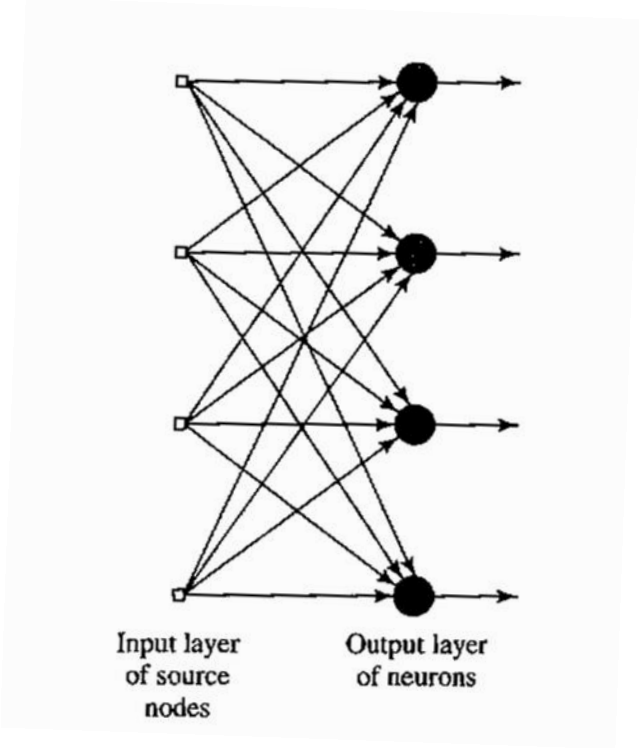

Figure 2-7 Feed forward network with a single layer of neurons

This is called a single-layer network where "single layer" refers to the output layer of computation nodes (neurons). The input layer is not counted because no computation is performed in that layer. In a single-layer neural network the output pattern (vector) is associated with an input pattern (vector) and the results are stored in the network according to the adjustments made to the synaptic weights of the network (Haykin,1994).

2.6.2 Multilayer feed forward networks

"The second type of feed forward neural has one or more hidden layers whose computation nodes are called hidden neurons or units. These hidden neurons or units form one or more hidden layers. Even though the network acquires a global perspective, it is able to obtain high order statistics locally because of its extra synaptic connections and its neural interconnectivity" (Churchland, 1992). "This becomes quite significant when the input layer is large. The source nodes in the input layer of the network transmits elements of the activation pattern (input vector), which constitute the input signals applied to the neurons (computation nodes) in the second layer (i.e. the first hidden layer). The output signal from each layer is used as inputs to the next layer, this cascading effect prevails for the remainder of the network. The output signals of the neurons in the output (final)

layer of the network results from the input vectors first layer. The architectural graph of Figure 2-6 illustrates the layout of a multilayer feed forward neural network for the case of a single hidden layer. The network architecture is referred to as a 10-4-2 network, i.e. it has 10 source nodes, 4 hidden neurons and 2 output neurons. A feed forward network with p source codes, h_1 neurons in the first hidden layer, h_2 neurons in the second layer and q neurons in the output layer will be called a p-h_1-h_2-q network. The network in Figure 2-6 is said to be fully connected since every node in each layer is connected to every other node in the adjacent forward layer. If any of the links (synaptic connections) are removed from the network, then the network is partially connected as illustrated in Figure 2-8" (Haykin, 1994). In a partially connected network each neuron in the hidden layer is connected to a local (partial) set of source codes that surrounds it and is most likely to have some interaction with them. Similarly the output layer is connected to a local set of hidden neurons. The number of source codes, hidden neurons and output neurons are the same in both Figures 2-6 and Figure 2-8. When comparing the two networks the locally connected network of Figure 2-8 has displays a specialized structure. This characterises the classification of the activation pattern. To illustrate this, an activation pattern of a time series (i.e. the sequence of uniformly sampled values of time-varying signal) has been included in Figure 2-8. This shows a spatial pattern over the input layer. Each hidden neuron responds to local changes of the source signal. A feed forward network (FFNN) computes an output pattern in response to an input pattern. Once the network is trained the output response to a given input pattern will be the same regardless of any previous network activity. This network type with fixed connection weights are also known as offline trained ANNs. This implies that the FFNNs do not demonstrate any real dynamics and they do not display any stability problems. Offline trained FFNNs are simplified to a single instantaneous nonlinear mapping. The same applies to online FFNNs where the connection weights are altered to suit the intention.

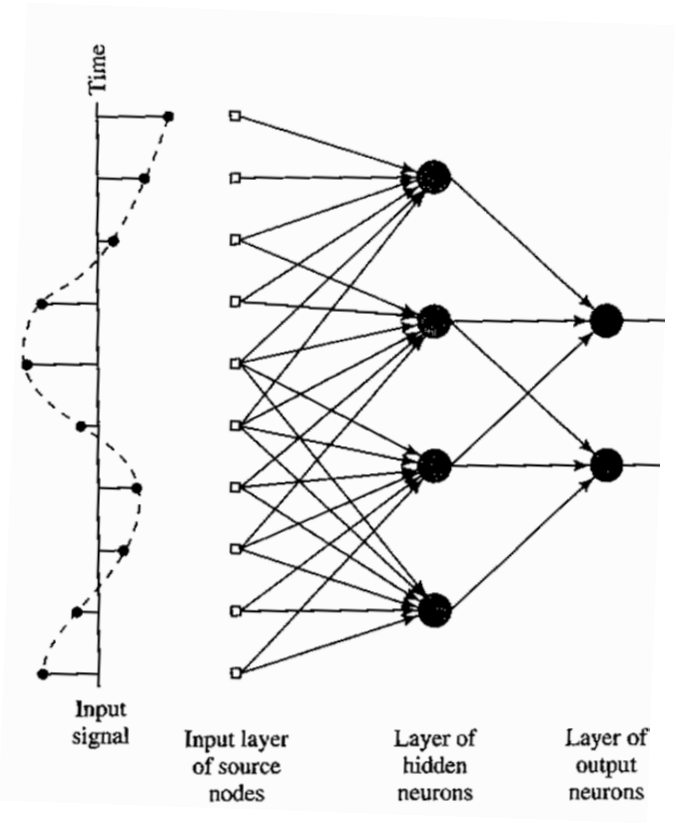

Figure 2-8 Partially connected feed forward networks (Haykin, 1994)

2.6.3. Feedback networks

A recurrent neural network differs from a FFNN in that it has one feedback loop. Recurrent networks may comprise a single layer of neurons with each neuron transmitting its output signal back to the neurons in the input layer as illustrated in Figure 2- 9. There is no self feedback loops in the network, i.e. no output from a neuron is fed back to its own input. The recurrent network in Figure 2-9 has no hidden neurons. Another type of recurrent network with hidden neurons is illustrated in Figure 2-10. Feedback is a result of both the hidden neurons as well as the output neurons. The learning ability of the network and its performance is largely affected by the feedback loop. The feedback loops require specific connectors composed of unit-delay elements(denoted by z^{-1}) which result in a non-linear dynamical behaviour because of the non-linear nature of the neurons. Non-linear dynamics has an effect on the storage function of a recurrent network.

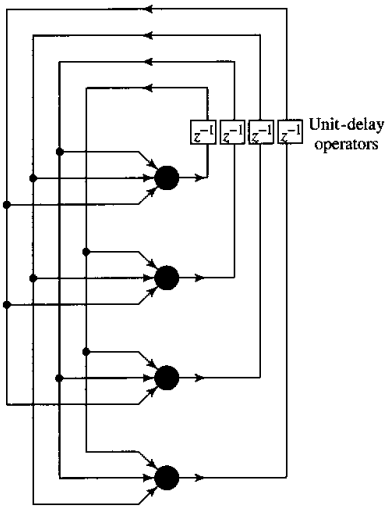

Figure 2-9 Recurrent network with no feedback loops and no hidden neurons (Haykin, 1994)

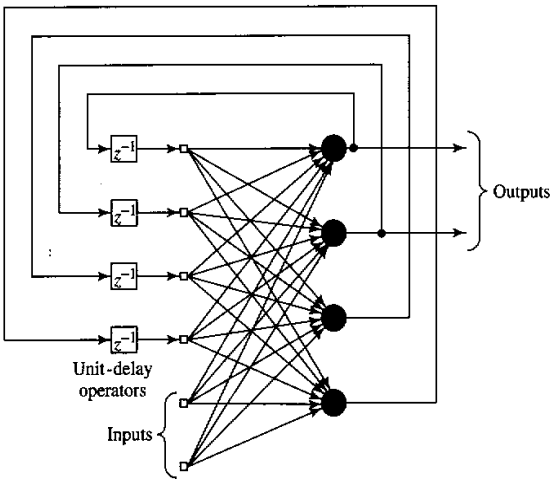

Figure 2-10 Recurrent network with neurons (Haykin, 1994)

2.6.4 Cellular networks

A cellular or lattice network consists of a one-dimensional, two-dimensional or higher-dimensional array of neurons with a corresponding set of source codes that supply the input signals to the array. The dimension of the array refers to the number of dimensions of space in which the graph lies. The architectural graph of Figure 2-11a depicts a one-dimensional lattice of 3 neurons fed from a layer of 3 source codes, Figure 2-11b depicts a two-dimensional lattice of 3-by-3 neurons fed from a layer of 3 source codes. Every neuron is connected to each other in the lattice. A lattice network is really a feed forward network with the output neurons arranged in rows and columns. In this form of local connectivity every cell is excited by its own signals and by signals flowing from its adjacent cells. Due to mutual interactions the processed signals propagate in time within the whole array of the lattice neural network.

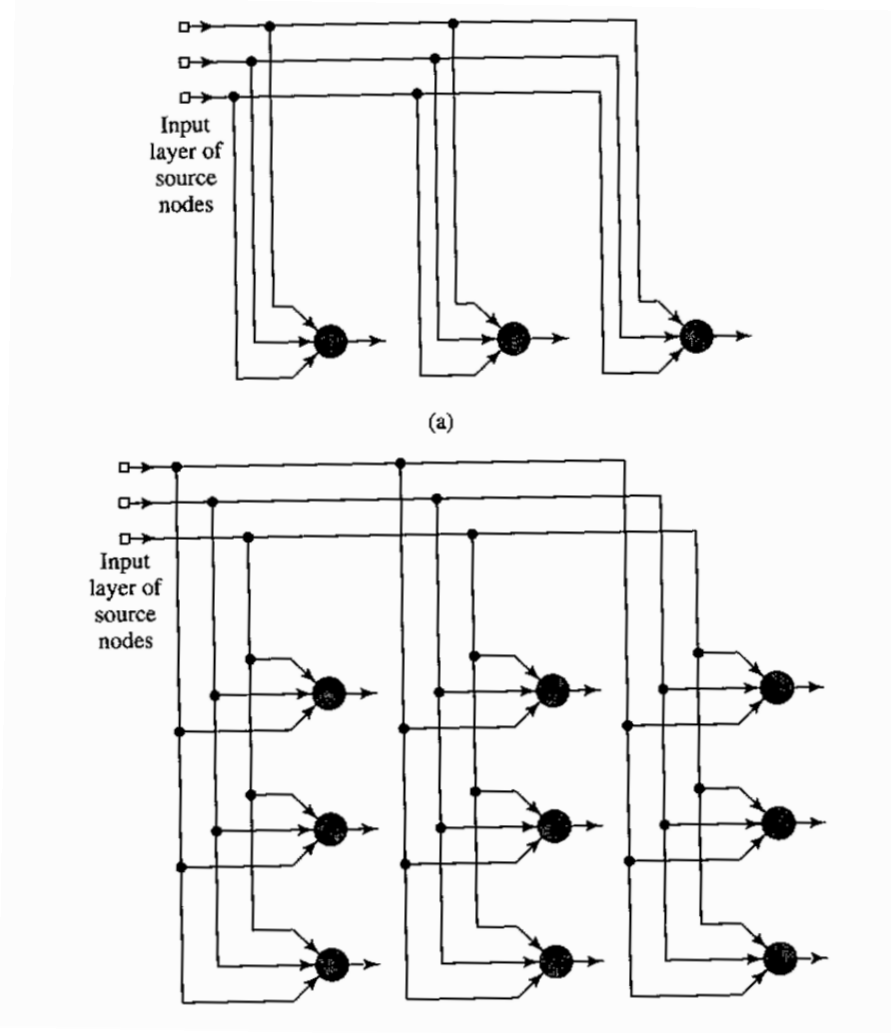

Figure 2-11 (a) One-dimensional lattice of 3 neurons. (b) Two-dimensional lattice of 3-by-3 neurons (Haykin, 1994)

Generally an ANN is characterised not only by its architecture but also by the type of the neurons used, i.e. by the learning (offline or online training) procedure and by the form (principle) of operation. ANNs can operate either as deterministic (signals have deterministic nature) or stochastic systems (signals and parameters or connection weights are changed randomly from time to time with same probability by some random amount). The artificial models (paradigms) of real biological neural networks developed so far are only simple and rather crude approximations of real biological structures. It is not very clear whether it is essential to model exact biological structures or whether only desired properties are sought using models which do not fully correspond to a real biological nervous system.

The following aspects have to be considered when selecting a neural network model: which network type can be applied to what kind of data, which network type does correspond to an algorithm traditionally known from statistics (in order to make use of research results in statistical literature), what do known neural network types have in common and what separates them, and what are the range and limits of the applicability of each network type. The characteristics of a general ANN that need to be specified are the network topology (structure), direction of information flow, computational characteristics of the individual processing elements or nodes and the training rule and methods used to adapt the network (DACS, 1992).

2.7. Feed forward multilayer artificial neural networks

This section describes the type of feed forward ANN which is of interest to the analyses in this thesis. The approach of treating the ANN as a “black box” technique, paying little attention to the mathematical details could lead to poor results. In this section, the architecture, notation and training of feed forward neural networks (FFNNs) will be discussed. FFNNs are a modern form of non-linear statistics. Many FFNNs can be equated with or directly compared to more classical methods like linear regression, logistic regression, nearest neighbour classification, etc. Neural networks must be analyzed in the same way classical statistical algorithms are analysed and compared (Dorffner, 1994). This type of network has an input layer, and one or more hidden layers of computation nodes. The input signal is transmitted through the network on a layer-by-layer basis.

Multilayer perceptrons have been applied successfully to solve a diverse range of problems by training them in a supervised manner with the use of the popular algorithm known as the error back propagation algorithm which is based on the error-correction learning rule. It is a generalisation of an equally popular adaptive filtering algorithm: the ubiquitous (ever-present or omnipresent) least-mean-square algorithm for the special case of a single neuron model. A multilayer perceptron trained with the back propagation algorithm can be used generally for non-linear input-output mapping. If p denotes the number of input (source) nodes of a multilayer perceptron, and q denotes the number of neurons in the output layer of the network, then the input-output relationship of the network defines a mapping from a p-dimensional Euclidean input space to a q-dimensional Euclidean output space, which is infinitely continuously differentiable. The capability of the multilayer perceptron from the perspective of input-output mapping is determined by the minimum number of hidden layers that provides an approximate realization of any continuous mapping. "The universal approximation theorem is an existence theorem in the sense that it provides the mathematical justification for the approximation of an arbitrary continuous function as opposed to exact representation. The theorem states that a single hidden layer is sufficient for a multilayer perceptron to compute a uniform approximation to a given training set represented by the set of inputs x_1, ……, x_p and a desired (target) output $f(x_1, \ldots\ldots, x_p)$. However the theorem does not say that a single layer is optimum in the sense of learning time on ease of implementation" (Haykin, 1994).

The error back-propagation process consists of two passes through the different layers of the network: a backward and a forward pass. An activity pattern (input) vector is applied to the sensory nodes of the network resulting in a signal moving through the network, layer by layer. The resulting outputs are the actual response of the network. The synaptic weights of the network are fixed during the forward pass. The error-correction rule adjusts this during the backward pass. The difference between the actual response of the network and the desired (target) response produces an error signal which is in turn propagated backward through the network against the direction of synaptic connections – hence the name "error back-propagation". Adjustments are made to the synaptic weights so that the actual response of the network approaches the desired response. The error back-propagation algorithm is also called the back propagation algorithm or simply, back-prop. Henceforth, it will be referred to as the back-propagation algorithm. The learning process associated with the algorithm is called back-propagation learning. This concept is illustrated in Figure 2-12.

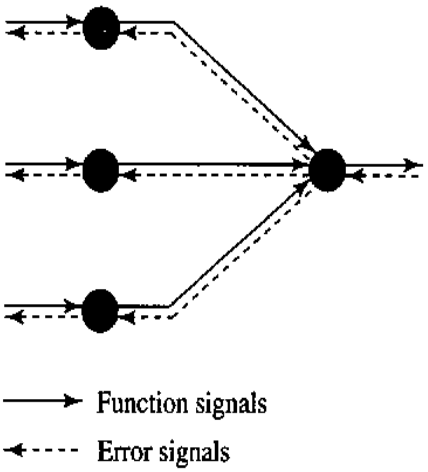

Figure 2-12 Forward propagation of function signals and back-propagation of error signals

A multilayer perceptron has the following three distinctive characteristics. The model of each neuron in the network includes a smooth (i.e. differentiable everywhere) non-linearity at the output end. The commonly used form of the non-linearity that satisfies this requirement is a sigmoidal non-linearity defined by the logistic function:

$$f(z_j) = \frac{1}{1+e^{-z_j}} \tag{2.7}$$

where z is the net internal activity level of neuron j and $f(z_j)$ is the output of the neuron. In the absence of non-linearities the input-output relation of the network can be reduced to that of a single-layer perceptron. One or more layers of hidden neurons are independent of the input or output of the network (Pineda, 1994). "These neurons in the hidden layer are tasked separately to learn complex patterns by looking for progressively more meaningful characteristics of the input patterns (vectors). The synapses of the network are responsible for the high degree of connectivity. If a change in connectivity occurs, the synaptic weights will correspondingly adjust. The above features and its ability to learn from known data during the training procedure, makes the multilayer perceptron gain its computing power. The above features also contribute to the lack of understanding of the network behaviour. Firstly, the theoretical analysis of the non-linearity and the number of interconnections is quite complex. Secondly, the learning process cannot be "seen" because of the neurons being within the hidden layers. In the learning process the important features of the input pattern have to be represented by the hidden neurons. This continues to become more complex as the learning process uses a much larger space of possible functions and decisions have to be made on how the input pattern will be represented" (Hinton, 1989).

"The development of the back-propagation algorithm represents a "landmark" in neural networks in that it provides a computationally efficient method for the training of multilayer perceptrons. It cannot be claimed that the back-propagation algorithm can provide solutions for all solvable problems, but it is fair to say that it has put to rest the pessimism expressed about learning in multilayer machines that may have been inferred" Minsky *et al* (1969).

2.8 Learning process or training

A neural network is able to learn from a known environment and through a series of learning steps is able to improve its performance. There is a measured variable which is used to compare new data continuously in order to improve performance. Learning comprises the use of iterative methods for adjustments which are applied to synaptic weights and thresholds. At the end of each iteration, the network becomes more knowledgeable. Haykin (1994) defines learning in the context of neural networks as follows: "learning is a process by which free parameters of a neural network are adapted through a continuing process of stimulation by the environment in which the network is embedded". Changes in parameter influence the type of learning.

The learning process is a sequential event that starts with stimulation by the environment which results in changes in the neural network and this leads to a new response by the neural network because of the changes that have occurred in its internal structure. A learning algorithm uses a set of well-defined rules for the solution of a learning problem. There is no specific learning algorithm for building neural networks but a "kit of tools" are available in the form of learning algorithms, each of which has its own set of guidelines. They differ from each other in the way in which the synaptic weight is formulated. Also important is the manner in which a neural network relates to its environment. A learning paradigm refers to the model of the environment in which the neural network operates. There are three basic classes of learning paradigms: supervised learning, reinforcement learning and self-organized learning. Supervised learning is performed under the supervision of an external "teacher". The teacher has information and knowledge of the environment that is represented by a set of input-output examples. "If both the teacher and the neural network are exposed to a training vector by virtue of the built-in knowledge of the environment, the teacher is able to provide the neural network with a desired or target response" (Hassoun, 1995). "Reinforcement learning uses a "critic" to advance through a trial and error process. Unsupervised learning is performed in a self-organised manner with no teacher or critic required for instructing synaptic adjustments in the network.

The environment is unknown to the ANN of interest. Network parameters are adjusted under the influence of the training vector and the error signal. The error signal is defined as the difference between the actual response of the network and the desired response" (Haykin, 1994). This adjustment is iteratively done step-by-step so that the neural network will ultimately "emulate" the teacher. The emulation is statistically presumed to be an optimum. The teacher transfers the knowledge of the environment to the ANN as best it can. When this condition is reached, the teacher can be dispensed with thereafter letting the ANN deal with the environment completely by itself, i.e. in an unsupervised fashion. This form of unsupervised learning is a form of error-correction learning which behaves like a closed feedback system with the unknown environment being external to the loop. Any function performed under the teacher‟s supervision is highlighted as a point on the error surface. The operating point has to move down successively towards a minimum point of the error surface as it improves performance. The minimum point may be local or global" (Haykin, 1994). A supervised learning system uses the gradient of the error surface. A vector that points in the direction of steepest descent is represented by the gradient of the error surface. This system uses an instantaneous value of the gradient vector, with the example indices being time. Examples of supervised learning algorithms include the least mean square (LMS) algorithm and the generalization known as the back propagation algorithm (BPA). The LMS algorithm uses a single neuron, whereas the BP algorithm uses a multi-layered interconnection of neurons. In the BPA the error terms in the algorithm are back-propagated through the network on a layer-by-layer basis. The BPA is more powerful in application than the LMS algorithm. Indeed, the BPA includes the LMS algorithm as a special case. A disadvantage of supervised learning is that without a teacher, a neural network cannot find solutions if the examples from the problem were not used to train the network. The BPA uses supervised learning and is currently the most commonly used algorithm for the design of multi-layer feed forward networks.

There are two distinct phases to the operation of back propagation learning (BPL): the forward phase and the backward phase. Signals are transmitted from the input layer through the network layer-by-layer, resulting in some response at the output of the network. The difference between the actual response and the desired target response is compared; error signals are then generated and accordingly propagated backwards through the network, hence the name "error back propagation". The aim of the backward phase is to allow the free parameters of the network to be adjusted so as to minimize the sum of squared errors. The synaptic weights (adjusted using the error correction rule) are adjusted to make the actual response of the network move closer to the desired response. The

BPA has a stochastic tendency to zigzag its way about the true direction to a minimum on the error surface. The learning algorithm is called the back propagation learning. The „back propagation" appears to have evolved after 1985. The basic idea was first described by Werbos (1974a) in his PhD thesis in the context of generated networks with neural networks representing a special case.

Back propagation learning (BPL) has been applied successfully to solve some difficult problems such as speech recognition from text, handwritten digit recognition and adaptive control. Unfortunately, BPL may be limited by its poor scaling behaviour. To understand this limitation, consider the example of a multi-layered feed-forward network consisting of L computation layers. The effect of the synaptic weight in the first layer on the output depends on its interactions with approximately f_i^L other synaptic weights, where F_i is the fan-in, defined as the average number of incoming links of neurons in the network. F_i or L, or both will increase as the size of the network increases. The network becomes computationally intensive and the time taken to train the network grows exponentially resulting in a learning process becoming unacceptably slow (Haykin, 1994).

A learning procedure is governed by the learning tasks which a neural network is required to perform. The following has been identified as learning tasks that befit a neural network in one form or another:

- Approximation: A non-linear input-output mapping can be described by the functional relationship $d = f(x)$ where the vector x is the input and the scalar d is the output. The function $f(\cdot)$ is assumed to be unknown. The aim is to design a neural network that approximates the nonlinear function $f(\cdot)$, given a set of examples denoted by the input-output pairs (x_1, d_1), (x_2, d_2), (x_N, d_N). This approximation is an example of supervised learning with x_i serving as the input vector and d_i serving the role of desired response, where $i = 1, 2, 3 \ldots N$.
- Association: The two forms are auto-association and hetero-association. In auto-association a set of patterns (vectors) stored by the neural network has to be repeatedly presented to the network. Subsequently, the network is presented with a partial description or distorted (noisy) version of the original pattern stored in it, and the task is to retrieve (recall) that particular pattern. Hetero-association differs from auto-association in that an arbitrary set of input patterns (vectors) are paired with another arbitrary set of output patterns (vectors). Auto-association involves the use of unsupervised learning, while in hetero-association supervised learning occurs.

- Pattern classification: In this task there are a fixed number of categories into which activations are to be classified. During the training session the set of input patterns are categorized according to its pattern. When a new pattern is presented to the network, it is an unseen pattern, but it belongs to the same population of patterns used to train the network. The task for the neural network is to classify the new pattern correctly. Pattern classification as described here is a supervised learning task. Pattern classification allows neural networks to construct nonlinear decision boundaries between the different classes in a nonparametric fashion, thereby offering a practical method for solving highly complex pattern classification problems.
- Prediction: Predicting is one of the most basic learning tasks. Given a set of *M* past samples *x(n-1), x(n-2)… x(n-M)* which are uniformly spaced in time, the requirement is to predict the present sample *x(n)*. Prediction may be solved using error-correction learning in an unsupervised manner in the sense that the training examples are drawn directly from the time series itself. Specifically, the sample *x(n)* serves the purpose of the desired response; hence, given the corresponding prediction $\hat{x}(n)$ produced by the network on the basis of the previous samples *x(n-1), x(n-2),…., x(n-M)*, the prediction error can be computed by the term $e(n) = x(n) - \hat{x}(n|n-1,\ldots,n-M)$ and the error-correction learning can be used to modify the free parameters of the network. Prediction can be viewed as a form of model building where the smaller the prediction error in a statistical sense, the better the network will serve as a physical model of the underlying stochastic process responsible for the generation of the time series.
- Control: Process control is a learning task of a neural network that is similar to the actions performed by the human brain. The human brain is a computer (i.e. information processor), and the outputs of this whole system are actions. In the context of control, the brain is living proof that a generalized controller can take full advantage of parallel distributed hardware that can handle many thousands of actuators (muscle fibres) in parallel, non-linearity and noise, and that can optimize over a long-range planning horizon.
- Beam-forming: Beam-forming is a form of spatial filtering, the purpose of which is to locate a target signal embedded in a background of additive interference. The adaptive beam-forming task operates in an unsupervised manner to “clean-up” background noise.

The learning task used in this study is prediction where the aim is to minimise the prediction error. As in the human brain, a neural network needs to be trained before it can be used. Training is achieved *via* the back propagation algorithm. The back propagation algorithm suffers from a slow

convergence property. The derivative of the error surface with respect to the weight is small in magnitude therefore the adjustment applied to the weight is small. A large number of iterations of the algorithm have to be computed to get a significant reduction in the error performance of the network. Also, the direction of the negative gradient vector (i.e. the negative derivative of the cost function with respect to the vector of weights) may point away from the minimum of the error surface, hence the adjustments applied to the weights may induce the algorithm to move in the wrong direction. Any procedures introduced to increase the rate of convergence must maintain the locality constraint that is an inherent characteristic of back-propagation learning.

2.9 Mathematical derivation of the back propagation algorithm

The derivation of the back-propagation algorithm is rather involved; therefore to ease the mathematical burden involved in this derivation, a summary of the notation is presented:

- "The indices *i, j* and *k* refer to different neurons in the network; with signals propagating through the network from left to right. Neuron *j* lies in a layer to the right of neuron *i*, and neuron *k* lies in a layer to the right of neuron *j* when neuron *j* is a hidden unit.
- The iteration n refers to the nth training pattern (example) presented to the network.
- The symbol $\varepsilon(n)$ refers to the instantaneous sum of error squares at iteration n. The average of $\varepsilon(n)$ over all values of *n* (i.e. the entire training set) yields the average squared error ε_{av}.
- The symbol $e_j(n)$ refers to the error signal at the output of neuron *j* for iteration *n*.
- The symbol $d_j(n)$ refers to the desired response for neuron *j* and is used to compute $e_j(n)$.
- The symbol $y_i(n)$ refers to the signal appearing at the output of neuron *j* at iteration *n*.
- The symbol $w_{ji}(n)$ denotes the synaptic weight connecting the output of neuron *i* to the input of neuron *j* at iteration *n*. The correction applied to this weight at iteration *n* is denoted by $\Delta w_{ji}(n)$.
- The net internal activity level of neuron *j* at iteration n is denoted by $v_j(n)$; it constitutes the signal applied to the non-linearity associated with neuron *j*.
- The activation function describing the input-output functional relationship of the non-linearity associated with neuron *j* is denoted by $\varphi_j(\cdot)$.
- The threshold applied to neuron *j* is denoted by θ_j; its effect is represented by a synapse of weight $w_{jo} = \theta_j$ connected to a fixed input equal to –1.
- The *ith* element of the input vector (pattern) is denoted by $x_i(n)$.
- The *kth* element of the overall output vector(pattern) is denoted by $o_k(n)$
- The learning rate parameter is denoted by η" (Haykin, 1994 and Hassoun, 1995).

A perceptron (neuron) is an information processing unit that is fundamental to the operation of a neural network. The perceptron is the building block in the neural net and has a number of inputs, $x_1, x_2, \ldots, x_p$, and one output y_k, where k indexes the associated perceptron. A basic element of a neuron model is a set of synapses or connecting links, each of which is characterized by a weight or strength of its own. Specifically a signal x_j at the input of synapse j connected to a neuron k is multiplied by the synaptic weight. In mathematical terms, a neuron k can be described by the following pair of equations:

$$u_k = \sum_{j=1}^{p} w_{kj} x_j \tag{2.8}$$

and

$$y_k = \varphi(u_k - \theta_k) \tag{2.9}$$

where w_{k1}, w_{k2}, w_{kp} are the synaptic weights of the neuron; u_k is the linear combiner output; θ_k is the threshold; $\varphi(\cdot)$ is the activation function; and y_k is the output signal of the neuron. The use of threshold θ_k has the effect of applying an affine transformation to the output u_k of the linear combiner of Figure 2.2 as shown by

$$v_k = u_k - \theta_k \tag{2.10}$$

The threshold θ_k is an external parameter of artificial neuron k. From

$$v_k = \sum_{j=0}^{p} w_{kj} x_j \tag{2.11}$$

and

$$y_k = \varphi(v_k) \tag{2.12}$$

a new synapse has been added to equation (2.11), whose input is

$$x_0 = -1 \tag{2.13}$$

and whose weight is

$$w_{k0} = \theta_k \tag{2.14}$$

The model of neuron k can now be reformulated as in Figure 2-13 (Haykin, 1994). The effect of the threshold is represented by doing two things: (1) adding a new input signal fixed at –1, and (2) adding a new synaptic weight equal to the threshold θ_k. Alternately $x_0 = +1$ and weight $w_{k0} = b_k$ accounts for the bias b_k. Both models are different in appearance, but they are mathematically

equivalent. The back propagation algorithm with supervised learning technique is used in this derivation. (David Black, 2005).

Consider one neuron *j*, where $d_j(n)$ denotes some desired response at time n. The corresponding value of the actual response of this neuron is denoted by $y_j(n)$ (David Black, 2005). The response $y_j(n)$ is produced by a stimulus (vector) $x(n)$ applied to the input of the network in which *j* is buried. The input vector $x(n)$ and the desired response $d_j(n)$ for neuron *j* constitute a particular example presented to the network at time *n*. The actual response $y_j(n)$ of neuron *j* is different from the desired response $d_j(n)$. The error signal is therefore defined as the difference between the target response $d_j(n)$ and the actual response $y_j(n)$ as defined by

$$e_j(n) = d_j(n) - y_j(n) \tag{2.15}$$

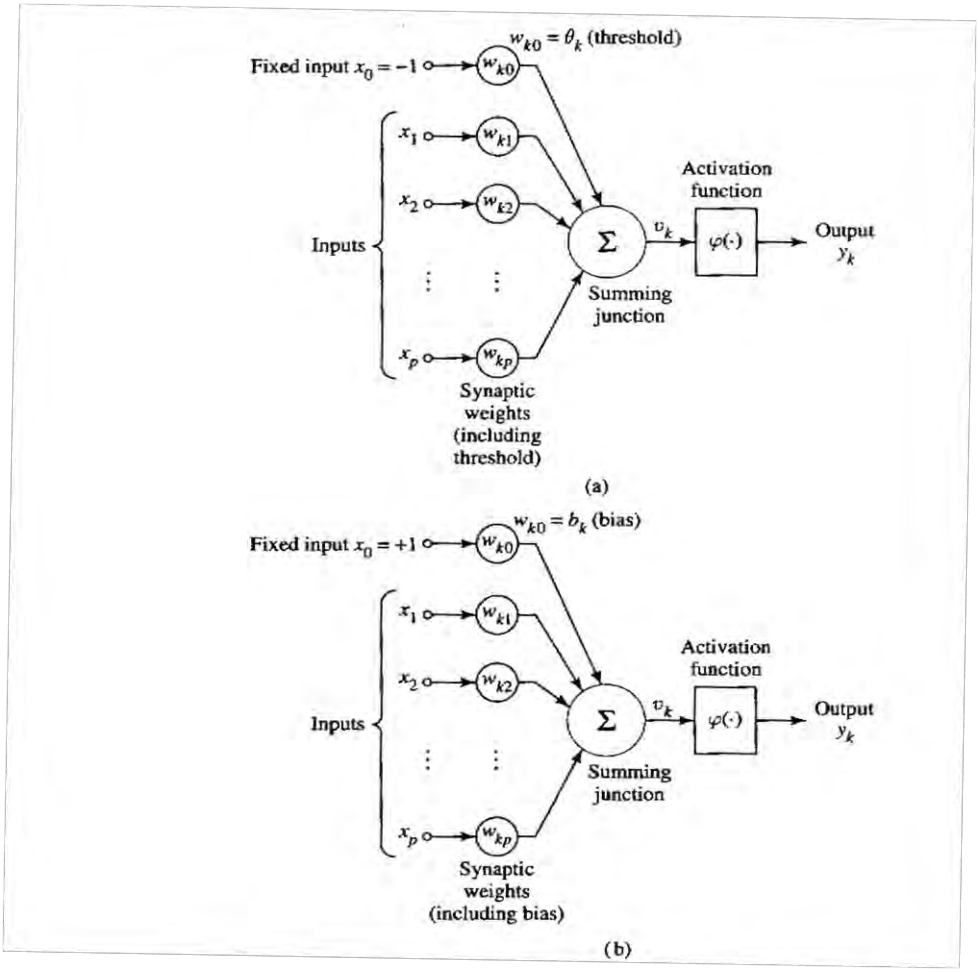

Figure 2-13 Non-linear model of a neuron with bias

The ultimate purpose of the error-correction learning is to minimize a cost function based on the error signal $e_j(n)$. The instantaneous value of the squared error for neuron j is defined as $\frac{1}{2} e_j^2(n)$. Correspondingly, the instantaneous value $\varepsilon(n)$ of the sum of squared errors is obtained by summing $\frac{1}{2} e_j^2(n)$ over all neurons in the output layer. These are the only "visible" neurons for which error signals can be calculated. The instantaneous sum of squared errors of the network is thus written as

$$\varepsilon(n) = \frac{1}{2}\sum_j e_j^2(n) \tag{2.16}$$

Consider Figure 2-14 (Haykin, 1994) which depicts neuron *j* being fed by a set of function signals produced by a layer of neurons to its left. The net internal activity level $v_j(n)$ produced at the input of the non-linearity associated with neuron *j* is therefore

$$v_j(n) = \sum_{i=0}^{p} w_{ji}(n) y_i(n) \tag{2.17}$$

where p is the total number of inputs (excluding the threshold) applied to neuron j. The synaptic weight w_{j0} (corresponding to the fixed input $y_0 = -1$) equals the threshold θ_j applied to neuron j. Hence the function signal $y_j(n)$ appearing at the output of neuron j at iteration n is

$$y_j(n) = \varphi(v_j(n)) \tag{2.18}$$

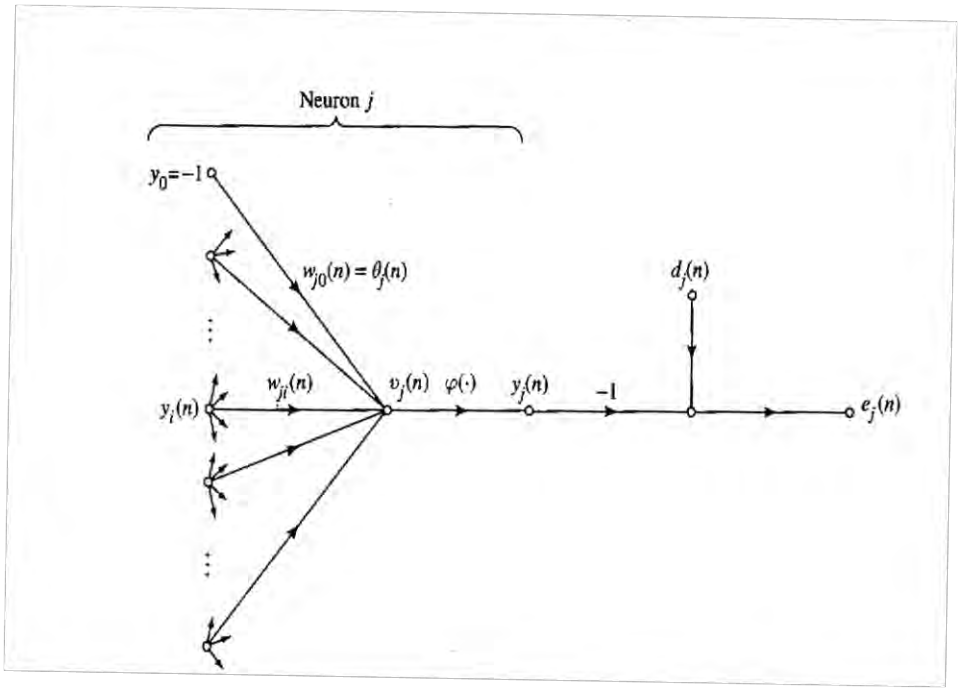

Figure 2-14 Signal flow graph highlighting the details of neuron j

The back propagation algorithm applies a correction $\Delta w_{ji}(n)$ to the synaptic weight $w_{ji}(n)$, which is proportional to the instantaneous gradient $\frac{\partial \varepsilon(n)}{\partial w_{ji}(n)}$. Applying the chain rule, the gradient becomes

$$\frac{\partial \varepsilon(n)}{\partial w_{ji}(n)} = \frac{\partial \varepsilon(n)}{\partial e_j(n)} \frac{\partial e_j(n)}{\partial y_j(n)} \frac{\partial y_j(n)}{\partial v_j(n)} \frac{\partial v_j(n)}{\partial w_{ji}(n)} \tag{2.19}$$

Differentiating both sides of equation (2.16) with respect to $e_j(n)$

$$\frac{\partial \varepsilon(n)}{\partial e_j(n)} = e_j(n) \tag{2.20}$$

Differentiating both sides of equation (2.17) with respect to $y_j(n)$

$$\frac{\partial e_j(n)}{\partial y_j(n)} = -1 \tag{2.21}$$

Next differentiating equation (2.18) with respect to $v_j(n)$

$$\frac{\partial y_j(n)}{\partial v_j(n)} = \varphi_j'(v_j(n)) \tag{2.22}$$

where the use of prime (on the right hand side) signifies differentiation with respect to the argument. Finally differentiating equation (2.17) with respect to $w_{ji}(n)$ yields

$$\frac{\partial v_j(n)}{\partial w_{ji}(n)} = y_i(n) \tag{2.23}$$

Hence, the use of equations (2.20) to (2.23) in (2.19) yields

$$\frac{\partial \varepsilon(n)}{\partial w_{ji}(n)} = -e_j(n)\varphi_j'(v_j(n))y_j(n) \tag{2.24}$$

The correction $\Delta w_{ji}(n)$ applied to $w_{ji}(n)$ is defined by the delta rule

$$\Delta w_{ji}(n) = -\eta \frac{\partial \varepsilon(n)}{\partial w_{ji}(n)} \tag{2.25}$$

where η is a constant that determines the rate of learning it is called the learning rate parameter of the back propagation algorithm. The use of the minus sign accounts for the gradient descent in weight space.

The above algorithm is modified to accommodate a new cost function $E(n)$ which is mathematically similar to $\varepsilon(n)$, but the parameter space pertaining to the new cost function $E(n)$ is assumed to consist of different learning rates. The cost function is defined as the instantaneous value of the sum of squared errors,

$$E(n) = \frac{1}{2}\sum_j e_j^2(n) = \frac{1}{2}\sum_j \left[d_j(n) - y_j(n)\right]^2 \tag{2.26}$$

Let $\eta_{ji}(n)$ denote the learning rate parameter assigned to synaptic weight $w_{ji}(n)$ at iteration number n. Applying the chain rule to E(n),

$$\frac{\partial E(n)}{\partial \eta_{ji}(n)} = \frac{\partial E(n)}{\partial y_j(n)} \frac{\partial y_j(n)}{\partial v_j(n)} \frac{\partial v_j(n)}{\partial \eta_{ji}(n)} \tag{2.27}$$

Substituting equation (2.25) into (2.17)

$$v_j(n) = \sum_i y_i(n)\left[w_{ji}(n-1) - \eta_{ji}(n)\frac{\partial \varepsilon(n-1)}{\partial w_{ji}(n-1)} \right] \tag{2.28}$$

Hence differentiating equation (2.28) with respect to $\eta_{ji}(n)$ and rewriting equation (2.18)

$$\frac{\partial v_j(n)}{\partial \eta_{ji}(n)} = -y_i(n)\frac{\partial \varepsilon(n-1)}{\partial w_{ji}(n-1)} \tag{2.29}$$

$$\frac{\partial y_j(n)}{\partial v_j(n)} = \varphi_j'(v_j(n)) \tag{2.30}$$

The partial derivative $\frac{\partial E(n)}{\partial y_j(n)}$ is evaluated. For the case when the neuron *j* lies in the output layer of the network, the desired response $d_j(n)$ is supplied externally. Differentiating equation (2.26) with respect to $y_j(n)$ results in

$$\frac{\partial E(n)}{\partial y_j(n)} = -(d_j(n) - y_j(n)) = -e_j(n) \tag{2.31}$$

where $e_j(n)$ is the error signal. Thus, using the partial derivatives of equations (2.29), (2.30) and (2.31) in (2.27), and then rearranging terms

$$\frac{\partial E(n)}{\partial \eta_{ji}(n)} = -\varphi_j'(v_j(n))e_j(n)y_j(n)\left[\frac{\partial \varepsilon(n-1)}{\partial w_{ji}(n-1)}\right] \tag{2.32}$$

The partial derivative $\frac{\partial \varepsilon(n-1)}{\partial w_{ji}(n-1)}$ on the right hand side of equation (2.32) refers to the cost function $\varepsilon(n-1)$ describing the error surface at time *n-1*; the differentiation is with respect to synaptic weight $w_{ji}(n-1)$. From equation (2.24) the factor $-e_j(n)\varphi_j'(v_j(n))y_j(n)$ equals the partial derivative $\frac{\partial \varepsilon(n)}{\partial w_{ji}(n)}$. Using this relation in equation (2.32), $\frac{\partial E(n)}{\partial \eta_{ji}(n)}$ is redefined simply as

$$\frac{\partial E(n)}{\partial \eta_{ji}(n)} = -\frac{\partial \varepsilon(n)}{\partial w_{ji}(n)}\frac{\partial \varepsilon(n-1)}{\partial w_{ji}(n-1)} \tag{2.33}$$

Equation (2.33) defines the derivative of the error surface with respect to the learning-rate parameter $\eta_{ji}(n)$, assuming that neuron *j* lies in the output layer of the network. It can be shown that this same formula also applies to a neuron *j* that lies in a hidden layer of the network. This implies that equation (2.33) applies to all neurons in the network. A learning-rate update rule can now be formulated to perform steepest descent on the error surface over the parameter space, where the parameter of interest is the learning-rate parameter $\eta_{ji}(n)$ as

$$\begin{aligned}\Delta\eta_{ji}(n+1) &= -\gamma\frac{\partial E(n)}{\partial \eta_{ji}(n)} \\ &= \gamma\frac{\partial \varepsilon(n)}{\partial w_{ji}(n)}\frac{\partial \varepsilon(n-1)}{\partial w_{ji}(n-1)}\end{aligned} \tag{2.34}$$

where γ is a positive constant, called the control step-size parameter for the learning-rate adaptation procedure. The partial derivatives $\frac{\partial \varepsilon(n-1)}{\partial w_{ji}(n-1)}$ and $\frac{\partial \varepsilon(n)}{\partial w_{ji}(n)}$ refer to the derivative (negative gradient) of the error surface with respect to the synaptic weight $w_{ji}(n)$ (connecting neuron i to neuron j), evaluated at iterations n-1 and n, respectively. There are two important observations regarding the learning rate. The first is when the derivative of the error surface with respect to the weight w_{ji} has the same algebraic sign on two consecutive iterations and the adjustment $\Delta\eta_{ji}(n+1)$ has a positive value. The adaptation procedure therefore increases the learning-rate parameter for the weight w_{ji}. Correspondingly, the back propagation learning along the direction will be fast. The second is when the derivative of the error surface with respect to the weight w_{ji} alternates on two consecutive iterations, the adjustment $\Delta\eta_{ji}(n+1)$ assumes a negative value. The adaptation procedure decreases the learning-rate parameter for the weight w_{ji} and correspondingly, the back propagation learning along that direction will be slow. The back propagation algorithm gives an "approximation" to the trajectory in weight space computed by the method of steepest descent. For small values of the learning-rate parameter η, there is a small change to the synaptic weights in the network for each iteration as it proceeds to the next. This will result in a smooth trajectory through space. For larger learning-rates the large changes in the synaptic weights lead to an unstable network (i.e. oscillatory). The learning rate can be increased without causing instability, by modifying the delta rule. A momentum term can be added to control the feedback loop around $\Delta w_{ji}(n)$. A value between 0 and 1 is normally chosen. "It can also be viewed as a way of increasing the effective learning rate in almost-flat regions of the error surface while maintaining a specified learning rate in regions of high fluctuations. One complete representation of the entire training set during the learning process is called an epoch. The learning process proceeds on an epoch-by-epoch basis until the synaptic weights and the threshold levels of the network stabilize and the average squared error over the entire training set converges to some minimum value" (Hassoun, 1995). The back propagation algorithm cannot, in general be shown to converge, nor are there well-defined criteria for determining the stopping point. A few considerations can be taken into account to terminate the weight adjustments. The following three criteria can be used to determine the convergence of the back propagation learning (Haykin, 1994). The Euclidean norm of the gradient vector reaches a sufficiently small gradient threshold, the absolute rate of change in the average squared error per epoch is sufficiently small or the algorithm is terminated at the weight vector w_{final} when $||g(w_{final})|| \leq \varepsilon$, where ε is a sufficiently small gradient threshold, or $\varepsilon_{av}(w_{final}) \leq \tau$, where τ is a sufficiently small error energy threshold. The architecture for the back propagation learning for both

the feed forward and backward phases of the computations involved in the learning process is presented in Figure 2-15.

