**IDENTIFYING STOCK WINNERS AND LOSERS:
A MULTILAYERED ARTIFICIAL NEURAL NETWORK
APPROACH**

by

Atul Srivastava, B.E.(Comp.Eng.)

A Thesis Submitted to
The Faculty of Graduate Studies and Research
In Partial Fulfilment of The Requirement for The Degree of

Master of Management Studies

School of Business
Carleton University
Ottawa, Canada

April 22, 1996
© Copyright
1996, Atul Srivastava

National Library
of Canada

Acquisitions and
Bibliographic Services Branch

395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale
du Canada

Direction des acquisitions et
des services bibliographiques

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file *Votre référence*

Our file *Notre référence*

The author has granted an irrevocable non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of his/her thesis by any means and in any form or format, making this thesis available to interested persons.

L'auteur a accordé une licence irrévocable et non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de sa thèse de quelque manière et sous quelque forme que ce soit pour mettre des exemplaires de cette thèse à la disposition des personnes intéressées.

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

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

ISBN 0-612-13862-3

Name ATUL SRIVASTAVA

Dissertation Abstracts International is arranged by broad, general subject categories. Please select the one subject which most nearly describes the content of your dissertation. Enter the corresponding four-digit code in the spaces provided.

FINANCE

SUBJECT TERM

0508

U·M·I

SUBJECT CODE

Subject Categories

THE HUMANITIES AND SOCIAL SCIENCES

COMMUNICATIONS AND THE ARTS

Architecture 0729
Art History 0377
Cinema 0900
Dance 0378
Fine Arts 0357
Information Science 0723
Journalism 0391
Library Science 0399
Mass Communications 0708
Music 0413
Speech Communication 0459
Theater 0465

EDUCATION

General 0515
Administration 0514
Adult and Continuing 0516
Agricultural 0517
Art 0273
Bilingual and Multicultural 0282
Business 0688
Community College 0275
Curriculum and Instruction 0727
Early Childhood 0518
Elementary 0524
Finance 0277
Guidance and Counseling 0519
Health 0680
Higher 0745
History of 0520
Home Economics 0278
Industrial 0521
Language and Literature 0279
Mathematics 0280
Music 0522
Philosophy of 0998
Physical 0523

Psychology 0525
Reading 0535
Religious 0527
Sciences 0714
Secondary 0533
Social Sciences 0534
Sociology of 0340
Special 0529
Teacher Training 0530
Technology 0710
Tests and Measurements 0288
Vocational 0747

LANGUAGE, LITERATURE AND LINGUISTICS

Language
General 0679
Ancient 0289
Linguistics 0290
Modern 0291
Literature
General 0401
Classical 0294
Comparative 0295
Medieval 0297
Modern 0298
African 0316
American 0591
Asian 0305
Canadian (English) 0352
Canadian (French) 0355
English 0593
Germanic 0311
Latin American 0312
Middle Eastern 0315
Romance 0313
Slavic and East European 0314

PHILOSOPHY, RELIGION AND THEOLOGY

Philosophy 0422
Religion
General 0318
Biblical Studies 0321
Clergy 0319
History of 0320
Philosophy of 0322
Theology 0469

SOCIAL SCIENCES

American Studies 0323
Anthropology
Archaeology 0324
Cultural 0326
Physical 0327
Business Administration
General 0310
Accounting 0272
Banking 0770
Management 0454
Marketing 0338
Canadian Studies 0385
Economics
General 0501
Agricultural 0503
Commerce-Business 0505
Finance 0508
History 0509
Labor 0510
Theory 0511
Folklore 0358
Geography 0366
Gerontology 0351
History
General 0578

Ancient 0579
Medieval 0581
Modern 0582
Black 0328
African 0331
Asia, Australia and Oceania 0332
Canadian 0334
European 0335
Latin American 0336
Middle Eastern 0333
United States 0337
History of Science 0585
Law 0398
Political Science
General 0615
International Law and Relations 0616
Public Administration 0617
Recreation 0814
Social Work 0452
Sociology
General 0626
Criminology and Penology 0627
Demography 0938
Ethnic and Racial Studies 0631
Individual and Family Studies 0628
Industrial and Labor Relations 0629
Public and Social Welfare 0630
Social Structure and Development 0700
Theory and Methods 0344
Transportation 0709
Urban and Regional Planning 0999
Women's Studies 0453

THE SCIENCES AND ENGINEERING

BIOLOGICAL SCIENCES

Agriculture
General 0473
Agronomy 0285
Animal Culture and Nutrition 0475
Animal Pathology 0476
Food Science and Technology 0359
Forestry and Wildlife 0478
Plant Culture 0479
Plant Pathology 0480
Plant Physiology 0817
Range Management 0777
Wood Technology 0746
Biology
General 0306
Anatomy 0287
Biostatistics 0308
Botany 0309
Cell 0379
Ecology 0329
Entomology 0353
Genetics 0369
Limnology 0793
Microbiology 0410
Molecular 0307
Neuroscience 0317
Oceanography 0416
Physiology 0433
Radiation 0821
Veterinary Science 0778
Zoology 0472
Biophysics
General 0786
Medical 0760

EARTH SCIENCES

Biogeochemistry 0425
Geochemistry 0996

Geodesy 0370
Geology 0372
Geophysics 0373
Hydrology 0388
Mineralogy 0411
Paleobotany 0345
Paleoecology 0426
Paleontology 0418
Paleozoology 0985
Palynology 0427
Physical Geography 0368
Physical Oceanography 0415

HEALTH AND ENVIRONMENTAL SCIENCES

Environmental Sciences 0768
Health Sciences
General 0566
Audiology 0300
Chemotherapy 0992
Dentistry 0567
Education 0350
Hospital Management 0769
Human Development 0758
Immunology 0982
Medicine and Surgery 0564
Mental Health 0347
Nursing 0569
Nutrition 0570
Obstetrics and Gynecology 0380
Occupational Health and Therapy 0354
Ophthalmology 0381
Pathology 0571
Pharmacology 0419
Pharmacy 0572
Physical Therapy 0382
Public Health 0573
Radiology 0574
Recreation 0575

Speech Pathology 0460
Toxicology 0383
Home Economics 0386

PHYSICAL SCIENCES

Pure Sciences

Chemistry
General 0485
Agricultural 0749
Analytical 0486
Biochemistry 0487
Inorganic 0488
Nuclear 0738
Organic 0490
Pharmaceutical 0491
Physical 0494
Polymer 0495
Radiation 0754
Mathematics 0405
Physics
General 0605
Acoustics 0986
Astronomy and Astrophysics 0606
Atmospheric Science 0608
Atomic 0748
Electronics and Electricity 0607
Elementary Particles and High Energy 0798
Fluid and Plasma 0759
Molecular 0609
Nuclear 0610
Optics 0752
Radiation 0756
Solid State 0611
Statistics 0463

Applied Sciences

Applied Mechanics 0346
Computer Science 0984

Engineering

General 0537
Aerospace 0538
Agricultural 0539
Automotive 0540
Biomedical 0541
Chemical 0542
Civil 0543
Electronics and Electrical 0544
Heat and Thermodynamics 0348
Hydraulic 0545
Industrial 0546
Marine 0547
Materials Science 0794
Mechanical 0548
Metallurgy 0743
Mining 0551
Nuclear 0552
Packaging 0549
Petroleum 0765
Sanitary and Municipal System Science 0790
Geotechnology 0428
Operations Research 0796
Plastics Technology 0795
Textile Technology 0994

PSYCHOLOGY

General 0621
Behavioral 0384
Clinical 0622
Developmental 0620
Experimental 0623
Industrial 0624
Personality 0625
Physiological 0989
Psychobiology 0349
Psychometrics 0632
Social 0451

The undersigned recommend to the Faculty of Graduate Studies
and Research acceptance of the thesis

**" IDENTIFYING STOCK WINNERS AND LOSERS:
A MULTILAYERED ARTIFICIAL NEURAL NETWORK APPROACH "**

submitted by Atul Srivastava, B.E.(Comp. Eng.)
in partial fulfilment of the requirements for
the degree of Master of Management Studies

Thesis Supervisor

Director, School of Business

Carleton University
Ottawa, Ontario

April 22, 1996

Abstract

The artificial neural networks represent an exciting new technology with wide scope for potential financial applications ranging from routine credit assessment operations to the driving of large scale portfolio management strategies. This study is an attempt to identify stock winners and losers using this technology. Two sets of attributes are used for the network training. The first set is modelled along the lines of attributes that are used by Value Line Investment Survey, whereas the second set of attributes is a combination of the first set and some attributes taken from the anomaly literature.