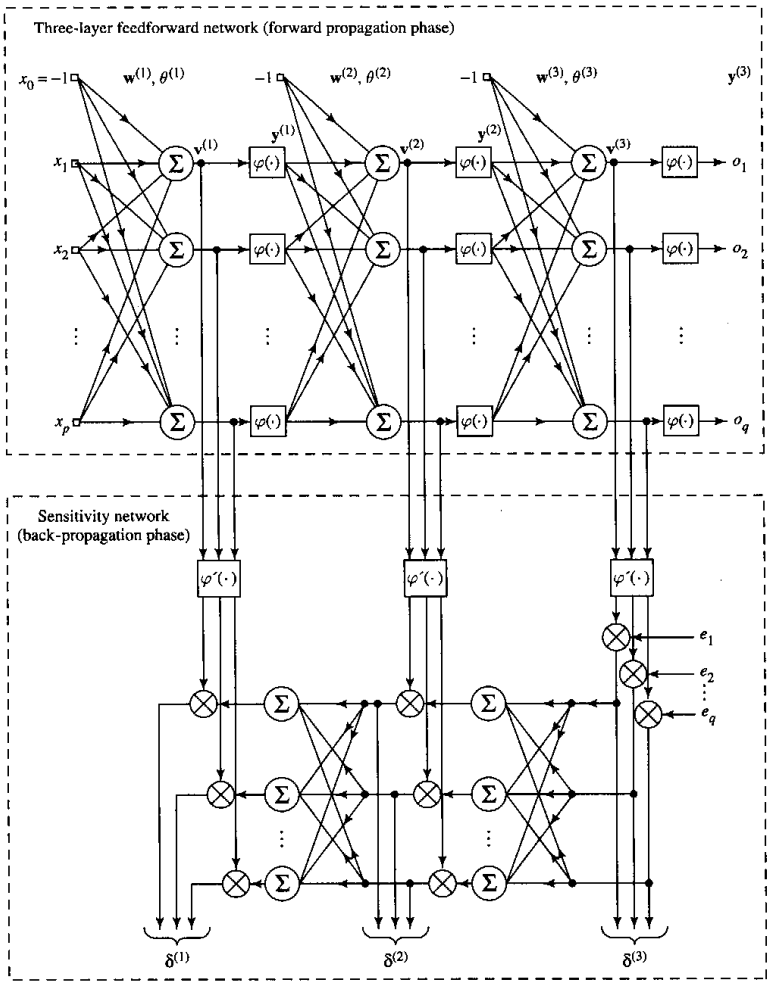

Figure 2-15 Architectural graph of a three-layered feed forward network and associated sensitivity network (back propagating error signals) (Haykin, 1994)

The multilayer network that is shown in the top part of Figure 2-15 accounts for the forward phase. The notations used in Figure 2-15 are as follows:

$w^{(l)}$ = synaptic weight vector of a neuron in layer l

$\theta^{(l)}$ = threshold of a neuron in layer l

$v^{(l)}$ = vector of net internal activity levels of neurons in layer l

$y^{(l)}$ = vector of function signals of neurons in layer l

$\delta^{(l)}$ = vector of local gradients of neurons in layer l

e = error vector represented by $e_1, e_2, \ldots, e_n$ as elements

The layer index *l* extends form the input layer ($l = 0$) to the output layer ($l = L$). $L = 3$ and is referred to as the depth of the network. The bottom part of the figure accounts for the backward phase, which is referred to as the sensitivity network for computing the local gradients in the back propagation algorithm. The pattern-by-pattern method of updating the weights is the preferred method for the on-line implementation of the back propagation algorithm. For this mode of operation, the algorithm cycles through the training data *{[x(n),d(n)]; n = 1, 2, ..., N}* as follows:

- "Initialisation: Starting with a reasonable network configuration, all the synaptic, weights and threshold levels of the network are set to small random numbers that are uniformly distributed.
- Presentations of training examples: The network is presented with an epoch of training examples. For each example in the set the following two sequences of forward and backward computations are performed.
- Forward computation: A training example in the epoch denoted by [*x(n), d(n)*] with an input vector *x(n)* is applied to the input layer of sensory nodes, and the desired response vector *d(n)* presented to the output layer of computation nodes. The activation potentials and function signals of the network is computed by proceeding forward through the network, layer by layer. The net internal activity level $v_j^{(l)}$ for neuron *j* in layer *l* is

$$v_j^{(l)}(n) = \sum_{i=0}^{p} w_{ji}^{(l)}(n)\, y_i^{(l-1)}(n) \tag{2.35}$$

where $y_i^{(l-1)}(n)$ is the function signal of neuron *i* in the previous layer *l*-1 at iteration *n* and $w_{ji}^{(l)}(n)$ is the synaptic weight of neuron *j* in layer *l* that is fed from neuron *i* in layer *l*-1. For $i = 0$, $y_0^{(l-1)}(n) = -1$ and $w_{j0}^{(l)}(n) = \theta_j^{(l)}(n)$, where $\theta_j^{(l)}(n)$ is the threshold applied to neuron *j* in layer *l*" (Haykin, 1994). A logistic function is used for the sigmoidal nonlinearity, where the function (output) signal of neuron *j* in layer *l* is

$$y_i^{(l)}(n) = \frac{1}{1 + \exp(-v_j^{(l)}(n))} \tag{2.36}$$

If neuron *j* is in the hidden layer (i.e., $l = 1$) then the following is set

$$y_j^{(0)}(n) = x_j(n) \tag{2.37}$$

where $x_j(n)$ is the *jth* element of the output vector *x(n)*. If neuron *j* is in the output layer (i.e., $l = L$), then the following is set

$$y_j^{(L)}(n) = o_j(n) \tag{2.38}$$

The error signal is computed

$$e_j(n) = d_j(n) - o_j(n) \tag{2.39}$$

where $d_j(n)$ is the *jth* element of the desired response vector $d(n)$.

- Backward computation: The δs (i.e. the local gradients) of the network are computed by proceeding backward, layer by layer. For neuron j in output layer L

$$\delta_j(n) = e_j(n) o_j(n)[1 - o_j(n)] \tag{2.40}$$

For neuron j in hidden layer l

$$\delta_j(n) = y_i(n)[1 - y_i(n)] \sum_k \delta_k^{l+1}(n) w_{kj}^{(+1)}(n) \tag{2.41}$$

The synaptic weights of the network are adjusted in layer l according to the generalized delta rule:

$$w_{ji}(n+1) = w_{ji}^l(n) + \alpha[w_{ji}(n) - w_{ji}(n-1)] + \eta\delta_j(n) y_i^{(-1)}(n) \tag{2.42}$$

where η is the learning rate parameter and α is the momentum constant.

- "Iteration: The computation is iterated by presenting new epochs of training examples to the network until the free parameters of the network stabilize their values and the average squared ε_{av} computed over the entire training set is at a minimum or acceptably small. The order of presentation of training examples should be randomised from epoch to epoch. The learning rate parameter and the momentum are adjusted (and usually decreased) as the number of training iterations increase (Haykin, 1994).

2.10 Generalization and overtraining

In back propagation learning, the synaptic weights are computed by loading (encoding) as many of the training examples as possible into the network. The goal is to get the network to generalize, i.e. when the input-output relationship computed by the network is correct (or nearly so) for the input/output patterns (test data) never used in creating or training the network. The term "generalisation" is adopted from psychology. In this model it is assumed that the test data are drawn from the same population used to generate the training data. The learning process, i.e. the training of the network may be viewed as a "curve fitting" problem. The network itself will be measured simply as a nonlinear input-output mapping. Generalisation then becomes the result of a good nonlinear interpolation of the input data. As the network is trained it uses interpolation because multilayer perceptrons with continuous activation functions generate output functions that are also continuous.

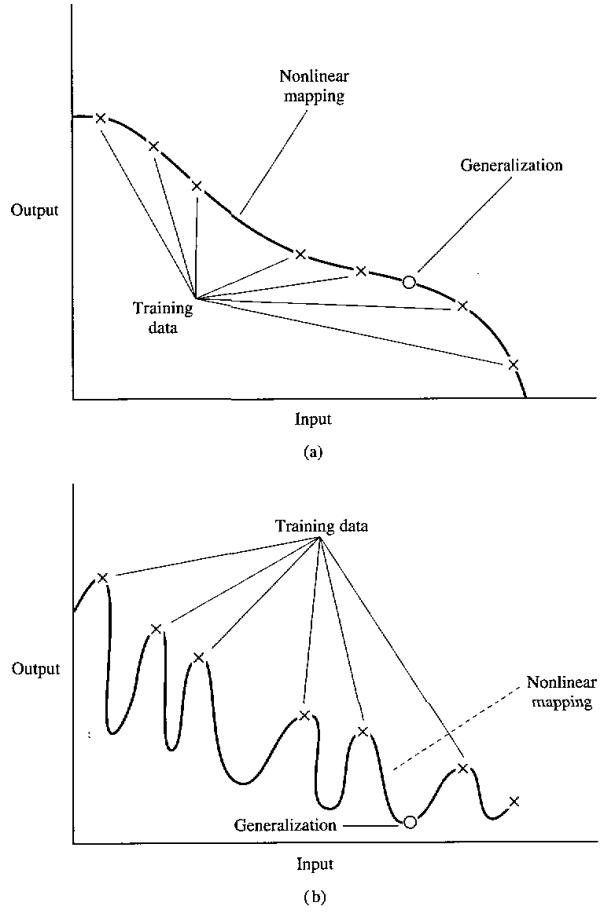

Figure 2-16 (a) Properly fitted data (good generalization) (b) Over-fitted data (poor generalization (Haykin, 1994)

Figure 2-16 illustrates how generalization may occur in a hypothetical network. The curve represents the nonlinear mapping of points learnt during the "training period." The points marked generalization are the result of the network interpolation. A neural network designed to generalise well will adapt accordingly even if the data is slightly different from the examples as illustrated in Figure 2-16 (a). When, however, a neural network learns too many specific input-output relations (i.e. it is over-trained), the network may memorise training data and therefore be less able to generalise between similar input-output patterns. Normally more than the required hidden neurons are loaded into a multilayer perceptron, resulting in unintended curves being stored in the synaptic weights of the network. This poor generalization due to memorisation is depicted in Figure 2-16 (b). "Memorisation" is essentially a "look-up table" which implies that the neural network is not "smooth," i.e. the input-output mapping computes a non-realistic view of the population being studied. There is the danger that the NN model will not learn the „general" behaviour of the system.

It may learn to mimic only the actual data set that is used for training. Instead of creating a model to predict outputs for new sets of inputs it learns to associate actual outputs with a specific set of input data (Haykin, 1994). When a new data set is presented to the model it incorrectly tries to apply the same associations from the training set resulting in inaccurate predictions for the outputs. This may occur if the data for the training does not have many records or if there is very little variation in the variables over the training period. This may also occur if there are too many hidden nodes in the proposed NN model. The model is created from the training set of data where too many weights are calculated to approximate the system and the „fit" of the model ends up being too tight.

Overtraining can be overcome using commercially available software which has built-in functions. Guidelines are presented for sensible predictions on the number of hidden nodes to use and the number of training iterations to perform. Overtraining can be monitored by observing the model error (MSE).

2.11 Size of training set

The size and efficiency of the training set, the architecture of the network, and the physical complexity of the problem influences the generalisation of a network. The latter cannot be controlled but the first two factors can be optimised for good generalisation by looking at two different perspectives. The architecture of the network (to be set by the system constraints) is fixed and the determination of the size of the training set is considered or the training size is set, and the best architecture configuration is sought to achieve good generalisation of the system.

Consider a pair $\{x,d\}$ with the input vector $x \in R^p$, and the desired output $d \in [-1,1]$. The network will act as a binary classifier. An epoch is defined as a sequence of examples drawn independently at random from some distribution D. Let f be a function from the space R^p into [-1,1], with $d = f(x)$. An error of the function f, with respect to the distribution D, is defined as the probability that the output $y \neq d$ for a pair (x,d) picked at random. Let M denote the total number of hidden computation nodes. Let W be the total number of synaptic weights in the network. Let N denote the number of random samples used to train the network. Let ε denote the fraction of errors permitted on test. The network will almost certainly provide generalisation, provided that the following conditions are met. The fraction of errors made on the training set is less than ε/2 and the number of examples, N, used in training is

$$N \geq \frac{32W}{\varepsilon'} \ln\left(\frac{32M}{\varepsilon'}\right) \tag{2.43}$$

Equation (2.43) provides a distribution free, worst case formula for estimating the training set size for a single layer network for good generalization. The term "worst case" is used because in practice there can be a huge numerical gap between the actual size of the training set required and that predicted by the criterion of equation (2.43). If the logarithmic factor is ignored, the number of training examples is, to a first order approximation, directly proportional to the number of weights in the network and inversely proportional to the accuracy parameter ε. In practice all that is needed for a good generalization is to satisfy the condition

$$N > \frac{W}{\varepsilon'} \tag{2.44}$$

Thus, for an error of 10 %, the number of training examples should be approximately 10 times the synaptic weights in the network (Haykin, 1994).

2.12 Cross validation

The essence of back propagation learning is to encode an input-output relation represented by a set of examples $\{x,d\}$ with a multilayer perceptron well trained in the sense that it learns enough about the past to generalise about the future. The MLP selection problem chooses from within a set of candidate model structures (parameterisations), the "best" one according to a certain criterion. A standard tool in statistics, known as cross validation, provides a guiding principle for the selection and analysis of the data. Firstly, the data set is randomly partitioned into a training set and a test set. The training set is further portioned into two subsets (Haykin, 1994), i.e. a subset used for the estimation of the model (i.e. training the network) and a subset used for evaluation of the performance of the model (i.e. the validation); the validation is typically 10 to 20 percent of the training set. The motivation in using this method is to validate the model on a data set different from the one used for the training or parameter estimation. The training set is used to evaluate the performance of a range of potential model structures, thereby choose the "best" fit. The model which displays the most efficient performing parameter values is then trained on the full training set. The generalization performance of the resulting network is measured on the test data. The training data set is partitioned into training and test subsets, in which case "overtraining" will show up as poorer performance on the validation set. Certain criteria have to be met to reach the stopping point when training of a network. The error performance can be used to determine the size of the

data set used in the training. Good generalisation can be achieved even if the neural network is designed to have too many parameters. This is possible if the training set is stopped at the point when the number of epochs is equal to the minimum point of the error performance curve on cross validation. The size of the learning rate parameter of a multilayer perceptron can be adjusted by the use of cross validation techniques. When the classification performance of the network on the cross validation is not reduced (typically, 0.5 %), the learning rate parameter has to be adjusted. Normally, a factor of 2 is used for the parameter reduction. The learning rate parameter needs to be adjusted until there is no change in classification performance on the cross validation set. The training of the network is stopped when that point is reached (Ripley, 1998). Some commercially available software like PREDICT, used in this study, have built-in cross validation techniques.

2.13. Limitations of the back propagation algorithm

The back propagation algorithm has emerged as the most popular supervised training of MLPs. It has two distinct properties: i.e. it is simple to compute locally, and it performs stochastic gradient descent in weight space (viz. pattern-by-pattern updating of the synaptic weights). Computations performed by a neuron are influenced solely by those neurons that are in physical contact with it. It is a first-order approximation of the steepest descent technique because it uses the gradient of the instantaneous error surface in weight space. The algorithm adjusts until it gets the true direction to a minimum on the error surface. This technique accounts for the slow convergence for the following reasons. The error surface is fairly flat along a weight dimension, resulting in a derivative of the error surface with respect to that weight being small in magnitude. Multiple iterations of the algorithm are essential to ensure a small change in the error performance. Alternatively, in a curved weight dimension, the derivative of the error surface with respect to that weight is large in magnitude. The adjustment applied to this weight is large and may cause the algorithm to overshoot the minimum of the error surface. The direction of the negative gradient (i.e. the negative derivative of the cost function with respect to the vector of weights) may point away from the minimum of the error surface; hence applied adjustments to the weights will induce the algorithm to move in the wrong direction. The above phenomenon causes the rate of convergence in back propagation learning to be relatively slow, thereby making it computationally expensive. The local convergence rates of the back propagation algorithm are linear, which is justified on the grounds that it uses the Jacobian matrix of the objective function (matrix of first order partial derivatives) which is almost rank deficient. This leads to local minima (i.e. isolated valleys) in addition to global minima. Since

back propagation is a hill-climbing technique, and may get trapped in a local minimum. An infinitesimal change in synaptic weights significantly increases the cost function.

The following "guidelines" should be considered for accelerating the convergence of back propagation learning through learning rate adaptation. "Every adjustable network parameter of the cost function should have its own individual learning rate parameter and each learning parameter should be allowed to vary from one iteration to the next. When the derivative of the cost function with respect to a synaptic weight has the same algebraic sign for several consecutive iterations of the algorithm, the learning-rate parameter for that particular weight should be increased. When the algebraic sign of the derivative cost function with respect to a particular synaptic weight alternates for several consecutive iterations of the algorithm, the learning-rate parameter for that weight should be decreased" Zhu *et al* (1994). Zhu *et al* (1994) suggest increasing the hidden neurons, lowering the learning rate and starting the training with different sets of random weights. To improve the performance, the learning rate must be as large as possible without leading to oscillations. However, the choice of learning rate may be limited by the nature of the error surface. If the error surface contains high frequency variations the learning rate must be reduced to prevent instability. Rumelhart *et al* (1986) recognised this predicament and suggested the use of a momentum term in the convergence term which is the same as equation (2.42). The momentum term effectively filters out the high frequency variations in the error surface. One of the less significant problems is that of symmetry breaking. Rumelhart *et al* (1986) noted that if all the weights start out at the same value and if unequal weights are necessary to obtain a solution, the network will never learn as it is used in proportion to the significance of the weights. Subsequently, identical error signals are propagated back, resulting in the updated set of weights having equal value. They concluded that the system is in fact starting out from a local minimum and will ultimately find a second local minimum. Also, that by starting the training with a set of small random weights, this problem does not surface.

2.14 Summary

The derivation of the BPA represents a "landmark" in neural networks in that it provides a computationally efficient method for the training of MLPs. BPL has been successfully applied to solving some difficult problems such as speech recognition from text, handwritten digit recognition and adaptive control. The adaptive gradient descent learning rule described above has been used by the software PREDICT in the building of the neural network models in this study. The applications of artificial neural networks are presented in the next chapter.

CHAPTER 3
REVIEW OF ARTIFICIAL NEURAL NETWORK APPLICATIONS

3.1 Introduction

"Neural networks provide a fundamentally different approach to material modelling and material processing control techniques in comparison to the statistical or numerical methods. This technique is applicable in many areas of engineering and has produced promising preliminary results in the areas of material modelling and processing. The main advantage of this approach is that there is no need to make priori assumptions about material behaviour even though in more complex neural networks modelling schemes one may take advantage of the knowledge of the process in network design. Although multi-layered neural network models cannot ensure a global minimum solution for any given problem, it is a reasonable assumption that if the network is trained on a comprehensive database with an appropriate representation scheme, the resulting model will approximate the entire mechanical laws which the actual material or process obeys. Well trained neural network models provide fast, accurate and consistent results, making them superior to all other techniques" (Forouzan *et al*, 2007). Aided by the continuous development in the field of computers and related equipment, artificial intelligence has been widely used to support decision-making. Theoretical and practical problems have contributed significantly to developments in statistics and computer science, but it is only recently that clinicians are using these methods in their day to day practice on a significant scale. Sometimes it includes the usage of neural networks. In the early stages computer models were used for clinical consultation only, with the aim of classifying the research into a range of possible focus areas. There is an enormous complexity that arises from the various clinical conditions, thus making any form of comprehensive analysis a monumental task (Lisboa, 2002b).

3.2 Use of ANN in engineering applications

The derivation of a mathematical model is dictated by the experimental system of interest and the types of experiments that will be used by the model for analysis. Only through comparisons between the model and the experimental data, can the interactions and correlations be found, tested and revised if necessary. A system has to be understood so that questions, previously unknown, can be now asked about the system, answers analysed and a precise description given on the proposed system (Goldstein, 2001). The outputs of a neural network correspond to the variables required for prediction and the inputs to the variables on which the prediction is based.

The partial derivatives of the error function with respect to the weights may be back-calculated from output to input throughout the network, a process called back propagation. A (local) minimum is found by updating the weights after every iteration by a small enough multiple of the derivatives. The system has to be well analysed to ensure that all inputs and interactions are taken into account.

When it comes to predictions statistical results seem to favour feed forward back propagation networks more than other types of neural networks, and non-linear models. Singh (2007) found that correlative coefficient, which is considered as a measure of accuracy, favoured the back propagation network. A statistical study was carried out on a ferrochrome production process to predict the effect of raw materials on the overall performance of the plant. In the initial stages of the study, the non-linear relationship between the raw material inputs and production capability index was used to predict by multivariate linear regression. Three different learning algorithms were developed using feed forward back propagation neural networks to improve the prediction accuracy (conjugate gradient descent, Levenberg-Marquardt optimisation and resilient back propagation). Trials were done with radial basis neural networks but there was no significant improvement in the performance prediction. The correlation between the predicted and actual values were 0.64 for multivariable linear regression, 0.70 for Radial Basis and 0.71 for feed forward neural networks with resilient back propagation. Comparisons were made between the use of statistical analysis, neural network structures and the results of production capability index calculations. In order to significantly affect the performance of the proposed model, the parameters affecting the production process and combination modelling was necessary. This multivariable problem is highly non-linear and becomes more complicated when the size of the network increases, thus reducing the performance. Resilient BPLA is especially used in pattern recognition problems to obtain data from multi-dimensional space.

The Levenberg-Marquardt algorithm is used to solve least squares method problems that are almost linear. In these cases as the number of process parameters increases the performance of the model is reduced. Zhu *et al* (2007) constructed a multi-layered feed forward back propagation artificial neural network (BPANN) using two training algorithms: the Levenberg-Marquardt algorithm and gradient descent to predict combustion efficiency of chicken litter in a swirling fluidised bed combustor (SFBC). BPANN is a generalisation of the Widrow-Hoff learning rule, or least mean squared errors (MSE) to multiple-layer networks and non-linear differentiable transfer functions. Application of the ANN was comprised of "data collection and preparation (including statistical experimental design and experiments), neural network training (including selection of ANN, structure, training and algorithm) and neural network evaluation, simulation, prediction and validation. The neural network consisted of one input layer, several

hidden layers and one output layer. The transfer function used for the hidden layers was a differentiable tan-sigmoid function and a linear function for the output" Zhu *et al* (2007). The combustion efficiency is the target value of the neuron in the output layer. Model building of the various ANNs was done with the aid of the neural network toolbox of Matlab. The results from the models, i.e. the network training and the prediction processes can be monitored both quantitatively and graphically. Two training functions, *viz*., traingd (gradient descent) and trainlm can be used in the training of BPANN. Trainlm is based on the Levenberg-Marquardt optimisation procedure. The updating of weights and biases in the direction of the negative gradient of (least mean squares errors) MSE results in the training of the model. This algorithm can be used to optimise the model by increasing the speed and its reliability. The two training functions were applied to the data with an equal learning rate of 0.05 and an equivalent ANN structure of 5 + 16 + 1. The results indicate that the Levenberg-Marquardt algorithm outperformed the gradient descent method. The training epoch (iterations) for the trainlm was 9 and 550 for the traingd. The ANN predicted line was within the 95 % confidence bounds, thus proving it to be a reliable model. To improve the reliability further, the model was used to test and verify data from similar experiments reported elsewhere. Convergence speed, number of hidden layers and number of neurons were finally chosen as the design parameters for the process. A trial-and-error method was adopted for selection of the final design values. For a mean squared error (0.2204) the Levenberg-Marquardt training algorithm outperformed the gradient descent method. When a model is developed new data has to be used to validate it. On validation the ANN approach was found to be most accurate and reliable for the prediction of combustion operation performance and for the overall optimisation of the combustion process. The development of this model has proved to be vital for the design and scaling up of SFBC.

Sterjovski *et al* (2007) built and validated a back propagation ANN "to predict the level of diffusible hydrogen deposited into weld material as a function of the welding process, atmospheric conditions and moisture already present in the consumables". A probabilistic ANN approach was adopted to investigate whether hydrogen was responsible for weld cracking. The software Neural Nets Prof II/Plus was used to build a back-propagation neural network model. As the model building progressed new parameters were introduced like the learning rule (algorithm used to relate all the neurons in the layers), learning coefficients (value that affects rate of learning) and transfer mode (prediction of output using the weighted sum). The delta learning rule was used in the model building and the error was calculated using the root mean square method. The derivative of the transfer function transforms the error which is then "back-propagated" to the previous layers where it is stored. The stored value gives an indication of the current error value for that layer. This back-propagating of the errors is progressive method which is stopped when the input layer is reached. The delta learning rule and sigmoid transfer

function formed part of Model 1‟s back propagation neural network. The model was used to predict the relationship between the levels of diffusible hydrogen with respect to a multitude of input parameters. Some of the variables were wire type (or classification), welding current and contact tip to work-piece distance. Model 1 was able to predict the diffusible hydrogen with a large amount of accuracy. Model 2 used a probabilistic ANN to solve the pattern classification problem. The neural network used a classification technique on the input data to separate the "crack" from the "no crack" groups. The levels of agreement for both sets of data were high. The prediction errors were not significant, with the root mean square error for all the data being 0.5 ml per 100g of deposited weld metal. Statistical analysis was used to further validate the models. ANNs were successful in predicting hydrogen diffusible content and hydrogen cracking susceptibility. The trends predicted by the models were consistent with actual experimental data (Sterjovski *et al,* 2007). NN tend to be flexible in nature in comparison to conventional regression analysis. One reason could be the ability of NN to perform interpolations using data in the same range of variables as that of the training (Dutta *et al*, 2007).

3.3 Current predictor methods used in medicine

The mortality due to various cancers can be significantly reduced when detected early. Any effective screening program that is available for the most common diseases, such as cervical or lung cancer, requires imaging of sampled cellular populations or radiographs for correct diagnosis. Various computer aided visualisation systems have been developed but clinical acceptance requires that the system be as good as or preferably better than the trained expert. The development of an automated diagnostic algorithm in the form of an expert system would have required the generation of an extensive rule base. This approach was thought to be too cumbersome and was discarded in favour of ANNs. Using the same data previously generated and analysed by other means, a set of hierarchical ANNs were assembled to create a system with a predictive accuracy almost equal to that of a human expert. The gradient descent method was used as it provides the most efficient weight changes for encoding the particular input-output mapping desired. The hypothesis that supervised ANN‟s (MLP using BPL) could classify cervical cells in to 4 categories with a precision comparable to that of a skilled cytotechnician was demonstrated. This approach successfully removed the crucial problem of classifying abnormal cells as normal (false negative). Implementation of neural networks for clinical use must first be optimised with respect to factors such as the composition of the training set, the network architecture, and the hierarchical assembly of separate neural networks specialised in specific sub-tasks. The advantage of separate neural networks allows for quicker convergence during training because the complexity for each sub-task was reduced. The most

important outcome was that the overall accuracy of the classification was increased (Mehdi *et al*, 1994).

A comparison of neural network and logistic regression models were done by Venkataraman (2004) to predict a medical outcome. The objective of the research was to prove that there are more accurate methods for pattern classification than conventional statistical models. The new technique to be tested for classification was the feed forward neural network model. In the medical field a reliable prediction of an outcome is vital for clinicians as it impacts on their diagnosis, treatment and overall care of a patient. A model was developed to predict the outcome of stress urinary incontinence surgery so that clinicians could gauge which of the patients benefited the most from the surgery (Venkataraman, 2004). "The artificial neural network was trained using 225 clinical sets with error back propagation and validated through independent testing of 200 records. 21 records were used to predict the categorical output value. Stepwise forward logistic regression was applied to the dependent variable. Stepwise forward selection starts with the best single regressor and then finds the best one to add to what exists. The next best set where all variables in each equation are checked to see if they remain significant after the new variable is then added" (Venkataraman, 2004). The same development and validation datasets were used for the logistic regression analysis to provide a comparison. The neural network model allows for flexibility as it allows the user to change the parameters within a model to increase accuracy. With conventional statistical methods specific correlations have to be used. "The neural network model had a higher sensitivity (78.3% versus 62.5%), specificity (75% versus 54.2%), area under the receiver operating curve (0.74 versus 0.72) and less error rate (22.5% versus 39.5%) compared to the logistic regression model. Both the positive (0.9084 versus 39.5%) and the negative predictive value (0.5217 versus 0.3138) for the neural network model was higher than the logistic regression model suggesting that the neural network model better fits the data" (Venkataraman, 2004). Neural networks are considered to be superior to logistic regression methods because they can find a multitude of relationships between variables which display characteristics of multi-dimensional non-linear functions. The parallel architecture of neural networks, a principal difference with the statistical method allows for this result. Overtraining may account for best results in a neural network but statistical methods are also prone to over fitting. The research has shown that the building of artificial neural networks for prediction of outcome for incontinence symptoms severity is advantageous to both clinicians and patients over time. Neural networks somewhat outperform logistic regression analysis and prove to be useful when adopting classification techniques for predictions.

Lately predictive analysis has been used in decision making for the allocation of relevant resources, quality assurance and improvement, research into health care delivery and comparison of health care centres. Logistic regression (LR) models can successfully be used to determine the relationship between multiple predictor variables and a dual outcome variable. LR models are unable to handle complex multivariable systems. ANN models are able to "learn" the variations in non-linear patterns when comparing the predictor variables to outcomes. This is done by adaptation, like humans do, when making decisions. Han *et al* (2000) used neural networks to predict the biochemical recurrence after surgery for prostrate cancer patients using pathologic and clinical data. The significance of Gleason scores with respect to biochemical recurrence was also studied. For 20 years the Gleason system has been used by pathologists to classify prostrate cancer into various groups. The system has helped in aiding clinicians to make a decision either individually, or combined with other parameters when choosing the adopted treatment regimes. The system is also used to predict the outcome after the prescribed therapy for prostrate cancer (Han *et al,* 2000). The data were randomly selected to obtain 50 % training, 20 % for testing and 30 % for validation. The ANN was an MLP because of its resilience and ability not to over-fit. The conjugate descent method was used in the training of known data. An optimised MLP structure was chosen to maximise the accuracy of the prediction and restrict large networks from over-fitting. A combination of 1000 architectures was investigated. The univariate Cox regression model found the surgical margin to be the parameter that most influenced a recurrence of the cancer. Overall, the accuracy of ANN was greater that the LR modelling when used in 3- and 5-year predictions for biochemical recurrence of radical retro-pubic prostatectomy (RRP). Previously, either traditional statistical methods were used to predict tumour recurrence after RRP, or a clinical diagnosis was carried out by specialists. LR is traditionally used to find a relationship between a response (output) variable and a set of input variables. ANN models have proved to be superior to LR, since the ANN is able to learn the complex relationships between independent and dependent variables. The development of ANNs is the result of the age old race to try and duplicate the fault-tolerance and infinite learning ability of biological systems. ANNs have repeatedly shown their high degree of accuracy when compared to other predictive systems, including LR, and particularly when there is non-linearity. Some ANNs can rank the variables to the extent that each influences the output. Han *et al* (2000). The ANN was chosen in this application because if its resilience and ability not to over fit. The accuracy of ANNs has led to its adoption in predicting a recurrence-free outcome for the patient population over LR. The Kaplan-Meier analysis and the Cox proportional hazards model, together with the ANN confirmed that the Gleason score should be maintained as an outcome predictor for patients after RRP. The adoption of ANN in the clinical decision making process for prostrate cancer is a fairly development. Clinicians can now use this model to determine the recurrence rate of cancer over a long-term period. In the treatment

of cancer a 5-year period is normally adopted to determine if a patient is in remission. Also for general application of the ANN there has to be inter-institutional validation. ANN models can also find use in increasing the accuracy of clinical outcomes when both pathologic and clinical data is available.

3.4 Neural networks in medicine

Increasing processor speeds and computer power have moved neural networks from an obscure topic to one where they offer practical uses in a wide range of fields. At their most basic they can be used as cheap, small programs (essentially as add-ins to commercial spreadsheets). Successful medical uses of neural networks include their use in pharmacokinetic or pharmocodynamic prediction (antibiotic peak and trough levels) and enhancement of diagnostic skills. In the clinical arena, the emergency room diagnosis of myocardial infarction and the radiologists‟ diagnosis of pulmonary embolism have been studied with favourable results. Neural network based prognostication in critical care has been shown to be superior to a conventional statistical approach. It has been found that enhanced performance by neural networks relative to humans is related to more appropriate weighting given by the network to common diagnostic factors, and also the empirical use by networks of diagnostic rules of thumb that have not been previously utilised. Neural networks have several limitations and the „black box‟ nature of their output is of major theoretical concern, i.e. the conclusions are generated without explanations. This criticism implies that conclusions should follow from hypotheses supported by data and rejects the role of pattern recognition in decision-making. The success of neural networks in medical decision-making, including a performance equal to that of radiologists in the diagnosis of pulmonary embolism and superior to that of emergency physicians in the diagnosis of myocardial infarction suggests that they are competent at the very least. Black box concerns may be overcome by sensitivity analyses where the effects of variables are assessed by their inclusion or exclusion.

In the past ten years clinicians and health institutions have been adopting artificial intelligence (AI) methods for use in medical diagnosis, treatment and related applications. Evidence is clearly shown by the availability of numerous medical devices with built-in AI algorithms on the market, together with a simultaneous flood of publications in medical journals. Superiority of neural networks has been proven by optimisation techniques which gives rise to cost-effective and flexible non-linear modelling of large data sets. Further, high accuracy which is vital in the medical field favours its adoption as additional support in clinical decision making. These models also enlighten the user by giving valid explanations, for example, use of rule extraction or sensitivity analysis (Lisboa, 2006b).

3.5 General survival studies

Numerous case studies in medical literature have highlighted that neural networks can significantly influence the modelling of cancer survival where previous research related the MLP with the Tumour, Nodes and Metastasis (TNM) clinical staging system prescribed by the World Health Organisation. A distinguishing feature of survival modelling when compared with conventional class discrimination is the unique factor of patient censorship. Censorship is defined as a follow-up loss without the event occurring. In survival studies of cancer patients, the event that the clinician finds important might be death due to breast cancer with a follow-up study done over a period of 5 years of recruitment. Patients lost during the 5 year period, from an address change or an event as serious as death from an unrelated course will be taken as censored. A patient may die from a natural, an unfortunate cause or it could be related to a weakened condition caused by the treatment, e.g. cardiac arrest, thus the original cancer was the real cause of the patient‟s death. All patients that survive the maximum follow up period are considered to be censored. Censorship is adopted because there is no precise method of determining what the health status of the patient would be like if they were still part of the study (Xin, 2006). Since the information is unavailable it is treated as missing, thereby removing censored patients from the case study. Also if ad-hoc techniques are employed, then the information is said to be biased. The element of bias occurs when simulated data is used in survival studies. MLP have been modified to include censorship. The reason for this is that assumptions are made when using traditional survival models that are linear in their parameters. The proportional hazards method, also called Cox regression, is traditionally adopted for use in medical statistical studies. MLPs have been used as extended versions of proportional hazards. They uphold the separation between the dependence on time and on patient specific vector of covariates, resulting in non-linear proportional hazard models. Time can be expressed as a covariate which serves as an input index to conditional hazard estimates from a single output unit. Neural network models of this type that are used for survival studies have proven to be stable when used for monthly studies where patients have follow-up periods of several years. As a consequence the proportionality of the hazards assumption is relinquished and ensures the fitting of non-linear effects. These models have been highlighted in case studies on patient mortality and disease recurrence consequent to surgery for breast cancer. The results were useful in that it could be practically applied by informing clinicians and patients when making decisions on types of treatment (Xin, 2006). Research carried out on breast cancer patients by Ripley (1998) indicated that the accuracy of relapse prediction did not change within staggered

time periods even though it was fixed. In the study of breast cancer there is no indication that neural networks have outperformed the proportional hazards model as in other forms of cancer. Both traditional and modern neural network models are tools to be used jointly to solve real world problems, especially in the study of survival rates of patients. It is recommended that neural networks be used to support clinicians in trying to facilitate patient performance rather than seeking a full recovery. It is better to reduce a patient‟s suffering with treatment than try to make radical changes which would be detrimental to the patient‟s health (Ripley, 1998). For neural network models to gain credibility by clinicians for patient evaluation and treatment, it is necessary to use data and systems from multiple units and centres that adhere to the gold standard of clinical evaluation, namely multi-centre Randomised Clinical Trials. This integrated method is globally considered as a key milestone for the acceptance of any medical decision support system. The reason being that it takes into consideration multiple factors that can influence patient outcome, i.e. it guards against inter-patient and inter-centre variability. It is more reliable to run patient trials rather than trials with a clinician. Patient trials ensure variations which can be analysed to give a true reflection of the prevalence and behaviour of a disease under study. The principles guiding survival modelling in comparison to other studies is largely challenged by the concept of censorship when dealing with patient trials. The term censorship is responsible for the validity of the outcome of the patients who have remained in the clinical trial (Lisboa, 2002b). This can be readily demonstrated by using simulated data to estimate survival, i.e., electronically generated data used to mimic actual results from tests done on patients. This can be used as an initial step to modelling survival, but ultimately real world data from actual patients have to be used to verify any mathematical models that are proposed.

Neural networks and traditional regression models are used extensively in the literature to aid with predictions of health status, e.g. of the outcome of injured adult patients. In contradiction, prediction of in-patient survival of injured children and teenagers have not been investigated or developed. The readily available predictive models for adult patients are logistic regression statistical models which are incapable of handling multivariable systems that are interdependent (DiRusso *et al*, 2001). DiRusso *et al* (2001) have proposed the use of ANN to predict the health status of paediatric trauma patients. An ANN was developed with the aid of an existing database of the National Paediatric Trauma Registry (NPTR) in the United States of America (USA). The data was used to develop and test a reliable model for evaluating the survival of injured children and adolescents. A standard feed forward back propagating neural network was chosen since its architecture was constantly used for predictive models. The architecture was chosen for its reliability and accuracy in pattern recognition of multivariable data where the interdependency is limited biological networks. It has also been proven to have a high accuracy for determining the outcome of treatment on adult trauma patients. The model consisted of 3 layers: an input

layer (with input variables), a single layer of hidden nodes, which are the algorithms that analyse the input data and the third layer which is the output variable. In this case study, it is the probability of death that lies in the range: range (0, 1). “The learning occurred rapidly with a computation time of approximately 8 minutes. The second ANN model was generated using a data search engine (Statistical Neural Networks, Statsoft, Tulsa OK). The search engine tested a large number of possible ANN structures and chose the models that performed the best based on any particular set of data. The choices of architecture included ANN type, number of hidden nodes, and activation function for each layer. In this way a tremendous number of possible outcomes were examined. This would have been prohibited if done singularly. However, computation time was extensive (approximately 155 hours of continuous computation time). Once the data was generated, testing new data took little computer time (usually less than 1 second)” (DiRusso *et al*, 2001). Interestingly, the search engine chose a feed-forward back-propagating network as suggested by the authors in the previous NN model. A standard logistic regression model was used for comparison. Missing data for both the randomly selected sample of 27 385 cases and the 8000 test cases were allocated the corresponding mode value as determined from the original 27585 cases. Some of the input variables “included age, sex, systolic blood pressure, heart rate, respiratory rate, intubation status, individual components of the Glasgow Coma Score (GCS), the New Injury Severity Score (NISS) or the Injury Severity Score (ISS), the Revised Trauma Score (RTS), and the Paediatric Trauma Score (PTS)” (DiRusso *et al*, 2001). The probability of death was the expected output. Discrimination and calibration was measured from the existing data. Since censorship is an issue the discriminatory power of the model is adopted to correctly classify survivors. The values were determined from the area under the receiver operator curves (ROC). The ROC curves are developed by plotting true-positive fraction versus the false-positive fraction. A ROC of 1 implies perfect discrimination, while a discrimination value of 0.5 is equated to a random model. Calibration is an indication of the reliability of the model to predict with a high degree of accuracy over the entire range of severity of injury. Of the 35 385 patients evaluated in the analysis, there were 1047 deaths (3% mortality rate). The numbers indicate that the model was trained with data predominantly from the survivors‟ data set, thus implying that the model will favour higher accuracy with these patients (DiRusso *et al*, 2001). In order for a model to be more reliable for future predictions, ideally the training data should be within the range of expected data for that system. This ensures that the proposed model can accommodate any new data, and that the prediction can be instantly validated. There were approximately 425 000 data points (12 input variables) from the database with 10% missing data. It can be expected that this model would not be too reliable in predicting the probability of death, which in turn affects the output for survival. Conversely, if there is a higher survival rate for paediatric trauma patients statistically then the model is a true indication of reality. The authors claimed that the proposed models

displayed excellent discrimination, i.e., the classification of survivors and non-survivors was considered to be reliable. From the figures above it can be noted that a bias does exist because the model would favour the survivors. A realistic prediction model should place all survivors in the lowest decile and all non-survivors in the highest. The two models ordered and predicted the surviving patients with a high degree of accuracy; however, both models were unreliable in the predictions where the patients had died. NISS, motor score of the GCS, and systolic blood pressure were the parameters that had the greatest impact on the developed models. The NPTR database collates data from most major trauma centres treating children in the USA, thus making it truly representative of the paediatric trauma population in general. The outputs from the proposed models can reliably be taken as a true reflection of the health status of the paediatric trauma population (Di Russo *et al*, 2001).

Most models use logistic regression for survival predictions. The use of LRs is widely accepted as a statistical tool thus making them reliable models, but there are limitations. Artificial neural networks are more robust modelling systems constructed of multi-layered equations that form algorithms. A patient‟s health status is dependent on various symptoms and the complexity of the medical condition is dependent on a patient‟s biological system, and this varies between patients. This complexity allows for the application of ANN in predicting outcomes in medical application, e.g. survival rate after liver transplantation, diagnosis of pulmonary emboli, myocardial infarction, breast cancer, etc. Statistics on cancer survival can be used if the length of the terminal period is known. It can be used to determine the period when the last stages of cancer occur thus allowing the clinicians and family to design and prepare for the palliative care stage. The terminal period is defined as: “the moment in the natural history of the cancer in which there is little likelihood that specialised oncological treatment will extend survival time, induce an objective response, or halt the progress of the disease; continuous use of health resources is required to alleviate the patient‟s symptoms; and any treatment is administered solely for palliative purposes (Llobera *et al*, 2000). “Terminal period” is generally defined as a survival of less than 6 months. Mean survival time is dependent on the definition of the terminal period. It is vital to know the exact period when the patient will be in the advanced stages of cancer. It is necessary as it impacts on the treatment plans and psychological well being of the patient and the family. The use of health and social resources also has to be adequately arranged for terminally ill patients. If the survival rate is incorrectly predicted then the patient may not get access to palliative care programmes. The survival time is usually overestimated, the reason being that clinicians and doctors give the patient and family hope for a longer survival period because it does have psychological implications (Llobera *et al*, 2000). But this can have a negative impact on the patient, where inappropriate treatment regimes are followed and the onset of palliative care is delayed. Various studies have been done to try and accurately identify

the factors that influence the calculation of the terminal period for cancer patients. A study was done by Llobera *et al* (2000) "to determine the duration of the terminal period of cancer, to assess the ability of clinicians to predict survival and to identify the indicators that might improve prognostic accuracy". Some of the key factors studied were demography, tumour site, dates of diagnosis, onset of the terminal period and death. The difference between the onset of the terminal period and date of death gives an indication of the terminal period. Clinicians have to be involved in the decision making of the terminal period. It needs to be reinforced that the model cannot be used in isolation. The human element, in this case the trained clinician, will always be the most reliable predictor, no matter the numbers that indicate the perfect fit of a model. The health professional"s inability to predict survival accurately has an effect on the patient both mentally and physically. Their decision will affect treatment stopping time and the onset of palliative care. The tendency has in the past been to overestimate. Only 5 % of the oncologists were accurate in determining the actual period correctly. Sensitivity and accuracy was greater. Despite the lack of accuracy the clinical decision of survival cannot be super ceded by a model. The following factors did not influence survival significantly, i.e., age, sex, marital status or tumour site. Parameters such as bed sores, deficient care; asthenia and anorexia; confusion and drowsiness; are all indications of the onset of the terminal period. It is more reliable to do prospective studies instead retrospective studies, because the information would be current and tangible in comparison to looking for past data which may contain missing points. Reliability on past data can also be another issue (Llobera *et al*, 2000).

The performance of neural network methods and Cox regression for censored survival data was carried out by Xiang *et al* (2000). If censorship is built into ANNs, then the outcomes (outputs) are dependent on whether an event occurred or not. If an outcome is dependent on time to an event, a method has to be developed for dealing with the effect of time on disease progression. If an event does not take place, then the output is considered to be (right) censored (e.g. patient has not returned for treatment, it is not known whether he is dead or alive). In the medical domain patient numbers and clinical data is usually limited. In order for mathematical modelling to be reliable there needs to be sufficient data for the model to be developed. Ideally if censored data is excluded from the model building it could lead to significant biases in predicting events or outcomes. Methods can be adopted to accommodate right-censored data in the model development and building. Cox regression analysis is a universal technique used to analyse censored data. The results of Cox regressions need to be related to the outcomes of ANNs. The study by Xiang *et al* (2000) does not indicate the data and type that was used to do the above analysis, it was only stated that simulated data and not actual clinical data was used. Their recommendations are that further research be carried out to evaluate calibration measures and that the recommended model uses actual clinical data from the risk groups. The effect of

data base size was also studied. Initially 200 cases were simulated. A random division of 100 training and 100 testing cases were used to build the first models. In the second phase of the model building 400 cases were simulated. This time a division of 200 training and 200 testing cases was used in the testing. Censoring was varied for the neural network model with an initial censoring rate of 20%. For the training data a 30%, 50% and 70% censoring rate was adopted. For the censored data, the neural networks outperformed the Cox regression models. This outcome implies that neural networks are adaptable in that they can deal with multiple parameters that are interdependent. For Cox regression analyses the user has to have the expertise to use the data in an appropriate manner and be able to make meaningful conclusions from the analyses. For a Cox regression the coefficients determine the probabilities of an outcome for some value(s) of the risk factor (e.g. odds ratios or relative risks). NN weights usually do not lend themselves to such interpretation. The models presented in the above research highlight a few models developed for patient survival. Overall the research showed that NN can reliably and effectively be used as methods for modelling right-censored data. The issue of performance does vary but it should be noted that the models are based on known data for a given system and is only applicable in the same system.

3.6 Health benefit from ANN in medical intervention

Clinical diagnosis is quite complicated because symptoms of unrelated diseases can surface thus complicating the clinician‟s diagnosis. The patient may display all the symptoms and clinical tests point to obvious and known diseases, but sometimes the diagnosis is poor or incorrect. Since diagnosis is related to the type of treatment that is dispensed to the patient, ultimately it affects the physician‟s prediction of prognosis. Feed forward multilayer neural networks with back propagation seem to give better diagnoses than the clinicians and traditional expert systems. Early diagnosis is very important in biomedical applications because of the benefits of immediate and correct treatment. Acute myocardial infarction (coronary occlusion) is one type of disease that physicians find difficultly diagnosing. The diagnosis can be difficult and different diagnostic methods have been studied. Together with patient history and clinical findings, the 12-lead ECG is still considered to be the traditional method. It is commonly available and is still considered to be the best method for the early diagnosis of myocardial infarction. 25% of the patients discharged from the hospital display ST elevations that were not detected. Over 80% of the patients admitted for acute myocardial infarction are sent home without having the ECG done or the diagnosis confirmed. Two separate case studies using neural networks for the diagnosis of acute myocardial infarction are presented. The first case study uses neural networks to predict acute myocardial infarction in the 12-lead ECG at a sensitivity greater than two sets of diagnostic protocols, and even better than an experienced

cardiologist. For the neural networks the sensitivity was 18.3% higher than a rule based diagnosis method with respect to a specificity of 95.2% (P<0.00001), and 10.5% between neural networks and cardiologist at a specificity of 86.3% (P<0.00001). A false classification as definite acute myocardial infarction was made in only 0.2% and 0.4% in the control group by the cardiologist and the rule based diagnosis method (Heden *et al*, 1997).

Attempts have been made to automate the diagnosis procedure but success rates have not been too different from the physician‟s conclusions. The second case study on diagnosis of myocardial infarction, using neural networks, show that the most successful attempt at finding an automated solution has been a detection rate of 88%. This value is close to the accuracy rate of practicing physicians. The incorrect diagnosis rate was 26%. This is an improvement on the 29% rate of physicians. A feed forward fully interconnected neural network with two hidden layers and a single output was trained to diagnose coronary occlusion. Each hidden layer had ten neurons. All neurons used a unipolar, sigmoidal activation. A back propagation algorithm was used in the training of data for the model building. 356 patients‟ was used in the training of the network. Of the 356, 236 were not diagnosed with a coronary disease and 120 did have the disease. Patients were randomly selected for training from subsets of half those considered to have the disease and the half that did not. The data for each patient was comprised of twenty variables that are normally predictive of acute myocardial infarction. These characteristics are age, sex, nausea and vomiting, shortness of breath, diabetes, hypertension, and angina. This is part of a series of 41 variables that is collected on all patients admitted to the emergency coronary care unit. Subsequently a method was followed to diagnose the infarction. The proposed network (model) was tested on the remaining 178 patients (118 non-infarctions, 60 infarctions) which was the unknown data. The network predicted a 92 % correct identification of the disease. The result was 96% for the correct prediction of the absence of the disease. Routine data was used to develop the models. The NN outperformed the clinicians, but the human element has to take precedence in the treatment centre. It should be noted that models are there as support for decision making (Hassoun, 1995).

A review of the health benefits of artificial neural networks in clinical diagnose was carried out. Initial studies in computer modelling were carried out to give clinicians some guidelines on possible diagnoses. Recent advances in computing power allows clinicians to be at the patient‟s bedside, while expert systems predict an outcome so that appropriate treatment regimes are adopted (Lisboa, 2002b). Support of decision making systems must be in line with medical problems. This is due to a clinician‟s reluctance to use electronic means to decide on the fate of a human being. Only if all patients can benefit with the aid of a generic category of clinical conditions, will the clinicians have faith in the model. In the real world some diseases are quite

rare and the general models that are developed for assessment and prediction are invalid for use. Clinicians are bound to treat every patient on an individual basis, which would make it unethical for them to follow a treatment regime based on the majority outcome of a predictive statistical model. Conventional numerical methods and Bayes profiling are unable to deal with the multiple combinations of disease presentation. The range of presentations with respect to a single disease and the extent of disease progression present a challenge. They are also unable to deal with the manner in which the disease progresses, i.e. if the same treatment is given to 2 patients, the results differ in outcome and the fact that it can also influence another ailment. It could also not explain the pathological and physiological states. If new data was introduced the model was unable to adapt to accommodate the deviations. Neural network modelling also has to take into account new data, but as long as the conditions are strictly followed then the results can be confidently applied to aid clinicians in medical decision making (Lisboa, 2002b).