The evidence presented in this study suggests that a blind, black box approach to artificial neural networks may not be very useful. Some pragmatism and careful knowledge of data is required for good results. A design that takes into consideration issues such as, sample size requirements, choice of attributes, presence of influential data points in the input data and use of classical techniques, such as multiple discriminant analysis, to enhance the network performance, may result in a potentially good system; it can even compare favourably with the classical methods, but with some caveats.

Acknowledgements

I would like to extend my sincerest thanks to my supervisor, Professor Vijay Jog for his on-going valuable help and guidance. I would also like to thank my other committee members - Professor Roland Thomas and Professor Wojtek Michalowski, whose suggestions and comments have been very helpful in shaping this thesis into its present form.

Special thanks to my brother and his family for their love and motivation throughout the entire MMS programme and this research.

I would like to take this opportunity to specially thank my wife, Namita for her patience and loving support that has been a constant source of inspiration during this research.

My parents also deserve my gratitude for their love and encouragement in all these years. It is to them that I dedicate this thesis as a token of my appreciation.

Table of Contents

Acceptance	ii
Abstract	iii
Acknowledgement	iv
Table of Contents	v
List of Tables	vii
List of Figures	viii
List of Appendices	ix
 Chapter 1: Introduction	 1
1.1 Background	1
1.2 The Proposed study	2
 Chapter 2: Literature Review	 5
2.1 Expert Systems	5
2.2 Decision Support Systems	6
2.3 Neural Networks	7
2.4 Business Failure Prediction	10
2.5 Debt Risk Assessment	12
2.6 Security Market Applications	14
2.7 Attribute Selection	23
2.7.1 Attributes Modelled along the lines of VL Ranking System	23
2.7.2 Attributes Taken From Anomaly Literature	26
 Chapter 3: Research Design	 28
3.1 Research Question	28
3.2 Research Model	28
3.3 Population	29
3.3.1 Data	29
3.4 Attribute Definitions	31
3.4.1 Current Year's Relative Earnings	31
3.4.2 Relative Earnings and Prices Compared to Past Eight Quarters	31
3.4.3 Price Momentum	32
3.4.4 Quarterly Earnings Momentum	32
3.4.5 Unexpected Quarterly Earnings	33
3.4.6 Stock Price Level	33
3.4.7 Firm Size	33
3.4.8 Price to Sales Ratio	34
3.4.9 Price to Book Ratio	34
3.4.10 Annualized Returns	35

3.5 Sample Selection Criteria	35
Chapter 4: Methodology	37
4.1 Creating Sample of Stock Winners and Losers	37
4.2 Creating Neural Network Model	41
4.2.1 Architecture of the Neural Network	42
4.2.2 Training	43
4.3 Investigating Applicability of ANN for Identifying Stock Winners and Losers	44
4.3.1 Preliminary Analysis	47
4.3.2 Main Analysis	50
Chapter 5: Empirical Results	54
5.1 Results of the Preliminary Analysis	54
5.1.1 Implications of the Preliminary Analysis	60
5.2 Results of the Main Analysis	60
5.3 Results of the Out-of-Sample Test	72
5.4 Comparisons with the Previous Studies	77
5.5 Conclusions	78
Chapter 6: Benefits, Limitations and Suggestions for Future Research	81
6.1 Benefits	81
6.2 Limitations	81
6.2 Suggestions for Future Research	82
References	84
Glossary	92

List of Tables

Table 1: Previous Studies in Business Failure Prediction	11
Table 2: Previous Studies in Debt Risk Assessment	13
Table 3: Previous Studies in Security Market Applications In General	15
Table 4: Previous Studies in Security Market Applications Similar to the Current Study	19
Table 5: Previous Studies on Performance of VL Ranks	25
Table 6: Previous Studies on Price Related Anomalies	27
Table 7: Distribution of Firms in TSE Industry Groups (quintile #1) and Winner and Loser (quintile #5) Portfolios Along With the Returns of the Corresponding Quintile	41
Table 8: Steps followed in the Preliminary Analysis	48
Table 9: Steps followed in the Main Analysis	51
Table 10: Mean Classification Accuracy off ANN Across 10 Random Samples, each for ATTR01 and ATTR02	54
Table 11: Results of Two-Group MANOVA on 1993 First Quarter for ATTR01	55
Table 12: Results of Two-Group MANOVA on 1993 First Quarter for ATTR02	56
Table 13: Mean Classification Accuracy of ANN and Discriminant Analysis for ATTR01 and ATTR02	58
Table 14: Classification Accuracy off Four Layer ANN for ATTR01 and ATTR02	60
Table 15: Mean Classification Accuracy of ANN and Discriminant Analysis for Samples with 80 Observations	61
Table 16: Mean Classification Accuracy of ANN and Discriminant Analysis for Samples with 100 Observations	62
Table 17: Mean Classification Accuracy of ANN and Discriminant Analysis for Samples with all Observations	63
Table 18: Change in Classification Accuracy of ANN and Discriminant Analysis for Samples with 80, 100 and all Observations	65
Table 19: Mean Classification Accuracy of ANN and Discriminant Analysis for Samples with 80 Observations for Five Different Random Samples	67
Table 20: Residual Statistics for the Third Quarter of 1993	68
Table 21: Residual Statistics for the Third Quarter of 1993 Without Influential Points	70
Table 22: Mean Classification Accuracy of ANN and Discriminant Analysis for Samples with 80 Observations Without Influential Data Points	71
Table 23: In-sample and Out-of-Sample Classification Accuracies of ANN and Discriminant Analysis for Sample with 20 Observations	74
Table 24: In-sample and Out-of-Sample Classification Accuracies of ANN and Discriminant Analysis for Sample with 60 Observations	75

Table 25: In-sample and Out-of-Sample Classification Accuracies of ANN and Discriminant Analysis for Sample with 80 Observations	76
Table 26: Comparison of Current Study with Previous Studies	77

List of Figures

Figure 3.1: Research Model	29
Figure 4.1: Creating Sample of Stock Winners and Losers	39
Figure 4.2: Network Architecture	42
Figure 4.3: The Entire Methodological Process	46
Figure 5.1: Sequence Plot of Mahalanobis Distance with Influential Data Points	69
Figure 5.1: Sequence Plot of Mahalanobis Distance without Influential Data Points	70

APPENDICES

Appendix A: Stock Groupings According to the Stock Guide Database	88
Appendix B: Overview of MATLAB Neural Network Toolbox	89
Appendix C: Learning in Neural Networks	90

1. INTRODUCTION

1.1 BACKGROUND:

In recent years, considerable interest has been shown in the application of new financial decision making tools emerging from different areas to decisions in corporate finance and investment fields. One of the areas that has received special attention is artificial intelligence which includes decision support systems, expert systems and neural networks. Various applications of these decision support systems have been suggested for corporate finance and portfolio managers. However, top level financial decision makers may have obtained very little direct benefit from systems mainly due to the highly unstructured information environments of financial decisions (Coats, 1988). The applications of these decision support systems for stock selection, for example, often requires continual review and refinement of rules as economic conditions evolve. However, each such refinement may require modifying the knowledge base, which is generally time consuming, costly and therefore leads to disillusionment.

In this context, Artificial Neural Networks¹ (ANN henceforth) are claimed to offer distinct advantages as they can find patterns in this highly unstructured information environment and thus, seem to be better suited for financial applications. Artificial neural networks can be defined as "highly simplified models of the human nervous system, exhibiting abilities such as learning, generalization and abstraction" (Hawley *et*

¹ Other synonyms used for artificial neural networks are artificial neural systems, neural networks, electronic neural networks, or simply neural nets. For an overview on neural networks, see Section 2.3.

al., 1990). A neural network not only accumulates, stores and recognizes patterns of knowledge based on experience, but it also constantly re-trains itself as the environment of examples evolves. While the concept of such systems is not new, only recently have technological advances in hardware and software technologies made artificial neural systems a viable alternative for many financial applications. Typical applications in finance have included financial simulation, prediction, evaluation, credit approval, assessing lending/bankruptcy risk, security asset portfolio management, pricing of IPOs, identification of arbitrage opportunities, technical analysis, fundamental analysis etc. (Hawley *et al.*, 1990). A typical neural network exercise for stock selection, for example, may involve training the network to understand the relationships between certain attributes that have a direct influence on future stock performance and the actual performance.

1.2 THE PRESENT STUDY:

This study attempts to identify stock winners² and losers³ using artificial neural network technology. The main reason for choosing ANN was because of its perceived potentials for such applications as shown in Hawley *et al.*, 1990. The ANN for this study is designed using MATLAB Neural Network Toolbox. The study, due to its exploratory and empirical nature, follows a set of sequential steps. The first step is to investigate the stability of the ANN each year from 1989-1993. For that purpose, the

² These are the stocks that provide superior returns, defined later.

³ These are the stocks that provide worst returns, defined later.

network is trained using two sets of attributes. The first set of attributes is modelled along the lines of the Value Line Investment Survey⁴ (VL henceforth) ranking system. The second set of attributes is a combination of the first set and some more attributes that are taken from the "anomaly literature"⁵. The study investigates the effect of varying sample size, architecture and the number of attributes on the performance of ANN. In addition to this, the study also compares the performance of ANN with a standard classical technique - linear multiple discriminant analysis. This comparison is necessary to evaluate the benefits and limitations of ANN over existing well known methodologies. The study uses data from TSE 300/Western and Stock Guide databases over a period from 1990-1993.

The results of this study should be of a significant interest to practitioners, who may be able to draw insights about the applications and limitations of ANN for portfolio management, to individuals who can use it to enhance their analysis of investment alternatives, and to academicians who will gain further knowledge about this new tool for financial decision making.

The remainder of the thesis is organized as follows: The next chapter provides a review of literature on artificial intelligence techniques for financial applications, chapter 3 outlines the research design followed by the methodology used for this study in chapter 4. Chapter 5 presents the empirical results and the conclusions of the study.

⁴ Value Line Investment Survey is one of the world's largest investment advisory services that provides a wide range of financial information on common stocks, including a weekly ranking of stocks into five groups based on projected relative price performance over the next twelve months. Value Line calls these ranks as timeliness ranks.

⁵ For details see, Section 2.7.2

Finally, chapter 6 presents the benefits and limitations of the study along with the suggestions for the future research.

2. LITERATURE REVIEW

Since this study attempts to apply ANN - a concept in the domain of computer science, to stock classification - a domain of financial decision making, it is necessary to review related concepts in both finance and computer science to put this study into context. Thus, the purpose of this chapter is to present a literature survey of artificial intelligence techniques for financial applications. To start with, a brief discussion on expert systems and decision support systems is presented in section 2.1 and 2.2 respectively. This is followed by a discussion on ANN in section 2.3. Sections 2.4, 2.5 and 2.6 present the research that has been done on the application of neural network technology for business failure prediction, debt risk assessment, and security market applications respectively. As this study uses attributes modelled along the lines of VL ranking system and from "anomaly literature", a brief literature review on each is presented in section 2.7 in order to provide necessary context for the choice of attributes.

2.1 EXPERT SYSTEMS:

In early 1980s, researchers became aware of the significance of Expert System technology-an application area emerging from the field of artificial intelligence, to business decisions. Two fundamental conclusions emerged from this research: (1) it is possible to simulate expert problem solving; and (2) this problem solving can be explained by the expert system technology. Examples of the use of such financial expert systems and decision support systems in commercial use in mid 1980's can be found in

Silverman (1987). However, other research on applying this technology to managerial problems, and financial problems in particular, demonstrated some shortcomings as well. These limitations arose for two reasons. First, expert systems are reasoning systems and are not particularly good at dealing with numbers and performing numeric computations. Second, they are not easily integrated into a decision making environment dealing with ill-structured decisions (for a discussion on expert system limitations, see Coats, 1988). Given such limitations, the usage of term "expert system" is now being replaced by less ambitious terms, such as "knowledge-based system" or "intelligent assistant"⁶.

2.2 DECISION SUPPORT SYSTEMS:

A second important concept coming out of the body of knowledge with regard to the use of computer systems for problem solving and decision making, is the concept of Decision Support Systems (DSS henceforth). The basic objective of a DSS is to provide information in a given domain of application by means of analytical decision models and access to databases, in order to support a decision maker in making decisions effectively in complex and ill structured tasks (Klein & Methlie, 1990). Decision support systems also have some drawbacks. One of the pertinent criticisms of the DSS concept was made by Winograd & Flores (1986). According to them, the potential problems with DSS are as follows: (1) orientation to choosing, (2) assumption of relevance, (3) unintended transfer of power, (4) unanticipated effects, (5) obscuring responsibility, and (6) false

⁶ This statement does not intend to undermine the capabilities of expert systems because even as a mere assistant to an expert decision maker, such a system can offer a variety of advantages over its human counterpart.

belief in objectivity.

2.3 NEURAL NETWORKS:

Another area of artificial intelligence that has gained a lot of popularity recently is ANN. An ANN is a type of system which tries to mimic the knowledge acquisition and organization skills of the human brain. It consists of an array of many simple processors called cells or **neurons**. Neurons, which are abstractions of biological neurons, are the basic processing elements of an ANN and could be simulated and trained using appropriate software. These neurons are interconnected with each other via links called **connections**. In a sense, it is simply a highly parallel computer with the difference in the way the knowledge is represented. Instead of encoding it as data and storing it in a memory bank, the ANN represents knowledge in the strength of connections between neurons. The strength of these connections is called their **weights** (like correlation coefficient in classical statistics). The neurons in an ANN are arranged in **layers**. Layers could be thought as groups of neurons arranged in hierarchical fashion. The group of neurons that receive inputs directly from outside the network is called **input layer**. The group of neurons whose output is passed to the world outside the network is called **output layer**. In addition to these two layers, there may be one or more intermediate layer(s) of neurons that neither receive inputs from outside the network nor sends outputs outside the network. This layer is called **hidden layer** and its purpose is to enhance network learning. The input to a neuron is the weighted sum of all the inputs coming into the neuron. This input is compared with the neuron's

threshold activation level. Whenever the value meets the threshold level, the neuron transmits its output to its neighbour. This output of a neuron depends on the **transfer function** of the neuron. A transfer function could be defined as a transformation that maps a neuron's inputs to its outputs. Another important component of an ANN is the learning mechanism. **Learning**⁷ is the process by which the network adjusts its weights to achieve some desired behaviour. This process may involve presenting an input information vector⁸ paired with a desired output or target vector⁹. The objective of this process is to let ANN discover the pattern underlying the relationship between the input and the target vector.

One of the important considerations while developing an ANN based application is the network architecture. While deciding the network architecture, two basic issues arise. The first is related to the choice of the number of hidden layers. The second issue is related to the number of neurons that should be there in the hidden layer(s). The literature provides very little guidance for addressing any of these issues. In order to decide the number of hidden layers, a single hidden layer is generally thought to be a good starting point because most of the problems can be solved with a single hidden layer (Casimir, 1993). For deciding the number of neurons in the hidden layer(s) there is no universal rule which can provide an answer. In general, the guidelines suggested by the studies indicate that a network with four neurons in the hidden layer performs reasonably well (Swales and Yoon, 1991 and Kimoto *et al.*, 1990).

⁷ For more information on learning, see Appendix C.

⁸ See glossary.

⁹ See glossary.