Neural networks can be used in various ways to aid in medical decision support. Patient data from hospital-based database or clinical laboratory systems are used in models to check for abnormalities. Each disease presents a set of symptoms used for diagnosis, but if there is a rare disorder it will not be easily picked up, therefore numerical methods can make a significant contribution by finding the abnormality. This early diagnosis technique can allow for specific treatment regimes to target the disorder. Neural networks are also being used in clinical environments for assessments and planning. The reason for using ANNs to model the complexity of disease presentations is to try and replicate inter-related memory functions. This is common when replicating events, even though it requires some rational judgement when complex tasks are presented. Technologically ANNs fall under the subject of non-linear statistical analysis. The introduction of NNs was to broaden the usage of an adaptable tool to accommodate non-specialist statisticians. If a methodology is based on information from data-bases, it is expected that reliable statistical methods will be applied to the system, in order for the results to have credibility. This also extends to the analysis and interpretation of the data, more so if it is to be applied in a clinical environment.

Several sectors of the medical domain have been proposed as ideal systems for the application of neural networks for diagnostic support. A typical example was applied to patients entering an emergency unit displaying symptoms of chest pains. By pre-screening the patients, those that were prone to Acute Myocardial Infarction (AMI) were identified and monitored, thus preventing a major AMI if the patient was sent home. Sensitivities and specificities from 80 % to 96 % were compared to a large study done on clinicians‟ performance, which gave an overall sensitivity and specificity of 88 and 71%, respectively. Most studies use current ANN methodologies, of which the multi-layer perceptron is the most widely used because of its „early

stopping‟ ability to prevent over-fitting. Computational methods are becoming popular for usage in clinical environments. These applications have been widely reported in the literature to be most suitable for drug dosing and preventive care. They are able to predict the harmful effects of prescription drugs, or be used to ensure that protocol is followed when vaccinations are dispensed and mammography tests are done. Computer aided evaluation of mammograms are currently being used to reduce the number of missed lesions by 50 %. This is currently being achieved without an increase in costs with respect to time, training and equipment.

3.7 Methodology for neural networks in medical applications

With regard to the methodological issues, clarification should be given on the framework and steps taken to ensure the robustness of the conclusions. Regularisation of the objective function with weight decay is a simple method than can be used to prevent over-fitting of the data. When performing statistical analyses the degree to which the model responds to changes in the value of the predictor variables is of interest. A review of neural network applications in oncology in the literature has highlighted the importance of accuracy. Frequently made mistakes have been identified in the application of ANNs, which includes conclusions of generalisation with respect to performance, insufficient data points for the training of large networks, inappropriate statistical guidelines and insufficient justification for the use of ANNs as an alternative for statistical rule-based models. In some studies the proposed survival models have not considered censorship thus introducing a bias into the outcome. In some studies irrelevant clinical factors which have minimal effect on treatment decisions have been used. In others tried and tested clinical principles cannot be validated by models that disregard the cause and effect relationship of treatment regimes. The literature is flooded with theoretical models which need to be investigated further, and especially with real clinical data before it can be adopted in a clinical trial. If sound clinical and mathematical principles are followed then the proposed models would have greater credibility and acceptance by clinicians (Schwartzer *et al,* 2000).

- Clarification of aim of study: Pragmatic studies provide suggestions for adoption of treatment strategies, thus the importance of pre-specifying the performance levels. Exploratory studies are aimed at formulating a new perspective on an existing condition. One way is to formulate a hypothesis about the relationships between covariates which can then be tested with conventional medical statistical methods. Another approach is to determine how the significant variables from the model output compares to the initial information obtained from the clinical setting (Schwartzer *et al,* 2000).

- Model design: The use of too few hidden nodes affects the variance of the predictions and introduces a bias because the model becomes too simple. A solution is to choose many hidden nodes while imposing a direct regularisation of the objective function, the easiest of which is in the form of the weight decay function. The selection of variables has to be well formulated, with a clear understanding of the covariates in relation to the predicted outcomes of the model. For exploratory studies the number of observations should be five times the existing (instead of selected) covariates. This method is adopted to prevent random variables from correlating with any given sequence of labels. The number of degrees of freedom is not easily determined for neural network models, but can be achieved by evidence approximation using factors related to the regularisation hyper-parameters (Schwartzer *et a,* 2000).

- Validation: Clinicians can only adopt and apply a procedure or system if it is well defined and clearly understood because they are liable for the results when used. In exploratory studies the improvement in diagnostic or prognostic accuracy by an adopted methodology is not as important as the reason for it being so (Schwartzer *et al,* 2000). No matter what the accuracy of the model is in statistical terms, doctors will be reluctant to use it without a clear demonstration of its capabilities. Relevant or real world clinical data is vital for training a model. The data used must be routinely acquired, readily available and most important reliable.

- Benchmarking: It is standard procedure to compare any new technology with an alternate method and it must be reasonably equivalent to a tried and tested method that it is replacing. In the case of MLP the obvious benchmarks are the multi-linear regression for regression. For classification it is the logarithmic regression (LogR) and the proportional hazards model is used for survival studies.

- Robustness: There are numerous factors and variations thereof that raise doubts about maintaining performance abilities from patient to patient. This includes within-patient variation, between-patient variation and instrumentation and protocol variations between clinical institutions. There has to be a clear distinction of the validation of data. Internal validation will use selected data to train and test the model (test sample), while new or unseen data (validation sample) will be used to determine the suitability or fit for a given system. Temporal validation uses data from the same clinical centre but at a later period than the testing and validation period. Data from other clinical centres are

considered for external validation, with the condition that it was not used in the model design. Retrospective data is normally used for modelling, while phase II exploratory studies use prospective data. Phase III clinical trials have to be applied across multiple centres or institutions with prospective data (Schwartzer *et al,* 2000).

- Comparative trials: **C**omparative trial design is necessary for the analysis of interventions but it is still lacking in success upon implementation. A controlled prognostic trial was carried out on 558 patients with acute abdominal pain. Results indicated that the clinician‟s personal diagnostic abilities changed extensively over a specific time period. Specifically, the number of appendices that perforated before operating was reduced form 36% to 4%. In another study 295 patients were monitored and the model was more accurate in its diagnoses than clinicians after a 2 month training period. Within 5 months of testing the model, the clinician‟s performance increased from 73% to 84%. Implementation of an ANN based decision support system in the UK has seen a reduction of 42% in survival rates from cervical cancer between 1987 and 1997. It should ne noted that understanding the clinical scenario is vital before implementation of new computer based methods for supporting medical decision making (Lisboa, 2002b).

3.8 Medical prognosis using artificial neural networks

In the field of medicine the establishment of prognoses and survival studies for individual patients are factors in the medical domain. Where diseases prevail for extended periods over several years, precise assessment of survival rates is vital. ANNs models have recently been adopted for predicting the various stages of a disease (Ohno-Machado, 1997a). The definition of medical prognosis revolves around predicting cure, complication, recurrence of disease, level of function, length of stay in health care facilities, or survival for a patient or group of patients. Hence, prognosis is a vital component of suitable patient care, treatment regimes and resource allocation (Ohno-Machado, 2001b). Once a diagnosis is confirmed, the clinician will indicate to the patient a probable survival scenario and map out a treatment strategy. Clinicians may find deviations from normal disease behaviour and coach patients to try alternate but aggressive treatments or a new clinical trial. The probability of survival can influence the patient‟s duration of stay in an intensive care unit (ICU). Usually there is a misunderstanding about whether the prediction of disease progression patterns can determine a patient‟s prognosis without treatment. This is almost always never the case, as any patient seen by a clinician will be treated accordingly unless the patient refuses therapy which is very rare and may occur if it is the final stages of a disease, where any palliative therapy is not going to lessen the end period. Any

prediction will be based on standard treatment for a particular disease or illness. Data collection for prognostic studies is quite a complex exercise and most reliable collections are obtained form randomised trials, usually carried out in multiple academic centres. In order for a model to gain credibility, analysis done by multiple centres and over a wide demographic area will be acceptable for implementation into a routine trial. Computers are now widely available and can do multiple tasks simultaneously at quite high speeds, saving time, thereby allowing an action based on an output from a model to be implemented as soon as possible. In the medical domain where time is vital, the earlier a diagnosis is made and the earlier treatment can start impacts on a higher survival rate and a shorter period for a patient to be healed, if possible. The onset of collecting structured data in electronic format (increasing availability in private practices, clinics and hospitals) and the ongoing improvement of computer methods (learning models), has given recognition to specific prognostic systems. Both general information about a patient‟s diagnosis and "patient specific information: demographics, past medical history, current treatments, clinic-related information such as personnel skills and available facilities, and, more recently, gene expression levels at the tissue level can be used in the model building (Ohno-Machado, 2001b). Amongst the many models available, there is a subset that uses data to "learn" and create a model that can be applied to new cases. Those related to "survival analysis" are gaining a wide recognition in the medical field. The models learn from actual cases, i.e. data collected over the duration of diagnosis and treatment to a curable stage or death. The proposed model is based on a real trend in the interactions between variables and gives a real output that can be expected in similar cases. There are no assumptions made in building the model, only relationships between the various parameters are determined. In essence they mimic reality. If there is a large enough set of data, the model becomes credible and can be justified in its implementation into a clinical trial for new patient‟s prognosis. This will not be a stand alone solution as the appropriately trained clinician is there to ensure that a reasonable prognosis according to his diagnostic abilities is obtained from the model. The clinician can then decide whether to use the output of the model to treat the patient or use his judgement from his classical training. Some current applications in the prognostic categories that have been implemented are the APACHE scores for estimating death in ICUs. The risk associated with heart disease can be determined by the Framingham risk model which uses logistic regression methods. There are also other type of models like the Glasgow scores for coma and the clinical stagings of cancer where the methods are evaluated manually and then compared with prognostic data obtained experimentally (Ohno-Machado, 2001b).

The different methods used for predicting survival have been limited in its use in real data sets from actual clinical trials. Artificial data sets are ideal for preliminary model building where control of data is possible. Real world clinical data when applied to a model will allow for

easier adoption. In real data sets, censoring of observations, noise, and missing data are normally present and hard to control. Any survival predictor model adopted has to be able to deal with these variations (Ohno-Macahdo, 1997a). ANN models have been used successfully for building prognostic systems and provide alternatives to the traditional survival analysis tools, such as the Kaplan-Meier estimate and the Cox proportional hazards methods. These traditional methods have been widely used in the medical field to identify markers of disease progression by selecting significant variables, rather than predicting survival for populations or individual patients. This method relies on the researcher‟s ability to select the appropriate variables thus lending a bias to the proposed model, whereas in the NN model the training of the model with real world data ensures that the given outcome portrays a realistic picture of the variables that affect the prognosis of the disease. In ANN there are no assumptions made on the distribution of the data. Outcome prediction is a complicated process in medicine, especially if independent estimations for the various time periods are merged in a meaningful survival curve. The AIDS epidemic has influenced communities with financial, psychological, sociological, and medical dilemmas. The results of a reliable model for the prognosis of AIDS progression will impact on patient‟s informed decisions, the clinician‟s treatment plan, the administrator‟s allocation of resources and the design of clinical trials by researchers. For specific diseases real data is not readily available in large numbers for use in neural network modelling. Simulations with artificial data sets can be used in developing relationships between variables, but gaining the trust amongst the readers of the study takes some convincing. (Ohno-Macahdo, 1997a) had obtained permission to use data from the AIDS Time-oriented Health Outcome Study (ATHOS). Patient information was collected from clinics and private practises in the Los Angeles and San Francisco Bay area. 588 records were selected with a range of variables, e.g. age, gender, race, laboratory results weight loss, medications, blood counts, etc. Because of the large number of variables, a grouping system was used with a combination to represent patients with similar characteristics. The data was used to relate the accuracy of the Cox proportional hazards model and a neural network with respect to prognosis of death resulting from AIDS. The leave-n-out technique was adopted for analysis, since the observation set was not large. The data was divided into 10 training sets and 10 different models were built for each of the two methods, *viz*. Cox regression and NNs. For the NN the output contained 4 binary variables with death in a particular year, i.e. death in year 1, death in year 2, death in year 3, and death after 3 years. The survival records from the database were used to categorise the patients into intervals. The output of the network estimated whether a patient died in a given interval. Of the 588 cases selected, 200 of them had various missing data. Results that represented the mean value or mode were substituted for missing data. The neural network had fewer data points because the Cox model was divided into four discrete categories. The areas under the receiver operating curves (ROC) for both the methods were not significantly different (Ohno-Macahdo, 1997a).

This cannot be considered to be a true comparison between the models as the information and format presented was not identical. For statistical purposes, it can be accepted for comparison. The Cox model was used to explain which variables were most important for the prognosis of AIDS. There are no guidelines as to what constitutes a good predictive model. Ideally the models outputs have to be compared to the performances of a clinician. Each of the methods have their advantages and disadvantages, therefore a combination of different models may be a more accurate tool, for example, the Cox model can be used to determine the significant variables which can then be applied to a neural network to provide prognostic distinctions.

Postoperative residual cicatrisation (PORC) after surgery is common and its detection has a high error rate. Laffey *et al* (2003) have hypothesised that a neural network would enhance prediction of PORC. Residual neuromuscular block after surgery may be a significant problem, even after the use of medium-or-short-acting agents. There is still insufficient proof that a peripheral nerve stimulator (PNS) can be used for the reduction of incidence in clinically significant postoperative residual cicatrisation (PORC). Human error in assessing PNS data is very common, possibly stemming from perceptual limitations. Errors of judgement in anaesthetic practice may also result from pressure of work. A train-of-four value that is greater than 0.7 has been suggested as a minimum criterion for safe tracheal estuation, as lower values may be associated with impaired airway protection. Higher train-of-four threshold values have been proposed, although they cannot be reliably assessed by the naked eye. Transducer or electromyography monitoring systems are unlikely to enter the clinical practice in the immediate future, suggesting that neuromuscular block will remain semi-quantitative. Several variables can predict successful reversal of neuromuscular block. However, the failure rate observed suggests that these are imperfect and insensitive in practice. Laffey *et al* (2003) tested the hypothesis that a neural network-based model would enhance the prediction of PORC when compared with human decision-making. To construct the predictive model, a feed forward, back propagation NN model was developed using a commercial NN software package Neuralyst 1.4 integrated with a spreadsheet program (Microsoft excel 98). A training algorithm an input layer, an output layer and a hidden layer with four nodes was used. The number of training cycles, and the learning rate (*viz*. the degree of weighting adjustment between cycles) was preset by the investigators. Two input variables that were indicative of successful recovery of neuromuscular block, i.e. degree of spontaneous neuromuscular recovery before antagonism and the time elapsed since administration of pharmacological antagonism was chosen. To ensure validity (specifically to prevent over learning) the network was trained a total of 40 times using these variables. The „leave out k" or jack-knife method which trained the neural network on all data except for a group (in this case 1 out of 40 patients was left out, i.e. for each run one patient at a time was left out). A pass through the data set using every individual in turn as a single test case

allowed comparison of the predictive powers of the training and test phases. The performances of the training and test phases of the artificial neural network were essentially identical, the network correctly classifying 38 of 40 during testing (χ^2=0.3, P=0.57). The test phase of the neural network performed with a sensitivity of 0.96 (25/26; (χ^2=44, P<0.001), relative to clinical performance and a specificity of 0.93 (13/14; P=0.54. Fisher"s test). In this small test group, with a high incidence of residual neuromuscular block, a simple neural network model was able to predict the likelihood of normal or abnormal neuromuscular function at the time of decision to perform tracheal extubation more accurately when compared to a clinician. This implies that even with restrictions and uncertain data, artificial neural network-based predictions of drug pharmcodynamic relations is the same, but sometimes it may outperform human assessment in this setting. In the above study the diagnostic variables used by the software were deliberately limited to two of the commonest ones used in clinical practice, instead of allowing all possible information to be used. The use of simple routinely measured variables for a neural network-based prediction is useful in estimating the likelihood of clinically significant residual neuromuscular block at the time of tracheal extubation with a significant performance improvement over that of the anaesthetic trainees from whose practice the data were derived.

If the predictions of multiple models are combined, then the overall accuracy will be greater than any of the individual models (Hayashi *et al*, 2002). The authors used a series of neural networks for predictions. Instead of taking the average of the predictions of individual networks, a more sophisticated model that combines the predictions from the first level of networks as inputs for a second level of neural networks. A second level has 16 input units which correspond to the outputs of the four groups of the first level of networks. Using a first level of neural networks, an accuracy of 79.75% was achieved compared to 83.47% for the addition of a second level of neural networks.

3.9 Previous neural network studies on survival of cancer patients

Statisticians have developed non-linear approaches such as splines, trees and local methods for survival analyses. NN models provide an alternative to these methods and offer an economical framework compared to that of splines. There has been a recent growth in interest in the use of NNs in the survival analysis from two sources: prognosis problems have reached NN researchers who work in the medical field and some medical statisticians have become interested in NN techniques.

Ripley (1998) presented a study on the applications of NN methods for the prognosis of breast cancer. The basis of the comparisons of the models is prediction of whether or not relapse occurred within 5 years of diagnosis. The initial treatment was to remove the tumour and this was followed by one or more of chemotherapy, radiotherapy or hormone treatment. The primary task that was investigated was the development of techniques to extract the maximum benefit from a minimal subset of possible patient information and to provide a quantitative method for use by clinicians. The task was essentially one of using data on prognostic factors (such as age, histology, size of tumour, types of treatment, etc.) at the time of tumour removal to predict the risk of recurrence. While the proposed treatment would not be used when making a prediction of risk for a patient, the treatments given to past patients had to be included in the models. This allowed for the identification of groups of patients who had received the same treatment - within such groups the effects of prognostic factors should not be confused with the effects of treatment. Access to treatment by patients was not allocated randomly thus therefore the effects from the analysis cannot be compared. There are two types of prognostic factors: those that give information about the patient such as age, and those giving information about the tumour and its growth rate, such as the size of the number of affected auxiliary lymph nodes. Predictions from available information can be limiting as patients with similar values on all factors had greatly varying times to relapse. This problem has been identified by analysis but it cannot be addressed by it. The reason for it may be that the risk of relapse depends on some unmeasured factor or is simply subject to great random variation. The other factor is the complexity of the human biological system which makes people unique in their own way, thus displaying a variance in effects of a particular disease even though both patients exhibit the same symptoms and physiological features. From a statistical point of view, censoring is the main factor. The time to relapse is unknown, only a minimum value can be deduced. This method is adopted since the patient did not return for treatment or was no longer available for some reason. The patient could have died from an unrelated cause, or it just means that there is no relapse.

Statistical modelling of the time of a specific event (survival time) has various applications in numerous fields of research. Regression models for determining survival time in biomedicine are used to determine the effect of treatment on patient outcome. Research on prognostic factors for any disease is to discover what risk each patient is exposed to and treat accordingly. In the past twenty years only certain basic qualitative variables were measured in cancer related studies. The complex nature of the biology of cancer has encouraged research into the various tumour features. Standard prognostic factors have been combined with biological parameters for quantitative analyses of blood or tissue tumour. Flexible regression tools which are based on feed forward artificial neural networks (FFANNs) have now been widely adopted for

predictions because of their non-linear nature and the sheer number of prognostic factors (Biganzoli *et al*, 2002b).

Survival studies are essential in the clinical environment as it gives clinicians opportunities to critique their method of diagnosis and plan treatment regimes accordingly. From a patient point of view it is vital as it gives them an indication of time available to do the things they need to do and to plan their remaining time with family members and dependents. Conventionally, the Kaplan-Meier (KM) non-parametric technique has been used for preliminary analyses of survival data. For each variable, e.g. males or females, a separate survival curve can be plotted to find the significance of the variable on the predicted outcome.

3.10 Model selection

Mathieson (1998) used classical models of survival in which linear predictors were replaced by non-linear predictors to model using neural networks. Seven models were compared, including log-logistic and log normal distributions, non-linear Cox regression, the fourth is an extension of the third varying over time, while the other three grouped survival time as a discrete variable. If hidden units are included in the network then a local maxima is obtained instead of a global one. The outputs from the neural networks were used to compute the probabilities for each of the models. The chosen networks were trained using different random starting weights and average results. Mathieson (1998) has shown that this method is advantageous when dealing with local maxima when fitting neural networks. The aim was to find the best number of hidden units and amount of weight decay for a particular model. The possible models were then trained several times on the same training data using different starting points and the probability (or density) which occurred in the likelihood was averaged for each patient individually (before taking the logs and summing). Once a model was chosen, multiple estimates of relapse before five years were averaged for every patient in the test data set to give a single result for each patient.

Ripley (1998) used different model types which were related to predictions of relapse within 5 years. The available data set was small, therefore it could not be divided into two sets (one for training and one for testing), so that a five-fold cross validation was adopted to ensure that all data points from the sets would be used at some given point, i.e. each set was used as a test set. The models were chosen to fit four-fifths of the data, while each fifth was kept back. The fifth kept back was used to predict the probability of relapse within five years. A cross validation technique was adopted to determine the hidden layer and amount of weight decay for each of the proposed models, as illustrated in Figure 3.1.

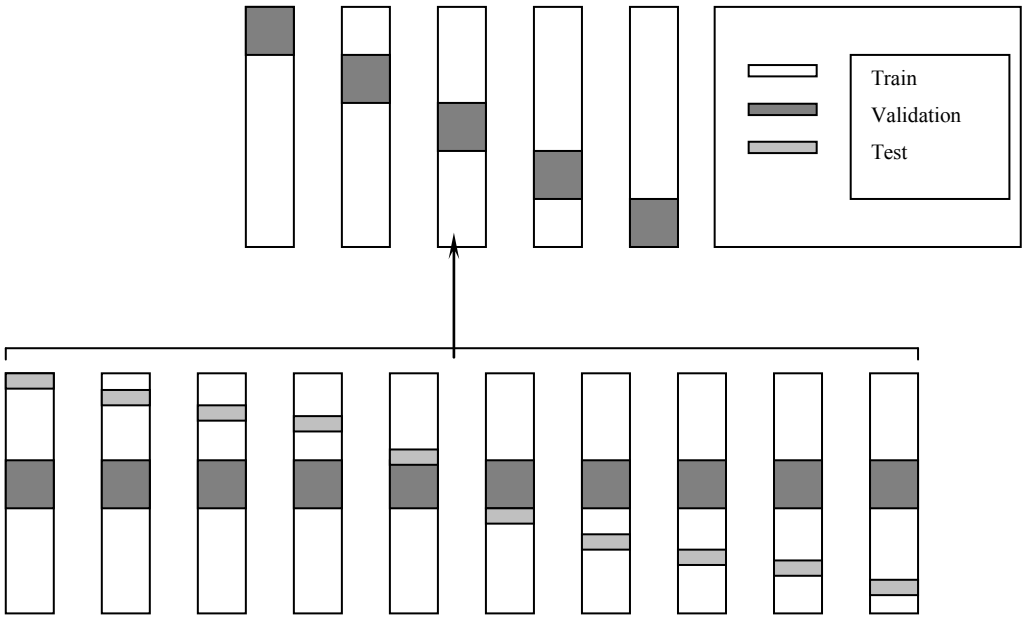

Figure 3.1 Nested cross-validation schema: ten-fold within five-fold

The second level of cross validation used ten divisions of data instead of five. The sum of the log probabilities on the validation data summed over the ten partitions was used to compare the models by. A scaling of variables was done to fall within the range 0 to 1: this is essential to equalise the effect of weight decay on each variable. Networks were fitted with the number of hidden units varying over 0 (a linear fit), 2, 6 and 10 or 20, depending on the number of output units in the model (10 for the multiple output models, 20 for the models with a single output). Weight decay parameters were varied according to the number of hidden units in the model, usually over the range 0.001 to 1, with reduction by a factor of 25 for the weights from the bias units in each case. The weight decay parameters were chosen so that in general the test statistic (-log probability) had a minimum within the range of values tested, for a fixed number of hidden units. As the weight decay was increased all the weights to the hidden units became very small and the fit (using the skip-layer units) was effectively linear. "Probabilities were calculated from the predicted values of the network for the patient in the current validation set. The probabilities were averaged over five fits from different random starting weights (with the same combination of size of hidden layer and amount of weight decay) and the sum of the logs of averages accumulated. The combination of the number of hidden units and weight decay which achieved the highest log probability sum over all the patients in the current training set was selected (viewed as the combination of ten validation sets). The corresponding model was fitted to the whole of the current training set which is 4/5 of the data. The fitting was repeated 25 times and the average probability of relapse within five years was found for each patient in

the current test set (1/5 of the data). There were 5 starting weights. For each weight there were five batches of the 4/5 data set for training therefore a total of 25 times in total" (Ripley, 1998).

3.11 Missing values

Ripley (1998) had many patients for whom information was missing. Excluding all those patients would have seriously reduced the size of the dataset, thus reducing the credibility of the results as a non-representation of the system under analysis. Two methods were used to cope with this problem: exclusion of the variable Chalkley vessel count (CVC) completely and use of the dataset of the patients with the known values for the above variable. The second approach was to fill in the missing values and retain all the patients. Two simple methods were used to fill in the missing values: regression on the known covariates and copying a value from the „nearest neighbour". The regressions were performed in sequence. The variable with the least missing values were regressed as a first pass. The fitted values were in turn used to regress to the second variable with the least missing values. The remaining missing values were then filled in by regressing on all available variables for each patient. An alternate set of data was produced using the nearest neighbour method, i.e. simply copying the value of the nearest neighbour for whom the value was available. Distance was defined as the square root of the average squared distance between available variables, after standardising all variables to have mean zero and variance 1. Only one neighbour was used for simplicity. From Ripley"s (1998) analysis it was found that the regression method was not particularly successful, but was better than the nearest neighbour technique. A small dataset with all patient information was analysed and compared to a full dataset where missing values were filled in. Both sets of data were then analysed separately using the proposed models. There was little difference between the results on the full dataset and those achieved on the small one. Sensitivity was found to be lower than specificity for all models.

3.12 Classification of data

Studies concerned with survival do not always have data in which time features at all. Sometimes the time interval of interest is so short that the analysis can be done, by simply classifying the data into two classes: event or no event. In this case standard classification techniques can be used as the question of censoring does not usually arise. Studies of survival have been classified in many ways but most commonly as: primary classification of studies whose methods predict probabilities of events occurring during one or more fixed time intervals (a „classification" problem), and those that predict a continuous quantity such as the distribution

of survival time or the mean survival time (a „regression" framework). Consider the example of a 5-year study of survival after a diagnosis or an operation, since the mortality rate is high in the post-operative phase. An assumption is that patients were monitored up to a period of 3 years with a few up to five years. Another assumption is that all censored patients are still alive up to the five year period, resulting in a downward bias. In exploratory studies the bias is used to justify any variations in outcomes and trends. In most studies the target is prediction of an event or outcome, not the explanation. Ravdin *et al* (1992) gave an example based on their research: 268 patients were monitored for 60 months and of these 213 had died. The Kaplan-Meier estimate of the survival probability at sixty months was 50 %, even though the status of censored patients is unknown.

3.13 Fitting networks

A bias due to patient-censoring and over-fitting of data are some of the problems associated with survival studies using neural networks. In a phrase borrowed from psychology, the aim is to achieve good generalisation. In comparison to classical statistics it is more difficult with NNs. In classical statistics model building starts from a simple model, which starts with fitting a linear model and then progressing to a quadratic or an interaction term. Testing occurs at each stage to check if there is an improvement between the models chosen. There is no analogue with NNs, since varying the number of hidden units can lead to some complicated models. Weight decay leads to good generalisation and ensures that the fitted function is smooth. This feature is only available in limited standard software for fitting NNs. Over-fitting of data can be overcome by using regularisation techniques, i.e. by penalising the minimand by a multiple of the sum of the squares of the weight.

$$\text{Minimand} = -\log \text{likelihood} + \lambda \Sigma w_{ij}^2 \qquad (3.1)$$

This process is known as weight decay. By changing the number of hidden layers and/or nodes, andλ, the degree of over-fitting can be controlled. Cross validation allows for a combination of data sets, with each being used for training and testing at a given time. A bias leads to a simple model, while the variance is a result of a flexible model. This trade-off has to be evaluated based on the raw data that is used in the modelling (Ripley, 1998).

There are various ways to use more than two intervals, based on three possible predictions, with three possible architectures for each case:

- Absolute probabilities: These are the probabilities for an event to occur in each of several non-overlapping classes, e.g. „death in year 1", „death in year 2," „death in year 3" and „survive 3 or more years". Three different network architectures can be used to model the

above system: a single network with multiple outputs, a single network having a single output (with the addition of a time period to the input variables), or multiple networks where there is one output per network. Censored patients could be ignored (with potentially substantial bias) or a death period imputed for those censored patients. Lapuerta *et al* (1995) used a network with four outputs for their final predictions which corresponded to death in one of three 40-month periods or survival for ten years. During training patients lost to follow-up were used in censored data by imputing a death period. Two separate networks were trained for death in periods 2 and 3. Variables of patients who did not return in period 1 (who were assumed to survive period 1) "were used as the input to the period 2 network, if that predicted death, death in period 2 was assigned, but if not, then the period 3 network was used to impute either death in period 3, or survival for three years" Lapuerta *et al* (1995). For multiple networks having a single output, the censored patients can be ignored or event times have to be imputed. Censored patients could be dealt with more correctly by altering the likelihood function used for training:

$$\text{Log}\ \{Pr\ (\text{death in year 3} \mid \text{x}) + Pr\ (\text{survive 3 or more years} \mid \text{x})\} \qquad (3.2)$$

It is possible, but rather computer intensive with standard software, to fit a network with time as an input variable and a single output (no study is known to have used this method). If a patient was used for the time interval during which they were observed, the practical size of the training set would grow. The predictions may differ from those obtained by a multiple output model since the smooth functions that can be fitted are different.

- Conditional probabilities: "The conditional probabilities are modelled as

$$\Pr\ (\text{die in the } i\text{th interval} \mid \text{survive first } I-1 \text{ intervals}, x) = g\ (\eta_i\ ) \qquad (3.3)$$

where g is usually the logistic function. Then a patient dying in the ith interval contributes $\log\{\ g(\eta_i\ )[1\text{-}g(\eta_{i-1})]\cdots[1\text{-}g(\eta_i\ )]\}$ to the log likelihood, and a patient lost to follow-up in that interval $\log\{[1\text{-}\ g(\eta_{i-1})]\cdots[1\text{-}\ g(\eta_i\ )]\}$, and from this the log likelihood L can be computed. The „scores" $\eta_1,\ldots.\eta_k$ is given by the output of a neural network with k linear outputs. This type of model is referred to as a „life-table" or discrete-time survival model. It is also known as a „chain-binomial" model" Lapuerta *et al* (1995). For one output the likelihood is undistinguishable from the absolute probability model, so standard software can be used to fit the model with time as the input variable and a single output unit. This

method was used by Biganzoli *et al* (1998) with neural networks. Censored patients introduce little bias as long as the training data contains multiple copies of the each patient. Standard software can be used if multiple networks with a single output are used, but here censored patients must be ignored or death times imputed. Standard software has been used with multiple outputs to fit this model. Some outputs were coded as „not-defined" and not allowed to affect the error function (Lapuerta et al, 1995).

- Cumulative probabilities: These are the probabilities that an event will occur in overlapping intervals (e.g. „death in year 1", „death in year 1 or 2" and „death in years 1, 2 and 3". Here multiple networks each with one output can be used. Standard software may be used but a bias will be introduced due to censoring. Theewun *et al* (1995) imputed death probabilities for censored patients using the proportion who had died out of those that were at risk at the time of censoring (this was done recursively from the longest survivors). Another approach is a single network with one output and time as an input variable, which allows for the use of standard software. But this is at the expense of an even larger dataset than the other two such models. Here the patients who have an event recorded must be entered for each time period considered, whereas in the other two cases they were not entered after the event had occurred. This technique was used by Ravdin *et al*, (1992 and 1994b).

One of the problems of assessing survival is the lack of good statistics for prediction. In traditional models the emphasis is usually to identify the deviation from the assumptions of the underlying model and selecting the best model with respect to the experimental data rather than evaluating prediction accuracy. The analysis of predictive ability is only commonly found in studies which compare neural networks to other techniques. Predictions are usually compared on the basis of sensitivity, the proportion of relapses correctly predicted, specificity, the proportion of survivors correctly predicted, and accuracy, the overall proportion correct. All of these are considered as estimates of the underlying population probabilities. The R test statistic can also be used as it is an indication of the accuracy of the fitting of the data to the model. This is a common statistical variable that is a standard output of most statistical and modelling software.

3.14 Problems with comparisons of NN models

Several studies cited above have claimed that their NN model outperformed a Cox regression and/or clinicians, but such conclusions need to be critically assessed. It is often claimed that Cox models cannot provide individual predictions or predictions of time of recurrence. The

production of survivor curve estimates requires some extra knowledge of standard of-the-shelf statistical software as it is provided with most packages. An estimate of the time to recurrence is most probably intended to represent a median (or mean) value: median survival times can be found from individual survivor curves. Statistical models used in comparisons are often not as effective as it could be. Comparisons are generally performed on small sets, or sets with small numbers of events, where greater accuracy can often be achieved by simply predicting no events at all (Ripley, 1998). For example, Bottaci *et al* (1997) validated their model by the use of a test set containing 92 patients, but only six events. The accuracy achieved by the network was 90%, whereas simply predicting no events would have achieved an accuracy of 92%. When a study is carried out there will always be limitations depending on the availability of the type and amount of data, censoring, missing values and other factors that may be beyond control. The focus and field of interest will also influence the assumptions made and the methods followed in the analysis. There will always be an amount of bias that will favour the researcher‟s hypothesis. The literature should therefore be used with caution and applied appropriately with justifications of choice.

3.15 Summary

Proposed factor(s) are investigated, usually using retrospective data to establish whether they are able to identify high and low risk groups for recurrence or survival. There is often appreciable disagreement between the conclusions of different studies of the same factor. There are several possible reasons for these discrepancies, apart from pure chance (Ripley, 1998). The method of assessing may not be sufficiently reproducible between laboratories and different methods may be used for the same factor. Once values are established, the analysis is done on the basis of a categorisation of a continuous variable, usually in groupings, e.g. low, medium and high. This can lead to loss of information and confusion where the definitions of groups differ between studies. With traditional models linear relationships are frequently assumed, but this may not be appropriate for all continuous variables (Altman *et al*, 1992). Statistical models may fit badly as it is unusual to find the model validated on a test set, or explicit checks of the assumptions to be made. The sample populations may vary and since the different factors are often highly correlated, their prevalence may influence the result. Adjuvant therapy may mask the effect of prognostic factors on recurrence. Patients that are perceived to be at high risk of recurrence may be given the most aggressive therapy. If the treatment is successful, the factors will appear to affect their rate of recurrence less. Some studies are performed on small datasets as the aim may have been to assess whether the factor is useful enough for routine measurement. For stepwise regression techniques, there are often several models that fit the data equally well. Such models are also often unstable. If the dataset is changed slightly the selection

procedure may yield quite different results. Different studies use different sets of „standard" factors against which to compare the new one.

Models which can be used to fit very flexible models would automatically improve the effectiveness of prognostic factor studies. The multilayered perceptron has established itself firmly as one of the most powerful and versatile general purpose classifiers in use today. The neural network classifiers also score well against human experts which underlines what is perhaps one of the most significant findings, namely, the widespread potential for application of pattern recognition techniques to problems which might appear at first sight to be completely unrelated (Lisboa, 1992). Intelligent decision systems need to be designed to fit real world data so that the outcomes can reliably be applied to a clinical trial. For the process to work the proposed application has to fit into the institutional management of data so that it can be readily available. Regular updating and validation of the models have to be carried out. Data management in the midst of censorship is a major issue when dealing with reliability. This is still quite challenging for the few neural network applications that have been applied to routine clinical use.

The introduction of computers into the medical domain has extended the reach of the physician"s intellect with regards to diagnosis and treatment. While this current scenario exists, artificial intelligence tools have not and will not replace general clinical consultation. Neural network models will act as a support for the clinician"s decision making. There have been extensive applications in the domain of medical instrumentation. Inference based decision support systems are being used on a large scale for routine clinical use like gene expression analysis and for prognosis of MFI. Some of them are based purely on statistical methods.

NNs may be considered as tools in the hand of the user. They do not work by themselves but will only perform as well as they are applied. Knowledge of the field of application is therefore as important as familiarity with the networks themselves. One way of including this knowledge is by data conditioning. Wherever prior knowledge is scarce, neural networks provide new techniques for analysing the raw data. These procedures are robust against noise, adaptive to non-linearities and potentially fast. Generally a single additional item of patient data such as age or a smoking record can aid clinicians when making complicated decisions. Very few of these models are adopted routinely to inform complex clinical decisions. Doctors never prognosticate, since they work in the present based on current symptoms but research on medical decision making show this to be untrue. Sometimes models are developed to publish journal articles that have no clinical relevance. The main reasons for rejecting prognostic models are reliability and absolute proof that support decisions about patient care, i.e. evidence of accuracy, generality

and effectiveness. Models need to be developed together with clinicians. They need to be involved in all aspects of the model building process. Statisticians working closely with doctors will have more credibility than those done in isolation as a solution to a mathematical or statistical problem. This example of teamwork and adherence to principles of accuracy and effectiveness will ensure that the proposed models can be easily implemented in a routine clinical trial, to support the clinician‟s diagnosis and treatment regime (Wyatt *et al*, 1995).

CHAPTER 4
REVIEW OF BLOOD AND HAEMATOLOGICAL MALIGNANCIES

4.1 Introduction

Fundamental to all biological sciences is the understanding of the molecular biology of cells. Research is this area is beneficial to basic science but also to the practical applications in agriculture, biotechnology and medicine. The ongoing research in the understanding of the cellular and molecular basis of human diseases has lent itself to being applied to various medical case studies. The similarities and variations between cells is the key factor to understanding cellular biology. Animal and plant cells have basic features that have been preserved over time. For example, all cells have DNA coding in their structure, are surrounded by plasma membrane, and follow the same interactions to produce energy. Lower order animal and plant cells such as bacteria, amoebas and yeasts are unicellular and are still able to independently self replicate. Higher order organisms are multi-cellular, and are co-ordinated according to a particular manner to form specialised cells. Each group of cells are able perform specific tasks at various locations within the organism. There are 200 different kinds of cells in the human body. Each has been developed to perform a specific function such as memory, sight, movement and digestion. Features and characteristics that are common between the various types of cells can be studied collectively studied to have an overall understanding of cellular biology. Knowledge gained from one cell is extrapolated and generalised so that other types of cells can be studied. Computational tools and scientific methods have an impact on progress made in the field of biology and medicine. Scientists are continuously being challenged to go beyond their boundaries (Cooper, 1996). Cell growth occurs *via* a cell cycle which is controlled by a variety of enzymes, kinases and cyclins. Genetic coding in the form of DNA determines whether cells will divide and form mature or immature cells. Leukaemia is a cancer of the blood and is confirmed when there are abnormal numbers of immature cells present in the bloodstream. On diagnosis of leukaemia, further tests are done to determine the specific sub-type of leukaemia: ALL, AML, CLL or CML which will aid clinicians in planning a treatment plan for a particular patient. Results of these tests, i.e. full blood count, differential, flow cytometry and cytogenetics are also used as indications for a patient‟s prognosis. Being able to predict a patient‟s mortality whether good or bad allows the patient to plan his life and allows clinicians to determine the course of treatment. In some institutions conventional

therapy is used for all patients, while other run clinical trials and established treatment programs for high risk patients.

4.2 The Eukaryotic Cell Cycle

There are two main classes of cells which are defined by the presence or absence of a nucleus. Prokaryotic cells (bacteria) have no nuclear envelopes. Eukaryotic cells contain a nucleus in which the genetic material and cytoplasm are separated. The cell is comprised of two major classes of macromolecules, *viz*., proteins and nucleic acids. Nucleic acids within the cells are able to self replicate without external factors. RNA (ribonucleic acid) is independently able copy itself via catalysis. Through evolution ordered reactions between RNA and proteins have produced the current genetic code which has taken the place of RNA as the genetic material. Self-reproduction is a fundamental characteristic of cells and for all living matter. A parent cell divides into two cells, to form daughter cells. The process of division is called a cell cycle. The daughter cells can in turn divide giving rise to a continuous replication of similar cells. A single parental cell can lead to a whole population, all duplicates of each other. Even though cells have the ability to grow and replicate, the actual mechanisms involved have only been discovered relatively recently. Cell division has to be carefully monitored and co-ordinated to ensure that method of cell growth and replication of the genetic code remains intact throughout this transformation. The understanding of the molecules behind these events is both relatively new and as yet incomplete.

The three main components of the cell cycle are the activating enzymes, cell division cycle kinases (CDC) or CDKs, their activating cyclins, and CDC/CDK inhibitors, collectively called CKIs. In eukaryotic cells a series of protein kinases conserved during the evolution stages from yeast to mammal control steps in the cell cycle. There are three main proteins that regulate the cell cycle, *viz*., cyclin dependent kinases (CDKs), the cyclins, and the cyclin dependent kinase inhibitors (CKIs). The CDKs, once activated phosphorylate other proteins, allowing them to perform at a specific stage of the cell cycle. The CDKs are regulated by cyclins which activate the CDKs only when they reach a critical concentration. The CKIs interact with the complex formed between a cyclin and a CDC/CDK in such a way that they inhibit the kinase activity and prevent the cell from progressing through the cell cycle. In higher order eukaryotes, the cell cycle is controlled by growth factors which are responsible for cell proliferation. Single cell development progresses according to the requirements of the organism as a whole. The proteins involved in the cell cycle have a similar purpose to ensure that after appropriate stimulation, the cell accurately and completely replicates it DNA before cell division. Each protein has a unique

role during specific stages of the cell cycle, so that the events occur in the correct sequence (Gillet *et al*, 2001).

Cells having defective checkpoints are advantageous when selection favours multiple genetic changes. Cancer cells are caused by defective cell cycle procedures. This has encouraged the simultaneous study of both the cell cycle and the disease. Cancer cells are often missing one or more checkpoints, which facilitates a greater rate of genomic evolution. Basic cell cycle regulation needs to be understood in order to understand the mechanisms that lead to haematological malignancies and the importance of tumour suppressor genes. Cell division consists of four co-ordinated processes: cell growth, DNA replication, distribution of the duplicated chromosomes to daughter cells, and cell division. A typical eukaryotic cell cycle is illustrated by human cells in culture which divide approximately every 24 hours. Some human cells may proliferate rapidly, such as the epithelial cells of the small intestine, but even the quickest of these will take 24 hours to complete a single cycle. The active cell cycle has been divided into four phases. Mitosis (M phase) was identified first because of its distinctive morphological stages. The other phases are collectively called the inter-phase. Cell mitosis is the final step of a cell cycle which is a defined program that can be divided into four phases: the G_1-, S-, G_2- and M- phases. Increased knowledge of the structure and function of DNA led to the term S (synthesis) phase. From the knowledge of the M and S phase it was apparent that these two phases could not just run from one to the other, and that there is a gap between the two. These breaks or gaps occur between rounds of mitosis and DNA synthesis. The cell cycle comprises of: Mitosis (nuclear division) and interphase. The process of mitosis begins with the division of the daughter chromosomes and ends with cell division (cytokinesis). Since mitosis and cytokinesis take approximately an hour, 95 % of the cell cycle time is actually the interphase, i.e. intermittent time between mitoses (Cooper, 1996).

4.3 Haematology

Haematology is the study of blood and its related disorders which are referred to as haematological malignancies. The average blood volume of an adult is 4 to 6L: women have 4 to 5L and men 5 to 6L. Blood has a pH between 7.35 and 7.45 and represents about 8% of the total body weight. It is composed of 55% plasma (the fluid portion) and 45% formed elements or cells. Of the 45% formed elements, approximately 44% of the cells are red blood cells (RBCs), whereas only 1 % is white blood cells (WBCs) and platelets (PLTs). Blood plasma contains 91.5% water and 8.5% solutes. The solutes consist of three different kinds of proteins: albumins (55%), globulins (38%), and fibrinogen (7%); other solutes are electrolytes, hormones, non-protein nitrogen compounds, nutrients and respiratory gases. The reference

values for the cellular elements are as follows: RBCs (4.2 to 5.4 x 1012 L^{-1} for females and 4.7 to 6.1 x 10^{12} L^{-1} for males), WBCs (4.8 to 10.8 x 10^{9} L^{-1}) and PLTs (150 to 350 x 10^{9} L^{-1}) in adults. The ranges will change with age, gender, geographic location, and health or disease (Harmening, 2002).

Blood cells generate in the bone marrow, which forms the central core of any bone structure. Externally bone is quite hard, while the core is sponge-like and red or yellow. Blood is composed of fluid (plasma or serum) and cells. Blood cells originate from „stem cells", which divide to form three important subsets in the marrow are: red blood cells (RBC) platelets and white blood cells (WBC). Red blood cells are responsible for the transport of oxygen throughout the body. Oxygen is used for respiration in cells. This process releases energy into the body for all functions, from walking to blinking. Platelets react to form clots if there is any leakage from blood vessels. This natural instinct prevents excessive bleeding when there is damage to a blood vessel by injury. White blood are responsible for protecting the body against foreign matter like bacteria, viruses and other foreign bodies such as a wood splint in your hand.

The blood in mammalian species includes a number of differential cell types essential for survival. Erythrocytes transport oxygen; platelets mediate blood clotting and support tissue integrity. Neutrophils, eosinophils, basophil granulocytes, and monocytes are vital too as a defence mechanism against bacteria, fungi, parasites and viruses. T-lymphocytes, natural killer cells, and dendritic cells all function as antigen-presenting cells and in cell-negotiated immunity. B-Lymphocytes are the source of antibodies. Multiple humoral and cellular factors control the amount of these cell types that circulate in the bloodstream and can be rapidly adjusted to meet the immediate need. Infection by a variety of micro-organisms results in almost immediate release of mature neutrophils from the marrow storage pool thereafter an increase in granulocyte and monocyte production until the infection is cleared. When a haemorrhage or acute haemolysis occurs marrow reticulocytes are signalled to be released. This results in increased red cell divisions until the desired level for a particular patient is obtained. Decreasing platelet number, acute anaemia, and tissue destruction or inflammation stimulates the formation and release of platelets. The control of T- cell and B-cell production is quite complex and occurs in response to immune stimuli, e.g. foreign antigens, and this control of increasing production also occurs within the various subsets of these cells (Harmening, 2002). Immature cells of any type are abnormal. Cells should be examined for abnormalities in the nucleus or cytoplasm.

4.3.1 Erythrocytes (red blood cells)

Red cells consist of plasma membrane surrounding a solution of proteins (haemoglobin) and electrolytes. A normal mature erythrocyte is a biconcave disc that is 7 to 8 μm in mean diameter and 1.5 to 2.5 μm thick: it has a mean volume of 90 fL (femtolitres). Red blood cells should be fairly uniform in size and relatively round in shape, with a small area of central pallor and no nucleus or inclusions (Harmening, 2002). The erythrocyte carries oxygen from the lungs to the tissues where it is exchanged for carbon dioxide. These cells are pliable or flexible and deformable, thus allowing them to change their shape in order to pass through the microcirculation to transport oxygen.

4.3.2 Platelets (thrombocytes)

Platelets are approximately 2 to 4 μm in diameter and vary in shape. Platelets contain particular molecules needed for haemostasis and are able to adhere, aggregate and supply a surface for coagulation reactions.

4.3.3 Leukocytes (white blood cells)

Myeloids and lymphoids form the basis of this group. Myeloid cells are referred to as polymorphs since their nuclei have variations in their shapes or granulocytes since they contain granules of chemicals used for combating bacterial invasion. The chemical composition dictates what they are called: neutrophils, eosinophils, basophils, lymphocytes and monocytes. Lymphoid cells are made up of lymphocytes, monocytes and plasma cells which can kill some viruses and cancer cells. These cells also produce antibodies to combat viruses and bacteria. Lymphoid cells are generated in the lymph nodes in specific parts of the body. Lymph nodes are distributed in the neck, armpit, and groin or inside the chest or stomach.