ANNs offer several advantages over the classical expert or knowledge-based systems. They are listed as follows:

- (1) An ANN does not require a pre-defined knowledge base.
- (2) Changes in the problem environment does not require reprogramming because the system automatically re-trains itself. Since the network learns in the new environments, training can occur continuously over its useful life, and can occur concurrently with the deployment of the network.
- (3) Since there are many processing neurons, each with primarily local connections, damage to a few neurons does not bring the system to a halt.
- (3) An ANN is able to perform reasonably well even if the information is incomplete, ambiguous (noisy), or partially erroneous.

Thus, ANNs represent an exciting new technology with wide scope for potential financial applications ranging from routine credit assessment operations to the driving of large scale portfolio management strategies. Some of these applications have resulted in dramatic increase in productivity. Moreover, recent innovations in the hardware and software technologies and improvements in our understanding of the strengths and weaknesses of ANNs, other forms of machine learning and human decision making processes, are now resulting in commercially successful systems like the one used by Sears Mortgage Corporation [see Smith (1991)].

In the context of financial applications which require predictions based on certain attributes, ANN has been applied in at least the following three categories:

- Business Failure Prediction
- Debt Risk Assessment
- Security Market Applications

Following sections discuss the research that has been done in these areas.

2.4 BUSINESS FAILURE PREDICTION:

Most of the research that has been done in this area has the principal focus of developing and modelling the bankruptcy prediction process using ANN approach. The task of predicting bankruptcy of a company can be posed as a classification problem: given a set of classes (bankrupt and non-bankrupt) and a set of input data vectors, the task is to assign each input data vector to one of the classes. In order to accomplish this process, robust and reliable model for bankruptcy prediction is required. Deriving an appropriate prediction model using classical statistical techniques require the assumption of a certain functional form for relating dependent attributes to independent attributes. When the assumed functional form is not correct, the statistical techniques merely confirm that, but does not predict the right functional form. ANNs do provide a more general framework for determining relationships in the data and do not require the specification of any functional form.

Some of the studies that have been done in this area are presented in Table 1 on the next page followed by their brief details.

Table 1: Previous Studies in Business Failure Prediction

Study (Year)	Architecture	Learning Algorithm	Results
Odom & Sharda (1990)	Two layer ¹⁰	Backpropagation ¹¹	81 % accuracy achieved.
Raghupati, Schkade, & Raju (1991)	Three layer fully connected	Backpropagation	86% accuracy achieved.
Coleman, Graettinger, & Lawrence (1991)	Two layer	Backpropagation	99 % accuracy achieved for predicting failed banks.
Tam & Kiang (1992)	Two & three layers	Backpropagation	Three layer n/w had lowest misclassification rate.
Salchenberger, Cinar & Lash (1992)	Two layer feed forward n/w	Backpropagation	Up to 92.5 % accuracy achieved.

Odom and Sharda (1990) performed an analysis of ratios using discriminant analysis and an ANN and subsequently compared their performances. Their results showed that for bankrupt firms, the network was able to achieve an accuracy of up to 81.5% compared to 70.4% for multiple discriminant analysis (MDA, henceforth). For the non-bankrupt firms, network achieved an accuracy of up to 85.7% compared to 89.3% for MDA. Raghupati, Schkade, and Raju (1991) used 13 financial ratios selected from the set of ratios, that have been proven useful and have traditionally been used for

¹⁰ Different authors count number of layers differently, but generally layers are counted by counting the number of intermediate layers. For example, a three layer network would have one input layer, two hidden layers and one output layer. For simplicity, same convention is adopted throughout this thesis.

¹¹ See Appendix C or glossary.

bankruptcy prediction, to train a two layered fully connected network using back-propagation learning algorithm. The network achieved an accuracy of 86%. Coleman *et al.* (1991) discussed an application developed by NeuralWare to predict banks that were certain to fail within a year. A set of eleven ratios were used as input to a two layer network that was trained using backpropagation. The network achieved an accuracy of up to 99% to predict failed banks. Tam and Kiang (1992) empirically showed that ANNs offer better predictive accuracy than discriminant analysis, logit, k Nearest Neighbour and ID3. The two layer ANN had the best performance with the lowest total misclassification rate. Salchenberger *et al.* (1992) developed a ANN to forecast probability of failure of thrift institutions using a set of five financial attributes as inputs. Results show that accuracy of up to 92.5% was achieved compared to 87.8 % for Logit.

These papers indicate that ANNs can not only be used in this area but they also compare favourably with the traditional techniques previously used.

2.5 DEBT RISK ASSESSMENT:

The research in this area mainly focuses on two areas - assigning ratings to the bond issues and mortgage risk assessment. The task of assigning bond ratings can be posed as a classification problem: given a set of classes and a set of input data instances, each described by a suitable set of features, assign each input data instances to one of the classes. Traditionally multiple regression analysis has been used but has shown limited success even after considering 35 attributes. As mentioned in the previous section, this

supports the premise that in the absence of well-defined domain models, the success of standard mathematical/statistical techniques is limited. On the other hand, mortgage underwriting at first glance, would seem to be an unlikely candidate for the application of ANNs. This is because this problem might seem to be best suited for a rule based system, as the methodology applied by underwriters is thought to be well defined. However, the difficulty in getting underwriting "experts" to specify the rules they use in making their more subtle decisions, and the high economic value of reducing the number of bad risks presently accepted by underwriters, make the application of ANNs potentially very valuable.

Table 2 presents some of the studies that have been done in this area followed by their brief details.

Table 2: Previous Studies in Debt Risk Assessment

Study (Year)	Architecture	Learning Algorithm	Results
Dutta & Shekhar (1988)	Two & three layers	Backpropagation	88.3% accuracy achieved with 2 layers.
Collins, Ghosh, & Scotfield (1988)	Nine coupled n/w's in 3x3 arrangement	?	Up to 96% accuracy achieved.
Surkan & Singleton (1990)	Two & three layers	Backpropagation	Up to 88% accuracy achieved.
Reilly, Collins, Scotfield, & Ghosh (1991)	Three layer RCE n/w	?	Up to 95% accuracy achieved.

Dutta and Shekhar (1988) used an ANN to assign bond ratings to different industrial bond issues. Ten financial attributes were used for predicting bond ratings. The performance of two-layered and three-layered networks were subsequently compared with regression analysis and was found that the network achieved an accuracy of up to 88.3% for two layered network compared to 64.7% for regression analysis. Collins *et al.* (1988) tried to mimic the underwriting skills of human mortgage insurance underwriters by training a RCE¹² network with an architecture of nine coupled networks in a three-by-three arrangement, on their judgements. The network achieved an accuracy of up to 96%. Surkan and Singleton (1990) demonstrated that ANNs with either single or pairs of hidden layers can provide up to 88% correct classification. Reilly *et al.* (1991) analyzed how well risk prediction has fared for the mortgage portfolio by training a three-layer feed-forward RCE network. The results of the study show that for a certain percentage of the applications (10%), it is possible to predict with 95% accuracy the loans, if granted, would go delinquent in payment.

2.6 SECURITY MARKET APPLICATIONS:

Another area of ANN usage is in attempting to extract nonlinear regularities from economic time series. This possibility of decoding previously undetected regularities in asset price movements has elicited a great deal of interest. The prediction of future stock prices involves the interaction of many attributes, making prediction very difficult and

¹² Restricted Coulomb Energy is an equilibrium model that shares the property of unrestricted storage density.

complex. Several studies have been done based on the fluctuations of the common stock prices as well as on the stock market performance. Based upon the similarities and dissimilarities with the current study, the research in this area can be classified into two groups. The first group comprises of the studies that have been done in this area in general, whereas the second group comprises of the studies that are similar to the current study. Table 3 summarizes the research that has been done in the security market application area in general followed by a detailed discussion of each study.

Table 3: Previous Studies in Security Market Applications In General

Study (Year)	Architecture	Learning Algorithm	Data	Result
Hlbert White (1988)	Three layer feed forward n/w	Backpropagation	IBM daily stock returns	Poor out of sample forecasting.
Kimoto <i>et al.</i> (1990)	Two layer hierarchical n/w	Backpropagation	Weekly index returns	Correlation of 0.991 achieved with teaching data.
Kamijo & Tanigawa (1990)	Four layer recurrent n/w	Backpropagation	Sharp price rises	Accuracy of 93.8% achieved.
Swales & Yoon (1992)	Three layer	Backpropagation	Qualitative information	Success rate of 77.5% achieved compared to 65% for MDA.
Olson & Chou (1993)	?	Backpropagation	Index returns	Poor out of sample forecasting.