4.3.3.1 Segmented neutrophil (filamented neutrophil, polymorphonuclear neutrophil)

In older children and adults there are 50% to 70% of mature granulocytes, which are also, called segmented neutrophils. The nucleus of the segmented neutrophil is separated into two to three (usually three) lobes with a narrow segment or filament connecting the lobes. Approximately 6% of neutrophils have one lobe (band), 35% have two lobes, 41 % have three lobes, 17% have 4 lobes and 2% have five lobes. Segmentation of the nucleus enables these cells to pass through an opening in endothelial lining cells of capillaries and to “home in” on selected prey (such as micro-organisms causing infection). Neutrophil secondary granules are

lysosomes that contain alkaline phosphatase. Mature neutrophils are approximately twice the size of normal erthrocytes.

4.3.3.2 Band neutrophil (non-segmented neutrophil non-filamented neutrophil)

Peripheral blood of healthy individuals contains 2% to 6% of the band neutrophils. Band neutrophils have a nucleus with a horseshoe shape in which the opposite edges of the nucleus being mostly parallel. These cells do not have a nucleus separated into lobes connected by a filament like the segmented neutrophil. The chromatin of the nuclear are clumped forming a dark mass at each pole where the pole is destined to be. The secondary neutrophil granules are small and evenly distributed. There can sometimes be a doubt when trying to differentiate between segmented and banded neutrophils. One has to decide whether the link between the lobes is narrow enough to be called a filament or wide enough to be identified as a band. In attempting to differentiate between a segmented and a band neutrophil, identification should not be made on a single morphological characteristic but on combined features. If there is a doubt regarding a borderline cell, the questionable cell should be placed in the mature category.

4.3.3.3 Eosinophil

Large, round, secondary, refractile granules that have an affinity for the acid eosin stain recognize these cells. The granules are spherical, uniform in size and evenly distributed. Normal adult peripheral blood contains 0 to 4% eosinophils. Normal blood eosinophils are about the size or slightly larger than neutrophils and have a band or a two-lobed nucleus with condensed chromatin; rarely does an eosinophil have three lobes. There is a diurnal variation in the percentage of circulating eosinophils, with an increase at night and a decrease in the morning.

4.3.3.4 Basophil

Although basophils constitute only 0 to 2% of normal blood cells, the large, abundant, violet-blue (or purple-black) granules aid in the immediate recognition of this cell. These granules vary in size from 0.2 to 1 μm and are visible above the nucleus as well as lateral to it thereby obscuring most of the nucleus. They are coarse and unevenly distributed, vary in number, shape, and colour, and are less numerous than esoinophil granules. Basophil granules are also water-soluble. In cells that are poorly fixed during staining, the centre of the granule may be washed away, leaving a small colourless cytoplasmic area. Basophils show a similar diurnal variation to that of eosinophils, increasing at night and decreasing in the morning.

4.3.3.5 Lymphocytes

Lymphocytes are the second most numerous cells in the blood, comprising 20% to 40% of the adult blood cells. Most lymphocytes are small varying from 7 to 10 μm and usually round with smooth margins. There are some intermediate sizes and some large lymphocytes but it is not a reliable basis for determining the age of metabolic activity of lymphocytes because there size varies with the thickness of smear. They tend to become spherical and small in thick areas of the smear, in the thinnest end of the smear lymphocytes may spread out and appear large. The margins of large lymphocytes are frequently indented by neighbouring erythrocytes causing them to have a serrated shape. Most lymphocytes do not have granules but some large lymphocytes may have a few well-defined granules that vary in size, are unevenly distributed, and can be easily counted. The diameter of the nucleus of a small lymphocyte in peripheral blood is slightly larger than, or the same size as a normal erthrocyte in the same microscopic field. The lymphocyte‟s nucleus in relation to its cytoplasm is large (N:C ratio is 4:1) and the nuclei are round or slightly indented. Nucleoli are present in some lymphocytes thereby making them capable of growth and replication.

The lymphoid progenitor cell is derived the haematopoietic stem cell. The common lymphoid progenitor cell can produce either T cells or B cells. The type produced depends on the location in the human body. T cells are produced in the thymus and B cells in adult bone marrow. Null cells, or third population cell also originate in the bone marrow, but the maturation sequence is unknown. T, B and null cells cannot be separately identified morphologically but can be distinguished functionally and by immunological markers that use target specific monoclonal antibodies. Lymphocytes proliferate and mature into fully functional immune cells in the primary lymphoid organs such as the thymus and bone marrow. Lymphocytes from the lymphoid organs contain lymph nodes, spleen, and mucosal tissues that communicate with antigen-presenting cells (APCs), phagocytes and macrophages with the immune response system.

About 10 % of the lymphocytes are much larger cells with more abundant cytoplasm and a reduced nuclear: cytoplasmic ratio. Prolymphocytes as seen in prolymphocytic leukaemia are relatively large cells in which several nuclei are visible. Lymphocytes with relatively small amounts of cytoplasm, large nuclei, less dense nuclear pattern and well defined nucleoli probably do not occur in healthy persons, except possibly in small numbers in the body of infants. In acute lymphoblastic leukaemia (ALL) they may be the prominent cell type. Sometimes, the cells, although clearly blasts, are quite small (microblasts) and do not exceed mature lymphocytes in size (Lewis *et al*, 2001).

4.3.3.6 Monocytes

In the thin areas of the peripheral blood smear, a monocyte measures about 15 to 18 μm and is larger than the mature neutrophil. Monocytes have abundant cytoplasm in relation to the nucleus (N:C is 2:1 or 1:1).There may be a varying number of prominent granules in addition to the small granules. Some may appear non-granular suggesting rapid turnover. Digestive vacuoles may be observed in the cytoplasm. In disease states, phagotised erthrocytes, nuclei, cell fragments, bacteria, fungi and pigment may be present. One of the distinctive features of the monocyte is the appearance of convolutions (like those in the brain) in the nucleus. Another characteristic is the lacy, often delicate chromatin network of intermingled fine strands with small chromatin clumps. Monocytes vary in shape with many cells being round while others reveal blunt pseudopods that are manifested of their slow mobility. These ameboid cells continue to move while the blood film is drying and become fixed before the cytoplasmic extensions are retracted. There are four helpful characteristic features of the monocyte: nuclear convolutions; lacy, often delicate chromatin; dull grey-blue cytoplasm; and blunt pseudopods. Kinetic studies have revealed that the half-life of monocytes in circulation ranges from 8 hours to 3 days before these cells enter tissues and are transformed into macrophages. Monocytes account for 2% to 9% of normal blood leukocytes.

4.4. Cell development

Blood cells are not born as mature cells. They develop in stages in the marrow from infancy, and childhood, to adolescence and adulthood. In acute leukaemias the maturity of their cells determines the disease subtype. The stem cell (parent of all blood cells) produces immature cells that are called blast cells, e.g. myeloblasts or lymphoblasts. Myeloblasts in turn will transform into promyelocytes, and eventually leading to myelocytes. Mature cells develop into granulocytes, such as neutrophil granulocytes (or neutrophils for short. Lymphoblasts evolve into lymphocytes and plasma cells (Cooper, 1994).

4.5 Analysis of composition of blood

Analysis of blood components is the most widely used form for diagnosis of various ailments and diseases. The primary step in assessing haematologic functions and the presence of disease is an examination of the cellular elements in the blood. Examination of the blood frequently gives important information that aids in the diagnosis of haematological disease and may suggest further tests to quantify the various components or specify the type of malignancy, e.g.

A low white cell count would indicate a leukaemia using the FBC (full blood count analysis) but flow cytometry analysis can specify the type of leukaemia, e.g. AML or CLL (Cooper, 1994). Understanding the morphology of cells allows for specialized techniques or methods to be used for analyses. The shape, size and content of the blood cell is used in the identification process, when analysing, via a combination of staining, and microscopic observations. It also influences reactions with commercial probes and monoclonal antibodies which allows for identification via assays and fluorescence absorption (flow cytometry).

4.6 Haematopoiesis

Haematopoiesis is the growth and generation of various types of cells found in blood. The haematopoietic system is characterised by a constant turnover of cells that continuously maintain a large number of erthrocytes, leukocytes and platelets. This massive cell population is distributed throughout the body via a complex network of tissues, organs, stem cells and regulatory factors. This network is responsible for the maturation and division of haematopoietic stem cells into the lineage-committed stages that transport oxygen and excrete carbon dioxide (RBCs). Fight infection (granulocytes), perform immune functions (lymphocytes), and maintain homeostasis, a process in which blood clots and bleeding is halted. The haematopoietic stem cell has the ability to divide infinitely and generate a continuous supply of cells. In addition it has the ability to differentiate into progenitor cells of lymphoid and myeloid lineages. The haematopoietic system is made up of the bone marrow, liver, spleen, lymph nodes and thymus. The organs and tissues are involved in the division, grow and death of cells. Haematopoiesis is evolved from the stem cells that support haematopoiesis, the progenitor cells that are committed to particular cell lines, and the regulatory factors (growth factors) to which the haematopoietic system responds. These features allow the system to respond to stimuli such as infection, bleeding, or hypoxia by increasing haematopoiesis with emphasis on the specific cell type that is desirable. Most descendants of the stem cells are committed to differentiate. This occurs through a series of steps or stages, each of which leads to a further restriction of lineage choice, until finally the descendant cells are limited to a single lineage. After commitment to a specific lineage, the progenitor cells continue to differentiate and mature into the terminally differentiated cells found in peripheral blood. The amplification of cell numbers that accompany the differentiation process is very large.

Haematopoiesis is sustained in a steady state as production of mature cells equals blood cell removal, e.g. when a person donates blood, mature cells would be produced to replace the blood drawn from person. Bone marrow haematopoietic activity consists of a stem cell pool

and a bone marrow pool with eventual release of mature cells into the peripheral blood. The macro-environment of the bone marrow consists of both morphologically unidentifiable multi-potential stem cells (MSCs) and uni-potential committed stem cells. The bone marrow pool can further be separated into two distinct regions: cells that are proliferating and maturing, and cells that are stored for later release into peripheral blood (Harmening, 2002).

4.7 Haematopoietic cell cycle kinetics

Stimulation by haematopoietic growth factors results in haematopoietic cells undergoing a continuous generative (G) cycle in which cells divide, differentiate or remain dormant. The bone marrow contains cell populations in all phases of cell development. The generative cycle is divided into five phases: G_0, G_1, S, G_2 and M phase (Harmening, 2002). The cells enter a resting or dormant phase (G_0) after dividing. It then enters the G_1 phase which is the post-mitotic rest phase and which directly precedes the deoxyribonucleic acid (DNA) synthesis phase. The cell proceeds into the synthesis phase (S) of active DNA synthesis where the DNA is duplicated. Thereafter the cell enters the premitotic rest period (G_2) as the cell readies itself for the mitotic period (M). During the final or M phase, there is cellular division of the chromosomes in the nucleus and the cytoplasm, resulting in two daughter cells. T_G is the cycle of one complete mitotic division. After final differentiation, the cell leaves the cycle as a non-dividing cell (G_{ND}).

4.8 Haematological disorders or malignancies

Cell mitosis is the final step in the cell cycle. A number of surveillance systems (or checkpoints) control the cell cycle and interrupt its progression when DNA damage occurs or when the cells fail to complete a necessary event. The cell cycle major checkpoints are: the DNA damage checkpoint, the spindle checkpoint, and the spindle pole body duplication checkpoint (Beutler *et al*, 2001). Cell death by apoptosis is the result of cell cycle checkpoint failure. However, small numbers of genetically altered cells may survive. When selection favours multiple genetic changes it is advantageous. Cancer or leukaemia is the result of one or more checkpoints that are missing. Mitosis is controlled by the M-phase and S-phase promoting factors (MPF and SPF). The key element of the SPF subunit is cdk (cyclin dependent kinase). The second component is cyclin B, which is synthesized in interphase and degraded in mitosis. Cdc2 interacts with cyclin B in mitosis, whereas the cdc2/cyclin A complex is formed before mitosis and is required for progression through the late G_2 phase. Thus the cyclins A and B are also called the mitotic cyclins since they are up-regulated in late G_2 and G_2/M and undergo proteolysis in M phase. The exit from mitosis is characterised by the abrupt ubiquitination and

subsequent degradation of cyclin B. Cells with defective cyclin B degradation mechanism or without mitotic cyclin B easily become aneuploid. The cyclin B/cdk2 checkpoint is very often defective in malignant cells leading to uncontrolled M-phase entry and aneuploidy. All cyclins share an approximately 150 amino acid region, called a cyclin box, which reacts with the cdks. The G_1 cyclins (C, D and E) and the mitotic cyclins (A and B) form distinct categories, although cyclin H and the type T cyclins fall outside these two major groups. In G_1 phase the most important substrate of the cdk4-cyclin D and cdc6-cyclin D complexes is retinoblast (rb). Deletions, mutations, and translocations of rb are common in various malignancies; homozygous deletions of the $p^{16INK4A}$ are even more frequent. Ectopic expression of both cyclin A and cyclin E restores rb hyperphosphorylation and causes cell cycle arrest in cancer cell lines.

The complex cell cycle network has its parallel in several oncogenes and tumour suppressor genes that influence carcinogenesis and tumour progression. The products of oncogenes, the oncoproteins, lead to or facilitate the transformation of a normal to a malignant cell. As a general rule if a mutation causes a functional loss of the gene product, and the recessive loss of function leads to uncontrolled cell division, the underlying gene is called a tumour suppressor gene.

4.9 Blood cells in leukaemias

The word leukaemia is defined from the Greek word *leukos* which means white and *haima* which means blood, resulting in the term white blood. The specific cancer type, leukaemia is a considered to be a malignant disease of the hemaetopoietic tissue. Abnormally grown cells replace healthy bone marrow components. Leukaemic cells are usually (but not always) present in peripheral blood and usually invade reticuloendothelial tissue, including the spleen, liver, and lymph nodes (Halbook, 2005). These cells may also overrun other tissues, infiltrating any organ of the body. Leukaemia ultimately causes death if left untreated. Abnormal and uncontrolled growth in leukaemia patients is similar to most cancers. In this case too many cells are detrimental to the patient. Leukaemic cells are underdeveloped white cells that are unable to perform their usual functions of combating infections. These cells circulate in the blood steam causing irregularities in various parts or organs (Harmening, 2002).

4.10 Symptoms and signs of leukaemia

Insufficient red blood cells for carrying oxygen for energy production will lead to weakness and fatigue. The patient becomes anaemic and pallid due to the lack of energy. There is easy

bruising and excessive loss of blood from the gums and nose. Other symptoms include skin rashes which tend to form spots. This develops because there are insufficient platelets for the clotting of small blood vessels. Frequent infections and fever will be due to fewer, normal mature white blood cells being available to fight infections. An accumulation of leukaemic cells in the bone marrow or around the brain can cause aches and pains in the bones or headaches. Glands or lymph nodes in the neck, groin or armpits can swell if leukaemic cells end up residing there. An enlarged liver or spleen from accumulation of leukaemic cells will result in a feeling of fullness in the stomach. This leads to a loss in appetite and unexplained weight loss. A person will not display all the symptoms, but maybe two or three of the above. Each of the above can be a symptom for other sicknesses and diseases. It is therefore essential that the physicians carry out the relevant tests for diagnostic purposes.

4.11 Diagnosis of leukaemia

The doctor will physically examine the patient and record the medical history to determine what the symptoms are. Lumps and tender spots in the areas of the lymph nodes and the spleen will be indicative of possible abnormal growth. Further testing has to be carried out to confirm the diagnosis. Blood samples are sent to haematologist full blood count analysis. A bone marrow sample may also be taken for analysis to confirm the type of leukaemia, if it exists. Imaging techniques like x-rays, CT scans (computerised tomography), MRIs (magnetic resonance screening), etc will show more of the internal physical detail, e.g. an enlarged spleen, liver or lymph node is seen best through a MRI scan. Spinal taps are done if there is confirmation of leukaemia. The spinal fluid that is sampled is circulated in the brain and along the spinal cord. Ideally it should not contain blood cells. If it does contain leukaemic cells then the treatment would have to be adjusted accordingly to cover the brain as well as the blood and marrow.

4.12 Causes of leukaemia

The prevalence of cancer is rising rapidly because of demographic changes brought about by globalisation. The disease impacts on patients and family, on health institutions and society at large (Albreht *et al*, 2008). Cancer develops as a result of genetic mutations that result in certain restrictions in the normal growth of the cell. These genetic alterations include point mutations, and chromosomal deletions, amplifications and translocations. The multitude of variations and combinations of these alterations complicates and challenges researchers, pathologists and oncologists who constantly seek to conquer this disease. Identifying molecular mechanisms that describe the genetics of a particular cancer proves to be a mammoth

undertaking. The genetics of cancer is not sufficient to classy the various types of cancer. Clinical data like patient history, tumour histology and tests for tumour markers are currently used for classification into the various types and subtypes of cancer. Treatment of cancer uses classification into the various types thus making it quite challenging to find the appropriate treatment for individual patients. Ideally oncologists should develop a treatment regime for each patient according to the clinical data and the symptoms displayed. This requires the knowledge of specific gene alterations so that clinicians can diagnose and give an individual prognosis for each patient (Ciro *et al*, 2003).

4.13 Types of Leukemias

Leukaemia cells do not cause death. It is the expansion of the neoplastic clone that leads to bone marrow failure, albeit at different rates in the various subtypes. Ongoing developments in haematology have led to greater accuracy following the discovery of cytochemical stains, and also an increase in the number and range of monoclonal antibodies. Specialist techniques for karyotyping have also helped clinicians to enhance their diagnostic acumen (Jacobs, 1997).

Leukaemia is generally considered to be cancer of the white blood cells. There are two broad groups which are classified according to the time period taken for the disease and the level of maturity of the cells. There are another two main types which are classified on the origin of the white blood cells. Cell maturity is used to distinguish between acute and chronic forms (http://member.rivernet.com.au). When the malignant cells are immature the leukaemia is classified as acute. When the cells are mainly mature, then it is described as chronic. First are the two types of chronic leukaemias that come on slowly and progress slowly, and then the acute leukaemia that comes on rapidly and progresses rapidly. Acute means that it occurs suddenly, and chronic means gradually. Leukaemias are further defined according to cell lineage as myeloid or lymphoid. This will depend whether it comprises marrow or lymphoid cells. The term myeloid (from “myelo”, Greek for marrow, and “eidos” which means form) encompasses granulocytic, monocytic, megakaryocytic, and erthrocytic leukaemias. Based on the above nomenclature the following four types of leukaemias exist: “chronic lymphocytic leukaemia or lymphatic (CLL), chronic myeloid leukaemia (CML) or chronic granulocytic leukaemia (CGL), acute lymphocytic leukaemia or lymphocytic or lymphoblast (ALL), acute myeloid leukaemia or myelocytic or myeloblast (AML)” (Harmening, 2002). Leukaemia appears to be inherited, although some individuals have a predisposition for acquiring this disease. For example, a group of people may be exposed to a carcinogenic substance but only

those that are predisposed to it will eventually obtain the disease. There is also an increased incidence of leukaemia in family members of leukaemic patients.

4.14 Acute leukaemia

The majority of patients with acute leukaemia display clinically unexpected onset of signs and symptoms of only a few weeks duration. Patients often seek medical attention because of weakness, bleeding abnormalities, or flu-like symptoms. These abnormalities reflect the failure of the bone marrow to generate sufficient numbers of normal cells and are caused by the production and accumulation of leukaemic cells in the marrow. Leukaemic replacement eventually results in marrow failure and the resultant life-threatening complications of anaemia, thrombocytopenia and other complications. The ratio of adult cases to children is 10:1 and males have a slightly increased incidence compared to females. ALL is more common in children and AML is more common in adults. 75% of childhood leukaemias are classified in ALL whereas nearly 80% of AML cases occur in adults. Cytogenetics analysis of leukaemic cells is a critically important addition to the standard classification of acute leukaemia. It is currently considered to be an essential component in the assessment of the newly diagnosed leukaemia patient, playing a major role in diagnosis, sub-classification, and selection of suitable therapy and monitoring the effect of therapy. Chromosomal abnormalities have been linked to the distinct forms of leukaemia (Harmening, 2002). Acute leukaemia, whether myeloblastic or lymphoblastic have been treated but with varying outcomes. Remission rates of cancer patients, especially children have progressively increased over the years with some groups achieving a rate of 100%. With non-specific drug regimes patients‟ responses vary, with some not responding at all, while others go into remission and then relapse. Survival rates have grown over the past few years with some categories having a rate of less than 50 % beyond 5 years (Beutler *et al*, 2001). After diagnosis of acute leukaemia the differentiation between AML and ALL is critical.

4.14.1 Acute lymphoblastic leukaemia

ALL is the most prevalent malignancy of children. Children make up 25% of all known cancers. ALL has its highest incidence in children of 1 to 5 years, with a peak at 3 to 4years. This peak is not seen in blacks, and as a result, ALL is more common in whites. In adults, ALL accounts for 20% of acute leukaemias. All is found to be more common among males than females. ALL is a heterogeneous disease with biologically and clinically distinct subsets. ALL is divided into B-ALL and T-ALL lineages. In adults the t(9;22) gene is more prevalent in the B-precursor group. HLA-DR, CD19 and CD10 in the B-lineage phenotype show excellent

prognosis. CD7 surface expression without the presence of CD4 or CD8 is an indication of traditional chemotherapy and early demise (Lee *et al*, 1999). In all cases follow up with cytogenetics has to be done to ensure that the correct subtype is identified and the appropriate treatment regime adopted. Correct diagnosis of a subgroup can lead to a longer survival time for the patient, as in some rare cases the therapy is non conventional. There are three subtypes of ALL (L1, L2 and L3). It is common in adults, but is sometimes confused with AML. The cells vary in size, quantities of cytoplasm, and characteristics of nucleoli. The L3 subtype is rare but it has to be identified because the treatment regime is specific to this subtype. Chromosome abnormalities are usually found in 90% children and 70% of adult ALL patients. The major chromosomal abnormalities in ALL are t(9;22)(q34;q11), t(12;21)(p13;q22), t(14;11)(q21;q23), t(1;19)(q23;p13) and the translocation involving chromosomal arm 8q24. The t(9; 22)(q34;q1) indicates a poor prognosis for both adults and children, while t(12;21)(p13;q22), indicates a good prognosis. "The most common B-lineage ALL are the B-cell markers (CD19, CD22) CD34 and CD10" (Kebriaei, 2003).

4.14.2. Acute myeloid leukaemia or myelocytic or myeloblast

If the patient is not treated immediately on diagnosis, there is very little chance of survival beyond 9 weeks. The effects of the chemotherapy and other drugs usually lead to death in this group of patients. The majority of patients with AML ultimately die of the disease or complications of treatment. The diagnosis of AML is demonstrated by approximately 30 % blast cells. AML and ALL can be distinguished by morphology and cytochemical reactions. Karyotypic analysis is vital for prognosis. "The subtypes that must be recognized because of the need for specific treatment include acute promyelocytic leukaemia (APL) which is the M3 subtype of AML and the L3 subtype or mature B-cell ALL" (Cripe *et al*, 1997). Most patients in this subgroup have disease relapse. A patient"s age at diagnosis, leukaemia cell karyotype and the status of the leukaemia, *de novo* or secondary are major points to be considered when treatment plans are established. Favourable prognostic factors dictate conventional therapy. Unfavourable prognostic factors do not improve with traditional therapy, neither is there any relief with an intensified treatment plan (Cripe *et* al, 1997).

4.15 Chronic leukaemia

Chronic leukaemia is either granulocytic or lymphocytic by nature. Allografting has been responsible for lengthy periods of stable and disease free survival. If the T-cell count drops dramatically then the relapse rate is significant. In lymphocytic leukaemia, cure is rare.

Preliminary data has proven that sometimes lymphocytic leukaemia cannot be identified even with the latest molecular diagnostic methods.

4.15.1 Diagnosis of chronic myelogenous leukaemia

Chronic myelogenous leukaemia (CML) is related to the cells found in the blood stream. 1 out of 100000 people per year is affected or 15 % of all leukaemias in adults. The immune system influences the course of CML. There are three major types of CML stages: chronic, accelerated, and blast. The chronic stage takes up the most time during which the cell counts grow steadily. The median survival time is generally 3 to 4 years and fewer than 30% of patients survive 5 years. The accelerated and blast phases may be a few months each. During this period there is a rapid increase in cell counts, followed by death of the patient (Moore *et al*, 2004). CML is primarily considered an adult leukaemia which occurs in adults between 30 and 50 years, however the disease can affect any age group, including the elderly, infants and toddlers. Although rare, when infants and toddlers are diagnosed, the disease is called juvenile CML and displays distinct haematopoietic, cytogenetic, and clinical differences from the adult type. CML accounts for approximately 20 – 25 % of all leukaemia cases with men being predominantly affected. Detection in early stages is difficult because some patients do not display any symptoms. When symptoms do appear, the most common are anaemia, weakness, fatigue, dizziness, headache and fevers (Harmening, 2002). A lymphocytic count greater than $5x10^9$ L^{-1} of small mature lymphocyte in the peripheral blood for more than a month is an indication of CML. Approximately 90 to 95% of patients with typical characteristics of CML carry the Philadelphia chromosome (Ph) in their leukaemic cells, and as a result its presence is virtually diagnostic of the disease. The Ph chromosome results from the aberrant conjoining of the proto-oncogene c-ABL from chromosome 9 with the break-point cluster region (BCR) gene on chromosome 22. This new fusion gene BCR/ABL is considered essential in the pathogenesis of CML (Harmening, 2002). Kappa or lambda light chain restriction is co-used for diagnosis. Immunophenotype as detected by flow cytometry includes co-expression of CD19 and CD 5 together with CD23. Patients display low expression of surface immunoglobulin (sIg) and accordingly, absence or low expression of CD79b. In the case of a lymphocytic count less than $5x10^9$ L^{-1}, a bone marrow biopsy may be needed to confirm the diagnosis, but it is also necessary to confirm whether a patient is in remission, as part of complete remission requires the absence of leukaemic involvement of the bone marrow. The Philadelphia chromosome, or Ph chromosome is a translocation involving chromosomes 9 and 22 [t(9;22)(q34;q11.2)]. The t(9;22) occurs in a pluripotent stem cell that gives rise to both lymphoid and myeloid lineage cells. The standard t(9;22) is identified in about 92% of CML patients.

4.15.2 Diagnosis chronic lymphocytic leukaemia

Chronic lymphocytic leukaemia (CLL) or small lymphocytic leukaemia is a hematopoietic neoplasm of B-lymphocytes (CD5+) found in the peripheral blood, bone marrow, and secondary lymph organs (lymph node and spleen). The chronic lymphoid leukaemia is made up of various stages: an initial stage where tumour cells are mainly small, with a low growth rate and high survival rate of the cell; and a transformation stage, with the recurrent incidence of extramedullary proliferation and a large number of immature cells. CLL is the most prevalent leukaemia amongst adults. CLL occurs in older adults with 90% of them being in patient‟s over 50 years. For patients below 40 years of age CLL is rare. (Inamdar *et al*, 2007). Unlike acute leukaemias, the signs and symptoms of CLL appear quite slowly. The time period for the asymptomatic phase of CLL is mostly inconsistent. The clinical course is relatively slow, but as the disease progresses, chronic fatigue, frequent or continual infections, and easy bruising are the consequences of anaemia, neutropenia, B-cell immunologic dysfunction, and thrombocytopenia (Harmening, 2002). Results from flow cytometry analysis reveal co-expression of CD19 and CD5, CD23 and CD20. This type of cancer is affects people over 65 years. Some cases younger patients have been discovered. Since DNA is a blueprint of human growth whether normal or abnormal, it is the most reliable method for analysing a patient‟s health status. Abnormalities of chromosomes 11 and 17 are indicates poor prognosis. The deletion of 13q arm is an indication of good prognosis. The clinical course of CLL is incessant in its growth regardless of the course of therapy. The diagnosis of CLL requires a continual absolute lymphocytosis of mature lymphocytes for diagnosis. This is possible only if other causes are absent. The diagnosis of CLL is established when the peripheral blood lymphocyte count is 10 x 10^9 or more cell/L (which is typically the case), lymphocyte infiltration of the bone marrow is more than 30% lymphocytes of all nucleated cells, and the circulating lymphocytes have a B-CLL immunophenotype. The co-expression of CD5, CD19, CD20 and CD23 is also necessary for the diagnosis of CLL (Harmening, 2002). The introduction of advanced therapeutic drugs and the introduction of monoclonal antibodies against CD20 or CD52 have improved rates of survival. (Glassman *et al*, 2005 and Ghia, 2007). “The overall median survival for CLL is currently 4 to 5 years; 50% of patients are living 5 years after diagnosis, while 30% have a 10 year survival” (Harmening, 2002). The disease in 20 % of patients proceeds quite fast on diagnosis leading to death within 1 to 2 years. The variance seen cannot be easily explained, but clinical and pathological data have been used to diagnose and find the patient‟s prognosis. It is also used to classify a range of stages and risk groups (Harmening, 2002). Various clinical phases and systems can be used to determine periods of survival. Data from genetics and cytogenetics are also used as prognostic elements. “These include immunoglobulin gene arrangements, the presence of trisomy 12, abnormalities of

13q14, deletions of 11q22, and abnormalities of 17q13. Cytogenetic study of CLL contributes important prognostic information" (Glassman et al, 2005). "Diploid karyotypes are said to be associated with the median survival of 15 years and the more complex karyotypes have a median survival of approximately 6 years. Trisomy 12 may be linked to an atypical morphology which is indicative of a more rapid progression and shorter survival time" (Neilson *et al*, 1997). The overall abnormality rate was 64 % in 100 patients with CLL. Detection of chromosome abnormalities from clinical data is vital in the diagnosis and prognosis for CLL patients and the clinicians who treat them. A high percentage of cells with chromosomal abnormalities, indicating high proliferative leukaemic cells, were associated with poor survival. All CLL patients must be considered for FISH studies, which should be done in conjunction with standard cytogenetic tests (Glassman *et al*, 2005). "Studies using FISH probes in CLL found the chromosomal abnormalities were more common than those detected using conventional cytogenetics and had a more different distribution" (Inamadar *et al*, 2007). This summative form of analysis will serve as a comprehensive diagnosis and prognosis indicator for clinicians. Some patients diagnosed with CLL may not be treated unless the signs and symptoms of the progressive disease appear. Therapeutic intervention takes place when major physical and clinical signs and symptoms identify advancing disease.

4.16 Prognosis and survival in acute leukaemia

The care of a patient with leukaemia displays many challenges. On diagnosis the relevant subtype should be identified, thereafter the treatment plan must be drawn up. Chemotherapy means chemical treatment and it is the mainstay of leukaemia treatment. Chemicals used are „cyto-toxic", meaning detrimental to the cells. All dividing cells are destroyed at a high rate. The particular chemicals used would be dependant on the type of leukaemia. If a patient does not react to a single dose of chemotherapy then it can be concluded that the prognosis will be poor. With therapy and the best available drugs most patients do not reach remission, although there are some exceptions (Cripe, 1997).

Prognostic factors help clinicians on estimating the benefit of a particular treatment. Estimating the likelihood of an outcome is dependent on the identification of prognostic factors. Its usefulness is in optimising treatment strategy but there are limitations when formulating decisions for patient treatment. In general prognostic factors identify a broad group of patients and may not apply to specific patients. Measurement that is non-standardised cannot be incorporated into a clinical trial. Prognostic factors normally relate to a single outcome like survival and the interactions with other factors are not usually studied thus making them unpredictable. The lack of effective alternate treatments makes it impractical in deciding which

patients will qualify for standard treatments. If the prognosis is good then those patients can obtain conventional therapy. Patients have to be counselled into accepting less conventional treatment which has the potential of increased morbidity and no definite evidence that there would be an improvement in the outcome (Cripe, 1997).

Overall survival offers a comprehensive estimate of the risk and benefits of therapy. The relevance of the end points is unclear from the patient‟s point of view. Overall survival may be important, since the onset of salvage therapy which has lead to successes in treatment of relapsed disease. If a treatment improves outcome, the patient needs to know if it is substantial enough to take on the potential increased risk. Patients need to participate in the decision making process since effects the combined treatments on individual patients is uncertain. Precise and truthful assessments of the benefits and risks need to be outlined to patients as part of the counselling process. The patient must be clear about all the risks involved in the process (Cripe, 1997).

The outcome of a study carried out by Perea *et al* (2005) revealed CD36 expression in AML patients to have a lower leukaemia free survival (LFS) rate and a higher relapse rate, irrespective of karyotype. A two year LFS rate was 34% for CD36+ patients and 55% for CD36- patients. CD36 and CD2 positivity and adverse karyotype correlated with a lower overall survival (OS) rate (Chang *et al*, 2004a). Raspadori *et al* (1997) found that “CD56 expressed most frequently detected in M2 and M5 AML patients is associated with lower CR rate and shorter overall survival”. Chang *et al* (2004) “correlated CD56 expression with a shorter overall survival in univariate analysis (P = 0.0262). However, neither the CR rate nor the duration of disease free survival was influenced by the expression of CD56”. Chang *et al* (2007b) investigated CD7 expression in patients with AML and normal karyotype. The AML chromosomal abnormalities that are presented on diagnosis are vital. Patients with a normal karyotype found with conventional testing make up the biggest subset of adult AML (approximately 50%). Since the clinical outcome of patients with a normal karyotype varies significantly, the outcome for overall survival is about 35-45% over a 5-year period. Supplementary CD markers are therefore essential to identify and categorise clinically the subgroups of AML patients with normal karyotype (Marucci *et al*, (2005) and Bienz *et al*, (2005)). Flow cytometric immunophenotyping is now used as a conventional method for the diagnosis and characterisation of AML. “Surface markers have been identified with prognostic significance including CD34, HLA-DR, CD7 and CD56 in AML patients in the context of cytogenetic abnormalities. The adverse impact of CD7 in AML has been controversial and in particular, little is known about the prognostic significance of CD7 expression in normal

karyotype AML" (Bene *et al*, 2005). Chang *et al* (2007b) assessed 185 adult AML patients with normal karyotype. CD7 expression was found to be a significant marker to for disease free and post-remission survival. Some of the parameters used in the study were age, gender, white blood count (WBC), FAB morphology and immunophenotype. Standard statistical methods were adopted. A logistic regression analysis was performed on the clinical o investigate he effect of remission rate. The Cox proportional model was adopted for disease free survival (DFS), overall survival (OS) and post remission survival (PRS). "Patients with WBC > 50 had a significantly lower CR (65% versus 87%) and shorter OS (median 10 months versus 34 months). CD34, HLA-DR, CD56 and CD7 were expressed in 57%, 83%, 8% and 37% of the cases, respectively. On univariate analysis, expression of CD34 and HLA-DR was associated with a lower CR (85% versus 60%, P = 0.0007 and 67% versus 89%, P = 0.019). CD7 expression on AML blasts ranged from 20% to 96%. Patients with CD7 expression were found to have a significantly poorer DFS and PRS than patients without CD7 expression (12 months versus 42 months for DFS, P = 0.005; 15 months versus 33 months for PRS, P = 0.013, respectively). CD+ patients had a median of OS of 12.7 months versus 23.7 months for CD7- patients, but the difference did not reach statistical significance (P = 0.18). CD7+ and CD7- patients had similar CR rates of 72% and 71%, respectively. On multivariate analysis that included significant variables identified in the univariate analysis (age, WBC, CD34, HLA-DR and CD7), CD7 expression was a significant predictor for a poor DFS and PRS, whereas high WBC was an independent risk factor for CR, OS, DFS and PRS. CD56 was not associated with a clinical outcome" (Chang *et al*, 2007b). With conventional chemotherapies malignant cells still proliferate, and few that are left behind are responsible for persistence or recurrence of the disease. CD56 in AML is a well known indicator of poor survival rates. A possible biological explanation is that it causes the stem cells to adhere to the niches in the bone. This still allows the cells to proliferate in this environment. Certain leukaemias can proliferate cancer cells to display the characteristics of stem cells, thus enabling them to withstand the toxic drugs. Their hibernation spaces in the bones provide a solid barrier to the chemicals (Nimer, 2008). Nimer (2008) varied the intensity of chemotherapy treatment and found that their therapies to destroy the cancer cells. For poor prognosis the only alternative treatment is allogeneic stem cell transplantation. 25% of AML patients with normal karyotype do not display signs of genetic changes; therefore the alternative is to understand the cell at a molecular level for each subtype. An improved understanding may lead to novel ways of target therapy. Individualising treatment thus becomes vital, but can only be achieved if large clinical trials are carried out on real data and predictions or outcomes are made with a relevant multivariate analysis of as many parameters as possible. Joint involvement of clinicians and statisticians can only make these clinician trials more reliable and easy to implement into normal treatment regimes. Untreated leukaemia reduces survival to less than 3 months. The outcome of several large clinical trials

done on AML patients reveals a median duration of CR of 9 to 15 months and an overall survival of 12% to 16% at 7 to 8 years. Only a few AML patients achieve DFS, thus making acute leukaemia unsatisfactory. Advani *et al* (2008) reported a 65% complete remission rate, with only 15-30% remaining disease free for five years. The patient characteristics used in the study were gender, history of antecedent blood disorder, cytogenetics, white blood count and c-kit MFI. c-kit is a tyrosine kinase receptor found by the mean fluorescent index (MFI)). They considered 22.6 months for surviving patients and found a median progression-free survival for all patients to be 10.7 months, i.e. 7.9 months for age less than 60 years and 12.6 months for age > 60 years. The median overall survival was 13.8 months, i.e. 8 months for age > 60 years and 18.5 months for age less than 60 years. Wahlin *et al* (1991) investigated newly diagnosed AML patients from a registry. He achieved a CR rate of 47% and a median survival time of 4 months when the overall numbers registered was used. For the subgroup without previous blood disorders and age below 60, the CR rate was 81%. A median survival time of 13.4 months was calculated for the subgroup

Remission is now a reality because of newly developed drugs for target therapy and technological advances in blood and karyotype testing. A new form of combination therapy is also followed on an individual patient basis. This has resulted in a significant impact, mainly for childhood acute lymphoblastic leukaemia (ALL). 70-80% of children treated with poly-agent chemotherapy have increased with a high success rate. Alternately, progress in adult acute myelogenous leukaemia AML is not so remarkable (Litzow, 2000). Yanada *et al* (2006) found that 70-90% of adult ALL patients attain complete remission with no effect on long term survivors. Despite all these improvements patients still relapse after remission. There is no simple answer to this phenomenon. Reasons are difficult to explain, but the complexity of the human body is one reason. Cells also build up resistance to mechanisms over time and are able to guard against the effects of chemotherapy drugs that leukaemic cells acquire during mutagenesis (Litzow, 2000). Cripe (1997) has indicated “that patients with the translocations involving chromosome 8 and 21, t(8,21) and chromosomes 15 and 17, t(15;17) and INV(16) should receive conventional therapy. Translocations involving chromosome arm 11q23 occurs in patients with ALL” and AML indicate poor prognosis. In AML the translocations most commonly involved are t(6;11), t(9;11) and t(11;19) with a likelihood of less than 5% for survival. CR or survival of patients beyond 2 months with standard treatments is not normally possible. Patients with these abnormalities are unlikely to achieve CR or survive more than 2 months with conventional therapy. Even though white blood numbers at diagnosis indicates poor prognosis, the second significant prognostic factor for both AML and ALL, is the time taken to remove blast cells. When this is achieved the patient is said to be in complete remission (Litzow, 2004b). It has been found that there is stricter compliance with paediatric

patients‟ protocols than with adult patients. Sometimes the treatments are delayed, and a reduction in dosage was also noted. A lower dosage leads to lower toxicity and lesser cells being targeted. This can be taken as an explanation for the higher relapse rate in adults when using the same protocols as children. Adolescent patients found greater success with childhood protocols than adult protocols. In order to ensure higher survival rates both patients and clinicians need to ensure that the protocols are strictly followed (Plasschaert *et al*, 2004). Cytogenetic change as the patient proceeds with a course of treatment. Poor prognosis based on cytogenetic findings at diagnosis, will be a good predictor for unfavourable changes as the disease progresses (Litzow, 2004b). Generally it has been found that there are no significant types of recurrent chromosome aberrations differences between men and women. There is however evidence that it is related to the patient‟s sex. These cases are too few to make a general statement. These findings have to be validated on a large scale, multi-centred system for it to be considered valid (Mrozek *et al*, 2001a). The prognostic importance of chromosomal classification is becoming necessary for the management of patients who display varying cytogenetic characteristics. For example, “patients having a hyperdiploidy of 50-65, the presence of Ph and abnormalities involving the chromosome band 11q23 are no longer given standard treatment” (Harrison *et al*, 1998). In the new Medical Research Council (MRC) childhood ALL treatment trial (ALL 97), patients are treated on certain protocols as per their cytogenetic results. Similarly in the MRC AML treatment trial (AML12) patients are being reassigned based on cytogenetic data with the appropriate treatment regime. Patients are classified under risk groups whether their prognosis is good or bad and accordingly managed (Harrison *et al*, 1998). Prognostic factors are treatment dependent; therefore a cytogenetic or molecular abnormality presenting an unfavourable prognosis with one treatment may be more favourable when an alternate treatment is implemented. “There is a continuous need for large prospective studies that will correlate karyotype with the appropriate genetic markers, gene expression profiles, immunophenotype, other biological parameters and clinical outcome in patients treated with both existing therapies and those receiving novel therapeutic agents” (Mrozek, 2004b).

Acute non-lymphocytic leukaemia (ANLL) is a rare disease among children and CML is rarer still. Of all new leukaemias that are diagnosed in patients less than 15 years of age, ANLL and CML accounts for 20-25%. The outcome for females was better than males. Monoclonal antibodies have now become available allowing, for immunological differentiation of haematopoietic cells and distinctive definitions for ANLL subtypes. Even though survival has significantly changed over the years; the prognosis remains bleak. Chemotherapy treatment regimes are responsible for only 30-40% of remission in children with ANLL. This conventional treatment has proven to be resistant in CML patients. It only works is

accompanied by allogeneic bone marrow transplantation (Gatta *et al*, 2001). Survival is also affected by the monumental task of trying to find a suitable bone marrow donor.

A study of the prognostic factors of any disease or ailment guides the clinician in predicting survival rate, recurrence of disease and outcome of treatment. Xin *et al* (2006) carried out a survival study and analysis of prognostic factors on acute promyelocytic leukaemia patients at a single centre. The study was aimed at long-term survival of patients with APL (AML-M3) and the rationale for using all-trans retinoic acid (ATRA), chemotherapy and arsenic trioxide (As_2O_3) in the treatment of newly diagnosed APL patients. The introduction of ATRA for induction therapy of APL, resulted in the complete remission rate (CR) of this disease has been demonstrated by clinical trials to be greater than 90%. There have been some complications known as ATRA syndrome in approximately 5-20% of the people studied. In 20 years the prognosis for APL has improved from highly fatal to highly curable. In some studies it has been proven that 75-85% of patients go into remission beyond 5 years. A mortality rate of 10% exists even though a combination treatment regime was followed. Some reports have shown that long-term usage of As_2O_3 could make relapsed patients go back into remission. Clinicians still face the dilemma of treatment strategy. Likely prognostic factors related to long-term survival was also investigated in the above study. Patients were grouped according to their treatment regime. A log-rank and a Cox-regression analysis were carried to identify the prognostic factors. The above study showed that CR rate was 82.88%, with an estimated 5-year of relapse free survival (RFS) of 80.9% and an overall survival (OS) of 71%. Mrozek *et al* (2001a) reported that "patients with inv(16) or t(16,16) and t(9;22)(q34;q11) frequently have high leukocyte counts, whilst those with t(15;7) have low leukocyte counts at diagnosis. Low platelet counts are common in patients with t(15;17). Platelets counts higher than in other patients with AML are observed at presentation or later in the course of their disease in patients with t(1;3)(p36;q21), inv(3)(q21q26) or t(3;3)(q21;q26)". Late achievement of remission (defined as 3-4 weeks from the start of treatment to remission or more than one course to achieve remission), an elevated white blood count over 30 x 10^6 L^{-1}, and high age (above 35 years) are together with BCR-ABL and t(4;11) recognised as adverse prognostic factors on diagnosis. Late achievement of remission is recognised as a high risk factor. A normocellular bone marrow with less than 5% blast cells is commonly used as a criterion for complete remission (Hallbook, 2005 and Mrozek, 2004b).

It is reported that chemotherapy is still vital in the remission stage. Various studies have shown that patients who received chemotherapy only results in a poor survival rate, i.e. a 5-year RFS of 26% and a median RFS of 22 months (Xin *et al*, 2006).

Coebergh *et al* (2001) found higher survival rates in European countries with easy access to well established treatment facilities. These centres provided "aggressive" treatments according to a protocol. Since chemotherapy and radiotherapy have adverse short and long-term effects, it is necessary that treatment protocols and palliative care has to be optimised. There is still ongoing research with new protocols continuously being tested and implemented. This trend can be adapted to other countries since growing technology linkages that are now being sought, and the global trend towards working together to eradicate fatal diseases has become a continuous mission for all health workers. A 5–year cancer survival study was carried out in France among adolescents (Desandes *et al*, 2006). AML is generally resistant to chemotherapy, thus the poor prognosis. "Treatment outcome in young adult patients with newly diagnosed AML has substantially improved over the past decade. Complete remission (CR) rates now range from 60 to 80 % with long-term survival in about 50 % of cases. Advances can be attributed to several factors related to supportive care which has allowed the safer use of more intensive chemotherapy, but also to improved treatment strategies. However, therapeutic failure remains a major concern" (Desandes *et al*, 2006). Prognostic factors generally differ with treatment. It has been found that if the biological characteristics of the disease is considered in a model, then it expected that they would show similar correlations to prognosis regardless of the type of induction therapy administered (Tavernier *et al,* 2003). Approximately 65-85% of AML and 80-90% of APL patients have been achieved with combination therapy. Once patients obtain initial remission, the optimal consolidation treatment still remains to be determined. A significant proportion AML patients relapse after chemotherapy and after autologous transplantation. Complete remission was prolonged in patients who received three different schedules of cytarubine consolidation therapy. It is important that this particular subgroup of AML be differentiated either cytogenetically or at the molecular level. This is vital to prevent toxic procedures. "The identification of genetically homogenous subgroups, with a favourable prognosis, may help to properly evaluate the impact on high-dose chemotherapy, ersus marrow ablative cyto-toxic treatments" (Biondi *et al*, 1996). Overall the findings of Biondi *et al* (1996) highlighted the need for large prospective studies in a centre where patients follow set treatment protocols. Standardisation, quantification, quality control and the proper assessment of sensitivity is vital for reliability. This will also allow for the results to be compared to other similar studies.

4.17 Prognosis and survival in chronic leukaemia

The clinical course of individual chronic lymphocytic leukaemia (CLL) patients is highly variable with survival rates ranging from months to decades. CLL may start out as a low grade form of the disease but eventually it will progress to an aggressive and fully blown disease.

Early diagnosis will lead to a prognosis that will dictate the course of treatment, thus catering for individual risk-adapted therapy. The disease in more than 50 % of patients usually proceeds quite rapidly in comparison to initial estimates. In the early stages of CLL it is a period of "watchful waiting" (Bockstaele *et al*, 2008). Early treatment will depend on the reliability of early prognostic markers, thus making it a crucial factor on diagnosis. Accurate prognostication is also essential for evaluation of novel therapies and treatment options like antibody-chemotherapy or autollogous and allogeneic stem cell transplantation. The survival of patients with CLL ranges from less than 1 to 2 years to more than 15 years (Moreno *et al*, 2008). Clinical staging is unable to predict whether a subgroup with early stage CLL will eventually advance to an aggressive disease leading to early death. Clinicians cannot use consultations and blood analysis only for treating individual CLL patients. Standard patient data that can be linked to prognosis are gender, age and performance status. Women patients with CLL display a longer life span that heir male counterparts. The reason for this phenomenon is not exactly clear from various studies. The influence of age is contradictory. Absolute overall survival is greater in young patients, but their relative survival is usually less than that of older patients. (Bockstaele *et al*, 2008). CD5+ monoclonal B cells displaying the distinctive phenotype of CLL can be detected by flow cytometry in 3.5% "healthy" individuals with normal blood counts. It was found to be prevalent in patients over 70 years. This was found in more than 7% of the individuals, and was found to be common in first-degree relatives of patients with CLL (Ghia, 2007). Studies have confirmed that CD38 independently affects prognosis of CLL patients. The actual value is still a controversial point. Some of the values initial estimates are 30% to 20%, and sometimes 7% and even less. The stability of the marker is also not well understood as the disease progresses, especially after the first stage of therapy. Most patients die with the disease but not from the symptoms of it. Complete remission is still not possible with CLL; if it does occur then it will be a rare event. Disease progression worsens with an increasing lymphocyte count, enlargement of lymph nodes and spleen, development of anaemia and thrombocytopenia, and auto-immune manifestations. 90% of cases of CLL occur in persons older than 50 years. The median survival time for this subgroup of patients is 10 years but is dependent date of diagnosis. CLL was previously managed in general practice, but not with the advent of new treatments patients are being referred to oncologists (Ghia, (2007). Inconclusive answers to these and many others related questions on leukaemias will not standard operating protocols. All current treatments have been based on clinical trials. The majority of prognostic markers are not included conventional treatment protocols. Since there has been reported studies showing the significance of cytogenetics analysis on diagnosis, and application of individually targeted treatments on higher survival rates, several of these markers need to be considered in prospective clinical trials. Eventually it may contribute to improved clinical management of leukaemia patients (Aouali *et al*, 2005).

4.18 Diagnosis of haematological disorders

Examination of the blood is central to the diagnosis and management of haematological diseases. In a few other disciplines the physician can make a specific diagnosis and monitor therapy with easily accessible tissue samples and readily available methodologies, many of which can be performed in a physician‟s office. Assessment can be done on the dominance of red blood cells, of the several types of leukocytes, and platelets (usually from automated particle counters) and blood can be examined for qualitative changes in appearance of red blood cells, leukocytes, and platelets. In addition the presence of marrow precursors, malignant cells and intracellular parasites can be used to diagnose specific diseases, gain insight into patho-physiology, and measure the response to treatment (Beutler *et al*, 2001).

But the diagnosis of a blood disorder is dependent on both clinical and laboratory evidence. Neither is sufficient on its own to draw a conclusion on the diagnosis of the patient. Clinical evidence includes the history of the patient‟s present illness (illness can be work related) and past history, age, occupation family history (considers hereditary disorders), racial origin (considers genetic disorders) and the result of a physical examination. Clinical evidence can only be gained by trained physicians and together with the laboratory evidence a conclusion is made on the patient‟s diagnosis. Laboratory evidence is derived from both haematological and non-haematological tests, e.g. biochemical tests, radiological examinations, etc. The results of such non-haematological tests are, however, not frequently of critical significance in arriving at a diagnosis, as when, for instance, a patient is found to have a high white blood cell count. The blood is examined in order to answer two principal questions: is the marrow producing sufficient numbers of mature cells in the haematopoietic lineages? Is the development of each haematopoietic lineage qualitatively normal? Quantitative measures routinely available from automated cell counters are generally reliable and provide a rapid and cost-effective way to screen for major disturbances of haematopoiesis. Morphological observation of the blood film is also necessary to confirm certain quantitative results and to investigate qualitatively abnormal differentiation of the haematopoietic lineages. The physician uses this to make a more focused assessment of the marrow or systemic disorders, which are related to the hematopoietic system. Results of a full blood test, i.e. screening for an abnormality is a starting block in the path to diagnosing a patient. The results can be used to make a tentative diagnosis, thereafter further specific test methods or procedures can be carried out to confirm an accurate diagnosis. Further investigation is necessary not to doubt the diagnosis but to elucidate the cause and mechanism of the patient‟s blood disorder or to classify it more precisely. The choice

of tests and procedures will depend on the preliminary results but facilities or resources, time and finance can be a limiting factor (Lewis *et al,* 2001).

4.19 Quantitative measures of haematopoietic elements in the blood

4.19.1 Full Blood Count

Most automated blood cell counters measure the red cell count, MCV (mean corpuscular volume) and haemoglobin concentration directly. All other red cell parameters, including the haematocrit are derived from these primary values. A well-mixed sample of blood diluted in an electrolyte solution is passed through a small orifice through which electrical impedance is measured which is correlated to the red cell count. Each cell causes a jump in impedance as it passes through the opening since it cannot conduct electric signal through its lipid membrane. Red cells are distinguished from platelets by the magnitude of the impedance signal which is proportional to cell size. In electronic instruments the haematocrit (hmr) proportion of blood volume occupied by erythrocytes) is calculated from direct measurements of the erythrocyte count and the mean corpuscular volume: (hmr (µl/100µl) = [RBC in millions per µl x MCV in fl] ÷ 10). Erroneously elevated MCV and decreased red cell counts can be observed when red cell antibodies are present and retain binding capability at room temperature, particularly cold agglutinins and in some cases of autoimmune haemolytic anaemia. This results in the red blood cells clumping and by affecting the accuracy of both RBC count and MCV; it also affects the result of the derived haematocrit. Haemoglobin is intensely coloured and this feature is used to estimate its concentration in the blood. Erythrocytes contain a mixture of haemoglobin, oxy-haemoglobin, carboxy-haemoglobin, meth-haemoglobin, and minor amounts of other forms of haemoglobia. To determine haemoglobin concentration in the peripheral blood, red cells are lysed and haemoglobin variants are converted to the stable compound cyanmethaemoglobin for quantisation by absorption at 540 nm. In automated blood cell counters, haemoglobin is accurately and directly measured. The haemoglobin level varies with age. After the first week or two of extra uterine life, the haemoglobin falls from levels of 17 $g.dL^{-1}$ to levels of 12 $g.dL^{-1}$ by two months of age. Thereafter the levels remain relatively constant throughout the first year of life. Automated blood counters measure the MCV directly using the Coulter principle in which the cross-sectional area of a non-conducting particle (i.e. any cell) in an electrolyte solution is proportional to the increase in electrical impedance as the particle passes through a constricted orifice. The mean corpuscular haemoglobin or the amount of haemoglobin per red blood cell (MCH) is calculated by the formula MCH ($pg.cell^{-1}$) = [haemoglobin in $g.L^{-1}$ divided red cell count in millions.μL^{-1}] x 10. The mean corpuscular haemoglobin concentration

(MCHC) or the concentration of haemoglobin in the red blood cell volume is calculated by the formula MCHC ($g.dL^{-1}$) = [haemoglobin ($g.dL^{-1}$) / haematocrit in μL per 100 μL] x 100. Another index, the red cell distribution width (RDW) is specifically designed to reflect the variability of red cell size. It is based on the width of the red blood cell volume distribution curve, with larger values indicating larger variability. Leukocyte counts are performed by automated blood counters on blood samples appropriately diluted with a solution that lyses the erythrocytes (e.g., acid or a detergent) but deserves leukocyte integrity. The normal differential leukocyte count varies with age. In the first few days after birth polymorphonuclear neutrophils are predominant, but thereafter lymphocytes account for the majority of leukocytes. This persists up to about 4 to 5 years of age when the polymorphonuclear leukocyte again becomes the predominant cell and remains so throughout the rest of childhood and adult life. Platelets are usually counted electronically by enumerating particles in the unlysed sample within a specific volume window (e.g., 2-20 fL). As with any laboratory test, the clinical use of these all blood cell parameters depends on the prevalence of disease and the clinical setting (Beutler *et al*, 2001).

4.19.2 Differential

Following the white blood cell number, the white cells are analysed to find the percentages of each white blood cell type by doing a differential leukocyte count. Automated methods for obtaining a leukocyte (Lee *et al*, 1998) have developed that has reduced the time and cost of performing routine examinations. However, this technology is incapable of identifying and classifying all types of abnormal or immature cells. Flow through automated systems collect and analyse data from large numbers of white blood cells to provide a differential count that has a high degree of precision. The total white blood count, as well as the neutrophil, lymphocyte, monocyte and eosinophil counts are enumerated in the myeloperoxidase channel. Atypical lymphocytes, blasts and plasma cells fall into the large unstained cells channel. To enumerate basophils a basophil-nuclear lobularity channel is used. Results are presented as percentages for each cell type.

4.19.3 Platelet counts

Platelets are counted in an automated haematology analyser once the red blood cells have been removed by sedimentation. The mean platelet volume (MPV), which has been correlated with several diseased states, is determined. In general, MPV has an inverse relationship with platelet number (Lee *et al*, 1998).

4.19.4 Flow cytometry

Optimisation and development of new bio-processing strategies require detailed information. This is achieved by using a range of analytical methods and tools to obtain data. The data provided is fundamental in that a single representative value for the whole population is obtained for each parameter, i.e. each cell is treated as an "average" microorganism. Segregated data, signifying different subpopulations of a single sample can offer more information and understanding than the average value that is generally given with most analytical methods. All cells in a sample may not be in the same metabolic or physiological state; therefore if all possible subpopulations could be detected and described with regards to metabolic activity then the data can be used to effectively optimize the bioprocess. Flow cytometry can be adapted to include analysis cell populations where segregated data is obtained for the subpopulations. Flow cytometry is a technology that provides rapid measurement (-metry) of physical characteristics of cells (cyto-) suspended in a moving fluid stream (flow). Single cells or particles are passed through a laser beam in a fixed fluid stream. The variables absorption, scattering and fluorescence are recorded as each cell passes through the laser beam. This technology has evolved to include both cellular characteristics such as size, membrane potential and intracellular pH, and the levels of cellular components such as DNA protein, surface receptors, and calcium. This information is easily related to the different cell characteristics and components. The result yields the variations of all data within the given cell population. These values for the segmented groups are more accurate than using average values for the whole population. Florescent technology was initially used in the field of medicine for oncology (e.g. for diagnosis of cancer, chromosomal defect diagnosis) and haematology. Applications of flow cytometry in the clinical and medical fields have been widely reported in the literature. Lately it is being applied in the field of biology, pharmacology, toxicology, bacteriology, virology, environmental sciences and bioprocess monitoring. There are commercially available state of the art equipment together with data acquisition software for data capture and analysis. The new method also reduces the turnaround time from sampling to analyses. (Rieseberg, 2001). Flow cytometry is used to measure protein expression which is an indication of good or bad prognosis. It is also used to detect multi-drug resistance and measure cell proliferation. Classification and assessment of prognostic markers are used by clinicians to confirm diagnosis of the subtypes of leukaemia (Stetler-Stevenson, 2003).

The development of monoclonal antibody technology has aided the classification of cell surface antigens on haematopoietic cells. The availability of virtually unlimited quantities of mono-specific typing reagents permitted the identification and study of previously unrecognised

lymphoid and myeloid-specific surface proteins. The rapid advances in the production of monoclonal antibodies and the development of multiple commercial sources under a variety of trade names and designations has led to the development of a standardized nomenclature for human leukocyte differentiation antigens termed the "cluster differentiation" (CD) nomenclature. In 1989, at a series of international workshops on human leukocyte differentiation antigens sponsored by the World Health Organization, monoclonal antibodies having similar reactive patterns with tissue, cells or molecules were assigned to a "cluster" and given a "cluster differentiation" (CD) number. CD numbers with a "w" indicate a provisional cluster that may or may not be promoted to full CD status at subsequent workshops. The current CD antigens for each cell lineage are quite extensive and are readily available in the literature. The use of monoclonal antibodies specific to cell surface markers (CDs) allows phenotypic characterization of cells in disease states. By using flow cytometry, cells labelled with monoclonal antibodies are sorted and enumerated to identify a specific population of cells. Cell markers have been identified on the surface of cells in the disease states such as acute leukaemia, autoimmune disease, and thromboembolytic disease. Cell markers have also been identified in the management of renal, cardiac and bone marrow transplantation. Diagnosis of disease states is dependent on clinical presentation, cytochemistry, and the study of the cells morphology, but flow-cytometry characterization of cells has added another dimension to disease classification. Monoclonal antibodies are used to characterize cells in acute leukaemias. The CD markers allow for differentiation of myeoblasts, lymphoblasts, monoblasts, megakaryoblasts, and erthroid ontogeny. Flow cytometric analysis of acute leukaemia is interpretive, combining the patterns of intensity of antigen expression to reach a definitive diagnosis (Rieseberg, (2001) and Harmening, (2002)).

.

4.19.5 Cytogenetics

DNA technology can be used to help diagnose and classify various types of cancer. This is possible since all cancers harbor genetic defects that are responsible for malignant transformation. Laboratory detection of cancer-associated genetic defects not only contributes to improved diagnosis of patients but also helps in some instances to formulate the most appropriate treatment, and help monitor the value of that treatment. Deoxyribonucleic acid (DNA) is the inherited component that encodes all the information required for the structure and function of cells. The transfer of information is *via* an intermediary substance called ribonucleic acid (RNA). DNA and RNA are collectively called nucleic acid. Nucleic acid analysis of patient samples is the basis for a new field of laboratory medicine called molecular

diagnostics. In no other discipline of laboratory medicine has this technology had a greater impact than in haematology, where it is used in the diagnosis of certain inherited, infectious and malignant forms of haematological diseases. The laboratory method to be implemented in this research is the polymerase chain reaction which is used to amplify particular segments of the DNA for ease of detection. To understand how this method is utilized, the structure of DNA and RNA has to be reviewed (Beutler *et al*, 2001).

Human DNA is composed of 46 chromosomes, each of which is a remarkably large molecule formed by two very long strands of nucleotides. All 46 chromosomes aligned end to end would comprise 3 billion nucleotide pairs long that would stretch over 2 m in length. The DNA strands are wrapped around histone proteins to form chromatin in the nucleus. All information necessary for life for an individual is encoded within the long strands of nucleotides. Nucleotides are the basic building blocks of DNA, comprising four different types of nitrogenous bases – adenine, guanine, thymine and cytosine – attached to a deoxyribose sugar and phosphate moiety. Hydrogen bonds bind the two strands of nucleotides that combine to form DNA. These bonds form between the nucleotides on one strand and the opposite strand. The rules of nucleotide pairing are – adenine in one strand can only bond with a thymine in the other strand, a guanine can bond only with cytosine. These rules ensure that the strands of DNA are complementary to each other (Beutler *et al*, 2001).

Encoded within the nucleotide sequences of DNA are functional units called genes that serves as templates for RNA transcription and protein transformation. All nucleated cells contain a full complement of DNA comprising that person‟s genome, but each cell expresses only a fraction of the approximately 100 000 different genes depending on the cell type and stage of differentiation. Current molecular diagnostic techniques are based on being able to identify a specific nucleotide sequence in DNA or RNA with the use of a “probe” that targets the specific sequence. A probe is a single-stranded segment of nucleic acid (either DNA or RNA) whose nucleotide sequence is complementary to the target sequence (either DNA or RNA). A probe binds to its target by hybridization which involves combining a probe and its target in a small tube (liquid phase) or a membrane or glass slide (solid phase). In both cases the probe is pre-labelled for detection and used as a marker for the target sequence. (Beutler *et al*, 2001).

Cytogenetic analysis provides pathologists and clinicians with a powerful tool for the diagnosis and classification of haemotologic malignant diseases. The detection of an acquired, somatic mutation confirms the diagnosis of a neoplastic disorder and excludes out a reactive hyperplasia or morphological changes due to toxic injury or vitamin deficiency. Specific cytogenetic abnormalities categorise homogenous subsets of various malignant diseases and

allows clinicians to predict their clinical course, and their probability of responding to particular treatments. In most cases, the prognostic information derived from cytogenetic analysis is independent of that provided by other clinical traits. Patients with favourable prognostic features benefit from standard treatment protocols with well known spectra of toxicities, whereas those with less favourable clinical or cytogenetic characteristics may be treated with more intensive or investigational therapies. The disappearance of a chromosomal abnormality present at diagnosis is used a significant indicator of complete remission of a patient following treatment. The reappearance of the aberration heralds relapse of the disease. Pre-treatment cytogenetic analysis can be useful in choosing among post-remission therapies that differ widely in cost, acute and chronic morbidity, and effectiveness.

The malignant cells in patients who have leukaemia or lymphoma would have acquired clonal chromosomal abnormalities. A number of specific cytogenetic abnormalities have been documented that are very closely, and sometimes exclusively, associated with morphological and clinically distinct subsets of leukaemia or lymphoma. The detection of one of these recurring abnormalities can be helpful in establishing the correct diagnosis, influencing selection therapy, and providing important prognostic information for both the clinician and the patient (Beutler *et al,* 2001). Even though all these indicators are considered vital for the clinicians and related health professionals, ultimately it is for the benefit of the patient‟s well being, to ensure the best, quickest and least painful course of therapy.

Cytogenetic analysis of malignant diseases is based upon the study of tumour cells. In leukaemia, the specimen is usually obtained by marrow aspiration and is either processed immediately (direct preparation) or cultured for 24-72 hours. When a marrow aspirate cannot be obtained, a marrow biopsy (bone core specimen) or a blood sample, for patients who have circulating immature or lymphoid cells, can often be studied successfully. Chromosome abnormalities are described according to the International System for Human Cytogenetic Nomenclature. The chromosomal complement is described firstly by the total chromosome number, followed by the sex chromosomes, then by numerical and structural abnormalities in ascending order. The observation of at least two cells with the same structural arrangement e.g. a translocation, deletion, inversion, or gain of the same chromosome, or of three cells confirming loss of the same chromosome, is considered confirmation for the presence of an abnormal clone. However, one cell with a normal karyotype is considered evidence for the presence of a normal cell line. If the cells show no alteration or nonclonal (single cell) abnormalities then it is considered to be normal. A single cell displaying a recurring structural abnormality is an exception to this rule. According to these results the specific karyotype is an indication of the malignant cells in that particular patient.

4.20 Summary

A mathematical model is dictated by the type of data and the conditions under which the data was measured and the system that it operates in. If a system is well understood then its behaviour can be predicted for a single case or for multiple cases by changing the conditions of the system. The purpose of a comprehensive quantitative model is to gauge the researcher's comprehension of the system under consideration. It makes one question the unknowns of a system, and asks relevant questions, and also to precisely describe the workings of the model. The model building exercise therefore becomes a valuable enterprise. The relationship between the theory and experimental system will lead to mathematical models becoming one of the essential tools for the study of prognosis and prediction of survival of diseased patients (Goldstein, 2001).

The collaboration of ANNs with clinical data has been proven to have significant advantages. Multidisciplinary collaboration can be effectively used to develop improved modelling strategies by allowing innovative techniques for the analysis of complex biomedical data. This supports the continuous effort for large prospective studies between genetic testing and disease prognosis. Hopefully cases with less common aberrations can also be identified to enable target specific treatment. The validity of the studies reported so far have be confirmed on a large scale, preferably as part of a clinical trial where real data can be analysed (Mrozek *et al*, 2001a). The prognostic importance of chromosomal is gaining significance. This is necessary to manage patients according to cytogenetic results. Clinicopathologic, cytogenetic and molecular features combinations have resulted in the division AML and ALL into further subgroups. Delineation of these subgroups and characterisation has lead to an improvement in prognosis on diagnosis. Clinicians are able to determine if chemical treatment and individualised therapy can prolong survival or lead to remission (Cripe, (1997) and Mrozek, (2004b)). "The unravelling of the molecular mechanisms underlying the pathogenesis of leukaemia has begun to reap benefits in the development of specific molecular therapies targeted at the molecular abnormalities of these disorders. While relapse in acute leukaemia remains a life threatening illness, the promise of molecularly targeted therapies in combination with immunotherapeutic approaches bring the hope that increased numbers of cures of these dreadful disorders can be realised" (Litzow, 2004). New agents for treatment have been researched and developed with great advancements in technical superiority being adopted at the sub-cellular level. Larry Chin from the University of Maryland Medicine, Department of Neurosurgery states as follows: "it is a privilege to be a physician at a time when molecular pathogenesis of disease is being unravelled. From the discovery of the structure of DNA to the human genome project, molecular biology over the past 40 years has revolutionized medicine"

(Jacobs, 1997). This quest is a continuous undertaking with limitless possibilities. It will take time for clinicians and joint researchers to relieve or cure the numerous cases seen everyday in clinics, hospitals and theatres throughout the world.

CHAPTER 5
NEURAL NETWORK MODELLING METHODOLOGY

5.1 Introduction

The methodology involves the use of four steps: collection of raw data, processing of raw data, the development of neural network models and the comparison of models. The aim of this study is to determine the survival rate of leukaemia patients and the prognostic factors that affects this outcome. A retrospective study was carried out on leukaemia patients from the Haematology Department at Inkosi Albert Luthuli Hospital (Ialch), Durban. Permission was requested from the Haematology department and was subsequently granted for this research. Ethical clearance has been obtained from the Bioethics Committee, University of Kwa-Zulu Natal for the use of the patient data. Patient confidentiality was maintained by using coded values. Data was obtained for the period from 2002 to 2008, but some patients had records from previous years in their files. A major part of this research was based on obtaining the patient data, analysing the results and determining of the relevant variables necessary for this study. The data collection was also a monumental task as the medical recording system is not well established. The processing of the data had to be in the proper format to be used as an input for the neural network modelling. The confidentiality of the patients has been retained by the use of coded numbers to represent patient names. The raw data was processed in Microsoft Excel into an appropriate format for input into the relevant software. Data was processed using the neural network software, PREDICT version 3.12 (Neuralware, 2003) to develop the proposed neural network models.

5.2 Data collection

Leukaemia patients attend a day clinic at the Inkosi Albert Luthuli Hospital. Patients are referred to from various hospitals throughout Kwa-Zulu Natal. When it is necessary, patients are admitted to the oncology ward for observation and in-patient treatment. All personal and clinical data is manually recorded in patient files either as hardcopies, electronically or a combination of both. Analyses of blood and marrow samples are carried out in the Department of Haematology laboratories at Ialch. Flow cytometry analyses were done at Ialch from 2003 to 2007. From 2007 to 2008 the flow cytometry analyses have been carried out by Johannesburg General Hospital, Johannesburg. Some results were from private hospitals and previous years and were recorded in the patient's files. Data was analysed and used as per diagnosis and according to the relevant dates. Chromosome analyses are done at Johannesburg General Hospital. The

following patient information was recorded: age, gender, type of leukaemia, full blood count, differential, flow cytometry, chromosome analysis, dates of births, dates of laboratory tests, treatment dates and health status. The health status for each patient was recorded in terms of whether they were in remission, on treatment, dead, or relapsed with treatment. All possible methods that recorded information were accessed to order to get all information of patients. There were hardcopy files available in the clinic, electronic access *via* the MEDICOM computer system to laboratory data, files available in the flow cytometry laboratory and Excel spreadsheets available with a summary of cytogenetic data and bone marrow analysis. If a value was missing then it has not been recorded or an analysis was not done for that parameter. For each patient their medical records were analysed to obtain their information. Every possible method was used to ensure that all available data for all the leukaemia patients at Ialch was accessed and recorded. During the collection of data the medical records and notes of the patients had to be read and recorded in terms of the available information. There was no standard procedure of recording this status in a file or electronically. In most cases there was a question mark surrounding this status. The clinician would look at the patients laboratory results, make a suggestion that the patient has relapsed, and put a question mark next to it. This status had to be confirmed with cytogenetics, which in many cases came back as unsuccessful samples, or else the clinicians did not qualify this diagnosis of relapse either manually or electronically in the patient's file. In order to maintain the integrity of this research these patients were incorporated into the treatment group. Since they were patients in the clinic they could be justified as part of the treatment group. An independent checker was used to verify the correctness of the data used in this research. The total number of patients that were accessible was 680. There may be some patients that are available electronically only but were not recorded for this research. There is no database available in the laboratory or the clinic that could have been used as a starting point for this research. The starting point was the files in the clinic and the files available in the flow cytometry laboratory. Of the 680 patients reviewed 70 had missing data and were found to be ineligible for this study. The input variables which were recorded from the patient's medical records were not explicitly recorded; hence it was safe to assume the missing values to be either negative or numerically having a value of zero. In clinical research missing data is accepted and patients are considered to be censored. Patients with missing data ideally should be excluded from the study. This method was not effective for this research data set as this would have drastically reduced the sample size. Since the use of censored data is acceptable in the medical domain, the patients were included. The variables age, race, gender, sex and type of leukaemia did not have any missing values. These variables had to be known in order for the patients to be considered for this study. These are the personal details that are recorded in the files and appear on each result sheet that is generated in the haematology laboratory. There are 38 input variables so in order to maintain the integrity of the

model, if more than 3 input variables were missing then that patient record was excluded from this study. This criterion was applied to all patients that were considered legible for this study. The average or mean was calculated for each parameter using the data from the 610 patients. The average value was used to replace the missing values for the appropriate parameters for each patient. Those patients whose current health status in October 2008 was known were included in the uncensored group. This group was used as a start-up group for the building of the neural networks since all available information was known. The remainder of the patients formed a second group of patients who were considered to be censored patients, i.e. their mortality is unknown. The patients in the censored group had a last test date, treatment or procedure recorded and thereafter no information existed in their medical records as to their status. These patients were included in some of the models as censored data. The censored patients had to be incorporated into the analysis to ensure sufficient numbers for this research.

5.3 Data processing

The following parameters were recorded for all patients included in this study: date of birth, age, gender, race, leukaemia type, date of diagnosis or tests, full blood count (haemoglobin, haematocrit, red cell count, mean corpuscular haemoglobin, mean corpuscular volume, mean corpuscular haemoglobin concentration, red cell width, platelet count, mean platelet volume and white cell count), differential (% of reticulocytes, neutrophils, lymphocytes, atypical lymphocytes, monocytes, basophils, eosinophils, band cells, metamyelocytes, myelocytes, promyelocytes and blasts) flow cytometry (CD3, CD4, CD5, CD7, CD 8, CD10, CD13, CD14, CD19, CD20, CD22, CD23, CD33, CD34, CD56, CD64, CD65, CD117, LC Kappa, LC Lambda, HLA-DR, mpo (myeloperoxidase), chromosome analysis, and their survival status (months alive). Explanations for the parameters used in this study have been explained in Chapter 3. An audit was done on the procedures followed when new patients visit the oncology clinic at Ialch. Patients are requested to give all their personal details and their medical history which is then recorded in files both manually and electronically. Blood and/or bone marrow samples are taken from patients and sent for testing as per the clinician's request. Tests include full blood counts, differential analysis, flow cytometry analysis and cytogenetic analysis. The various results were studied and this result base was the start of the patient list. In order for a variable to be considered for this study it had to be available for all the patients. The only cytogenetic result that was widely available for all patients was the chromosome analysis. The results for FISH, PCR and other abnormalities were not available for all patients because it was not requested for the sample was insufficient or results were inconclusive. After a review of all data for all patients in the clinic a list of possible variables were drawn up. A comprehensive list

was built where results for each patient was recorded. Thereafter an elimination procedure was adopted according to the number of patients that had all data for all the variables. In an ideal situation each patient would have had a value assigned for each variable where a result is sought, but this is not so for the initial list of patients. Initially there were 52 variables recorded for each patient. The raw data obtained from the medical files for all patients initially considered for this study is recorded in Table A-1 (Appendix A). Patient data was studied to determine which patients would qualify for this study. The following variables had to be eliminated from the study due to a multitude of missing values for the majority of the patients: reticulocytes, atypical lymphocytes, basophils, eosinophils, band cells, metamyelocytes, myelocytes, promyelocytes, CD64, CD65 and CD117. A final input of 38 variables was used in building the neural network. Some of the main variables used are given an expanded explanation below:

- Age: the patient's date of birth was used to determine their age. It was recorded in years at the time the laboratory analyses were carried out, and the diagnosis of the patient was confirmed. The patient's age varied from a few months to < 90 years. The response to treatment varies according to age and especially between younger and older patients. All patients < 20 are considered as children in a clinical environment. 20-30 is considered to be young adults and > 30 as normal adults. Patients are termed "older" if they are > 60 years old. Patients ages ranged from 0 to >90. The patients were divided into 5 age groups, with each group being assigned to a value between 1 and 5 as shown in Table 5-2.
- Gender: both male and female patient data was used, thus the two options of 0 for a female and 1 for a male as shown in Table 5-2.
- Race: there are four broad race groups that are classified according to the South African population registry: blacks, indians (or asians), whites and coloureds. This status is recorded in the patient's file. This system was used for classification into the race groups and denoted as codes 0-3 according to Table 5-2.
- Type: the oncology department at Albert Luthuli Hospital treats leukaemia patients from all hospitals in the Kwa-Zulu Natal region. Patients are referred to the hospital to have their diagnosis confirmed by specialised laboratory tests like flow cytometry and cytogenetics. The data was divided into four types of leukaemia: diagnosis of AML, ALL, CML and CLL. Even though there is a differentiation between the subsets of the leukaemias, for purposes of this study only the four variations will be used. The majority of the patient records only confirmed the main type. Since the specific type was not known for all patients, the above grouping was adopted. The final type chosen was as per the date of diagnosis and the laboratory tests confirming the type of leukaemia. The final coded system is given in Table 5-2.

- Full blood count: When a blood sample is obtained a standard analysis is carried out to determine the following variables: haemoglobin (13.5-15.5 $g.dL^{-1}$), haematocrit (37-52%), red cell count (4-5x10^{12} L^{-1}), mean corpuscular haemoglobin (27-32 pg), mean corpuscular volume (78-99 fL), mean corpuscular haemoglobin concentration (30-35 $g.dl^{-1}$), red cell width (11.5-14.5%), platelet count (150-450 x 10^{9} L^{-1}), mean platelet volume (7.4-10.4 fL) and white cell count (4-11 x 10^{9} L^{-1}). The expected ranges when the laboratory analyses are carried out are within brackets near each variable. For each variable the maximum and minimum values were recorded and the frequency of values in various ranges was studied. The final adopted ranges and their codes for each variable are tabulated in Table 5-3.
- Differential count: neutrophils, lymphocytes, monocytes, and blasts. These variables are represented as % therefore the range would be 0-100. For each variable, different groupings within the range were checked for the frequency of the patient results. The final ranges and their codes are shown in Table 5-5.
- Flow cytometry: These parameters were recorded from the flow cytometry files and laboratory database. CD markers were used as a diagnosis for the type of leukaemia together with the cytogenetic analysis. Values were recorded as % as determined from the graphs obtained from the flow cytometer. Cells were labelled with monoclonal antibodies and sorted and enumerated to identify a specific population of cells. Depending on their light scatter characteristics, cells were chosen for analysis according to their FSC. The final codes are given in Table 5-4.
- Cytogenetics: all data was analysed in Johannesburg General Hospital. Data was accessed from a cytogenetic database compiled by the haematology laboratory at Ialch. If the sample was insufficient or the analysis was not successful then an unknown result was returned. There were also results returned with a normal karyotype. Missing values were allocated 46, XX for females and 46, XY for males. The following notation was used to denote the cytogenetic data: normal male pattern, normal female pattern, addition, deletion, t(1,12), t(1,16), t(2,14), t(4,6), t(8,12), t(8,21), t8,22), t(9,17), t(9,22), t(10,11), t(11,17) and t(15,17). If cytogenetics were not done for patients then the normal karyotype for the male and female pattern was used. AML patients with a normal karyotype on standard cytogenetic examination constitute the largest subset of adult AML (approximately 50%) (Chang *et al*, 2007). AML patients comprise almost 1/3 of the patients in this study. Patients coded under the number 7 form a small number, therefore they were grouped in this way as shown in Table 5-6.
- Survival time: the diagnosis date is taken as the time of the flow cytometry analysis, blood count analyses and the diagnosis of type of leukaemia. All patients were not part of the initial start date of the study, which was 01 January 2002. Patients were entered

into the study as they were diagnosed. The diagnosis date is the most critical date recorded as it is linked to the dates of the analyses of the appropriate laboratory tests, the patient's age and it is the date of diagnosis of the type of leukaemia. The time taken for the health status of the patient, i.e. dead, alive with treatment, remission, relapse is also taken from the time of diagnosis to the event. All laboratory data was taken from the date of diagnosis. There is a two to three day difference between some tests as the requests were made when patients were seen by the clinician. The actual tests were done a day or two from the request date. The time recorded as survival time was used to determine whether the patient had survived over the period under study. For the dead patients approximately one month was subtracted from the number of months calculated according to the method described above. Since the study is based on survival, it can be considered that the patient was alive at least a month before their death. Data had to be used from 2002 to 2008 or else there would not have been sufficient patient numbers in order to carry out this study. Some patients had recorded tests dating prior to 2002 and had to be incorporated into this study to ensure sufficient numbers for the mathematical modelling. Survival time was taken as the current status date minus the diagnosis date. The time was recorded in months and related to their status, i.e. dead, alive and in remission or alive and on treatment. Patients who had a last data recorded with no further entry prior to October 2008 in the hardcopy file or computer were considered to be censored (assumed alive or considered as an event that has not occurred). For example, if a patient had a last test or visit recorded in June 2007, then the patient was considered to be censored, as it is not known whether the patient is dead or alive. Death can also occur from other causes like car accidents. Patients in remission with a last test date have been censored, since it is uncertain whether they are alive or not. Once in remission some patients did not go back to the hospital for further consultations. Patients also move away from an area or country and may seek medical help elsewhere if they relapse or need any medical help, either related to the leukaemia or not. This data was not coded but used as calculated in months.

5.4 Missing values

The average for each variable was calculated using all patient data, both censored and uncensored. If a value was missing for a particular patient the average value replaced the missing value. Value for age, sex, gender and type was all recorded according to the files, i.e. no missing values. Those that had missing information were eliminated form the study. The values used for the replacement of other variables with missing information are tabulated in Table 5-1.

Variable	Average Value	Variable	Average Value
CD 3	16	rcc	3.15
CD4	8	hmb	9.09
CD5	27	hmr	27.2
CD7	8.9	mcv	88.8
CD8	8.8	mch	29.1
CD10	13	mchc	32.9
CD13	29	rcw	17.3
CD14	4.7	pc	119
CD19	34	mpv	9.74
CD20	13	wcc	67.9
CD22	2.8	np	20.3
CD23	11	lc	34
CD33	19	mnc	5.93
CD34	21	blasts	46.7
CD56	4.3	hladr	46.4
kappa	21	mpo	7.87
lambda	13		

Table 5-1 Average values of variables

5.5 Coding data

Data was categorised into ranges and coded for input into the software for the building of the neural networks. The coded system is tabulated in Tables 5-2 to 5-6. An analysis was done on the range of values that were recorded for each parameter and an appropriate division and allocation of a coded value was used. Some ranges were easily done, e.g. male and female allocated 1 and 0 respectively while others like chromosome analysis had to be coded according to the translocation and the types of leukaemias where it is used for diagnosis. A neural network consists of an input layer of variables, a hidden layer or layers which denote the actual values for each parameter being studied and an output layer which predicts or classifies as per the model chosen. The hidden layers will become congested if the actual value for each parameter is used. If a small model is designed with a few parameters having one or two values in the range then it is possible to use actual data. As the number of parameters and the values in the individual ranges increases the hidden layer becomes more complicated. It also increases computational time, as the software will take longer to process the data. As the patient numbers

increase with a multitude of variables, so does the number of units in the hidden layer of the neural network. There is also a limitation on the software as it will not be able to process this large number of variables, with each having numerous values in their range. In order to make the system less complicated each parameter is coded into ranges of values for the inputs as this allows for shorter and less complicated network architecture. An analysis was done on the ranges by initially determining the minimum and maximum values for each variable. The frequency of values in the various subsets of the range of values was studied for each variable. Each variable was then divided into the appropriate subsets and coded values allocated to them. The format of all variables was chosen as numerical. All variables were based on laboratory results and numeric results returned. Since the range of values was quite extensive and since there was a large number of patients the numeric format or categorical data was adopted.

Variable	Ranges	Codes
age (years)	0-20	0
	20-40	1
	40-60	2
	60-80	3
	> 80	4
gender	female	0
	male	1
race	black	0
	indian	1
	coloured	2
	white	3
type of leukaemia	all	0
	aml	1
	cll	2
	cml	3

Table 5-2 Patient specific information coded system

Variable	Ranges	Codes
red cell count (rcc)	0-1.99	1
	2-2.99	2
	3-3.99	3
	4-4.99	4
	5-5.99	5
haemoglobin (hmb)	0-5	0
	5-10	1
	10-15	2
haematocrit (hmr)	1-10	0
	10-20	1
	20-30	2
	30-40	3
	40-50	4
mean cell haemoglobin (mch)	15-30	0
	30-45	1
mean cell volume (mcv)	< 70	0
	70-80	1
	80-90	2
	90-100	3
red cell width (rcw)	10-15	0
	15-20	1
	20-25	2
	25-30	3
platelet count (pc)	0-200	0
	200-400	1
	> 400	2
mean platelet volume (mpv)	5-7.5	0
	7.5-10	1
	10-12.5	2
white cell count (wcc)	0-50	0
	50-150	1
	150-300	2
	300-500	3
	> 500	4
mean cell haemoglobin concentration (mchc)	25-30	0
	30-35	1

Table 5-3 Full blood count coded system

Variable	Ranges	Codes
CD3	0-25	0
CD13	25-50	1
CD33	50-75	2
CD34	75-100	3
CD8	0-20	0
	20-40	1
	> 40	2
CD20 CD22 CD23 CD33 CD3 CD56 myeloperoxidase (mpo)	0 > 0	0 1
CD5 CD7 CD9 CD10 CD17 CD19 LC Lambda LC Kappa HLA-DR	0-50 50-100	0 1
CD4	0-10	0
	10-20	1
	20-30	2
	30-40	3
	> 40	4

Table 5-4 CD markers coded system

Variable	Ranges	Codes
monocytes (mc)	0-10	0
	10-20	1
	20-30	2
	> 30	3
neutrophils (np)	0-25	0
lymphocytes (lc)	25-50	1
blasts	50-75	2
	75-100	3

Table 5-5 Differential count coded system

Variable	Ranges	Codes
chromosome	46, XX	1
	46, XY	2
	addition deletion []	3
	t(9;22) t(9;17) t(9;11)	4
	t(8;21) t(8;12) t(8;22)	5
	t(5;17)	6
	t(1;1) t(1;16) t(1;19) t(2;14) t(4;6) t(12;21) t(10;11) t(11;17) t(4;17) t(17;19) t(11;17) t(4;11) other	7

Table 5-6 Chromosomes coded system

5.6 Development of neural network models

The neural networks were developed using the software PREDICT version 3.12. The software package allows for variation in the parameters that are used to build neural networks. The various options like learning rate or coefficients, learning rule, weight decay, number of hidden layers and type of transfer functions was varied to obtain the most efficient neural network for that particular data set. The objective of varying the various parameters is to minimise the error function in the neural network.

The following models are proposed based on an analysis of the patient data where the health status is known (uncensored) and that which is unknown (censored) for the parameters under study.

- Case study 1: patients having all data available based on the parameters chosen for this study over a 24 month period, i.e. it is known that they survived at least 24 months. Some had died and some have survived beyond the 24 months.
- Case study 2: patients having all data available based on the parameters chosen for this study over a 36 month period, i.e. it is known that they survived at least 36 months. Some had died and some have survived beyond the 36 months.
- Censored case study: censored and uncensored patients whose survival status is known for sure (uncensored) and those known for a specific time period (censored) were used as input for this model. Some patients were dead, some were still on treatment and the rest were censored. The large number of censored patient data had to be incorporated into his study to ensure sufficient patient numbers for this research.

5.6.1 Genetic algorithm

The software PREDICT was used to develop the neural networks. This software uses a feed forward neural network with an architecture of 1 input layer, 1 hidden layer and 1 output layer. It uses the gradient descent back propagation learning algorithm for training the network. The mathematical procedure for the back propagation learning algorithm used by the software PREDICT is illustrated in Figure 2-15 and the methodology explained in equations 2.35 to 2.42. The input variable selection employs a genetic algorithm. In the case of input variable selection, the individuals are a set of input variables. Different initialisations will yield different variable sets. The algorithm starts off with small sets of variables, and successful groups of variables are retained within the system. This is then used by the algorithm to select larger sets of variables if necessary. Smaller variable sets are usually preferred over larger variable sets. Set refers to the

index of the individual in a current population. Patience is a mechanism which is responsible for the convergence of the genetic algorithm. The patience factor changes by 1 if the tolerance is not attained. When the patience factor surpasses certain number (4 by default), the algorithm is terminated. The outcomes of the analyses gave the parameters that were most significant. Each model built can be thought of as an expert who uses a different set of criteria (the selected variables) to make its decision. The "Flashcode" component converts the completed multi-layer perceptron model into C, FORTRAN or Visual Basic code. (Neuralware, 2001).

A frequency value is given as an output for each variable used as an input to the model, whether accepted or not. The value indicates the frequency of occurrence of the field or the transform in the final population of the variable selection genetic algorithm. For example, a frequency of 1 for the variable "age" shows that it is chosen 100% of the time and is therefore considered very important to the model. The significant parameters or inputs that were used to build the initial model were then used further to build a second neural network. Since the variable was mathematically rejected from the model it implied that it was not significant for the particular data set used. Its elimination can therefore be justified, resulting in a new neural network with only the accepted variables. This method was expected to improve the efficiency of the models. For the neural network model eliminating the non significant parameters reduces the complexity of the model and the efficiency at the same time. There is a trade off between the number of significant variables to be used in the model (complexity) and computing power. This technique needs to be optimised to meet the latter requirements.

5.6.2 Network type and architecture

The feed forward back propagation or better known gradient descent network type (equations 2.35 to 2.42) was used for modelling the survival and prognosis for leukaemia patients. The difference between the network output and the target is treated as an error to be minimised. The back propagation algorithm is "offline" in the sense that training and normal operations occur at different times. In PREDICT the weight adjustment is performed by back propagation. The weight adjustments are done independently for every architecture chosen. Pattern selection is arbitrarily selected and distributed to the network. In PREDICT the weight updates default for the hidden layer and for the output layer is 0.005. If varied it is recommended that the weight of the output layer be at least of the order 10 less than the hidden layer.

5.6.3 Network parameters

The following training parameters were used to get the minimum possible error. Some options are a default setting in the software PREDICT and does not allow for any changes.

- Batch: The synaptic weights are updated for every pass through the allocated training data. This method is used since it minimizes the total error. There is no alternative in PREDICT.
- Learning rules: The learning rule gives an indication of how the connection weights are changed during the learning process. The gradient descent back propagation learning rule is adopted for this study.
- Momentum: The initial momentum parameter is specified for the gradient descent algorithm. The user is allowed to change as necessary. A high learning rate may cause instabilities; therefore the momentum term is added to prevent this. This value can be changed by the user in the gradient descent mode and is illustrated in equation (2.42).
- Learning Rate: The initial value of the learning rate for the gradient descent algorithm is set automatically. The learning rate affects the speed of the training. This value needs to be "watched" as it can cause instability if it is too large. The back propagation error modifies the weights of the nodes at a specific rate. This rate is referred to as the learning rate or learning coefficient defined in equation (2.42). In PREDICT the default learning rate is 100 but it can be changed by the user.
- Hidden layers: The hidden layer is made up of network neurons or nodes (units). A single hidden unit is a function of the weighted sum of the inputs. The function used in the algorithm is the activation function. The weights are determined by the estimation algorithm. If there is a second hidden layer, then each hidden unit in the second layer is a function of the weighted sum of the units in the first hidden layer. The activation function is the same in both hidden layers. The number of units in each hidden layer can be entered into the software program or it can be left to the program to be determined automatically by the estimation algorithm (Haykin, 1995).
- Activation function: It is a non-linear function that transfers the sum estimated for each neuron into a possible output value. The activation function also serves to link the weighted sums of units in each layer to the values of units in the succeeding layer. "The hyperbolic tangent function has the form: $\gamma(c)=\tanh(c)=(e^{c}-e^{-c})/(e^{c}+e^{-c})$. The data is transformed to the range (–1, 1). When an automatic architecture selection is used, this is the activation function for all units in the hidden layers. The sigmoid function has the form: $\gamma(c) = 1/(1+e^{-c})$. It takes real-valued arguments and transforms them to the range (0, 1)" (Haykin, 1995). The transfer functions used were sigmoid and

hyperbolic tangent (*tanh*) in PREDICT. The outputs are transformed and include the exponential, power, log and linear transfer functions. The general form of a continuous transformation is:

$$f = s_0(s_i x + o_i) + o_0 \quad (5.1)$$

Where f: is a continuous function

s_i, o_i: implement an inner scaling of the raw data to map it to an optimal sub-domain of f

s_0, o_0: implement an outer scaling so that y lies within a suitable range for the neural net (Haykin, 1994).

Each transform function in the data analysis table is identified by its continuous function f which can be any one of the following shown in Table 5-7.

Transformation	Definition
linear	identity function
log	natural logarithm function
exp	exponential
pwr2	square function
rt2	square root function
inv	inverse function
tanh	hyperbolic tangent function
logical	logical function (0/1)
rlogical	reverse logical function (0/1)

Table 5-7 Abbreviation for transfer functions

- Output layer: The output layer is made up of the variables for prediction.
- Partition: This function is used to split the data. The data id split according to the user's choice. The input data is split into training, testing and validation as per the limitations set out. The relative number (ratio) of cases randomly assigned to each sample (training, testing, and validation) can be specified by the user in PREDICT. The % box is used to input the split that the user chooses. PREDICT has a default allocation of data which 70% training and 30 % testing. It also allows for further testing of new data in addition to the above 30%. This verification for the new data set can only be used after the initial model has been trained and tested as per the 70%-30% split. The data is randomly selected into the training and testing sets. The user can change this ratio, but a balance has to be created where there is sufficient data to train and test. The training, testing and validation sets can be manually selected by the user. The output results are

based on a built-in three-fold cross validation which PREDICT has as a default setting. The partitioning of data into training, testing and validation was done with the default settings in PREDICT and manually. Manual selection was done as per the selected partition, but each value in the data set was manually tagged as a train, test or validation data point. An example is shown in Table D-1. The Excel interface is use to do the manual selection before the models are built.

- Seed value. For a particular set of data, the training and testing set allocations are fixed by this method to ensure repeatability of analysis, i.e.to reproduce the same randomised results in the subsequent analyses, the same initialisation value for the seed value is set as an input before each run of the neural network analysis procedure. A randomly selected 7 digit number of 9191972 was used for this research.
- Cross validation: The aim was to compare the different model types on the basis of their predictions of survival. Since the amount of data was not large enough to split into a training batch and a testing batch, a five-fold cross-validation technique was adopted in this study to ensure that all patients could be used in the training and testing (Ripley, 1998). This manual technique has been explained in Chapter 4. For each proposed data set and corresponding model a fifth was kept back in turn, the models fitted to the remaining four-fifths and predictions of the survival made for each patient in the fifth held back. This method would give a realistic picture when the true model is chosen. Each of the five groups was used as the test group at least once. This manual cross validation technique is different to the 3-fold cross validation method adopted by the software PREDICT. The software takes a set of variables and divides the data according the required partition, e.g. 70% training and 30% testing. This is done three times with the allocation of the partition and testing changing for each model built with this data set. The final result that is produced as an output is the model with the best statistical performance, e.g. R value and the accuracy.

5.7 Summary

A major part of this research was based on the collection of patient data, the determination of the relevant variables necessary for this study and analysis of the results. The data also had to be extensively processed into an appropriate format for use an input to the software used for the neural network modelling. Data was recorded from the patient files, average values used to replace missing values and data sorted to check which patients were eligible for this study. Patient information was transformed from the raw data available in the patients' files to the numerical format that is required by the software program. A coded system was used and all

data was of the type categorical. During the building of the networks various combinations were used to train, test and validate the models obtained. The patient raw data is presented in Appendix A to C. The results of the final models are presented in Appendices D to F. The Visual Basic Code for all the final proposed models is in Appendix G.

CHAPTER 6
RESULTS AND DISCUSSION

6.1 Introduction

Currently there are no studies that have been carried out on predicting survival of leukaemia patients by any mathematical methods, either generally or per type of leukaemia. The prediction method adopted in this study aims to provide a robust and accurate method for predicting survival of leukaemia patients for both censored and uncensored patient data. The aim of this research was to find out how effective neural networks can be in modelling leukaemia prognosis and to determine the factors that have the most influence. The raw data was collected and processed as explained in the methodology in Chapter 5. The data was grouped into uncensored and censored patient groups and then used to build the neural networks. These groups were further subdivided into the subtypes of leukaemia, i.e. ALL, AML, CML and CLL, and new models built. The final model for each data set is illustrated with the relevant confidence intervals. Tables are presented for the prognostic factors that have been used in the building of the neural network models and the level of importance is denoted as a frequency value. The following statistical analysis has been done on all patient groups used and is indicated in Table 6-1. The percentages were calculated on the number of patients that were eligible for each group. A 95% confidence interval has been used in all statistical analysis in this study.

6.2 Fitting neural network models

The models were fitted as described in the methodology in Chapter 5. The full data set was analysed and patients were grouped according to a survival period of two years, a three year survival period and a third group of all patients, i.e. censored and uncensored. In an initial analysis of all patient data there were 84 deaths in the 610 patient group (censored). There were 77 deaths in the first 24 months, 6 deaths in the period 2-5 years and 1 in the 6^{th} year. The general trend is that if a patient survives at least two years with treatment then the likelihood of remission and or a survival rate beyond 5 years is much greater. The availability of reliable data (uncensored) for the first three years has prompted the decision to do a two year survival, a three year survival and a full data set (censored) grouping for the building of the neural networks for prediction of survival. Each of the above groups was then divided according to the type of leukaemia and the neural network model

was determined for each type. The results of the fitting procedures are presented and the outputs compared for a particular group of patients.