One of the first papers by White (1988) proposed that it is conceivable that previously undetected regularities exist in the historical asset price data and that such regularities may yet persist. His paper illustrates the search for such regularities using an ANN using the case of IBM daily common stock returns. Daily data from CRSP tapes, for a period of 1,000 days for the training set & two samples of data for 500 days before & after the training period for testing set, were used. To investigate the possibility that ANN can detect nonlinear regularities inconsistent with the simple efficient markets hypothesis, the author trained a three layer feed-forward network with five input and five hidden nodes. Backpropagation was used as the network learning algorithm. The trained network was tested using a pre and post sample of 500 days. The correlation between the predicted and the actual returns for the presample and the postsample periods were 0.0751 and -0.0699 respectively. This suggests that the ANN was not able to achieve good out-of-sample prediction power. However, the results are very much driven by the fact that the scope of the study was very limited in use of number of attributes for training purpose. Better network prediction power may have been obtained by elaborating the network to accommodate additional inputs, e.g. volume, stock prices etc.

Kimoto *et al.* (1990) designed a stock buying and selling prediction system with the inputs consisting of several technical and economic indexes. The prediction system consisted of four ANNs that were trained to learn the relationships between various technical and economic indexes and the timing for when to buy and sell stocks. The authors suggested that stock prices are determined by time-space patterns of economic

indexes such as foreign exchange rates and interest rates. The prediction system used a moving average of weekly average data of each index for minimizing influence due to random walk. The timing for when to buy and sell used as teaching data was weighted sum of weekly returns. The prediction was done by simulation while moving the objective learning and prediction periods. The output of the prediction system was the average output from the networks. The basic architecture of the networks used in the system consisted of two layers. These layers were completely connected to form a hierarchical network. Prediction was done for 33 months of data and results show that buying and selling according to the prediction system made a greater profit than the buying and holding. As the timing for when to buy and sell stocks is not linear, so classical statistical methods are not effective for creating a model. The results also show that the ANN learned data well enough to show a very high correlation coefficient of 0.991 between the predictions and teaching data compared to 0.543 for multiple regression analysis.

Kamijo and Tanigawa (1990) used an ANN to recognize the beginning of sudden stock price rise from the time sequence of the stock prices. A recurrent ANN with four layers was used for this purpose. Sixteen stock price patterns, when prices suddenly rose, were identified and the network was trained by presenting these patterns in a random sequence. After the training, it was attempted to recognize sixteen patterns from the past three years and it was confirmed that the patterns were accurately recognized in 15 out of 16 experiments, which gives an accuracy of 93.8%.

The prediction of stock price performance involves the interaction of many

attributes making prediction very difficult and complicated. In addition to the quantitative information, qualitative information, while not as easily interpreted, may also have an effect on investment value. Swales and Yoon (1992) used the president's letter to shareholders to differentiate between stocks that will provide superior future returns from those that will provide poor future returns. A content analysis was done on the president's letter and nine themes were identified. A data set was constructed by recording the frequency of allusion to each theme and the percentage of the letters devoted to it. The nine themes formed the input to a non-linear nine-attribute network with 9 nodes in the input layer, 3 nodes in the hidden layer, and 2 nodes in the output layer. Each output node represented one of the two possible outcomes - good stock price performance or poor stock price performance. Performances of network and MDA were subsequently compared and was found that ANN model did better with a mean success rate of 77.5% compared to 65% for MDA.

Olson and Chou (1993) examined the cross correlation patterns in index returns for three major U.S. stock exchanges. The nonlinear relationships were investigated using an ANN and regression analysis and their performances were subsequently compared. The hypothesis examined was that if cross correlations are measured with a lag from the large stocks on the NYSE to the smaller AMEX and OTC stocks, lagged NYSE returns and lagged squared NYSE returns should help predict future AMEX and OTC returns, but not vice versa. Five regression models were constructed to account for seasonality, significance of exchange's own lagged returns etc. The equivalents of these regression models were estimated using an ANN model with a single layer using

the backpropagation algorithm. The results show that both regression and ANNs have little out of sample prediction power. These results are consistent with White (1988). As recommended by White (1988), the authors also agree that networks should be trained and evaluated using profits and losses rather than the forecast error.

The studies that have been done in the area of security market applications that are similar to the current study are summarized in Table 4 on the next page followed by their details.

Table 4: Previous Studies in Security Market Applications that are Similar to the Current Study

Study (Year)	Architecture	Learning Algorithm	Data	Results
Wong, Wang, Goh & Quek (1992)	Three Layer	Backpropagation	Accounting Variables	Results satisfactory.
Kryzanowski, Galler & Wright (1993)	?	Boltzman Machine	Accounting Information	Accuracy of 72% achieved for +ve/-ve returns.

Wong, Wang, Goh, and Quek (1992) proposed an intelligent stock selection system, a combination of an ANN and an expert system, in order to encourage financial managers and researchers to recognize the enormous potential of an integrated system that combines merits of individual artificial intelligence tools. To test this system, data on beta, price-to-book ratio, three-year average price/book, price-to-earnings ratio, asset turnover, three-year average asset turnover, earnings growth, dividend discount model

value, profit margin, three-year average profit margin, and current-year return was collected for 800 companies for a period of three years. The network was trained using backpropagation learning algorithm and its task was to mix various types of information supplied to it in order to generate a value that can be used to rate stock performance. The test case demonstrated that a correlation of 0.84 was obtained between the actual and predicted returns which suggests the possible usefulness of such systems.

Kryzanowski, Galler and Wright (1993) examined the ability of ANNs using historical and current accounting and macroeconomic data to discriminate between the stocks providing superior future returns and those providing inferior future returns. The ANN was trained to learn the relationship between firm's stock return one year in the future and the most recent four years of financial data on the firm and its industry as well as data on seven macroeconomic attributes. Five industry ratios were used as benchmarks for analyzing each firm, and 14 firm specific financial ratios were used for analyzing the trends of each firm. The network was trained using pattern classification algorithm of the Boltzman Machine¹³, which employs a technique for stochastic optimization, called simulated annealing¹⁴. Results of their study shows that network achieved an overall accuracy of 72% for the two and three category classification.

Compared to the previous studies, the present study provides some significant enhancements to our understanding of the performance of artificial neural network for identifying stock winners and losers. Key areas such as, robustness of ANN to different sample sizes, choice of attributes and presence of influential data points in the sample,

¹³ See Glossary

¹⁴ See Glossary

that are not addressed by previous studies, were identified as a result of the literature review and are investigated in the present study. In addition to this, the study also presents a comparison of the performance of ANN and classical discriminant analysis and investigates the out-of-sample classification power of ANN.

Two important conclusions can be drawn from the literature review done thus far. First, it provides the necessary justification for the use of ANN for this study as the ANN seems to provide favourable and in certain cases, better results when compared to the classical methods.

Second, it is evident that training a network based on returns data as the only input parameter may not be sufficient and a significant out-of-sample explanatory power can not be achieved. Some of the limitations of the studies that have employed this approach are outlined as follows.

- The network is generally very simple and limited in the number of inputs. Additional inputs should have been used to achieve better prediction power. Any elaborations may have to be supported by massive infusion of data for training.
- As in White (1988), the optimization methods used were essentially local. Although the final weight values were determined as giving the best performance over a range of different starting values, there is no guarantee that a global maximum was found. Therefore, methods such as simulated annealing would be preferable.
- The networks for market trading purposes should be evaluated and trained to maximize trading profits rather than to minimize forecast error.

On the other hand, the studies that have used other information in addition to the returns data, were able to achieve excellent results in favour of ANNs. Thus, it is evident from this discussion that the choice of input parameters is critical for good network performance and a rationale is necessary for choosing attributes for network training. In order to establish a rationale for the choice of attributes used in this study, a literature survey was done and is summarized in the following sections.

2.7 ATTRIBUTE SELECTION:

The purpose of this section is to provide a rationale for the choice of the attributes used in the study. Two sets of attributes are used for training the network. The first set of six attributes is modelled along the lines the VL ranking system. These attributes are referred to as attribute set ATTR01 and are outlined in section 2.7.1 along with a discussion on the performance of VL timeliness ranks which is directly related to the predictive power of these attributes. The second set of ten attributes is a combination of ATTR01 and four attributes that are taken from the "anomaly literature". These attributes are referred to as attribute set ATTR02. These attributes are outlined in section 2.7.2.

2.7.1 Attributes Modelled Along The Lines Of VL Ranking System :

Value Line Investment Survey is the world's largest published advisory service that has been providing a range of information on approximately 1700 U.S. securities since 1965 and on 70 Canadian securities since 1991. The information they provide includes performance predictions by assigning timeliness ranks from 1 to 5 to the securities in their universe, with rank 1 associated with excellent performance in the next 12 months and rank 5 associated with worst performance in the next 12 months. The security rankings are based upon a filter rule which purportedly utilizes following criteria.

- Current year's relative earnings
- Relative earnings and prices in comparison to past ten years.
- Price momentum.

- Quarterly earnings momentum, and
- Earnings "surprise".

The first suggestion that VL timeliness ranks might be successful came in 1965, the year of the first VL contest¹⁵. During the contest, the average performance of the stocks in the VL groups were in the right order of 5%, 3%, -1%, -5% and -8% for ranks 1, 2, 3, 4, and 5 respectively (Black, 1973). Table 5 on the next page presents a summary of previous studies done in this area followed by their brief details.

¹⁵ Value Line used to stage a contest as a promotional device for its services offered. As a by product, it provided an unusually good opportunity to test the stock selecting ability of a large number of people. The contest required self-selected entrants to submit a certain list of stocks out of the master list of stocks followed and ranked by Value Line.

Table 5: Previous Studies on Performance of VL Ranks:

Study	Year	Event Window	Benchmark Period	Data	Approach
Shelton	1967	---	---	Contest Results	VL Contest
Kaplan & Weil	1973	---	---	Contest Results	VL Contest
Black	1973	---	---	Weekly	CER
Holloway	1981	---	---	Daily & Weekly	Buy & Hold Portfolio
Copeland & Myers	1982	26 week test pd., 13 weeks rank change pd.	52 or 26 weeks; following 26 weeks test pd.	Weekly	Market Model & CAAR
Stickel	1985	-50 to +50	+51 to +290	Weekly	CAAR
Huberman & Kandel	1987	---	---	Daily & Weekly	Constructi on of size based positions within each quintile

Shelton (1967) found that VL ranks outperformed the average of the contestants' stocks by 17%. Black (1973) also found that monthly rebalanced portfolios of stocks with ranks 1 and 5 had a risk adjusted rates of return of +10% and -10% respectively. Portfolio 1 had an annual turnover rate of 130%. In 1981, Holloway showed that an equally weighted portfolio of all securities ranked 1 by VL at the beginning of the year and held for a calender year outperforms a similarly held market portfolio even after transaction costs are taken into account. Subsequently in 1982, Copeland and Myers found statistically significant six month negative abnormal returns of about 3% associated

with securities ranked 5. Stickel (1985), extending the analysis of Copeland and Myers, found that alternative rank changes have differential impact upon security returns. Huberman and Kandel (1987) found that first and second ranked firms outperformed fourth and fifth ranked firms even after controlling for firm size.

Thus, the discussion provides sufficient evidence to support that VL timeliness ranks have been doing very well and are quite accurate. This in turn suggests that the attributes that are used by the VL ranking system are related to the firm performance and thus, may possess predictive power.

2.7.2 Attributes Taken From Anomaly Literature:

The attributes that are taken from anomaly literature are outlined as follows:

- Stock Price Level
- Market Capitalization
- Price to Sales Ratio, and
- Price to Book Ratio

The choice these attributes is directed by the existing evidence of their significant anomalous behaviour as found by Li (1994). The literature shows that portfolios created on the basis of these attributes systematically earn significant abnormal returns relative to an appropriate performance benchmark, either the prediction of the CAPM, or the performance of a carefully selected control portfolio. Table 5 presents a summary of study results regarding individual price related anomalies as outlined in Li (1994).

Table 6: Previous Studies On Price Related Anomalies:

Study (Year)	Type of Anomaly*	Direction of Effect	Magnitude (t - statistics)
Blume & Husic (1973)	P	-	0.0204 (4.1)
Basu (1977)	PER	-	0.27% (3.73)
Banz (1981)	MC	-	1.01 (3.07%)
Rosenberg, Reid, & Lanstein (1985)	PBR	-	0.36% (5.7)
Wilson (1986)	PCR	-	0.230 (3.87)
Senchack & Martin (1987)	PSR	-	0.31%
Bhandari (1988)	PDR	-	0.0013 (3.84)
Li (1994)	P PER MC PBR PCR PSR PDR	- - - - - - -	0.84% (2.82) 1.1% (3.94) 0.4% (1.35) 0.26% (1.06) 0.52% (1.94) 0.23% (0.91) 0.01% (0.04)

* Where P=price per share, PER = price to earnings ratio, MC = market capitalization, PBR = price to book ratio, PCR = price to cash flow ratio, PSR = price to sales ratio, PDR = price to debt ratio.

Thus, the evidence does seem to suggest that anomalous stock return behaviour is associated with the above mentioned price related attributes and investment strategies may be derived from these price related anomalies.

3. RESEARCH DESIGN

3.1 RESEARCH QUESTION:

The research question that this study addresses is:

"Is it possible to identify stock Winners and Losers using Artificial Neural Networks ?"

In addition to this, the study also addresses some sub-questions that are outlined as follows.

- Q1.** How would the Artificial Neural Network perform compared to the Classical Discriminant Analysis ?
- Q2.** How robust is ANN to the the sample size and the choice of attributes ?
- Q3.** How robust is ANN to the presence of influential data points in the sample ?

These research questions are investigated by training an ANN using a specified set of attributes as explained in Section 3.4 and subsequently testing network performance.

3.2 RESEARCH MODEL:

The research model for the present study along with the hypothesised relationships is presented in Figure 3.1 on the next page. The definitions of the attributes are discussed in Section 3.4.

Figure 3.1: Research Design

3.3 POPULATION:

The population for the study includes all publicly traded firms listed on the Toronto Stock Exchange.

3.3.1 Data:

3.3.1.1 Data Sources:

The study uses secondary data that is commercially available in computer readable form.

(a) Stock Returns Data:

The study uses monthly stock returns data from TSE/Western database. The database consists of monthly stock returns for most of the traded companies listed on the Toronto Stock Exchange. The monthly data available from this database consists of historical monthly returns over a period from 1950 to date.

(b) Quarterly Earnings Data:

The data for quarterly earnings is obtained from the Stock Guide database Ver 4.9. This is a comprehensive database consisting of data on financial ratios, financial statements, corporate profiles, quarterly data on revenues, and earnings & earnings per share for about 1,000 firms listed on Toronto Stock Exchange or Montreal Stock Exchange over the period from 1985 to date.

(c) Price Data:

As the actual price data at the end of each quarter are not available in Stock Guide directly, prices are calculated using the year-end prices from Stock Guide and monthly returns data from the TSE/Western database. Quarterly prices are calculated starting from the last but one month of the year and then continuing the process backwards until the first month of the year. The model for the price is derived from the stock return definition of the TSE/Western database.

(d) Consumer Price Index:

The data for the consumer price index (CPI) are obtained from The Bank of Canada Reviews. It is used to deflate all the dollar values to 1989 dollars.

3.4 ATTRIBUTE DEFINITIONS:

3.4.1 Current Year's Relative Earnings:

The current year's relative earnings is modelled as follows.
where

$$RELER_{it} = \frac{E_{it}}{(\sum_{i=1}^{n_t} E_{it}) / n_t}$$

$RELER_{it}$ = Relative Earnings for firm i for quarter t .

E_{it} = Earnings for firm i for quarter t .

n_t = # of firms in the indus. to which firm i belongs in quarter t .