Variables	Parameter	2-year Case study	3-year Case study	Censored Case Study
gender	% males	58	60	59
	% females	42	40	41
race	% blacks	67	63	67
	% indians	23	27	23
	% coloureds	2	1	1
	% whites	8	9	9
type	% ALL	43	47	42
	% AML	37	34	27
	% CLL	17	17	29
	% CML	2	2	2
age	% 0-20	41	44	36
	% 20-40	28	26	25
	% 40-60	20	20	21
	% 60-80	10	9	17
	% > 80	1	1	2

Table 6-1 Statistics for patient data

A default learning rule was adopted for moderately noisy data, moderate data transformation and comprehensive variable selection. The adaptive gradient learning rule uses back-propagated gradient information to guide an iterative line search algorithm. This general learning rule or algorithm is explained in Chapter 2 with a detailed explanation of the methodology and illustrated by the final equations 2.35 to 2.42. The statistical outputs and transformations of the model building process are produced in Excel format in Appendix D. The final programming codes in Visual Basic are presented for the final proposed models (Table 6-48) in Appendix G. For each set, „training" occurs on all records within the set and a composite test score is calculated across all validation sets. Thereafter an 80-20 and 90-10 partition was applied to all models. The favoured model was chosen to build the final model by excluding the outlying patients, i.e., outside the calculated confidence interval band. The justification for this removal is that there are always exceptions to a rule and especially with diseases where there is uncertainty in the recovery success rates of all types of patients. The neural network model was then applied to the final data set to

predict the survival for the particular group of patients. Each data set was also tested by retraining with different values of the learning rate and weight decay. Only the final models are presented. The accuracy (ACC) tolerance is set at 0.2 representing 20% of the range of the output, i.e. the percent of predicted output values lying within 20% of their corresponding target value outputs. The default weight decay is 0.005 and the default learning rate is 100.

The option "keep last network" was used since it allows larger networks to be favoured over smaller ones. This is favoured if accuracy is more important than generalisation, as is in this study. The result of the network format in PREDICT yields an architecture label which indicates (left to right), i.e. units in input layer-units in hidden layer(s)-units in output layer (e.g. 28-8-1). The number of units (or sometimes called nodes) in the input layer is equal to the number of transformations of input variables. Due to PREDICT'S data analysis, transformation, variable selection and algorithms, the number of input units in general will not correspond to the number of input data fields in a training data record. The architecture built is based on the training of the data specific to the group, thus the mathematical output is based on the variables of the group that have the most contribution to forming this "pattern" or mathematical model. Variables that are prognostically insignificant are mathematically rejected during the training phase. Accepted variables are rated according to the frequency, a value between 0 and 1 (1 corresponding to 100%), indicating the importance of the variable in the neural network model building process. Final models are built based on a unique set of accepted variables that are then transformed to produce the required output. PREDICT produces an architecture for a unique set of data, thus the difference in the architecture when the partitions are changed from 70% for training to 90% for training for a given group. The R value is a measure of the linear correlation between the real world target and the real world model output. Perfectly correlated data have an R value of 1. Anti-correlated outputs have a value of -1 and uncorrelated data have an R value of 0. The R value is dependent on the problem domain, e.g., in stock market data which is very noisy a value of 0.15 or 0.2 is considered good. The real test of the effectiveness of a model can only really be gauged by comparing it with other models on previously unseen data. The confidence interval (CI) corresponds to an error bar around the output. The confidence interval implies that 95% of the model predictions lie within the range around the target output values that bounded by the confidence intervals (dashed lines on all graphs). Accuracy (ACC) is the fraction of times the real world target is "close" to the real world prediction, where, for this test, "close" is defined to be 20% of the output range. CI, ACC and the R value were used to determine the best model for a particular data set. A Transform table gives an indication of the variables acceptance or rejection for the proposed model. A table for each model indicating the

accepted and rejected variables used in the building of the neural network model are produced. The table includes the equations used to transform each variable and the frequency of use in the building the model. There is one output for each model and that is "survival". "I" at the beginning of a value in the Transform tables (Appendix D to F) indicates that the variable has been used as an input to the model and "A" means it was rejected during the data analysis. The output "V" implies that the variable was rejected during the variable selection process. This means that the field could potentially have been used in the model (and may be used in a different model) but was not part of the set of input variables chosen for that particular model. A frequency (f) value is given as an output for each variable used as an input to the model, whether accepted or not. The value indicates the frequency of occurrence of the field or the transform in the final population of the variable selection genetic algorithm. For example, a frequency of 1 for the variable age shows that it is chosen 100% of the time and is therefore considered very important to the model. For rejected variables the range of f was from 0 to 0.89. For the accepted variables f ranged from 0 to 1. There is no indication from the Transform tables (Appendix D to F) that a limiting value or range of f will indicate whether a value is accepted or rejected for a specific model. The graphs presented represent the final models obtained for each of the data sets. Actual survival data is compared to the survival predicted by the neural network models. The dashed lines represent the 95% confidence interval range obtained from the statistical analysis produced by PREDICT, i.e. 95% of the predicted data lies within the confidence interval represented by the dashed lines bordering the 45^0 line on the graphs.

6.2.1 2- year case study

A data set of 235 was used to predict survival for a 2-year period as this information was known for all the patients in this group (uncensored). The neural network models were run by varying the learning rate, the weight decay and the learning rule. The partitioning for training, testing and validation was varied between 70% and 90% for the training data with the remainder used for testing. The full data set, i.e. training and testing was used for validation initially. The data was also manually (M) partitioned into a training (T), testing (S) and validation (V) set. A summary of the results based on the output of the confidence intervals in the various models is given in Table 6-2. The learning rate (LR) was varied between 80 and 150, the weight decay (WD) was varied between 0.005 and 0.001 and the learning rule chosen was the gradient descent learning rule. The confidence interval was used in selecting the best model. The CI value is an indication of the interval between the actual and predicted value that is within 95 % of the target value. The model chosen for this data

set has a 90-10 partition, an R value of 0.87 for the training and 0.65 for testing. A weight decay of 0.0005, gradient descent learning rule and a learning rate of 100 were the default values. Prediction accuracy of 0.89 and 0.58 was obtained for the training and testing set, respectively. The projected architecture was 24-18-1, i.e. 24 inputs, 18 hidden units, and 1 output layer where the sigmoid transfer function was used for the output layer. There are 18 variables shown in Table 6-3 that have been used to build the model with 24 transformations of these inputs. The remainder of the 38 variables were mathematically rejected by the model. In the initial analysis phase each data set or group of models were changed to see the effect of varying the allowable parameters, i.e. LR, WD and partitions. The best models were obtained with the 70-30, 80-20 and 90-10 partitions and in all cases PREDICT˜S default settings of LR=10 and WD=0.005 gave the narrowest confidence interval. The summary in Table 6-2 illustrates this effect. The two 90-10 (M1) and 90-10(M2) models are based on a manual allocation of the training, testing and validation data points. From row 6 to row 12 in Table 6-2, the WD and LR was varied to see if there was any change in the accuracy of the model. The model which gives the best prediction of survival for the 2-year case study has the network architecture of 24-18-1, with an R value of 0.87. For all subsequent data sets and groups only the best model in the 70-30, 80-20 and 90-10 partitions will be presented. The graphs in Figure 6-1 to 6-3 illustrate the predicted survival for the 70-30, 80-20 and 90-10 partitions.

Partition	Network	LR	R		WD	ACC		CI (months)	
			T	S		T	S	T	S
70-30	21-11-1	100	0.78	0.45	0.0005	0.78	0.52	18	27
80-20	16-2-1	100	0.55	0.53	0.0005	0.62	0.64	24	24
90-10	24-18-1	100	0.87	0.65	0.0005	0.89	0.58	14	15
90-10 (M1)	22-8-1	100	0.55	0.2	0.0005	0.61	0.62	23	22
90-10 (M2)	20-7-1	100	0.63	0.4	0.0005	0.69	0.25	21	33
90-10	24-2-1	120	0.72	0.46	0.0005	0.76	0.58	19	28
90-10	24-2-1	80	0.71	0.48	0.0005	0.72	0.58	20	28
90-10	24-7-1	150	0.74	0.44	0.0005	0.76	0.58	19	30
90-10	24-2-1	150	0.69	0.45	0.0001	0.72	0.5	20	28
90-10	24-9-1	120	0.88	0.52	0.0001	0.87	0.62	14	29
90-10	24-23-1	80	0.86	0.59	0.0001	0.85	0.57	14	27
90-10	24-8-1	100	0.85	0.51	0.0001	0.84	0.54	15	29

Table 6-2 Summary of models for 2-year survival

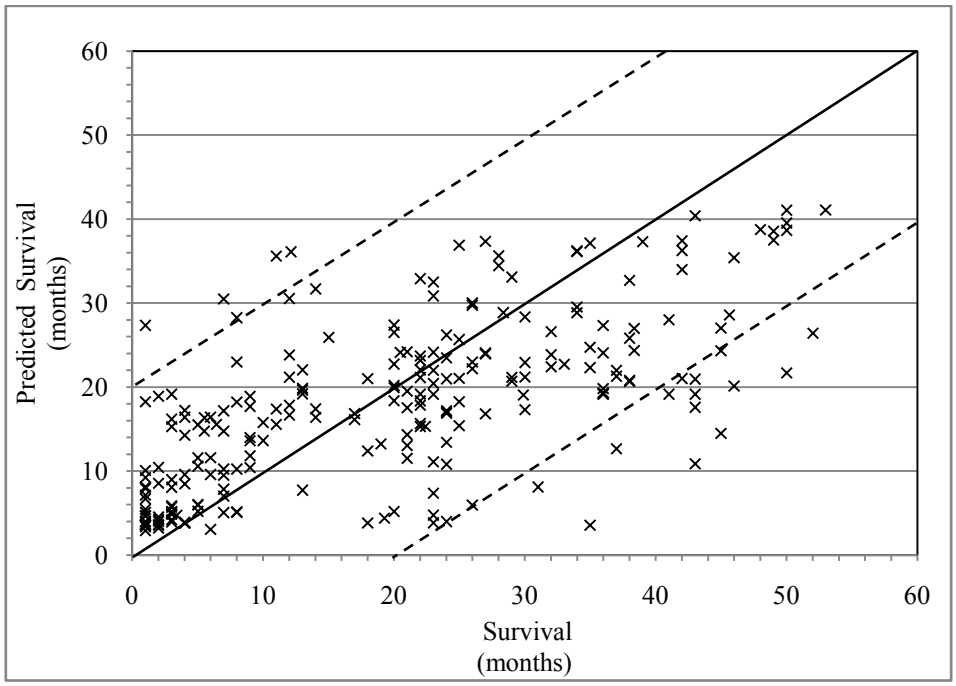

Figure 6-1 Predicted 2-year survival (21-11-1, 70-30)

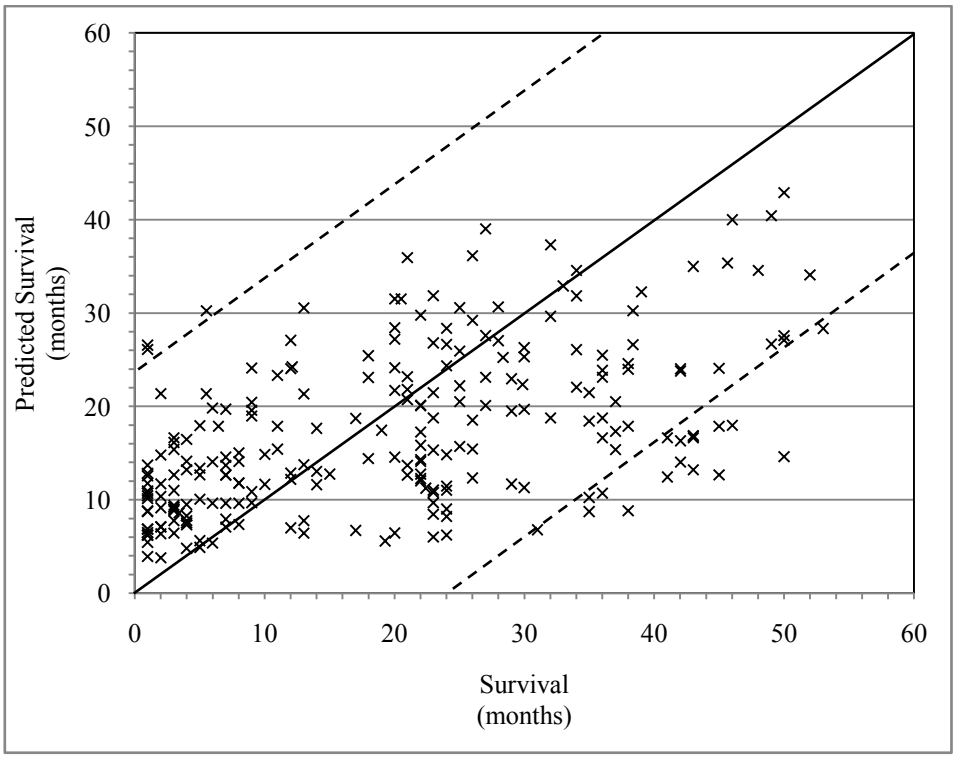

Figure 6-2 Predicted 2-year survival (16-2-1, 80-20)

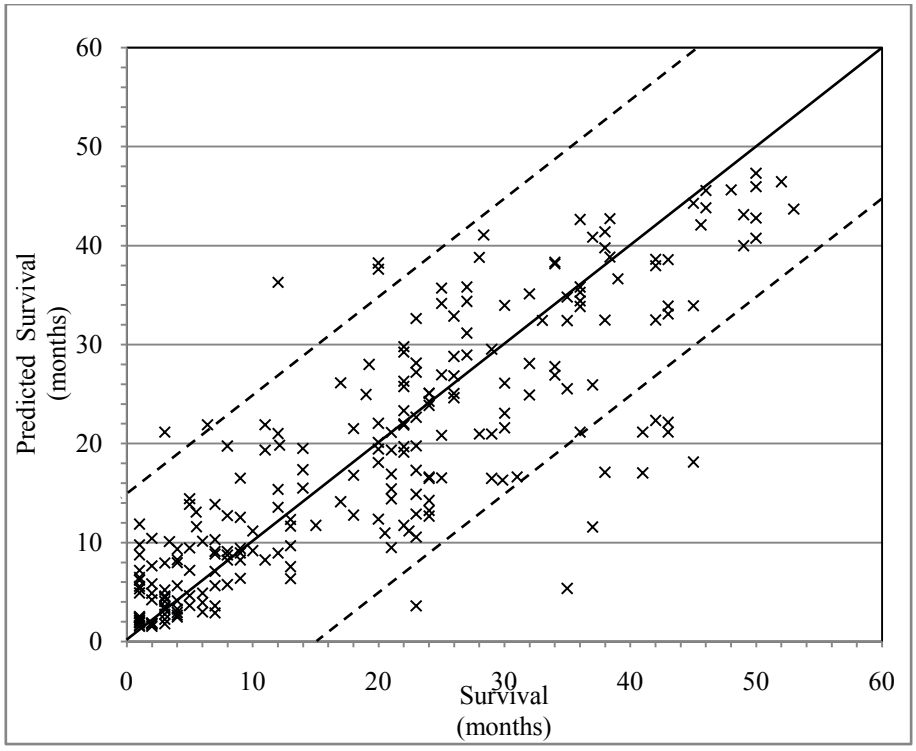

Figure 6-3 Initial predicted 2-year survival (24-18-1, 90-10)

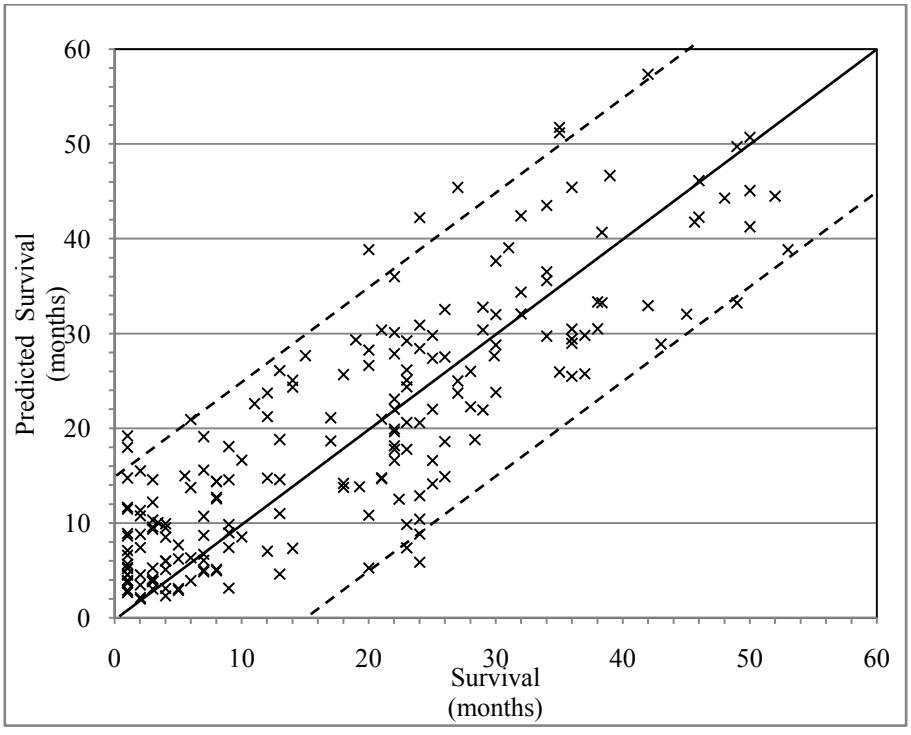

Figure 6-4 Final predicted 2-year survival (19-9-1, 90-10)

The patients whose survival is over-predicted in Figure 6-1 are in the 0-40 age group, of type acute leukaemia with normal karyotype or one addition or deletion. It is an expected result as the prognosis for survival of children and young adults with acute leukaemia and normal karyotype is quite high (Cripe, 1997), thus the model predicts this well. Those patients whose survival is under-predicted are in the age group >40 with karyotype from normal to translocations on chromosome 9. Older patients with these abnormalities have a poor prognosis, thus justifying the predicted survival by this model. In Figure 6-2 and 6-3 the trend for the patients in the relevant age groups and karyotypes were similar with regards to the predictions. These results indicate that even though the partitions were changed, the trends for the predictions are the same and are consistent with the revised literature (Cripe, 1997). The proposed model is well representative of the actual patient data resulting in a well trained neural network that can reliably used for predictions of new patients who fall into this category.

The statistical analysis table and the input variables Transforms table are shown in Table D-3, Appendix D. All variables with an "I" at the beginning have been used to determine the neural network model. The transfer function transforms for the input variables used in the model and their importance as indicated by the frequency are given below in Table 6-3. Some variables were rejected as inputs to the model building process. If these rejected variables are excluded when building a new neural network for the same data set then one would expect similar results for both the models since they were already found to be mathematically insignificant in the building of the initial neural network model. A new model was built that excluded the variables that were discarded in all three partitions. This was compared to the results in Table 6-2. The new model is a combination of accepted variables for all three partitions used therefore it can be expected that there will be a small variation in the results as can be seen from results in Table 6-4. The first two outputs in Table 6-4 are based on the accepted variables and the third is the favoured model from Table 6-2. The 70-30 model gives a better prediction than the 90-10 modelled with only the accepted variables. This may be due to the difference in the number of hidden units determined when the network was built. Overall the 90-10 (R value = 0.87 and network 24-18-1) partition from Table 6-2 gives the best prediction and is therefore the proposed model for the data set used for the 2-year survival analysis.

There are outlying values, i.e. predicted values outside the range of the confidence interval as can be seen in Figure 6-3. There will always be patients whose health status cannot be rationally explained. Patients can display similar symptoms and have similar laboratory results but their health

status or survival cannot be predicted exactly. There are numerous cases were clinicians are unable to explain how a person can have a disease at one stage and be cured further down the line, i.e. they are unable to find a specific or scientific reason for this phenomenon. All patients have had similar symptoms and similar laboratory results, yet there is a difference in the accuracy of their predictions as can be seen in Figure 6-3. Since there are some points that lie outside the boundaries they will be left out of the new model. The latter reasoning can be used as justification for leaving out the outlying points. The revised model is illustrated in Figure 6-4. The recommended model for this data set is the 90-10 partition summarised in Table 6-4. This model has the poorest confidence interval of 15 months for the training set and 13 months for the testing set, i.e. 95% of the predicted survivals lie in the bounded area in Figure 6-4. An accuracy of 1 indicates that (all) 100% predicted values are within 20% of the actual target values thus making this model the preferred one for this data set for predicting the survival of all leukaemia patients (uncensored) in 2 years. Since the model mathematically excluded some variables, those tabulated in Table 6-3 according to their frequency can be accepted as having the most significance for this data set. Since the data set is made up of the four types of leukaemia and the prognostic factors are normally used in diagnosis to specify the type of leukaemia, this table gives an overall summary of the factors that have a general impact on all types of leukaemia. This data set was further divided into the 4 types of leukaemia and separately analysed to predict a specific leukaemia and to investigate the prognostic factors for each type. The results for each type of leukaemia are presented in section 6.2.2.

Cytogenetic analysis of leukaemic cells is a critically important addition to the standard diagnosis and classification of leukaemia. It is currently considered to be an essential component in the assessment of a newly diagnosed leukaemia patient, playing a major role in diagnosis, sub-classification, selection of suitable therapy, and monitoring the effect of therapy (Harmening, 2002). The variable for chromosomes was excluded from all the models. A possible reason could be that in this study the various abnormalities were only grouped into 7 classes (of which 2 were of normal karyotype) for analysis compared to the multitude of chromosome abnormalities that exist and are used to diagnose the specific type of leukaemia. Further analysis is required with regards to specific cytogenetic testing and for the actual results to be incorporated into the proposed models.

Variable	Transformation (s)	Frequency
age	tanh	0.92
race	linear	0.79
	log	0.57
type	linear	0.56
mcv	linear	0.44
	pwr2	0.11
mchc	rlogical	0.92
rcw	linear	0.79
pc	inv	0.35
mpv	linear	0.37
wcc	linear	0.47
np	inv	0.97
mnc	linear	0.64
CD4	linear	0.61
	inv	0.56
CD13	logical	0.89
CD20	logical	0.85
CD22	logical	0.52
CD33	logical	0.49
CD56	logical	0.89
mpo	logical	0.47
survival	rt2	output

Table 6-3 Prognostic factors for 2-year survival

Partition	Network	R		ACC		CI (months)	
		T	S	T	S	T	S
70-30	21-10-1	0.84	0.58	0.83	0.56	15	24
90-10	18-2-1	0.70	0.58	0.67	0.50	20	26
90-10	24-18-1	0.87	0.65	0.89	0.58	14	15
Final 90-10	**19-9-1**	**0.83**	**0.91**	**1**	**1**	**15**	**13**

Table 6-4 2-year survival models based on accepted variables

For the 2-year survival group there were 77 deaths resulting in a survival rate of 67%. The average or mean survival calculated from the existing patient data is indicated in Table 6-5. The actual value

was calculated for the known survival period for all patients in the 2-year survival group and compared to the predicted mean survival for each of the proposed models. The average value is calculated from the predicted values of the model for each patient. The 90-10 partition predicted 18.77 months which is closest to the actual mean survival value of 19.56 months. The comparison of survival rates for leukaemia in general is not available in the current literature since all studies are done specifically for a type of leukaemia. A model for prediction of survival for all types of leukaemia is still useful as it indicates the main prognostic factors that affect leukaemia patients in general. In this study a multitude of all possible parameters that were available were used to build the neural networks. This approach therefore gives quite a comprehensive mathematical analysis of the variables that have the most prognostic significance to predicting survival generally in leukaemia patients. This can be used by clinicians as an initial start to testing all patients for these prognostic factors when they display the relevant symptoms. This can be used as a baseline start to the diagnosis of the patient‟s specific type of leukaemia. This will then give the clinician an overall impression of the patient‟s prognosis regardless of the type of leukaemia. If these are standard testing procedures for all patients then resources can be allocated in advance for all patients and costs can be minimised as this type of routine work can be easily managed.

Partition	Actual Mean Survival (months)	Std Dev	CI (months)	Predicted Mean Survival (months)	Std Dev	CI (months)
70-30	19.56	14.5	0.06	17.93	10.05	0.04
80-20	19.56	14.5	0.06	17.30	8.57	0.03
90-10	19.56	14.5	0.06	18.77	12.75	0.05
Final 90-10	**18.28**	**14.1**	**0.06**	**19.57**	**13.11**	**0.06**

Table 6-5 Mean survival for 2-year model

A five-fold cross validation technique (Ripley, 1998) was applied to the above data set. The cross validation procedure resulted in the selection of five „optimal‟ combinations of hidden units and weight decay for each model type. A five-fold-cross-validation was applied resulting in a grouping of 47 X 4 and a 48 subset. Each model was run by leaving out a subset. There were 5 models produced so that all the data was used in both the training and testing. A partition of 90-10 with the default learning rate of 100, weight decay of 0.0005 and gradient descent learning rule was applied to all the models. The results are displayed in Table 6-6.

Network	R		ACC		CI (months)	
	T	V	T	V	T	V
20-12-1	0.91	0.29	0.92	0.27	11	34
21-15-1	0.91	0.21	0.91	0.49	12	33
16-4-1	0.91	0.25	0.95	0.27	11	38
20-3-1	0.92	0.004	0.96	0.25	9	43
16-21-1	0.89	0.33	0.92	0.54	11	34
mean	**0.91**	**0.22**	**0.93**	**0.36**	**10.80**	**36.40**
std dev	**0.01**	**0.11**	**0.02**	**0.12**	**0.98**	**3.72**
CI (months)	**0.00**	**0.00**	**0.00**	**0.00**	**0.03**	**0.10**

Table 6-6 Cross validation for 2-year survival

The results in Table 6-6 indicate that the confidence interval of the validation data is quite large in comparison to the training data but is consistent in all five models. The data for analysis is limited in this study and this can be a possible reason for the wide confidence interval obtained in the validation models. If the model is applied to a clinical trial there must be a substantial number of data points and it should also incorporate multi-centred institutions to obtain a wide range of patients.

6.2.2 Two year survival based on type of leukaemia

The group of patients in 6.2.1 were divided into the specific types of leukaemia. Since the CLL group was too small to build a separate model (PREDICT requires a minimum of 20 data points to build a neural network model therefore it was grouped with the CML data set). In order to validate this model, new patient data must be added to this data set and the model retrained if this proposed method is to be used for future predictions. The other two groups comprised the ALL and AML patients. The 70-90 % partition with default LR, WD and gradient descent rule similar to 6.2.1 was applied in building the neural networks for each of the subdivided groups. The effect of varying the LR and WD did not have any effect in improving the models.

6.2.2.1 ALL

A summary of the network results are given in Table 6-7 and the proposed model based on the narrowest confidence interval is illustrated in Figure 6-5. There is a definite improvement in the

predictions for the individual sub-types of leukaemia, as can be seen from the improvement of the CI from 15 months to 9 months. It also confirms that each sub-type of leukaemia has a separate set of prognostic factors which influences survival. Neural networks are trained with data to form a model (or pattern) that is representative of all the data in a given data set. For the general leukaemia models in 6.2.2 four model patterns were incorporated to produce a general trend, thus the large confidence interval in comparison to the ALL model. The patients who lie outside the 95% confidence interval all have a normal karyotype with a single addition or deletion with the majority in the 0-20 age group. ALL is the most frequent malignancy of children in the 0-20 age group. Survival rates of childhood leukaemias are much higher than in adults. Children are able to withstand the effects of chemotherapy and respond well to conventional therapy. Adults are more prone to the toxicity of the drugs administered during chemotherapy leading to a lower success rate in this type of treatment. Some assumptions were made for patients that had missing or uncertain values for their cytogenetic results, i.e. they were replaced with a normal karyotype. This could account for the prediction of the 0-20 age group patients to be lower than the actual rate. The predicted survival is 20.99 months compared to the actual mean survival of 20.79 months. A 70-90 % remission rate for ALL has been reported by various research groups Yanada *et al* (2006), Xin *et al* (2006) and Hallbook (2005). The model has predicted age to be a major prognostic factor for ALL as is in the revised literature. As explained above, survival of ALL patients is dependent on their age, i.e. children and young adults respond well to treatment and have higher survival rates. Remission rates amongst children have steadily increased over the years with 100 % certainty in certain groupings.

Partition	**Network**	**R**		**ACC**		**CI (months)**	
		T	**S**	**T**	**S**	**T**	**S**
70-30	21-2-1	0.76	0.29	0.75	0.42	22	35
80-20	18-8-1	0.84	0.73	0.75	0.68	18	23
90-10	17-20-1	0.92	0.87	0.92	0.86	12	16
Final 90-10	**17-20-1**	**0.95**	**0.95**	**0.95**	**0.95**	**9**	**9**

Table 6-7 Summary of models for 2 year survival of ALL

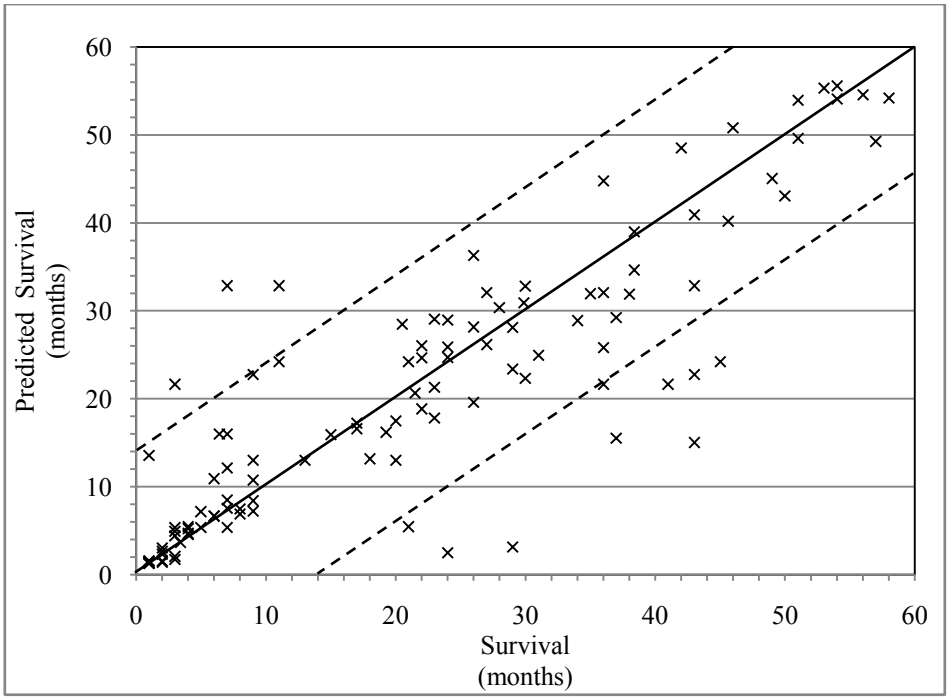

Figure 6-5 Initial predicted 2-year survival for ALL (17-20-1, 90-10)

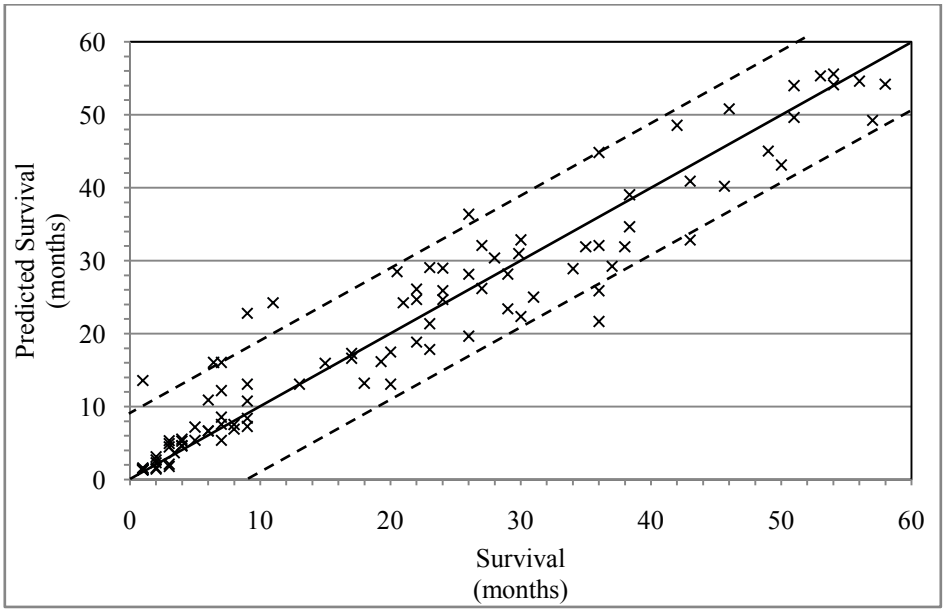

Figure 6-6 Final predicted 2-year survival for ALL (17-20-1, 90-10)

Cytogenetics is a critical component that is essential in the assessment of newly diagnosed leukaemia patients and this is confirmed by a frequency of 0.98 in Table 6-8. Chromosome abnormalities in ALL are divided into those that have a poor or good prognosis but other blood

results and expression of specific markers also influences a patient"s prognosis. A frequency of 0.98 implies that it has a significant role in the building of the neural network. There are no reported studies in the literature which show that race has an effect on ALL survival, but this model does. If applied to a clinical trial it should be confirmed by adding more patient data and having the model retrained. This can be taken a step further where HLA-typing which is specific to a race type, can be used as an added variable to see if this phenomenon is as per the predicted outcome for this model. The most common B-line-age ALL are the B-cell markers CD10, CD19, CD22 and CD34. This model has rejected the markers CD10 and CD19, but has included markers CD3, CD13, CD20, CD22, CD34 and CD56. In ALL lymphocytes are the predominant cell types (Lewis *et al*, 2001) as prognosed in this model with a frequency of 0.93. The variables "neutrophil" (f = 1.00) and "CD13" (f = 0.9) should also be further confirmed in the training of an extended model as explained for the variable "race". The wide confidence interval obtained for the validation models can be attributed to the reduced number of data points when the main group was subdivided.

Variable	Transformation (s)	Frequency
age	inv	1.00
race	exp	0.97
mchc	rlogical	0.40
mpv	linear	0.67
np	linear	0.39
	inv	1.00
lc	linear	0.93
	tanh	0.77
mnc	logical	0.09
CD3	log	0.44
CD13	logical	0.90
CD20	logical	0.44
CD22	logical	0.09
CD34	rlogical	0.32
CD56	rlogical	1.00
kappa	logical	0.23
chromo	linear	0.98
survival	rt2	output

Table 6-8 Prognostic factors for ALL 2-year survival

Partition	Actual Mean Survival (months)	Std Dev	CI (months)	Predicted Mean Survival (months)	Std Dev	CI (months)
70-30	21.48	18.71	0.11	18.82	13.87	0.08
80-20	21.48	18.71	0.11	19.59	15.92	0.10
90-10	21.48	18.71	0.11	20.69	17.56	0.11
Final 90-10	**20.79**	**18.61**	**0.01**	**20.99**	**17.73**	**0.01**

Table 6-9 2-year mean survival for ALL

Network	R		ACC		CI (months)	
	T	V	T	V	T	V
24-18-1	0.95	0.01	0.96	0.23	9	56
23-8-1	0.97	0.03	0.98	0.31	6	54
19-22-1	0.93	0.15	0.91	0.4	12	47
13-20-1	0.93	0.29	0.93	0.89	12	38
17-13-1	0.96	0.18	0.96	0.45	9	41
mean	**0.95**	**0.13**	**0.95**	**0.46**	**9.60**	**47.20**
std dev	**1.53**	**1.84**	**1.53**	**1.73**	**2.69**	**17.19**
CI (months)	**0.04**	**0.05**	**0.04**	**0.05**	**0.08**	**0.48**

Table 6-10 Cross validation for 2-year ALL

6.2.2.2 AML

Partition	Network	R		ACC		CI (months)	
		T	S	T	S	T	S
70-30	19-6-1	0.77	0.54	0.92	0.75	18	25
80-20	21-7-1	0.85	0.73	0.92	0.86	13	18
90-10	19-11-1	0.98	0.94	1.00	0.97	6	9
Final 90-10	**19-11-1**	**0.97**	**0.97**	**1**	**1**	**7**	**7**

Table 6-11 Summaries of models for 2-year survival for AML

The final recommended model has architecture of 19-11-1, accuracy of 1 and a confidence interval of 7 months. Once again this is a great improvement on the 15 months (training) CI for the general 2-year model. The division into subtypes is once again proving to be a better option than using the whole data set as 6.2.1. The factors age, % lymphocytes and the CD34 count have had the most influence on the building of the neural network model. According to the revised literature for AML, CD5, CD34, age, HLA-DR, white cell count and cytogenetics are used as indicators for diagnosis and prognosis (Chang *et al*, 2004a). Their study also categorised AML patients into three risk groups according to age, white cell count, cytogenetics and CD34/HLA-DR expression. This grouping may be used to develop strategies based on risk of individual patients. As can be seen from Appendix D, Table D-13 for the 2-year AML group, the white cell count has a frequency of 0.60, which implies that even though it was rejected in this model in favour of the other variables, it may become significant if more data points are added and the model rebuilt. HLA-DR has a frequency of 0.52 and chromosome 0.74 thus confirming the significance according to conventional factors used for diagnosis and prognosis of AML patients. Diagnosis and prognosis based on cytogenetics depends on the type of abnormality, e.g., t(9,22) patients with AML have a poor prognosis while t(8,21) and t(15,17) patients display a favourable prognosis. Cytogenetics is considered one of the key factors affecting prognosis but other factors like blast count and flow cytometry are used concurrently to determine the specific type of leukaemia. A predicted mean survival of 16.34 months compared to the actual mean survival of 16.86 months was obtained. The predicted value compares very well as can be seen from the accuracy of "1" for the model. According to the literature (Wahlin *et al*, (1991) and Chang *et al*, (2007b)) AML patients with normal karyotype have a 35-45% survival rate and a median survival time of 13.4 months for patients 60 years and younger. An overall 60-70% survival rate has been reported by various authors. There were 77 deaths in the first two years, resulting in a survival rate of 69.6% thus comparing well with the literature.

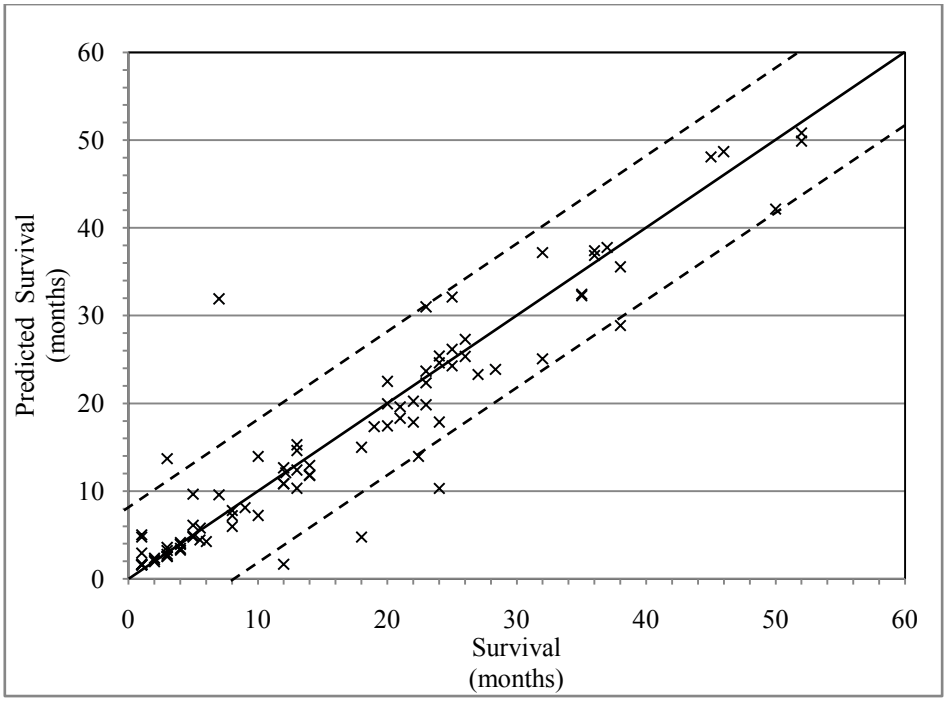

Figure 6-7 Initial predicted 2-year survival for AML (19-11-1, 90-10)

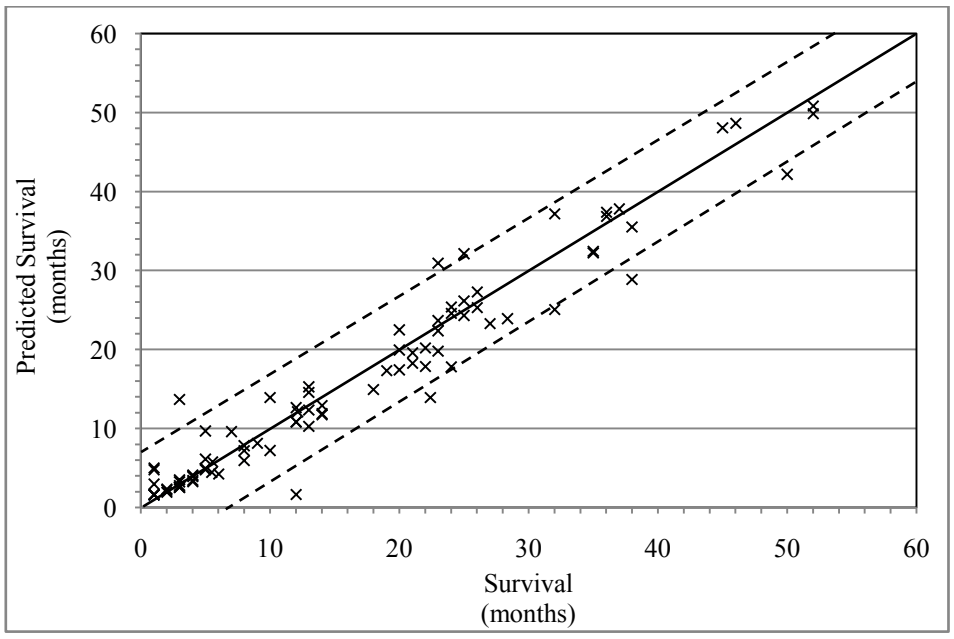

Figure 6-8 Final predicted 2-year survival for AML (19-11-1, 90-10)

Variable	Transformation (s)	Frequency
age	pwr2	0.98
race	log	0.79
hmb	linear	0.57
hmr	linear	0.72
mcv	pwr2	0.89
mch	logical	0.52
mchc	rlogical	0.87
mpv	power	0.63
np	linear inv	0.76 0.61
lc	linear	0.96
CD4	linear	0.21
CD7	logical	1.00
CD14	logical	0.79
CD33	rlogical	0.25
CD34	rlogical	1.00
kappa	logical	0.09
hlad	logical	0.52
chromo	pwr2	0.74
survival	inv	output

Table 6-12 Prognostic factors for 2-year survival for AML

Partition	Actual Mean Survival (months)	Std Dev	CI (months)	Predicted Mean Survival (months)	Std Dev	CI (months)
70-30	16.85	14.27	0.09	14.85	10.95	0.07
80-20	16.85	14.27	0.09	15.19	11.16	0.07
90-10	16.85	14.27	0.09	16.32	13.58	0.09
Final 90-10	**16.86**	**14.45**	**0.1**	**16.34**	**13.64**	**0.09**

Table 6-13 2-year mean survival for AML

Network	R		ACC		CI (months)	
	T	V	T	V	T	V
17-4-1	0.96	0.38	0.98	0.35	8	52
22-10-1	0.93	0.18	0.98	0.53	6	49
19-2-1	0.83	0.15	0.87	0.41	14	43
20-16-1	0.93	0.1	0.98	0.47	11	36
22-14-1	0.97	0.3	0.98	0.41	7	48
mean	**0.92**	**0.22**	**0.96**	**0.43**	**9.20**	**45.60**
std dev	**1.66**	**1.95**	**1.65**	**1.87**	**3.39**	**17.50**
CI (months)	**0.05**	**0.05**	**0.05**	**0.05**	**0.10**	**0.49**

Table 6-14 Cross validation for 2-year survival for AML

6.2.2.3 CML-CLL

Partition	Network	R		ACC		CI (months)	
		T	S	T	S	T	S
70-30	21-3-1	0.98	0.83	1	0.88	6	17
80-20	21-3-1	0.93	0.77	0.98	0.88	12	19
90-10	21-3-1	0.98	0.96	1.00	0.96	6	8
Final 90-10	**24-12-1**	**0.98**	**0.98**	**1.00**	**1.00**	**6**	**6**

Table 6-15 Summaries of models for 2 year survival for CML-CLL

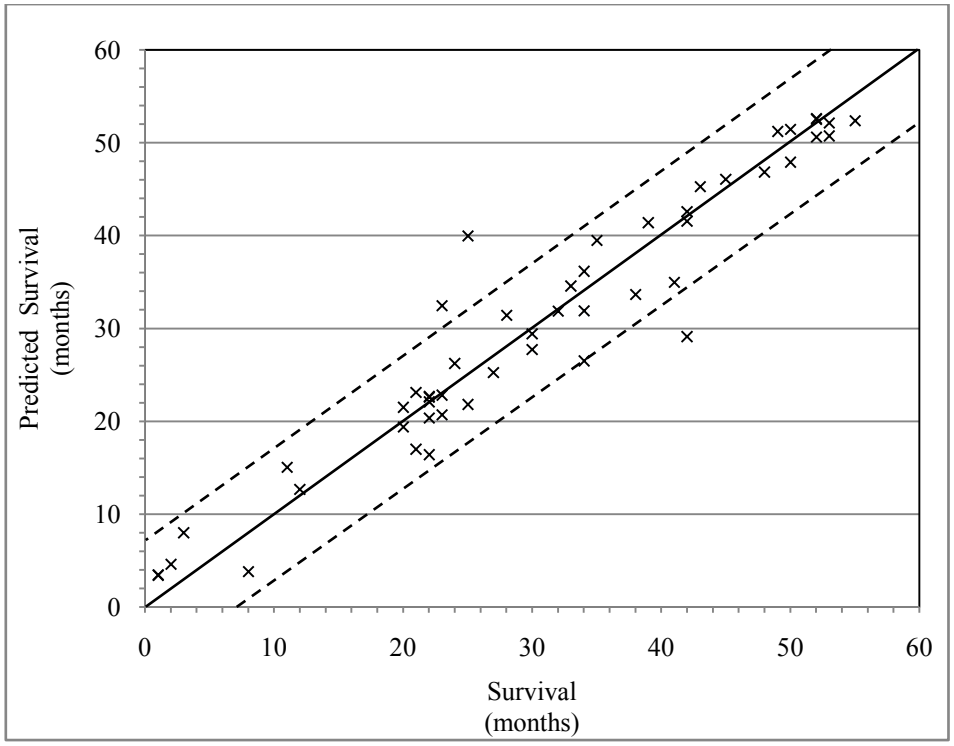

Figure 6-9 Initial predicted 2-year survival for CML-CLL (21-3-1, 90-10)

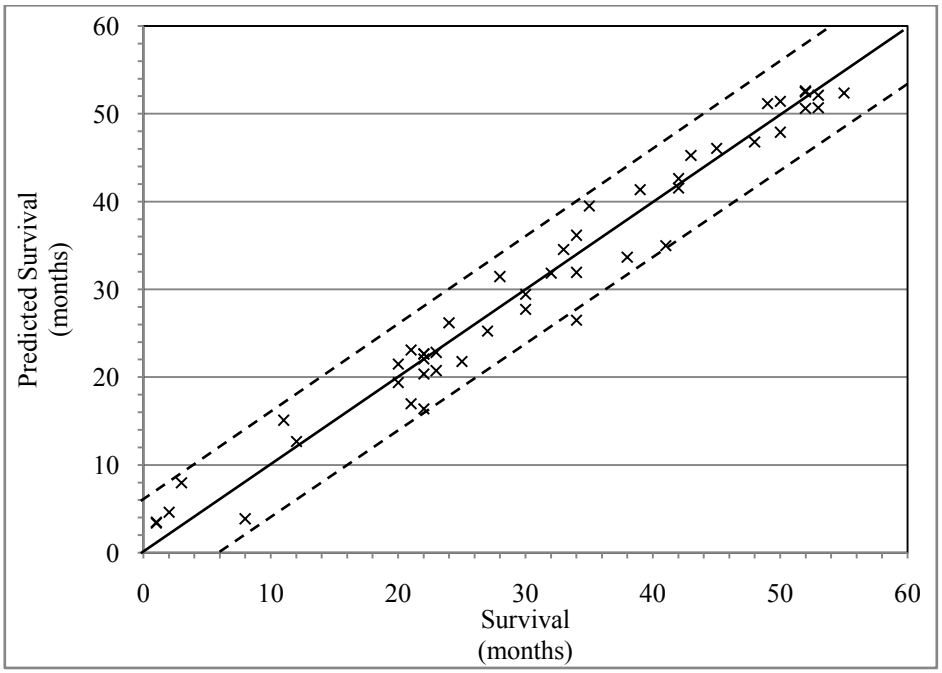

Figure 6-10 Final predicted 2-year survival for CML-CLL (24-12-1, 90-10)

The outlying patients in figure 6-9 that are over predicted are in the 20-40 age group and the patients under-predicted are from the age group > 80, all having a normal karyotype. This result is consistent with the trend that younger patients have a greater probability of survival than older patients. These patients could have had some abnormal karyotype that was not detected because of insufficient samples or inconclusive results, thus the difference in prediction from the other patients in this subgroup. The final model in figure 6-10 predicts the survival for CML and CLL patients with a confidence interval of 6 months and accuracy of 1. The mean survival rate is higher than both the AML and the ALL group. In chronic leukaemias the patients have the disease for long periods of time before symptoms are noticed. Even when diagnosed many live a normal life with minor symptoms. Those patients that have more severe symptoms usually lapse into an accelerated path to an acute leukaemia which eventually leads to death. This result is in keeping with a study done by Chase *et al* (2001) where a survival rate of 28 months was determined for various sub-groups of chronic leukaemias. Moore *et al* (2004) reports a median survival time of approximately 3 years.

According to the literature CD5, CD19, CD 23 together with kappa and lambda light restriction are used for the diagnosis of CML and this is confirmed by the prognostic factors in Table 6-16 where all have a frequency > 0.5. CD 5, CD19, CD20 and CD23 are usually revealed in flow cytometry for CLL patients. Since this model is a combination of CML and CLL, prognostic factors are jointly predicted for both groups. The model has revealed that race (f = 92), platelet count (f = 0.99), CD3 (f=1) and CD34 expression (f = 0.92) are the most important variables in the prognosis of both CML and CLL, with age and monocytes also having a significant effect on the model.

Partition	**Actual Mean Survival (months)**	**Std Dev**	**CI (months)**	**Predicted Mean Survival (months)**	**Std Dev**	**CI (months)**
70-30	30.72	14.99	0.13	31.5	13.5	0.11
80-20	30.72	14.99	0.13	31.2	12.7	0.11
90-10	30.72	14.99	0.13	31.7	14.6	0.13
Final 90-10	**30.76**	**15.37**	**0.14**	**30.55**	**15.01**	**0.14**

Table 6-16 2-year mean survival for CML-CLL

This subgroup consisted of only 50 patients, therefore although the training accuracy was quite high (0.97-1) the validation shown in Table 6-18 was extremely low. The validation models training and testing was done on 40 patients and validated on 10 patients. This model would have to be updated with new uncensored CML-CLL patient data to confirm its reliability in predicting survival.

Variable	Transformation (s)	Frequency
age	linear	0.73
race	linear	0.85
	inv	0.92
type	logical	0.52
sex	rlogical	0.14
hmr	exp	0.05
mcv	linear	0.25
pc	linear	0.99
	pwr2	0.82
np	tanh	0.4
mnc	logical	0.89
blast	log	0.43
CD3	linear	1.00
	log	0.99
CD4	linear	0.95
CD5	logical	0.70
CD7	logical	0.19
CD19	logical	0.50
CD22	logical	0.72
CD33	logical	0.33
CD34	logical	0.92
CD56	rlogical	0.05
kappa	logical	0.45
lambda	logical	0.91
survival	pwr2	output

Table 6-17 Prognostic factors for 2-year CML-CLL

Network	R		ACC		CI (months)	
	T	V	T	V	T	V
21-12-1	0.98	0.15	1	0.12	4	69
19-23-1	0.99	0.07	1	0.25	2	59
18-2-1	0.94	0.033	0.97	0.4	10	58
20-21-1	0.99	0.24	1	0.57	2	53
19-4-1	0.99	0.52	1	0.75	2	40
mean	**0.98**	**0.20**	**0.99**	**0.42**	**4.00**	**55.80**
std dev	**1.64**	**1.97**	**1.64**	**1.88**	**3.13**	**22.79**
CI (months)	**0.05**	**0.06**	**0.05**	**0.05**	**0.09**	**0.64**

Table 6-18 Cross validation for 2-year survival for CML-CLL

6.2.3 3-year case study

The 3-year survival group consisted of 223 patients where all the relevant information was pre-processed as per the methodology in Chapter 5. The final recommended model had an R value of 0.86 and accuracy of 0.79. The 2-year final model seems to give a better prediction based on its R value of 0.83, accuracy of 1 and confidence interval of 15 months. The variables age, race, mean corpuscular haemoglobin concentration, platelet count, mean platelet volume, white cell count, neutrophils, CD13, CD20 and CD33 were found to be common prognostic factors in both the final 2-year and 3-year models. The confidence interval of 19 months is also worse than the 15 months for the 2-year final model. A difference in the results can also be due to the fewer number of patients in the 3-year group. The limitation once again is the availability of uncensored patient data, thus resulting in the low numbers used in the modelling process. The actual mean survival of 23.54 months was greater than the 18.28 months for the 2-year model. This is expected as the time period is greater. The CI for the predicted mean survival was 0.08 months for the 2-year final model and 0.07 months for the 3-year final model. Once again there are no results on survival of combined leukaemia groups in the literature but as explained above the prognostic factors can be used for initial testing and diagnosis of all leukaemia patients when they display the relevant symptoms.

Partition	Network	R		ACC		CI (months)	
		T	S	T	S	T	S
70-30	24-6-1	0.63	0.38	0.63	0.58	29	36
80-20	22-5-1	0.63	0.40	0.61	0.64	29	36
90-10	21-6-1	0.80	0.79	0.77	0.61	22	25
Final 90-10	**21-6-1**	**0.86**	**0.88**	**0.79**	**0.82**	**19**	**18**

Table 6-19 Summary of models for 3-year survival

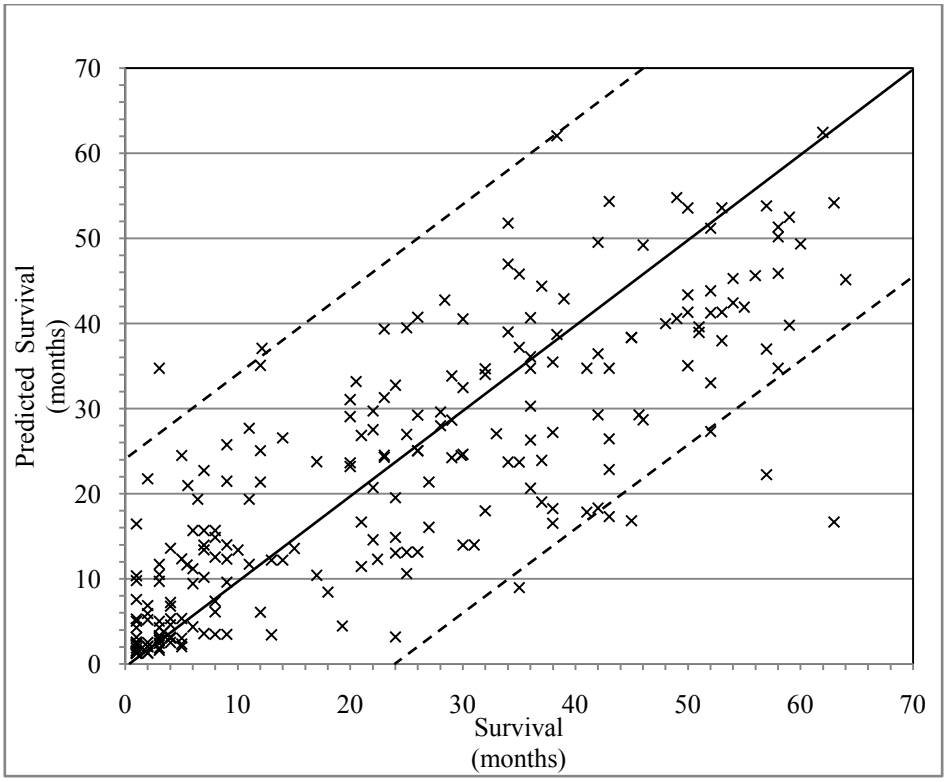

Figure 6-11 Initial predicted 3-year survival (21-6-1, 90-10)

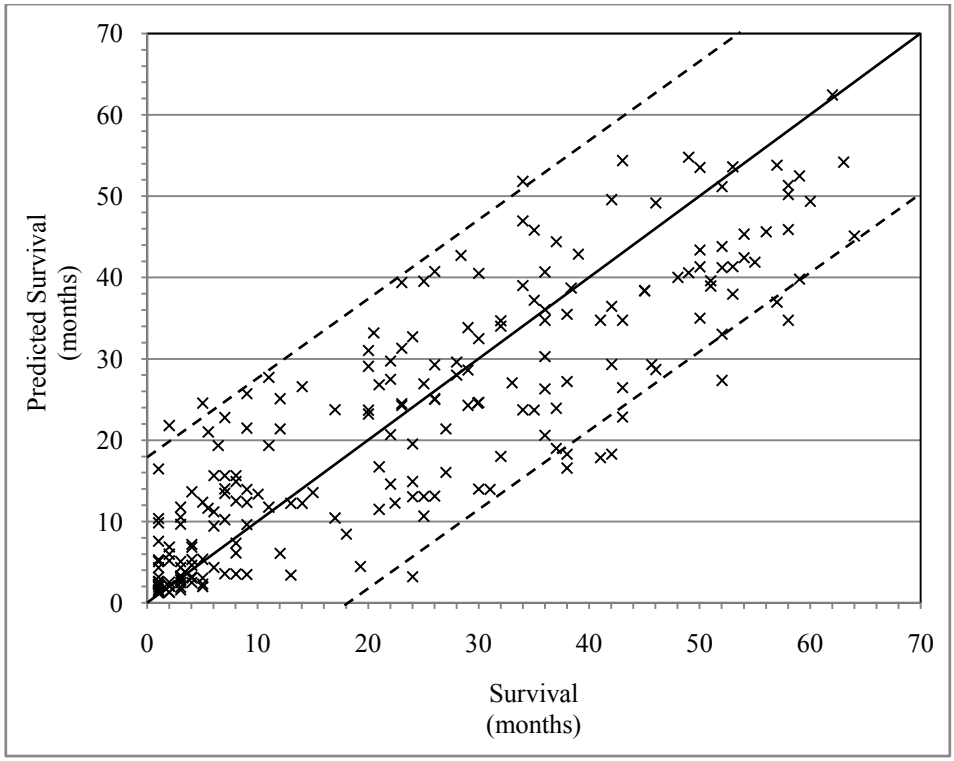

Figure 6-12 Final predicted 3-year survival (21-6-1, 90-10)

Variable	Transformation (s)	Frequency
age	linear	0.58
	tanh	0.93
race	linear	0.25
	log	0.99
rcc	tanh	0.16
mchc	rlogical	0.82
pc	linear	0.84
mpv	linear	0.98
wcc	inv	0.4
np	linear	0.87
	tanh	0.59
lc	linear	1.00
CD5	logical	0.94
CD8	tanh	0.75
CD10	logical	0.16
CD13	logical	0.98
CD14	logical	0.49
CD20	logical	0.92
CD33	logical	0.61
CD34	logical	0.18
lambda	logical	0.18
survival	rt2	output

Table 6-20 Prognostic factors for 3-year survival

Partition	Actual Mean Survival (months)	Std Dev	CI (months)	Predicted Mean Survival (months)	Std Dev	CI (months)
70-30	23.97	18.86	0.08	20.60	11.3	0.05
80-20	23.97	18.86	0.08	20.83	12.4	0.05
90-10	23.97	18.86	0.08	22.49	15.7	0.07
Final 90-10	**23.54**	**18.74**	**0.08**	**22.26**	**15.73**	**0.07**

Table 6-21 3-year mean survival

Network	R		ACC		CI (months)	
	T	V	T	V	T	V
17-21-1	0.92	0.04	0.94	0.42	13.00	64.00
22-9-1	0.88	0.11	0.87	0.44	17.00	48.00
15-10-1	0.92	0.07	0.88	0.37	14.00	52.00
18-8-1	0.96	0.18	0.96	0.30	10.00	55.00
19-4-1	0.90	0.06	0.89	0.58	17.00	36.00
mean	**0.92**	**0.09**	**0.91**	**0.42**	**14.20**	**51.00**
std dev	**0.03**	**0.06**	**0.04**	**0.10**	**2.95**	**10.25**
CI (months)	**0.00**	**0.00**	**0.00**	**0.00**	**0.08**	**0.29**

Table 6-22 Cross validation for 3-year survival

6.2.3.1. 3-year ALL

The grouping in the sub-types has again improved the modelling results. The results in this model are similar to the 2-year model for ALL, except for an R value of 0.97 in this model. The CI of 9 months is the same but the 3-year model has more outlying patients as can be seen in Figure 6-14, thus making the 2-year ALL model more reliable. The common prognostic factors between the 2- and 3-year models are age, race, mean haemoglobin concentration, neutrophils, lymphocytes, CD22, CD56 and chromosomes.

Partition	Network	R		ACC		CI (months)	
		T	S	T	S	T	S
90-10	19-9-1	0.95	0.88	0.91	0.87	12	18
Final 90-10	**19-9-1**	**0.97**	**0.97**	**0.95**	**0.95**	**9**	**9**

Table 6-23 Summary of models for 3-year ALL

Variable	Transformation (s)	Frequency
age	linear	0.95
race	inv	0.72
rcc	linear	0.56
mchc	logical	0.98
pc	pwr2	0.49
wcc	linear	0.12
np	inv	0.79
lc	linear	0.68
blast	linear	0.97
CD4	linear	0.19
CD7	logical	0.8
CD8	tanh	0.97
CD22	logical	1.00
CD56	rlogical	0.59
hlad	logical	0.56
mpo	logical	0.34
chromo	linear	0.39
survival	ln x/(1-x)	output

Table 6-24 Prognostic factors for 3-year ALL

Partition	Actual Mean Survival (months)	Std Dev	CI (months)	Predicted Mean Survival (months)	Std Dev	CI (months)
90-10	23.63	19.25	0.12	23.16	18.09	0.11
Final 90-10	**23.21**	**19.17**	**0.12**	**22.64**	**18.50**	**0.11**

Table 6-25 Mean survival for 3-year ALL

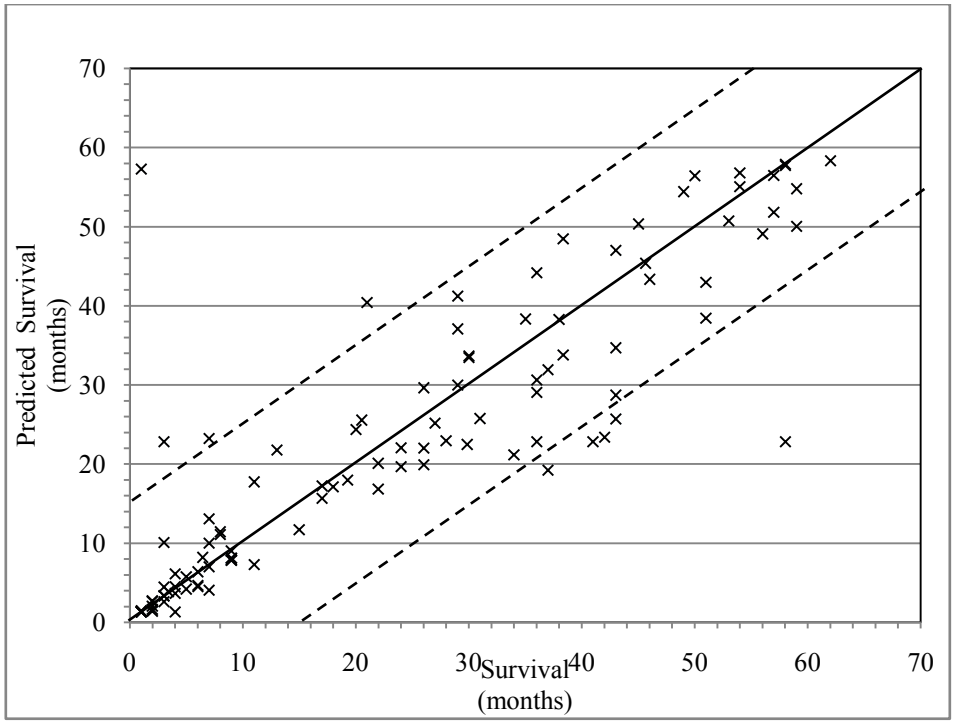

Figure 6-13 Initial predicted 3-year survival for ALL (19-9-1, 901-0)

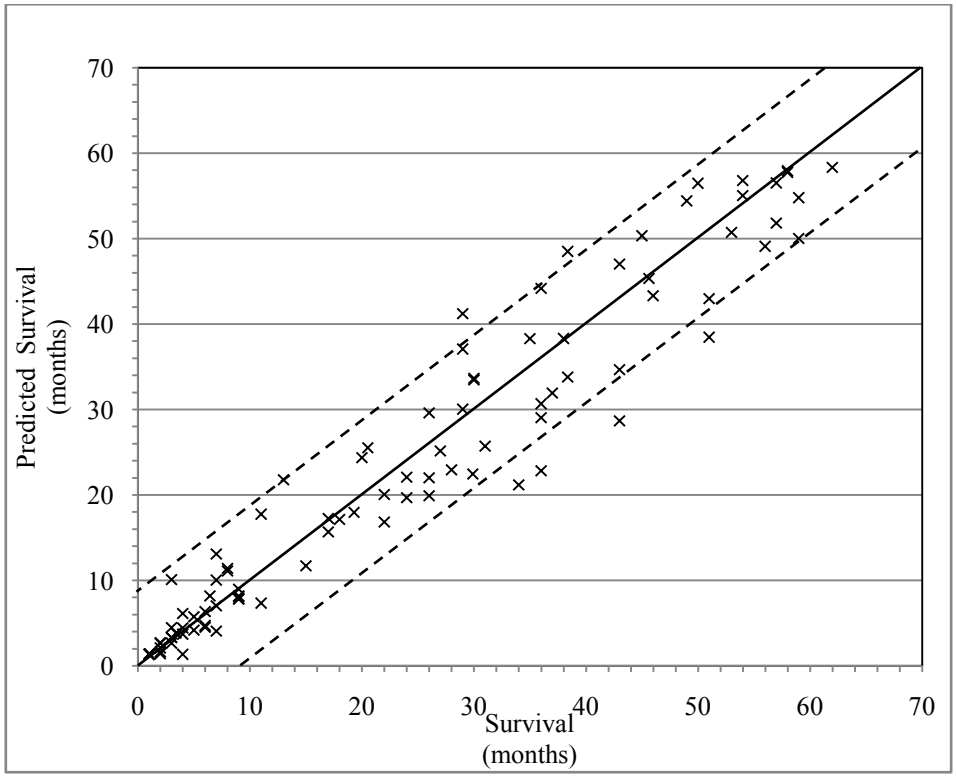

Figure 6-14 Final predicted 3-year survival for ALL (19-9-1, 90-10)

Network	R		ACC		CI (months)	
	T	V	T	V	T	V
17-5-1	0.99	0.20	1.00	0.43	6.00	50.00
14-11-1	0.93	0.30	0.93	0.62	14.00	50.00
12-13-1	0.83	0.11	0.77	0.33	21.00	56.00
18-6-1	0.94	0.52	0.96	0.30	13.00	51.00
17-10-1	0.95	0.20	0.91	0.43	12.00	61.00
mean	**0.93**	**0.27**	**0.91**	**0.42**	**13.20**	**53.60**
std dev	**0.06**	**0.16**	**0.09**	**0.13**	**5.36**	**4.83**
CI (months)	**0.00**	**0.00**	**0.00**	**0.00**	**0.15**	**0.14**

Table 6-26 Cross validation for 3-year ALL

6.2.3.2. 3-year AML

Partition	Network	R		ACC		CI (months)	
		T	S	T	S	T	S
90-10	20-12-1	0.99	0.95	1	0.97	4	10
Final 90-10	**20-12-1**	**0.99**	**0.99**	**1**	**1**	**4**	**4**

Table 6-27 Summary of 3-year models for AML

A final model with 20-12-1 architecture is proposed. The R value of 0.99 corresponding to an accuracy of 1 implies that this model had sufficient data for training to give a true prediction of the actual survival. A CI of 4 months will give clinicians a clear indication whether to treat and how to treat the patient. If a patient is deemed to be terminally ill with no response to a treatment regime then the treatment is stopped and palliative care is prescribed. This means that the clinicians ask the patients‟ families or care givers to make their final days as comfortable as possible while all treatment is terminated. Patients treated for AML should show a favourable response by the 4th month of treatment with some patients going into remission. A CI of 4 months will now allow clinicians to make important decisions for a shorter period of time. The following prognostic factors were derived for both models: age, haemoglobin, mean corpuscular haemoglobin, mean corpuscular haemoglobin concentration, mean platelet volume, neutrophils, lymphocytes, CD7, CD34 and chromosomes. The variable age, neutrophils, lymphocytes and chromosomes have also been found

to be significant factors for predicting survival. This is in keeping with the explanation given in 6.2.2.1 about the factors used for the diagnosis of AML. The mean predicted survival has a CI of 0.12 compared to 0.09 in the 2-year model, but this would be the favoured model as the CI of 4 months will be favoured by clinicians for the treatment and care of their patients.

Variable	Transformation (s)	Frequency
age	linear pwr2	0.73 0.97
hmb	linear	0.67
hmr	linear	0.54
mcv	pwr2	0.43
mch	logical	0.48
mchc	rlogical	1.00
rcw	pwr2	0.68
mpv	pwr2	0.29
np	linear	0.95
lc	linear	0.88
CD7	logical	0.97
CD8	logical	0.51
CD13	logical	0.65
CD20	logical	0.37
CD22	logical	0.02
CD34	rlogical	0.85
mpo	logical	0.35
chromo	linear pwr2	0.77 0.43
survival	log	output

Table 6-28 Prognostic factors for 3-year AML

Partition	Actual Mean Survival (months)	Std Dev	CI (months)	Predicted Mean Survival (months)	Std Dev	CI (months)
90-10	18.29	16.61	0.12	18.35	16.69	0.12
Final 90-10	**17.19**	**16.60**	**0.12**	**17.03**	**15.96**	**0.12**

Table 6-29 Mean survival for 3-year AML

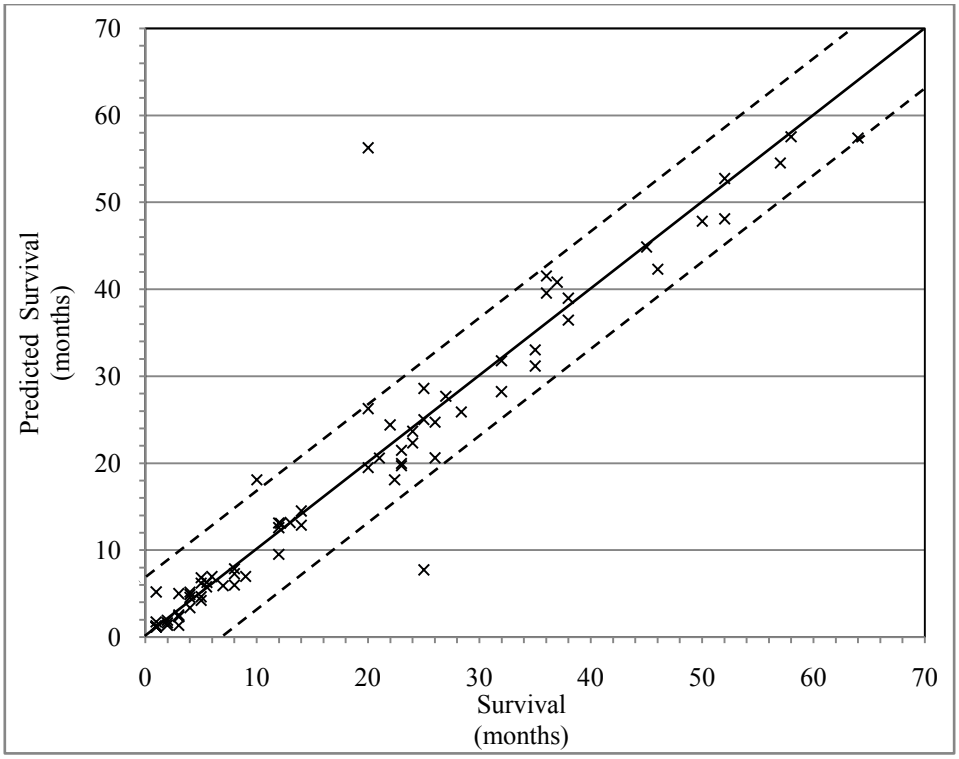

Figure 6-15 Initial predicted survival for 3-year AML (20-12-1, 90-10)

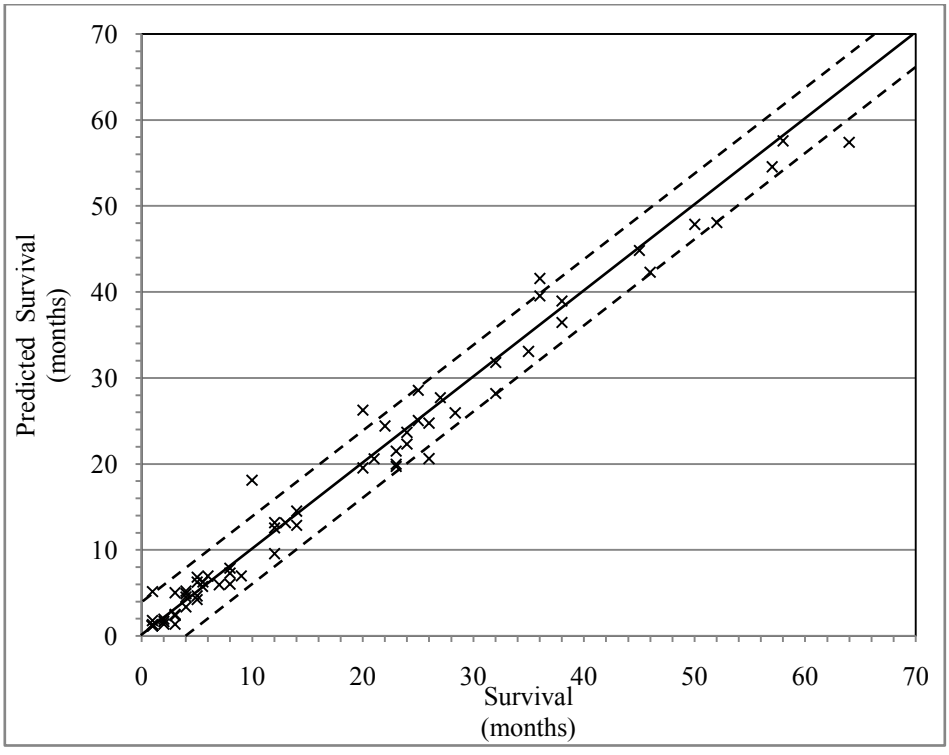

Figure 6-16 Final predicted survival for 3-year AML (20-12-1, 90-10)

Network	R		ACC		CI (months)	
	T	V	T	V	T	V
21-17-1	0.98	0.21	0.98	0.40	6.00	61.00
22-9-1	0.98	0.11	1.00	0.30	6.00	61.00
25-22-1	0.96	0.09	0.96	0.47	9.00	42.00
18-4-1	0.98	0.25	0.98	0.27	6.00	60.00
21-8-1	0.96	0.70	0.96	0.67	9.00	39.00
mean	**0.97**	**0.27**	**0.98**	**0.42**	**7.20**	**52.60**
std dev	**0.01**	**0.25**	**0.02**	**0.16**	**1.64**	**11.10**
CI (months)	**0.00**	**0.01**	**0.00**	**0.00**	**0.05**	**0.31**

Table 6-30 Cross validation for 3-year survival for AML

6.2.3.3. 3-year CML-CLL

The model proposed has architecture 17-7-1, R value of 0.97 and accuracy of 1. When compared to the 2-year model the CI of 6 months is favoured when compared to the 8 months proposed by this model. This group was made up of 43 patients. The accuracy of 1 proves that the model has been trained on representative data and can predict quite well as illustrated in Figure 6-18, with only 2 patients outside the confidence bands. Once again for validation of this model for use in future predictions, more data points from new uncensored patients need to be added and the model retrained. The common prognostic factors determined by both models are age, race, haematocrit, platelet count, CD3 and CD13. The 2-year model found a total of 9 CD markers to be significant for this leukaemia subtype while only three were deemed to be important in this model. The variable lambda normally used in diagnosis was used as an input in both models. The favoured model for CML-CLL group would be the 2-year model. This is a combination of two leukaemia subtypes therefore this model would have to be extended with data from new uncensored patients to see if the prediction trend is the same, or else more uncensored data collected and the models trained separately. A study by Chase *et al* (2001) revealed a mean survival of 28 months. The predicted mean survival compares well with the actual value of 30.76 months. The predicted mean survival of 30.55 months is quite high, but realistic since patients with chronic leukaemia live longer with the symptoms, sometimes without any adverse effects. Chronic means long term but there are cases that rapidly deteriorate and transform into acute leukaemias with a very poor prognosis for survival.

Partition	Network	R		ACC		CI (months)	
		T	S	T	S	T	S
90-10	17-7-1	0.97	0.92	1	0.95	8	13
Final 90-10	**17-7-1**	**0.97**	**0.97**	**1**	**1**	**8**	**8**

Table 6-31 3-year survival models for CML-CLL

Variable	Transformation (s)	Frequency
age	tanh	0.87
race	linear	0.74
	exp	0.99
hmr	linear	0.18
	pwr2	0.80
mcv	tanh	0.59
mch	logical	0.97
mchc	rlogical	0.96
pc	pwr2	0.98
mpv	linear	0.83
np	linear	1.00
lc	linear	0.57
blast	logical	0.97
CD3	exp	0.91
CD13	logical	1.00
lambda	logical	0.88
chromo	exp	0.92
survival	inv	output

Table 6-32 Prognostic factors 3-year CML-CLL

Partition	Actual Mean Survival (months)	Std Dev	CI (months)	Predicted Mean Survival (months)	Std Dev	CI (months)
90-10	35.12	17.38	0.17	34.60	16.75	0.16
Final 90-10	**35.23**	**17.79**	**0.17**	**35.78**	**16.25**	**0.16**

Table 6-33 Mean survival for 3-year CML-CLL

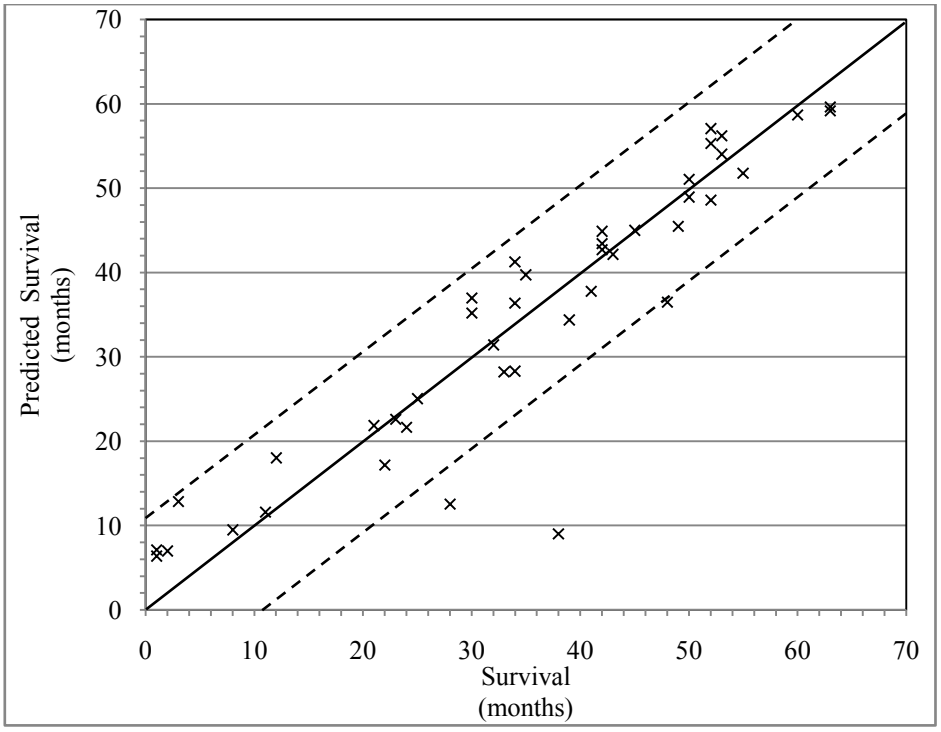

Figure 6-17 Initial predicted survival for 3-year CML-CLL (17-7-1, 90-10)

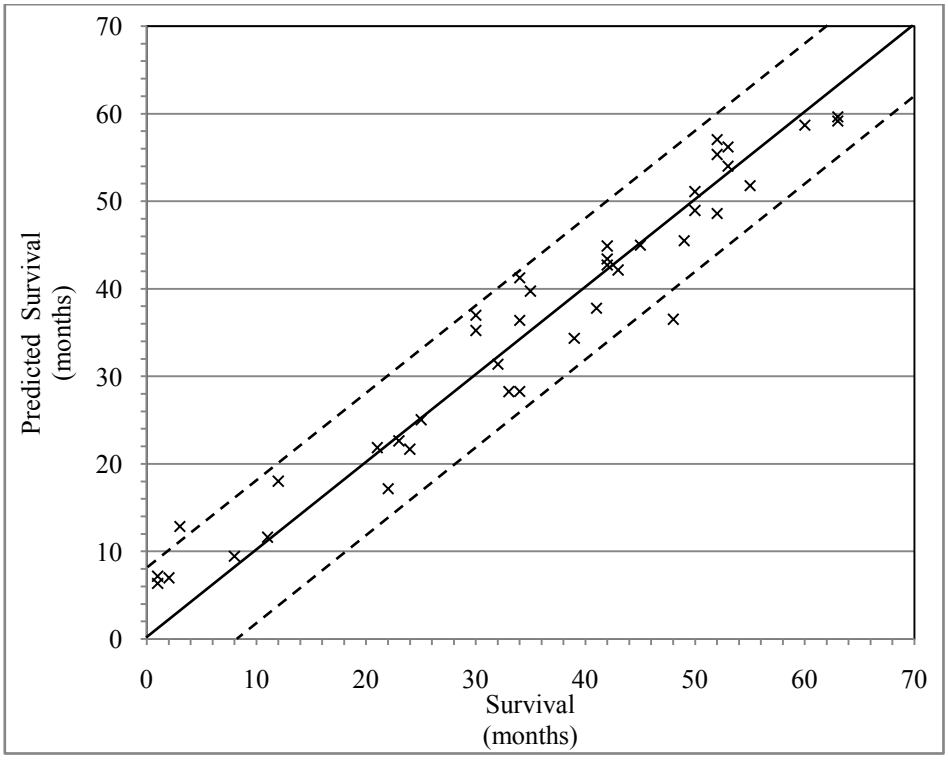

Figure 6-18 Final predicted survival for 3-year CML-CLL (17-7-1, 90-10)

Network	R		ACC		CI (months)	
	T	V	T	V	T	V
11-6-1	0.99	0.01	0.99	0.27	0.80	54.00
11-8-1	0.98	0.44	1.00	0.28	0.50	38.00
mean	**0.99**	**0.23**	**1.00**	**0.28**	**0.65**	**46.00**
std dev	**0.01**	**0.30**	**0.01**	**0.01**	**0.21**	**11.31**
CI (months)	**0.00**	**0.01**	**0.00**	**0.00**	**0.01**	**0.50**

Table 6-34 Cross validation for 3-year survival for CML-CLL

There were only 43 patients in the CML-CLL group and PREDICT has requires a minimum of 20 data points for running a model. The data set was divided into two and compared, thus the large difference in the values.

6.2.4. Censored case study

All 610 patients who were eligible for this study was used to form a group which included censored patients. Censored patients have to be incorporated into medical statistical analysis since the trend for patients to leave a hospital or institution is a common occurrence all over the world. If these patients are left out then the proposed results become biased and in most cases censored patients make up about 70-80 % of most studies (Xiang *et al*, (2000) and Ripley, (1998)). As can be seen this is quite a large number, and if excluded then the results do not become a true indication of the actual analysis in any centre or institution. Similarly in this study, the censored patients made up 61% of the initial group whose medical diagnosis and laboratory results were known. Their survival period was known up to a point where treatment or laboratory analysis was carried out. Thereafter patients may have moved to different areas or medical institutions, some defaulted on treatment, some passed on and then there were those for whom reasons are unknown. A neural network model including censored data was determined for the whole group and then subdivided into each sub-type of leukaemia for further analysis.

Partition	Network	R		ACC		CI (months)	
		T	S	T	S	T	S
70-30	18-3-1	0.38	0.28	0.78	0.75	32	33
80-20	17-3-1	0.29	0.40	0.76	0.75	33	32
90-10	15-9-1	0.41	0.37	0.77	0.73	31	32
Final 90-10	**18-3-1**	**0.61**	**0.51**	**0.94**	**0.89**	**12**	**14**

Table 6-35 Censored models

The final model has an R value of 0.61, but an accuracy of 0.94 making it reliable to use for predictions. A CI of 12 is much better than the 19 months for the 3-year survival and the 15 months for the 2-year model. Even though the data is censored the model has comparable results. Also the number of data points are almost double that of the 2-year and 3-year groups (uncensored). For more robust modelling a large number of data points will give a better representation of a data set, thus making the censored model as reliable as the two uncensored models. Uncensored models should be periodically updated with new patient data to ensure that the predictions are as reliable.

As can be expected the illustrated results in Figure 6-19 and 6-20 show an overall effect of under prediction of survival. Mathematically this is true as the censored patients had a shorter than actual recorded survival value which contributes to the low survival rate for all leukaemia patients. This trend is learnt in the training of the neural network and is projected in the prediction results. In the revised model the prediction is much better as the outlying patients were removed. There are still a number of outlying patients in the improved model in Figure 6-20 but any further reduction in outliers will reduce the number of data points, thus reducing the reliability of this censored model. Some of the outlying patients (censored) eliminated from Figure 6-19 have survived beyond 4 years and these patients are most likely to be in remission and do not need any more treatment or a few may have passed on. This is a combined group of censored and uncensored patients and the points have not been distinguished on the graph. The actual mean survival for the censored group is much higher then the predicted value. The actual mean survival is quite low at 6.86 months, but understandable because 61 % of the data is censored. Many patients have been in the treatment system for a few months as can be seen by the large clusters between 0 and 12 months in both Figures 6-19 and 6-20, thereafter some may be remission and others unknown. If the model is applied in a clinical trial more uncensored patient data are needed to improve on the initial model and only if reasonable results are obtained can it be used for prediction of survival for new patients. The limitation of uncensored patient data in this study is a major contributing factor to the poor

prediction results of some data sets. Unavailability of uncensored data is not a unique phenomenon to this study as this occurs in most hospitals and institutions worldwide.

The variables age, race, platelet count, markers CD5, CD13 and lambda were found to be common prognostic factors for both the uncensored groups and this censored group. Cytogenetics play a key role in diagnosis of leukaemia and is confirmed in this model where chromosomes (f = 0.87) was found to be a significant factor. The difference is that there are a large number of data points that were used in the training of this model compared to the uncensored models, thus resulting in the significance of the variable "chromosomes". Since this is a general model for all leukaemias the data will have to be subdivided to improve the prediction accuracy.

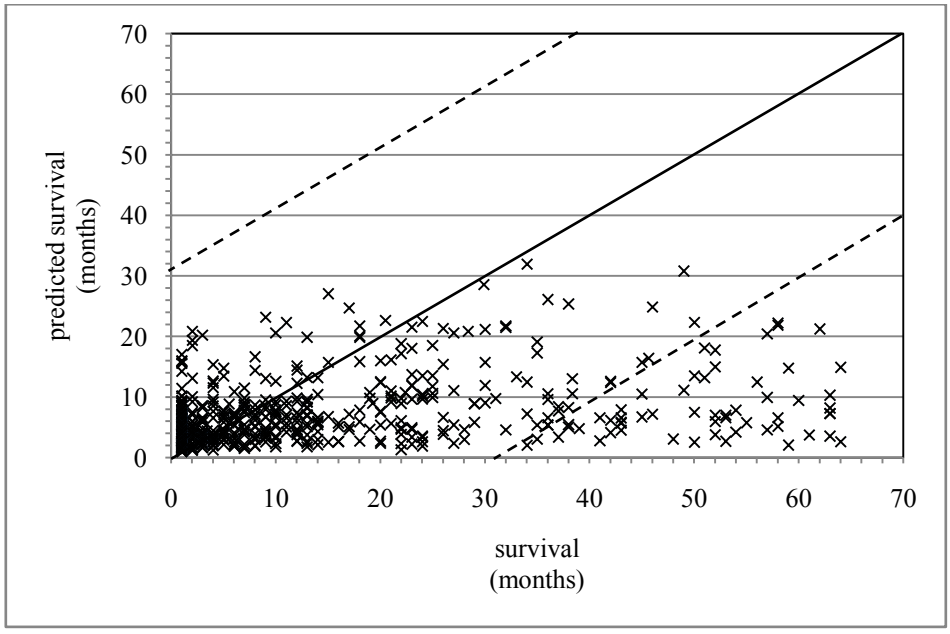

Figure 6-19 Initial predicted survival for censored group (15-9-1, 90-10)

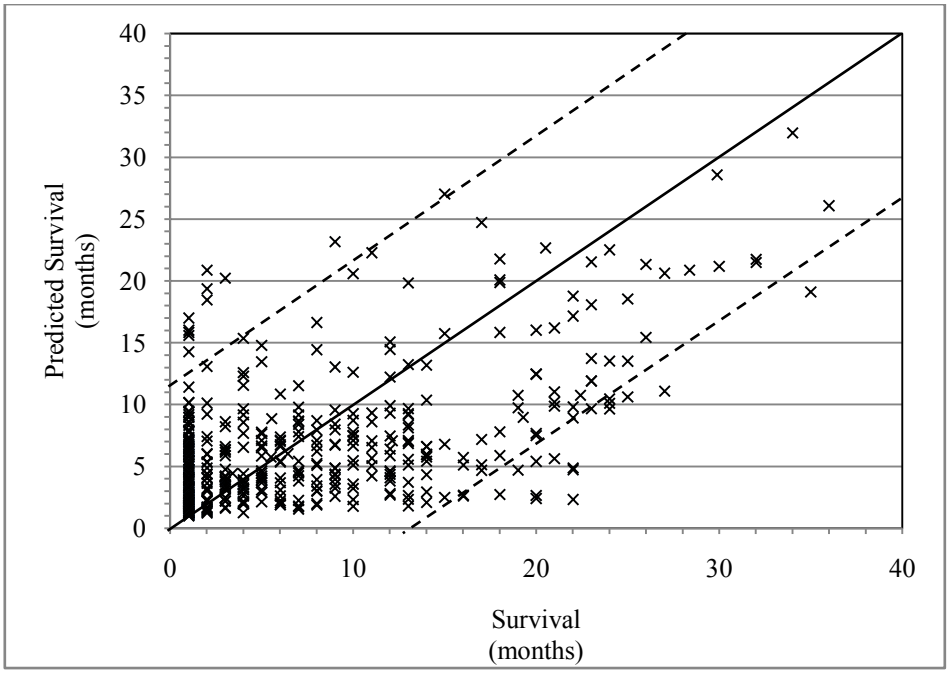

Figure 6-20 Final predicted survival for censored group (18-3-1, 90-10)

Partition	Actual Mean Survival (months)	Std Dev	CI (months)	Predicted Mean Survival (months)	Std Dev	CI (months)
70-30	13.15	28.77	0.07	6.13	24.62	0.06
80-20	13.15	28.77	0.07	6.17	24.69	0.06
90-10	13.15	28.77	0.07	7.55	25.28	0.06
Final 90-10	**6.86**	**23.07**	**0.06**	**6.26**	**22.50**	**0.06**

Table 6-36 Censored mean survival for all leukaemias

Variable	Transformation (s)	Frequency
age	linear	1.00
race	linear	0.99
	log	0.73
rcc	linear	0.80
hmb	linear	0.61
pc	linear	0.26
mnc	inv	0.53
CD5	logical	0.31
CD13	logical	0.94
CD19	logical	0.05
CD23	logical	0.61
CD56	rlogical	0.99
lambda	logical	0.75
chromo	linear	0.87
survival	log	output

Table 6-37 Prognostic factors for censored group

Network	R		ACC		CI (months)	
	T	V	T	V	T	V
18-8-1	0.44	0.20	0.83	0.43	23.00	54.00
23-2-1	0.47	0.36	0.74	0.87	31.00	19.00
19-3-1	0.38	0.09	0.74	0.84	33.00	22.00
23-4-1	0.34	0.33	0.73	0.80	33.00	23.00
18-20-1	0.69	0.23	0.81	0.64	24.00	37.00
mean	**0.46**	**0.24**	**0.77**	**0.72**	**28.80**	**31.00**
std dev	**0.14**	**0.11**	**0.05**	**0.18**	**4.92**	**14.61**
CI (months)	**0.00**	**0.00**	**0.00**	**0.01**	**0.14**	**0.41**

Table 6-38 Cross validation for censored group

6.2.4.1. ALL

There is an improvement in the R value of the general censored model from 0.61 to 0.76 but the CI is greater in this model at 16 months compared to 12 months in the above model and it is also greater than the 2-year and 3-year uncensored models. The common prognostic factors are age, race, mean corpuscular haemoglobin concentration, neutrophils, lymphocytes, chromosomes, and the CD markers CD7, CD8 and CD56. The high lymphocyte count and marker CD22 used for the diagnosis of ALL is common to all 3 ALL models. Age and race have proved to be significant prognostic factors in all three ALL models. The actual mean survival for this group is 11.01 months compared to 23.21 months for the 3-year ALL model and 20.79 months for the 2-year model. The predicted mean survival is 9.55 months compared to 20.99 for the 2-year model and 22.64 for the 3-year model. This is a large difference compared to the other two models but the data is more reliable for the previous two ALL models, thus more acceptable than the censored prediction. This low actual mean survival and predicted survival is acceptable as the censored model has a large number of patients with a short survival period due to the uncertainty in their true survival time. This is illustrated in the graphs where there are many patients in the 0-15 month range in both Figures 6-21 and 6-22, and since some of it is censored it also contributes to the low mean survival for ALL patients in this group. The mean survival CI is 0.01 months for the 2-year model compared to 0.11 months for the 3-year model, thus making the 2-year ALL model the favoured model for this leukaemia sub-type.