3.4.2 Relative Earnings And Prices In Comparison To Past Eight Quarters:

These attributes are modelled as follows.

$$RELER8_{it} = \frac{E_{it}}{\sum_{t=t-1}^{t-8} \left[\frac{\sum_{i=1}^{n_t} E_{it}}{n_t} \right]}$$

$$RELPRICE_{it} = \frac{P_{it}}{\sum_{t=t-1}^{t-8} \left[\frac{\sum_{i=1}^{n_t} P_{it}}{n_t} \right]}$$

where

$RELER8_{it}$ = Relative Earnings compared to past eight quarters for firm i for quarter t .

$RELPRICE_{it}$ = Relative Prices compared to past eight quarters for firm i for quarter t .

E_{it} = Earnings of the firm i for quarter t .

P_{it} = Closing price of the firm i for quarter t .

n_t = # of firms in the indus. to which firm i belongs in quarter t .

3.4.3 Price Momentum:

The price momentum models are based on the premise that carefully studying historical prices can provide some indications of the future trend of a security's prices. Such models attempt to identify stocks that appear to be riding a wave. As the wave begins to weaken, the model identifies new stocks that appear to be riding the next wave. As outlined in Brush (1986), price momentum or "relative strength" models vary from simple to quite sophisticated. The study uses one such simple model outlined in Forester(1995) and is explained as follows.

$$PMRANK_{kt} = (.4) R_{t-1} + (.2) R_{t-2} + (.2) R_{t-3} + (.2) R_{t-4}$$

where

$PMRANK_k$ = Price Momentum for the firm k for the quarter t .

R_{t-1} = Return for the most recent quarter

R_{t-2} = Return for the quarter before R_{t-1}

R_{t-3} = Return for the quarter before R_{t-2}

R_{t-4} = Return for the quarter before R_{t-3}

and

$$R_i = \prod_{j=1}^3 (1 + R_{ij}) - 1$$

$$\forall \quad i = 1..4 \text{ month}$$

where

i = quarter

j = month in quarter i

3.4.4 Quarterly Earnings Momentum:

The quarterly earnings momentum is modelled as follows.

$$QEMRANK_{it} = (.4) E_{t-1} + (.2) E_{t-2} + (.2) E_{t-3} + (.2) E_{t-4}$$

where

$QEMRANK_{it}$ = Quarterly earnings momentum for firm i in quarter t .

E_{t-1} = Earnings in the most recent quarter

E_{t-2} = Earnings in the quarter before E_{t-1}

E_{t-3} = Earnings in the quarter before E_{t-2}

E_{t-4} = Earnings in the quarter before E_{t-3}

3.4.5 Unexpected Quarterly Earnings:

The unexpected quarterly earnings model is defined as the absolute value of the difference between this quarter's EPS and EPS from the same quarter last year. The model assumes that investors simply expect this quarter's earnings to equal EPS of the same quarter last year. It is represented as follows.

$$UE_{it} = \frac{EPS_{it} - EPS_{it-4}}{EPS_{it-4}}$$

where

UE_{it} = Unexpected earnings for firm i in quarter t .

EPS_{it} = Earnings per share for firm i for the most recent quarter.

EPS_{it-4} = Earnings per share for firm i for the same quarter last

year.

3.4.6 Stock Price Level:

The stock price level is defined as the closing common stock market price per share at the end of each quarter.

3.4.7 Firm Size (Market Capitalization):

This is defined as the product of the closing common stock price per share at the

end of each quarter and the average number of shares outstanding during that quarter.

It is modelled as follows.

$$MC = P \times n$$

where

MC = Market capitalization.

P = Closing price per share.

n = Average no. of shares outstanding.

3.4.8 Price to Sales Ratio:

The price to sales ratio is defined as the ratio of quarterly stock price to quarterly revenue per share. It is modelled as follows.

$$PSR = \frac{P}{REV}$$

where

PSR = Price to sales ratio.

P = Closing price per share.

REV = Total revenue per share.

3.4.9 Price to Book Ratio:

The Price to book ratio is defined as the ratio of quarterly stock price to quarterly book value per share. It is modelled as follows.

$$PBR = \frac{P}{BV}$$

where

PBR = Price to book ratio.

P = Closing price per share.

BV = Book value per share.

3.4.10 Annualized Returns:

This attribute is constructed in order to capture the performance of sample firms over the next 12 months from the time when it is calculated. These returns are used as a representative of the firm performance one year in the future. It is also used to calculate the class for the sample firms at a given month in the year.

$$ANR_{it} = \frac{\prod_{t=t+1}^{t+12} (1 + R_{it})}{12}$$

where

ANR = Annualized return for firm i in month t .
 R = Return for the firm i in month t .

3.5 SAMPLE SELECTION CRITERIA:

To qualify for the inclusion in the sample of this study, a firm must meet the following criteria¹⁶.

- (1) The firm must be listed on the Toronto Stock Exchange.
- (2) The firm must have 10 out of 12 months returns data available on the TSE/Western database for any year in the test period.
- (3) The book values, earnings at the end of fiscal year, and earnings per share must be available for the firm.
- (4) The firm must not belong to Gold and Precious Metals industries, Mining and

¹⁶ These selection criteria are essentially the same as those outlined in Li (1994).

Exploration without Production, Utilities, or Financial Services and Insurance¹⁷.

- (5) The firm must not be engaged in mining and integrated mining¹⁸.
- (6) The firm must have a traded year end closing price above \$1.0 in 1989 dollars¹⁹.
- (7) The firm must have total annual revenue above \$1.5 million in 1989 dollars.
- (8) The firm must have a year-end total asset base larger than \$3 million in 1989 dollars.

Criterion 1 is used to restrict the sample only to the Toronto Stock Exchange. Criteria 2,3, and 4 ensure that meaningful data are available for the analysis of this study and also filters out missing data. Criterion 5 excludes the firms in those industry groups whose attributes may not be easy to interpret. For example, gold and precious metals are excluded as their stock performance may be related to the speculative movement of the underlying commodities. Mining without production is excluded because of the low or negative earnings nature of the business. Utility industries are excluded because of their highly regulated and predictable environment. Financial services are excluded because they tend to have high debt financing. Criteria 6, 7, and 8 impose size restrictions on the sample. It also eliminates penny stocks as they do not behave the same way as larger stocks. To take rate of inflation into account, all the size restrictions are deflated to the 1989 dollar value.

¹⁷ They must not belong to the groups #13, 14, 77, 79, 90, 92, 95, 96, and 97 of the Stock Guide database. The industry groupings according to Stock Guide are presented in Appendix A.

¹⁸ The firm must not belong to groups #10 and 11 of Stock Guide database.

¹⁹ The \$ signs throughout this proposal refers to Canadian dollar.

4. METHODOLOGY

4.1 CREATING A SAMPLE OF STOCK WINNERS AND LOSERS:

In order to construct the training and testing set, portfolios of winners and losers are constructed each year during the specified sample selection period. As noted later, the quintile 1 portfolio of stocks is termed "winners" and the quintile 5 portfolio of stocks is termed "losers". Portfolios are formed based upon the industry groupings. Following steps are used to construct portfolios at the beginning of each year. These steps are also presented in Figure 4.1 on the next page. The process is repeated from 1989 to 1993.

- (1) For each year, at the beginning of the year, classify all sample firms according to their industry groups.
- (2) Get monthly returns and other information from TSE/Western and Stock Guide databases for the sample firms, and calculate values for the attributes as explained in Section 3.4.
- (3) Calculate annualized returns for each firm.
- (4) Within each industry group, identify quintile points using the cross-sectional distribution of annualized returns of the sample firms. Use these points to group stocks into five quintile portfolios within each industry. Thus, portfolio 1 consists of stocks belonging to the highest return quintile and portfolio 5 contains stocks belonging to the lowest return quintile within each industry group. Portfolio 1 and 5 represent the stock winner and loser portfolios, respectively.

The cut-off points for winners and losers were compared for all industry groups across the period of the study from 1990 to 1993. This was done in order to find out if there is an overlap between winners and losers from different industry groups. No overlapping was found between the two groups in the sample that was created.