Partition	**Network**	**R**		**ACC**		**CI (months)**	
		T	**S**	**T**	**S**	**T**	**S**
70-30	22-16-1	0.68	0.30	0.82	0.70	24	33
80-20	20-6-1	0.56	0.29	0.78	0.72	28	33
90-10	19-21-1	0.72	0.38	0.79	0.65	23	33
Final 90-10	**19-21-1**	**0.76**	**0.83**	**0.83**	**0.88**	**16**	**15**

Table 6-39 Censored models for ALL

Variable	Transformation (s)	Frequency
age	tanh	0.96
race	linear	0.71
	log	1.00
mcv	linear	0.57
	tanh	0.71
mchc	rlogical	0.95
rcw	linear	0.85
np	linear	0.61
	exp	0.17
lc	linear	0.73
mnc	inv	0.87
blast	rt	0.19
CD3	log	0.57
CD7	logical	0.71
CD8	exp	0.91
CD20	logical	0.35
CD23	rlogical	0.75
CD56	rlogical	0.34
chromo	tanh	0.75
survival	log	output

Table 6-40 Prognostic factors for censored ALL

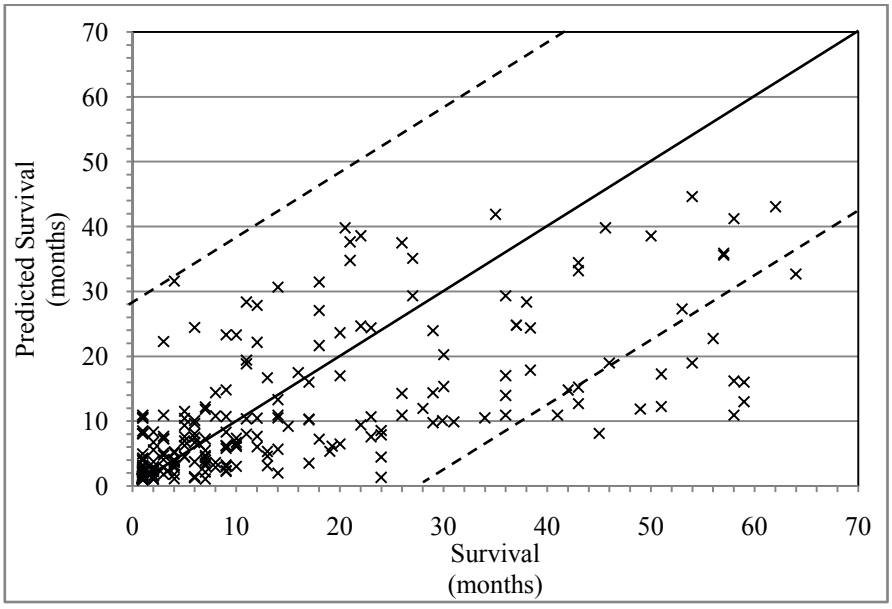

Figure 6-21 Initial predicted survival for censored ALL model (19-21-1, 90-10)

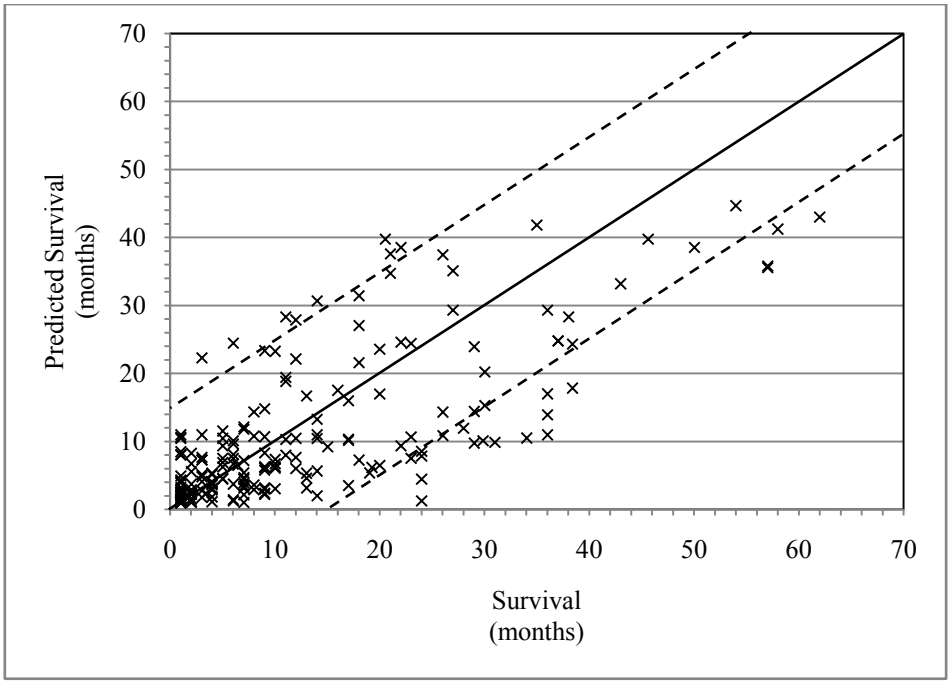

Figure 6-22 Final predicted survival for censored ALL model (19-21-1, 90-10)

Partition	Actual Mean Survival (months)	Std Dev	CI (months)	Predicted Mean Survival (months)	Std Dev	CI (months)
70-30	13.85	22.00	0.09	8.58	17.27	0.07
80-20	13.85	22.00	0.09	9.40	17.93	0.07
90-10	13.85	22.00	0.09	10.21	18.52	0.07
Final 90-10	**11.01**	**19.32**	**0.08**	**9.55**	**17.94**	**0.07**

Table 6-41 Mean survival for censored ALL

6.2.4.2. AML

The R value of 0.89 is lower than the 2-year model (R = 0.97) and the 3-year model (R = 0.99), with the CI of 4 months being the smallest in comparison to 7 months for the 2-year model and 10 months for this censored AML model. The prognostic factors common to all three models are age, race, haemoglobin, haematocrit, mean corpuscular volume, mean corpuscular haemoglobin, mean corpuscular haemoglobin concentration, and mean platelet volume, neutrophils, lymphocytes, markers CD7, CD34 and CD56, together with chromosomes. Age, white cell count, CD34 and

cytogenetics are used in the diagnosis of AML. The other diagnostic factors CD5 was considered by any of the models but HLA-DR (f = 0.52) was found to be significant for the 2-year model. Figure 6-22 illustrates the low prediction cluster in the 0-15 month range. The actual mean survival is 10.9 months while it is 16.86 for the 2-year model and 17.19 for the 3-year model. Once again the uncertainty resulting in short known survival times have contributed to this phenomenon. There are a few outlying patients who fall in the 18 month confidence interval but they have been retained for the model so that there is sufficient data for the building of the neural network. Since the data is censored these outlying patients can be expected. Overall the 3-year AML is the favoured model because of its low CI, accuracy and reliability of patient data.

Partition	**Network**	**R**		**ACC**		**CI (months)**	
		T	**S**	**T**	**S**	**T**	**S**
70-30	21-4-1	0.57	0.48	0.79	0.79	24	24
80-20	28-0-1	0.56	0.56	0.81	0.79	24	24
90-10	23-15-1	0.91	0.81	0.94	0.91	11	16
Final 90-10	**23-15-1**	**0.89**	**0.89**	**0.94**	**0.94**	**10**	**10**

Table 6-42 Censored AML model

Variable	Transformation (s)	Frequency
age	linear	0.81
	pwr2	0.93
race	linear	0.89
rcc	pwr2	0.25
hmb	linear	0.63
	pwr2	1.00
hmr	linear	0.93
mcv	linear	0.75
mch	logical	0.03
mchc	rlogical	0.17
pc	linear	0.15
mpv	pwr2	0.65
np	log	0.54
lc	exp	0.38
mnc	linear	0.32
blast	linear	0.65
CD5	logical	0.12
CD7	logical	0.54
CD13	rlogical	0.96
CD14	logical	0.93
CD20	logical	0.85
CD34	rlogical	1.00
chromo	log	1.00
survival	log	output

Table 6-43 Prognostic factors for censored AML

Partition	Actual Mean Survival (months)	Std Dev	CI (months)	Predicted Mean Survival (months)	Std Dev	CI (months)
70-30	11.9	13.64	0.07	7.51	7.16	0.04
80-20	11.9	13.64	0.07	7.66	7.57	0.04
90-10	11.9	13.64	0.07	10.76	11.43	0.05
Final 90-10	**10.9**	**12.11**	**0.06**	**10.30**	**10.86**	**0.05**

Table 6-44 Mean survival for censored AML

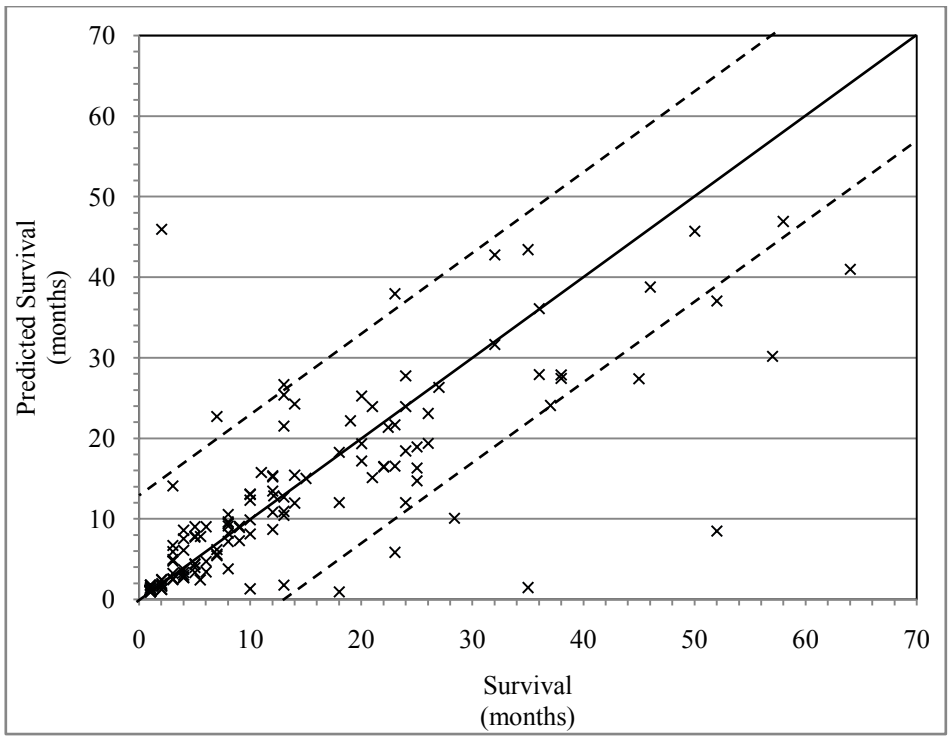

Figure 6-23 Initial predicted survival for censored AML model (23-15-1, 90-10)

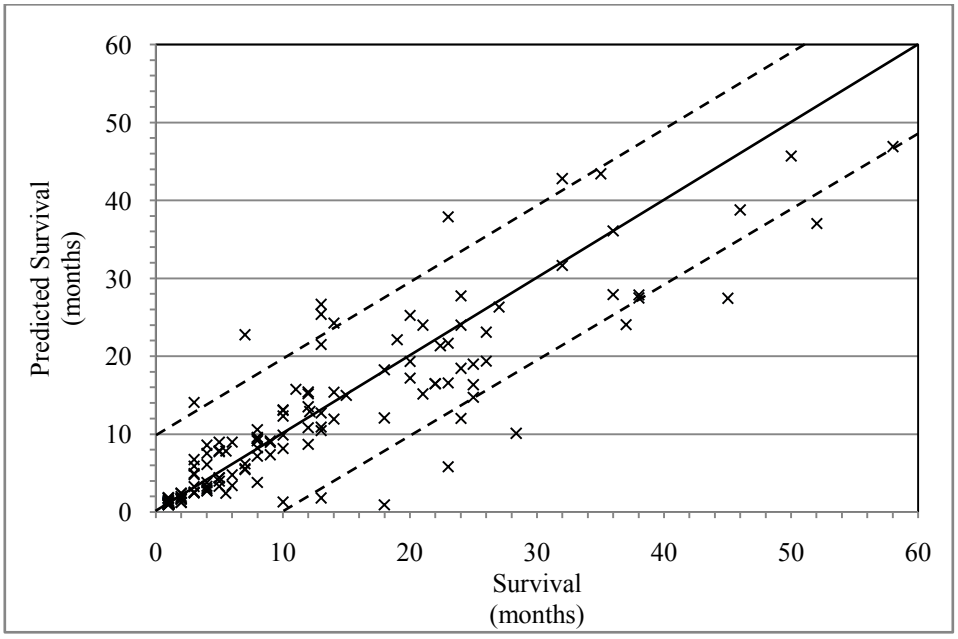

Figure 6-24 Final predicted survival for censored AML model (23-15-1, 90-10)

6.2.4.3. CML-CLL

The R value in this model is 0.96 (CI = 9 months, ACC = 0.98) compared to 0.98 (CI = 6 months, ACC = 1) for the 2-year model and 0.97(CI = 8 months, ACC = 1) for the 3-year model. The common prognostic factors are age, race, type, haematocrit, mean corpuscular volume, platelets, and the marker CD3. Age and race have proved to be significant prognostic factors in all three CML-CLL models. The actual mean survival for this group is 11.53 months compared to 35.23 months for the 3-year CML-CLL model and 30.76 months for the 2-year model. The predicted mean survival is 10.66 months compared to 30.55 for the 2-year model and 35.78 for the 3-year model. This is a large difference compared to the other two models but the data is more reliable for the previous two CML-CLL models, thus more acceptable than the censored prediction. This low actual mean survival and predicted survival can be expected as there are many patients with a lower than actual survival period which is due to the uncertainty of their true survival time. This is illustrated in the graphs where there are many patients in the 0-15 month range in both Figures 6-25 and 6-26. The mean survival CI is 0.14 for the 2-year model compared to 0.17 for the 3-year model and 0.09 for this model. The favoured model for the CML-CLL sub-type is the 2-year CML-CLL model because of its low CI and overall accuracy in its predictions.

Partition	**Network**	**R**		**ACC**		**CI (months)**	
		T	**S**	**T**	**S**	**T**	**S**
70-30	23-10-1	0.76	0.24	0.84	0.72	28	37
80-20	22-5-1	0.46	0.43	0.78	0.68	36	36
90-10	21-12-1	0.92	0.82	0.93	0.90	14	20
Final 90-10	**21-12-1**	**0.96**	**0.95**	**0.98**	**0.97**	**9**	**9**

Table 6-45 Censored CML-CLL models

Variable	Transformation (s)	Frequency
age	linear	0.93
race	log	0.72
type	logical	0.79
rcc	rt	0.89
hmr	linear	0.71
mcv	linear tanh	0.04 0.4
mchc	rlogical	0.1
pc	linear	0.83
wcc	linear inv	0.5 1.00
np	linear	0.25
lc	pwr2	0.33
mnc	inv	0.32
blast	pwr2	0.72
CD3	log	0.88
CD4	log	0.85
CD8	linear	0.33
CD10	logical	0.23
CD13	logical	0.88
CD33	logical	0.94
survival	log	output

Table 6-46 Prognostic factors for censored CML-CLL

Partition	Actual Mean Survival (months)	Std Dev	CI (months)	Predicted Mean Survival (months)	Std Dev	CI (months)
70-30	13.62	22.01	0.11	7.18	15.24	0.07
80-20	13.62	22.01	0.11	5.44	14.24	0.06
90-10	13.62	22.01	0.11	11.47	19.58	0.09
Final 90-10	**11.53**	**20.15**	**0.09**	**10.66**	**18.84**	**0.09**

Table 6-47 Censored mean survival for CML

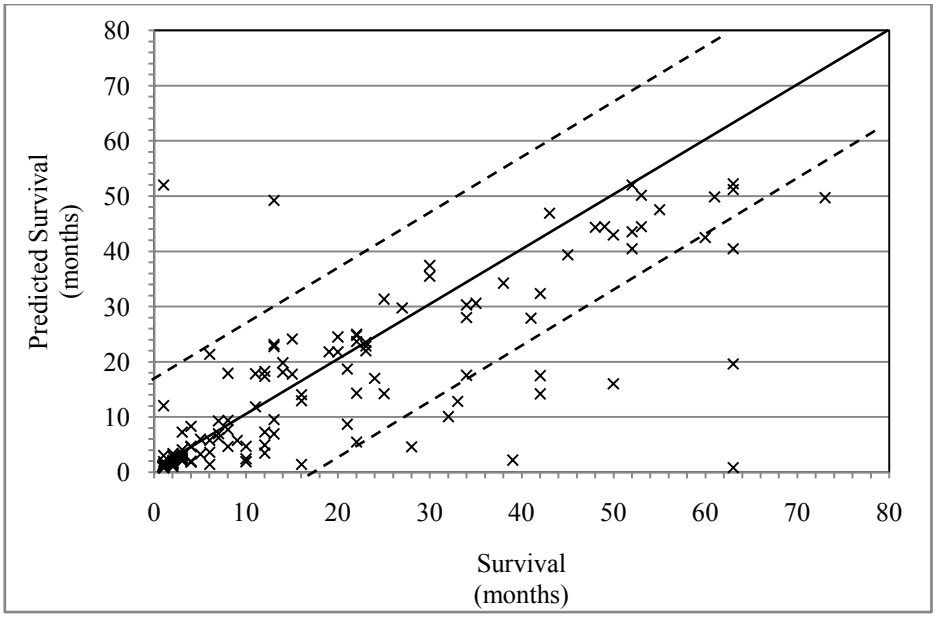

Figure 6-25 Initial predicted survival for censored CML-CLL (21-12-1, 90-10)

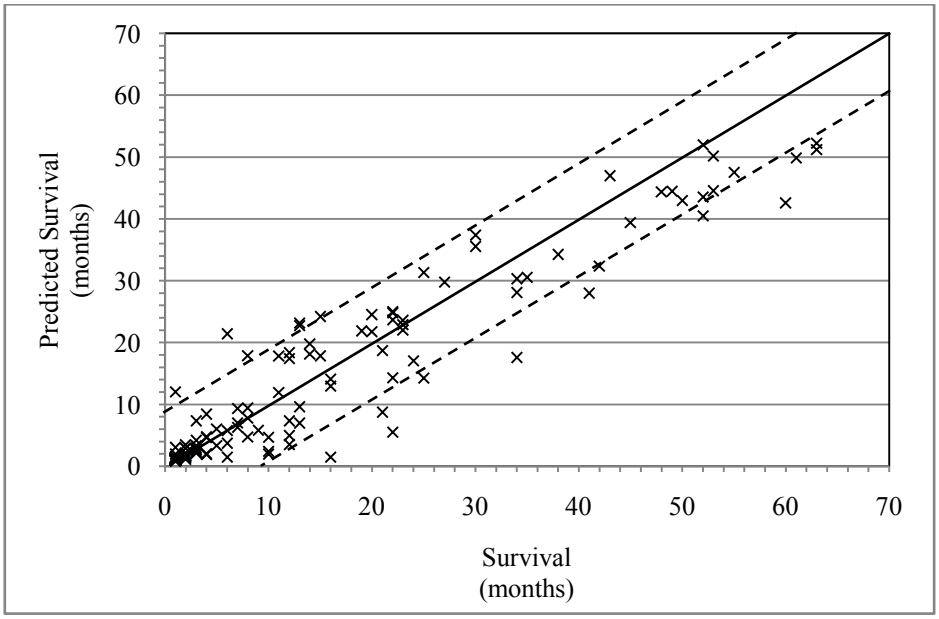

Figure 6-26 Final predicted survival for censored CML-CLL (21-12-1, 90-10)

6.3 Limitations in this research

The patient data recorded for this study was assumed to have all information correct, i.e. no incorrect diagnoses. All data was checked by an independent checker. Some patients‟ cytogenetic results were not confirmed as the sample was insufficient or the test was not successful, thus the normal male and female karyotype was used. The replacement of missing values by the average value does not give a true reflection of the group under study. Since records of patients are generally limited and there is also the censored group to consider, any additional patients can only give the study more credibility. It was necessary to include patients with missing values. This inclusion thus lent its measure of error into the model. Treatment protocols usually used are based on conventional therapy for all low risk patients and specific medication for high risk patients based on their blood and marrow analyses. There is a treatment protocol for each type of leukaemia. Similarly diagnosed patients are given the same drugs which are administered as per body mass index and can therefore not be used as a factor. In this research the general protocols for treatment of leukaemia have been assumed. It is not a variable in this study, but it should be. Some patients do not stick to the schedule, they leave the treatment program or default and some do not take their medication according to the prescription which can affect their survival. The treatment has to be monitored and witnessed in order to confirm that the patient has taken the prescribed dosage. There is no certainty that the “take home” medication is adhered to or even whether, if taken, it is the correct dosage. Therefore even though a treatment plan is set out for the patient, in order for a study to be valid there has to be unconditional proof that it was received as per the schedule. The models presented show the variables that are used to build the neural network, it does give any indications whether the individual factors affecting the patient‟s prognosis is favourable or not.

6.4. Summary

The predicted survival in the final models was well within the 95 % confidence bounds in many of the proposed models which indicate that the prediction is a good fit. The summary results are based on the training (T) data were the proposed models were all based on the 90-10 partition, LR=100, WD = 0.005 and the adaptive gradient descent learning rule was used for the genetic algorithm in PREDICT. The main prognostic factors listed in Table 6-48 below are based on those having a frequency > 80%, where frequency denotes the importance of the variable to the final recommended models.

Data Group	Network	R	ACC	CI	Prognostic Factors
2-year case study	24-18-1	0.87	0.89	14	age, mchc, np, CD13, CD20, CD56
2-year ALL	17-20-1	0.95	0.95	9	age, race, np, lc, CD13, CD56, chromo
2-year AML	19-11-1	0.97	1.00	7	age, mcv, mchc, lc, CD7, CD14, CD34
2-year CML-CLL	24-12-1	0.98	1.00	6	race, pc, mnc, CD3, CD4, CD34, lambda
3-year case study	21-6-1	0.86	0.79	19	age, race, mchc, pc, mpv, np, lc, CD5, CD13, CD20
3-year ALL	19-9-1	0.97	0.95	9	age, mchc, blast, CD8, CD22
3-year AML	20-12-1	0.99	1.00	4	age, mchc, np, lc, CD7, CD34
3-year CML-CLL	17-7-1	0.97	1.00	8	age, race, hmr, mch, mchc, pc, mpv, np, blast, CD3, CD13, lambda, chromo
Censored case study	18-3-1	0.61	0.94	12	age, race, rcc, CD13, CD56, chromo
censored ALL	19-12-1	0.76	0.83	16	age, race, mchc, rcw, mnc, CD8
censored AML	23-15-1	0.89	0.94	10	age, race, hmb, hmr, CD13, CD14, CD20, CD34, chromo
censored CML-CLL	21-12-1	0.96	0.98	9	age, rcc, pc, wcc, CD3, CD4, CD13, CD33

Table 6-48 Final recommended neural network models for leukaemia

For the combined groups the 2-year model has a higher accuracy and lower CI of 14 months thus making it the favoured model for this dataset. Since reliability of patient data is quite low after two years, this model can be periodically updated with new patient data, thereby maintaining a current working model for clinicians to use on an ongoing basis. The models performed better when predicting survival for the individual leukaemias. This is a logical outcome as each leukaemia is morphologically different, thus the analyses of blood and marrow samples will strongly affect the patient‟s survival. For ALL patient data the 3-year model has the best R value, making it the preferred choice for predicting survival. The 3-year AML is the favoured model because of its low CI, accuracy and reliability of patient data. The favoured model for the CML-CLL sub-type is the 2-year ALL model because of its low CI and overall accuracy in its predictions.

CHAPTER 7
CONCLUSIONS AND RECOMMENDATIONS

This thesis contains contributions to the development of neural network models for survival analysis of leukaemia patients. The prognostic factors that affect survival have also been determined by the neural networks. The comparisons of models were based on using combined groups of leukaemia patients and comparing them with individual groups of the sub-types of leukaemia, i.e. ALL, AML, CLL and CML. The performance of the 2-year model at predicting survival was consistently better than the other models on the basis of predicting as close as possible to the real data. Thus for prediction of survival over two years for the combined leukaemias and the sub-group CML-CLL, the best strategy would be to use the 2-year model. For the ALL and AML groups the 3-year models would be the favoured models. The chosen models also give an indication of the prognostic factors which contribute the most to the survival of the patient. These prognostic factors will aid clinicians in drawing up individual treatment strategies. The model should be fitted to the whole chosen dataset, with the number of hidden units (architecture) and the amount of weight decay chosen by cross validation. More patients will be treated at the hospital and more data will be received. The model should be retrained periodically on the updated information. The software code for the final proposed models are presented in Appendix G. This code can be used to predict survival of new patients if accepted into a clinical trial.

The results for factors that affect prognosis can be used to divide patients according to their diagnosis into multiple risk groups. This grouping may be used to develop strategies based on risk of individual patients, e.g. CD 34 expression can be used as a target for treatment as pharmacological companies are producing drugs that target these markers. This individualised treatment can only lead to a prolonged lifespan for each patient treated in this manner. This type of treatment may be more expensive initially but eventually once it becomes routine then only can it benefit future patients diagnosed with leukaemia. Accuracy in prediction of survival is vital. Over prediction of a patient's survival would put the high risk patients into the low risk group which usually receives conventional treatment that is not always successful for the high risk patients. Cytogenetics is normally used to determine the alternate therapy for the high risk groups. If their survival is correctly predicted they can be given the appropriate alternate therapy or become part of clinical trials for new treatment protocols. If there is an under prediction then the low risk patients may be categorised as high risk and put onto alternate therapies and clinical trials which normally are quite costly and time consuming as their has to be more manpower, a

higher degree of monitoring and sometimes specialised equipment and analysis may be required. Conventional therapy will apply to a low risk group and the cost would be minimised as it is a routine protocol. The right prediction of survival is therefore vital for all patients in terms of the individual knowing about their mortality and at the same time for the management and clinical staff who must have the proper patient management systems, allocate the relevant resources and ensure adequately trained clinicians are available to implement all appropriate protocols for the relevant risk groups.

The differing treatment patterns among the patients on whom the model must be based has other implications: firstly, past treatments to the extent they have been successful would have reduced the apparent prognostic effect of the factors and, secondly, new treatment practices may invalidate the model. Any such new treatment would only slowly be incorporated as sufficient follow-up became available. Studies are ongoing in cancer therapy. Xin *et al* (2006) carried out a survival study and analysis of prognostic factors on acute leukaemia patients. They looked at induction with ATRA and As_2O_3 rather than other induction therapy and found that the interval of less than 60 days to complete remission were found to be favourable factors for both overall survival and relapse-free survival. Changes of this type imply that care must be exercised when applying the models to new real world data. While the proposed models would be best to use for this prediction problem, this comparison has its limitations: there is no information about the survival experience, and would need retraining if, for example, the probability of survival in the first year became of interest. The type of treatment and the variables for each treatment have not been incorporated into these models. The current models have been based on the patients receiving conventional therapy for leukaemia based on their diagnosis.

Treatment plans based on selection of prognostic factors will depend on adequate requests for tests on diagnosis and for the analysis to be done promptly. Cytogenetics is one of the key factors that determine categorisation into the leukaemia sub groups. These tests should be standard tests on diagnosis for all patients so that they can be put into the appropriate risk group which will establish their course of treatment.

The clinical records need to be explicitly updated and all patient results should be available both in hardcopy files and electronically. Since survival studies are generally done over a number of years and normally start after the prescribed period there needs to be adequate information on patients for the relevant period under study. Well documented clinical records are essential in any research study and especially so in clinical trials where retrospective studies are generally adopted.

In order to use the data as input to the relevant software the range of values had to be scaled down to a limited range. The frequency of values in each of the subsets in the range was used as a guideline in selecting the scaled down values for each variable. The choice of replacement values could have an effect on the outcome of the models. The number of subsets of ranges can be increased for each variable. This allows for more variance among the data points, thus the proposed model would have trained over a wider range for each variable. This would ensure that new data points are well represented when the model is tested on these unknown values. Alternate methods could be tried by varying the scaled down values into various subsets and testing which system gives the best for a group of patients. A method of using probabilities of data being in multiple ranges should also be researched. The accuracy can then be used to compare the models for final adoption.

The prognostic factors predicted by the models in this study need to be validated with more uncensored patient data. The models should also be discussed with the clinicians to ensure that medically the mathematics has made the appropriate predictions which are of value to the patients. Thereafter these prognostic factors can be used as an initial screening for both diagnosis and survival prediction which will determine the course of treatment. Prognostic factors have been proposed based on the modelling results obtained in this study. Whether these factors contributed to a good or bad prognosis was not the aim of the study and neither could it be determined from the results obtained. These factors now need to be further researched to determine their effect on survival and whether it is favourable or not.

New prognostic factors should be fully standardised and their prognostic value validated in large prospective clinical trials before being used in routine clinical practice. Implementation in clinical trials using all possible prognostic parameters would be impractical if not impossible. Only the most important prognostic parameters – namely those with independent prognostic value which identifies a significant proportion of patients and those therapeutically relevant should be retained in the future, with clinical stages being used as benchmarks. Treatment decisions based on biological parameters can only be justified within clinical trials. Patients not included in a clinical trial should continue to be treated with conventional therapy.

This study should incorporate other centres and institutions in the country to get a larger dataset that will then provide sufficient numbers of uncensored patient data for a more reliable neural network that can then be used in a clinical trial together with the expertise of clinicians.

REFERENCES

Advani, A.S., Rodriguez, C., Jin, T., Jawde, A., Saber, W., Baz, R., Kalaycio, M., Sobecks, R., Sekeres, M., Tripp, B. and His, E., 2008. Increased C-kit intensity is a poor prognostic factor for progression-free and overall survival in patients with newly diagnosed AML. *Leukaemia Research*. 32:913:918.

Albreht, T., McKee, M. Alexe, D.M. Coleman, M.P. and Martin-Moreno, J.M., 2008. Making progress against cancer in Europe 2008. *European Journal of Cancer*. 44:1451:1456.

Altman, D.G., Lausen, B., Sauerbrei, W. and Schumacher, M., 1994. Dangers of using "optimal" cutpoints in the evaluation of prognostic factors. *Journal of National Cancer Institute.* 86(26):829:835.

Aouli, N. Eddabara, L. Macadre, J. and Morjani, H., 2005. Immunosuppressors and reversion of multidrug-resistance. *Critical reviews in Oncology/Hematology*. 56:61:70.

Bain, B.J., 2001. Overview. *Best Practice and Research Clinical Haematology*. 14(3):463:477.

Bene, M. C., 2005. Immunophenotyping of acute leukaemias. *Immunology Letters*. 98:9:21.

Beutler, E., Coller, B., Lichtman, M., Kipps, T. J and Seligsohn, U., 2001. *William's Hematology*. 6th Edition. New York: McGraw-Hill.

Biez, M., Ludwig, M., Mueller, B.U., Leibundgut, E.O., Ratschiller, D. and Solenthaler, M., 2005. Risk assessment in patients with acute myeloid leukaemia and a normal karyotype. *Leukaemia Research*. 11:1416:1424.

Biganzoli, E., Boracchi, P. and Marubini, E., 2002b. A general framework for neural network models on censored survival data. *Neural Networks.*15:209:218.

Biganzoli, E., Boracchi, P., Matiani, L. and Marubini, E. (1998a). Feed forward neural networks for the analysis of censored survival data: a partial logistic regression approach. *Statistics in Medicine*. 17:1169:1186.

Biondi, A. and Rambaldi, A., 1996. Molecular diagnosis and monitoring of acute myeloid leukaemia. *Leukaemia Research*. 20(10):801:807.

Black, D., 2005. *Neural network gaze tracking using web camera*. MSc thesis. Linkoping: University of Linkopings.

Bottaci, L., Drew, P.J., Hartley, H.E., Hadfield, M.B., Farouk, R., Lee, P.W.R. Macintyre, I.M.C., Duthie, G.S. and Monson, J.R.T., 1997. Artificial neural networks applied to outcome prediction for colorectal cancer patients in separate institutions. *The Lancet*. 350;469:472.

Chang, H., Salma, F., Yi, Q.L., Patterson, B., Brien, B. and Minden, M.D., 2004a. Prognostic relevance of immunophenotyping in 379 patients with acute myeloid leukaemia. *Leukaemia Research*. 28:43:48.

Chang, H., Yeung, J., Brandwein, J. and Yi, Q.L., 2007b. CD7 expression predicts poor disease free survival and post-remission survival in patients with acute myeloid leukaemia and normal karyotype. *Leukaemia Research*. 31:157:162.

Chase, A., Humtly, B.J. and Cross, N.C.P., 2001. Cytogenetics of chronic myeloid leukaemia. *Best Practice and Research Clinical Haematology*. 14(3):553:571.

Churchland, P.S. and Sejnowski, T.J., 1992. *The Computational Brain*. Cambridge: MIT Press.

Ciro, M., Bracken, A.P. and Helin, K., 2003. Profiling cancer. *Current Opinion in Cell Biology.*15:213:220.

Coebergh, J.W., Pastore, G., Gatta, G., Corazziari, I., Kamps, W. and the EUROCARE Working Group, 2001. Variation in survival of European children with acute lymphoblastic leukaemia, diagnosed in 1978-1992: the EUROCARE study. *European Journal of Cancer*. 37:687:694.

Cooper, G.M., 1996. *The Cell: A Molecular Approach*. Washington D.C.: ASM Press.

Cripe, L.D., 1997. Adult Acute Leukaemia. *Current problems in Cancer.* 21(1):4:64.

Cronje, M., Murdoch, M. and Smit, R., eds., 2003. *Reference techniques-Harvard method and APA style*. Johannesburg: Rand Afrikaans University.

Data and Analysis Centre for Software (DACS), 1992. *Technical report: Artificial neural Networks Technology*. Rome.

Desandes, E., Lacour, B., Sommelet, D., Danzon, A., Delafosse, P., Grosclaude, P., Mace-lesech, J., Maarouf, N., Marr, A., Raverdy-Bourdon, N., Tretarre, B. and Velten, M., 2006. Cancer survival among adolescents in France. *European Journal of Cancer*. 42:403:409.

DiRusso, S.M., Chahine, A.A., Sullivam, T., Risucci, D., Nealon, P., Cuff, S., Savino, J. and Slim, M., 2001. Development of a model for prediction of survival in pediatric trauma patients: Comparison of artificial neural networks and logistic regression. *Journal of Pediatric Surgery*. 37(7):1098:1104.

Dorfnner, G., 1996. Neural network for time processing. *Neural Network World*. 6(4):447:468.

Dunwoodie, R., 2001. *Inverse internal model control of an ethylene polymersiation reactor using artificial neural networks*. MSc. Durban: University of Natal.

Dutta, P. and Pratihar, D.K., 2007. Modeling of TIG welding process using conventional regression analysis and neural network-based-approaches. *Journal of Materials Processing Technology*. 184:56:68.

Farouzan, S. and Akbarzadeh, A., 2007. Short communication: Prediction of effect of thermo-mechanical parameters on mechanical properties and anisotropy of aluminum alloy AA3004 using artificial neural networks. *Materials and Design*. 28:1678:1684.

Gatta, G., Luksch, R., Coleman, M.P., Corazzi, I. and the EUROCARE working group. 2001. Survival from acute non-lymphocytic leukaemia (ANLL) and chronic myeloid leukaemia (CML) in European children since 1978: a population based-study. *European Journal of Cancer*. 37:695:702.

Ghia, P., Ferreri, A.J.M. and Caligaris-Cappio, F., 2007. Chronic lymphocytic leukaemia. *Critical Reviews in Oncology/Hematology*. 64:234:246.

Glassman, A.B. and Hayes, K.J., 2005. Shorter communication: The value of fluorescence in situ hybridization in the diagnosis and prognosis of chronic lymphocytic leukaemia. *Cancer Genetics and Cytogenetics.* 158:88:91.

Goldstein, B., Faeder, J.R., Hlavacek, W.S., Blinov, M.L., Redondo, A. and Woksy, C., 2001. Modeling the early signaling events mediated by FcεRI. *Molecular Immunology*. 38:1213:1219.

Hallbook, H., 2005. *Acute lymphoblastic leukaemia in adult patients: Studies of prognostic factors, treatment results and in vitro cellular drug resistance.* Uppsala: Uppsala University.

Han, M., Snow, P.B., Epstein, J.I., Chan, T.Y., Jones, K.A., Walsh, P.C. and Partin, A.W., 2000. A neural network predicts progression for men with Gleason score 3+4 versus 4+3 tumors after radical prostatectomy. *Urology*. 56(6):994:999.

Harmening, D.M. ed., 2002. *Clinical Hematology and Fundamentals of Hemostasis.* 4th edition. Philadelphia: FA Davis Company.

Harrison, C.J. and Secker-Walker, L.M., 1998. The importance of cytogenetics and associated molecular techniques in the management of patients with leukaemia. *Clinical Oncology.* 10:255:261.

Hassoun, M.H., 1995. *Fundamentals of artificial neural networks*. Cambridge: MIT Press.

Hayashi, Y. and Setiono, R., 2002. Combining neural network prediction for medical diagnosis. *Computers in Biology and Medicine*. 32:237:246.

Haykin, S., 1994. *Neural networks: A comprehensive foundation*. New Jersey: Macmillan College Publishing.

Heden, B., Ohlin, H., Rittner, R. and Edenbrandt, L., 1997. Acute myocardial infarction detected in the 12-Lead ECG by artificial neural networks. *Circulation*. 96(6)1798:1802.

Hinton, G.E., 1989. Connectionist learning procedures. *Artificial Intelligence*. 40:185:234.

Hirama,T. and Koeffler, H.P., 1995. Role of the Cyclin-Dependent Kinase Inhibitors in the Development of Cancer. *The Journal of the American Society of Hematology.* 86(3):841:854.

Hornik, K., Stincombe, M and White, H., 1989. Multilayer feed forward networks are universal approximators. *Neural networks,* 2:359:366.

Inamdar, K.V. and Bueso-Ramos, C.E., 2007. Review article: Pathology of chronic lymphocytic leukaemia: an update. *Annals of Diagnostic Pathology*. 11:363:389.

Jacobs, P., 1997. Myelodysplasia and the leukaemias. *Disease a Month*. 43(8):509:585.

Jennings, C. D. and Foon, K. A., 1997. Recent Advances in Flow Cytometry: Application of the diagnosis of Hematologic Malignancy. *Blood: The Journal of the American Society of Hematology.* 90(8):2863:2892.

Kebriaei, P., Anastasi, J. and Larson, R. A., 2003. Acute lymphoblastic leukaemia: diagnosis and classification. *Best Practice and Research Clinical Haematology*. 15(4):597:621.

Laffey, J.G., Tobin, E., Boylan, J.F. and McShane, A.J., 2003. Assessment of a simple artificial neural network for predicting residual neuromuscular block. *British Journal of Anaesthesia*. 90(1):48:52.

Lapuerta, P., Azen, S.P. and LaBree, L.,1995. Use of artificial neural networks in predicting the risk of coronary artery disease. *Computers and Biomedical Research*. 28:38:52.

Lee, G.R., Foerster, J., Lukens, J., Paraskevas, F., Greer, J.P. and Rodgers, G.M., eds., 1999. *Wintrobe's Clinical Hematology*. Volume 1. 10th edition. Baltimore: Williams and Wilkins.

Leukaemia. Available from http://member.rivernet.com.au/balehirs/Bishyp7cancerleukaemia.html. Accessed 15 June 2008.

Leukaemia. Available from http://www.startoncology.net. Accessed 15 June 2007.

Lewis, S.M., Bain, B.J. and Bates, I., 2001. *Dacie and Lewis Practical Haematology.* 9th edition. London: Churchill Livingstone.

Lisboa, P.J.G., (ed)., 1992a. *Neural networks: current applications*. Liverpool: Chapman and Hall.

Lisboa, P.G.J., 2002b. A review of evidence of health benefit from artificial neural networks in medical intervention. *Neural networks*. 15:11:39.

Lisboa, P.G.J. and Taktak, A.F.G., 2006c. The use of artificial neural networks in decision support in cancer: a systematic review. *Neural Networks*. 19:408:415.

Litzow, M.R., 2000. The therapy of relapsed acute leukaemia in adults. *Blood Reviews*. 18:39:63.

Llobera, J., Esteva, M., Rifa, J., Benito, E., Terrasa, J., Rojas, C., Pons, O., Catalan, G. and Avella, A., 2000. Terminal cancer: duration and prediction of survival time. *European Journal of Cancer*. 36:2036:2043.

Marucci, G., Mrozek, K. and Bloomfield, C.D., 2005. Molecular heterogeneity and prognostic biomarkers in adults with acute myeloid leukaemia and normal cytogenetics. *Current Opinions in Haematology*. 12:68:75.

Mathieson, M.J., 1998. *Ordinal models and predictive models in pattern recognition*. PhD thesis. Oxford: University of Oxford.

Medical dictionary. Available from http://www.wikapaedia.com. Accessed 16 December 2008.

Mehdi, B., Stacey, D. and Harauz, G., 1994. A hierarchical neural network assembly for classification of cervical cells in automated screening. *Analytical Cellular Pathology*. 7:171:180.

Minsky, M.L., 1969 and Lapert, S.A., *Perceptrons*. Cambridge: MIT Press.

Moore, H. and Li, N.K., 2004. A mathematical model for chronic myelogenous leukaemia (CML) and T cell interaction. *Journal of Theoretical Biology*. 227:513:523.

Moreno, C. and Montserrat, E., 2008. New prognostic markers in lymphocytic leukaemia. *Blood reviews*. 22:211:219.

Morris, A.J., Montague, G.A. and Willis, M.J., 1994. Artificial neural networks: Studies in process modeling and control. *Transactions of the Institute of Chemical Engineering.* 72(A):3:19.

Mrozek, K., Heinonen, K. and Bloomfield, C.D., 2001a. Clinical importance of cytogenetics in acute myeloid leukaemia. *Best Practice and Research Clinical Haematology*. 14(1):19:47.

Mrozek, K., Heerama, N.A. and Bloomfield, C.D., 2004b. Cytogenetics in acute leukaemia. *Blood Reviews*. 18:115:136.

Narotram, N.K., 1999. *Development of Artificial Neural Networks for the Modelling and Control of Chemical Processes*. MSc Thesis. Durban: University of Natal.

Neilson, J.R., Auer, R., Whie, D., Bienz, N., Waters, J.J., Whittaker, J.A., Milligan, D.W. and Fegan, C.D., 1997. Deletions at 11q identify a subset of patients with typical CLL who show consistent disease progression and reduced survival. *Leukaemia*. 11:1929:1932.

Neuralware, 2003. *Neural networks PREDICT:User guide*. Pasadena: Neuralware.

Nimer, S.D., 2008. Is it important to decipher the heterogeneity of "normal karyotype AML"? *Best Practice and Research Clinical Haematology.* 21(1):43:52.

Ohno-Machado, L., 1997a. A comparison of Cox proportional hazards and artificial neural networks for medical prognosis. *Computers in Biology and Medicine*. 27(1):55:65.

Ohno-Machado, L., 2001b. Modelling medical prognosis: Survival analysis techniques. *Journal of Biomedical Informatics.* 34:428:439.

Parenthetical referencing (n.d.). Available from http://en.wikipedia.org/wiki/Harvard_referencing (accessed 28 October 2008).

Perea, G., Domingo, A., Villamor, N., Palacios, C., Junca, J., Torres, P., Llorente, A., Fernandez, C., Tormo, M., Queipo de Llano, M.P., Bargay, J., Gallart, M., Florensa, l., Vivanos, P., Marti, J.M., Font, L., Berlanga, J., Esteve, J., Bueno, J., Ribera, J.M., Brunet, S., Sierra, J. and Nondedeu, J.F., 2005. Adverse prognostic impact of CD36 and CD2 expression in adult de novo acute myeloid leukaemia patients. *Leukaemia Research*. 29:1109:1116.

Pineda, F.J., 1988. Generalisation of back propagation to recurrent and higher order neural networks. *Neural Information Processing Systems*. D.Z.Anderson edition. New York: American Institute of Physics. 602:611.

Plasschaert, S.L.A., Kamps, W.A., Vellenga, E., de Vries, E.G.E. and de Bont, E.S.J.M., 2004. Tumour review: Prognosis in childhood and adult acute lymphoblastic leukaemia: a question of maturation? *Cancer Treatment Reviews*. 30:37:51.

Raspadori, D., Lauria, F., Ventura, M.A., Rondelli, D. Visani, G., de Vivi, A. and Tara, S., 1997. Incidence and prognostic relevance of CD34 expression in acute myeloid leukaemia: analysis of 141 cases. *Leukaemia Research*. 21:603:607.

Ravdin, P.M. and Clark, G.M., 1992. A practical application of neural network analysis for predicting outcome of individual breast cancer patients. *Breast Cancer Research and Treatment*. 22:285:293.

Rieseberg, M.; Kasper, C.; Reardon, K.F. and Scheper, T., 2001. Flow Cytometry in Biotechnology. *Applied Microbiology and Biotechnology*. 56:350:360.

Ripley, R., 1998. *Neural network for breast cancer prognosis.* PhD thesis. University of Oxford.

Rumelhart, D.E., Hinton, G.E. and Williams, R.J., 1986. Learning representation by back propagating errors. *Nature.* 323:533:536.

Schwartzer, G., Vach, W., & Schumacher, M., 2000. On the misuse of artificial neural networks for prognostic and diagnosis classification in oncology. *Statistical Medicine*. 19:541:551.

Shepstone, N.M., 1997. *An investigation into the feasibility of using artificial neural networks to control turbogenerators.* MSc thesis. Durban: University of Natal.

Singh, V., Tathavadkar, V., Rao, S.M. and Raju, K.S., 2007. Predicting the performance of submerged arc furnace with varied raw material combinations using artificial neural networks. *Journal of Materials Processing Technology*. 183:111:116.

Sterjovski, Z., Pitrun, M., Nolan, D., Dunne, D. and Norrish, J., 2007. Artificial neural networks for predicting diffusible hydrogen content and cracking susceptibility in rutile flux-cored arc welds. *Journal of Materials Processing Technology*. 184:420:427.

Stetlet-Stevenson, M.S., 2003. Flow cytometry in lymphoma diagnosis and prognosis: useful? Bets Practice and Research Clinical Haematology. 16(4):583:597.

Takahashi, H., Murase, Y., Kobayashi, T. and Honda, H., 2007. New cancer diagnosis modeling using boosting and projective adaptive resonance theory with improved reliable index. *Biochemical Engineering Journal.* 33:100:109.

Taktak, A., Antolini, L., Aung, M., Boracchi, P., Campbell, I., Damato, B., Ifeachor, E., Lama, N., Lisboa, P., Setzkorn, C. Stalbovskaya, V. and Biganzoli, E., 2006. Double-blind evaluation and benchmarking of survival models in a multi-centre study. *Computers in Biology and Medicine*. doi: 10.1016/j.compbiomed.2006.10.001.

Tavernier, E., Le, Q.H. Elhamri, M. and Thomas, X., 2003. Salvage therapy in refractory acute myeloid leukaemia: prediction of outcome based on analysis of prognostic factors. *Leukaemia Research*. 27:205:214.

Theeuwun, M., Kappen, B. and Neijt, J., 1995. Neural network analysis to predict treatment outcome in patients with ovarian cancer. *In Proceedings Session 5, Medicine, International Conference on Artificial Neural Networks.* Paris.

University library guide to Harvard style of referencing. 2008. Available from http://libweb/anglia/ac.uk/referencing/harvard.htm. Accessed 12 September 2008.

Van Bockstaele, F., Verhasselt, B. and Philippe, J., 2008. Prognostic markers in chronic lymphocytic leukaemia: a comprehensive review. *Blood reviews*. doi: 10.1016/j.blre.2008.05.003.

Venayagamoorthy, G.K., 1998. *Implementation of a continually online trained artificial neural network control for a turbogenerators*. MSc Thesis. Durban: University of Natal.

Venkataraman, A., 2004. *Comparison of neural network and logistic regression model to predict medical outcome*. MSc thesis. Cincinnati: University of Cincinnati.

Wahlin, A., Hornsten, P. and Jonsson, H., 1991. Remission rate and survival in acute myeloid leukaemia: impact of selection and chemotherapy. *European Journal of Haematology*. 46:240:247.

Warner, B. and Misra, M., 1996. Understanding neural network as statistical tools. *The American Statistician.* 50(4):284:293.

Werbos, P.J. (1974a). *Beyond regression: new tools for prediction and analysis in the behavioral sciences.* PhD thesis. Cambridge: University of Harvard.

Werbos, P.J., 1990b. Back propagation through time: what it does and how to do it? *Proceedings of IEEE.* 78(10):1550-1560.

www. member.rivernet.com.au/balehirs/Bishyp7cancerleukaemia.html. Accessed 15 June 2007.

www.neurosurgery-umms.com. Accessed 15 June 2007.

Wyatt, J.C. and Altman, D.G., 1995. Commentary: Prognostic model: clinically useful or quickly forgotten. *British Medical Journal*. 311:1539:1541.

Xiang, A., Lapuerta, P., Ryutov, A., Buckley, J. and Azen, S., 2000. Comparison of the performance of neural network methods and Cox regression for censored survival data. *Computational Statistics and Data Analysis*. 34:243:257.

Xin, L., Wan-jun. S., Zeng-jun, L., Yao-zhong, Z., Yun-tao, L., Yan, L., Chang-chun, W., Qiao-chuan, L., Reb-chi, Y., Ming-zhe, H., Jian-xiang, W. and Lu-gui, Q., 1994. A survival study and prognostic factors analysis on acute promyelocytic leukaemia at a single centre. *Leukaemia Research*. doi: 10.1016/j.leukres.2006.07.028.

Xu, L., Xing, J., Wei, S., Zhang, Y. and Long, R., 2007. Optimization of heat treatment technique of high-vanadium high-speed steel based on back propagation neural networks. *Materials and Design*. 28:1425:1432.

Yanada, M., Jinnai, I., Takeuchi, J., Ueda, T., Miyawaki, S., Tsuzuki, M., Hatta, Y., Usui, N., Wada, H., Morii, T., Matsuda, M., Kiyoi, H., Okada, M., Honda, S., Miyazaki, Y., Ohno, R. and Naoe, T., 2006. Clinical features and outcome of T-lineage acute lymphoblastic leukaemia in adults: A low initial white blood cell count, as well as a high count predict decreased survival rates. *Leukaemia Research*. doi:10.1016/j.leukres.2006.08.004.

Zhu, S., Lee, S., Hargrove, S.K. and Chen, G., 2007. Prediction of combustion efficiency of chicken litter using an artificial neural network approach. *Fuel*. 86:877:886.

Zhu, Y.H., Nagamune, T., Endo, I. and Linko, P., 1994. Neural networks in state variable prediction during start-up of chemostat culture for ethanol production by *Saccharomyces Cerevisisiae* yeast. *Transactions of the Institute of Chemical Engineers*. 72(C):135:142.