Figure 4.1: Creating Sample of Stock Winners and Losers

Once the winner and loser portfolios are constructed, the attribute values for both attribute sets, ATTR01 and ATTR02, are calculated and the training examples are constructed. For each training example, the attribute values form the input to the network whereas the class²⁰ of the stock based upon the annualized return, form the desired output. Table 7 presents the distribution of sample firms in different TSE industry groups along with the number of firms in the winner and loser portfolios and their corresponding quintile returns.

²⁰ The class of a stock could be either a winner or loser. The class has a value of 1 for stock winners and has a value of 0 for stock losers.

Table 7: Distribution of firms in TSE Industry Groups and Winner(Quintile #1) & Loser(Quintile #5) Portfolios Along With the Average Returns for Quintile #1 and #5 :

	1990	1991	1992	1993
Total number of firms in different TSE industry groups				
TSE Group #1	6	3	-	-
TSE Group #3	15	22	22	28
TSE Group #4	5	8	9	4
TSE Group #5	9	11	16	19
TSE Group #6	24	26	26	33
TSE Group #7	7	8	4	8
TSE Group #8	3	2	4	5
TSE Group #9	-	1	1	1
TSE Group #11	5	5	11	12
TSE Group #12	12	10	11	15
TSE Group #14	4	6	3	14
Total firms in Quintile #1	47	52	54	67
Total firms in Quintile #5	43	50	53	63
Total	90	102	107	130
Average quintile returns for Quintile #1 and #5				
Av. Quintile #1 Ret	12%	13%	10%	17%
Av. Quintile #5 Ret	-5%	-3%	-4%	-2%

4.2 CREATING THE ANN MODEL:

The ANN that is used for this study is designed using MATLAB Neural Network Toolbox²¹.

²¹ See Appendix B for an overview on MATLAB ANN Toolbox.

4.2.1 Architecture of the ANN:

A two layer (one input, one hidden and one output layer) fully connected feedforward network architecture is chosen for this study. As the study uses two sets of attributes, the architecture for each set differs slightly with the difference in the number of neurons in the input layer. For the attribute set ATTR01, the network has six neurons in the input layer; four neurons in the hidden layer and one neuron in the output layer. The six neurons in the input layer correspond to each of the six attributes modelled along the lines of VL ranking system. The single output neuron represents either a winner or a loser. This architecture is shown in Figure 4.2.

Figure 4.2: Architecture

The architecture for the attribute set ATTR02 would have ten neurons in the input layer, four neurons in the hidden layer and one neuron in the output layer.

4.2.2 Training:

The network has to be trained before it can be used for classifying stocks into winners and losers. The present study uses supervised learning²² using backpropagation network learning paradigm to train the network. Backpropagation is chosen as it has almost universally become the standard network paradigm for modelling, forecasting, and classification (Casimir, 1991). Supervised learning involves network learning by trying to duplicate behaviour specified by a set of *training examples*, where each training example consists of input data and the corresponding desired output. This causes the network to adjust its weights in order to produce desired outputs. In the classical statistical context, this process is equivalent to parameter (i.e. weight) estimation. The changes in the values of weights in the network depend upon how close the network output is to the desired output. The backpropagation algorithm for learning in feedforward networks uses mean squared error and gradient descent methods. The goal is to determine weights and biases $\{w_{i,j}\}$ for the network that minimize the mean sum squared error. The algorithm begins by randomly assigning small initial weights, say in the range $[-0.1, 0.1]$. The reason for this is to break the symmetry so that various intermediate cells can take on different roles. Next, a training example is chosen and gradient with respect to this single example is then computed. Computing the gradient

²² For more information on network learning, see Appendix C.

involves a forward pass over the network to compute the cell activations, followed by a backward pass that computes the gradients. Once the gradient is computed, a small step²³ is taken to update the weights. The weights are updated according the following model.

$$W_{i,j}^* = W_{i,j} + \rho \delta_i u_j$$

where

$W_{i,j}^*$ = The new weight for the cell u_i .

$W_{i,j}$ = Weight for the cell u_i .

ρ = Step size.

δ_i = Gradient.

u_j = Activation²⁴ for the cell i .

The process continues until changes become sufficiently small, i.e., the algorithm converges.

The next section presents a detailed description of the procedure that is followed for determining the applicability of ANN for identifying stock winners and losers.

4.3 INVESTIGATING APPLICABILITY OF ANN FOR IDENTIFYING STOCK WINNERS AND LOSERS:

The first and the most important step in approaching a problem like this involves testing the stability of the network. The present study also starts with a similar approach. As ANN is a relatively new tool for financial applications, the analysis for the present study is carried out as a two step process. In the first step, a preliminary analysis is

²³ The gradient descent algorithm contains a constant, $\rho > 0$, that governs how big the step size is made at each iteration. This is called step size.

²⁴ Cell activations may be discrete, taking on values $\{0,1\}$ or $\{-1, 0, 1\}$, or it could be continuous, assuming values in the interval $[0,1]$ or $[-1, +1]$.

performed on only one quarter out of the sixteen quarters in the entire period of the study. It is mainly aimed at understanding the performance of ANN in the context of the present study. The preliminary analysis also investigates the effect of varying the sample size, number of attributes and architecture on the performance of ANN. The second step involves conducting the main analysis over the entire period of study. Two different strategies (explained later in Section 4.3.2) are used for investigating the performance of ANN and comparing it with the performance of discriminant analysis. The main analysis also investigates the performance of ANN on hold-out samples and its robustness to the presence of influential data points in the input data set. The entire methodological process of study is presented in Figure 4.3 and is explained as follows.

Figure 4.3: The Entire Methodological Process

4.3.1 Preliminary Analysis (Step #1):

Before starting the main analysis, some preliminary analysis is done in order to develop a better understanding of the performance of ANN in the context of the present study. The main objectives of the preliminary analysis are outlined as follows. (1) To investigate the robustness of ANN to the sample size and to test the sample size that is required for good network training. (2) To find out the classification power of the attribute sets ATTR01 and ATTR02 (see Section 2.7 for explanation of the attribute sets). This is important because ANN might be a very good technique but if the underlying attributes don't have any relation to stock winners and losers then the performance of ANN will not be good. (3) To evaluate the impact of expanding the size of the attribute set on the classification accuracy of ANN. (4) Since ANN is an alternative to classical discriminant analysis, especially for this study, it is important to compare the performance of ANN with its classical counterpart, discriminant analysis. Since the principles behind the two techniques are quite different, this comparison would demonstrate the relative classification accuracy of the two techniques in the context of the present study.

The rest of the section outlines the steps that are followed in the preliminary analysis. As mentioned in earlier sections, the present study focuses on the quarterly data over a period from 1990 to 1993; a total of 16 quarters. Out of these 16 quarters, the first quarter of 1993 is arbitrarily chosen for the preliminary analysis. The steps that are then followed are summarized in Table 8.

Table 8: Steps followed in the Preliminary Analysis:

STEP #	WHAT	WHY
1.	Samples with varying sample size used.	To investigate the sample size requirements for training.
2.	Multiple random samples for each size are constructed.	To test robustness of the classification accuracy of ANN.
3.	Train ANN for both attribute sets.	To demonstrate the performance of ANN for both attribute sets.
4.	Perform two group MANOVA.	To ensure that the two attribute sets can differentiate winner and loser groups using statistical methods.
5.	Use step-wise discriminant analysis.	To identify the attributes that are redundant from each attribute set.
6.	Delete the redundant attributes and re-train ANN.	To see the impact of the redundant attributes on the classification accuracy of ANN.
7.	Compare the results of discriminant analysis and ANN.	To show the relative classification accuracy of the two techniques.
8.	Test the architecture.	To check if the results are architecture specific.

The steps followed in the preliminary analysis are explained as follows. (1) Samples with varying sample sizes (20, 40, 60, 80 and 130 observations per sample out of a sample of 130) are used for the analysis. Using such samples allows for testing the robustness of the ANN's classification accuracy to the sample size and also investigates the kind of samples that are required for network training. (2) Ten random samples, each for a sample of a given size (except for the sample with 130 observations) as mentioned in the first step, are generated for both attribute sets ATTR01 and ATTR02. Using multiple random samples for each size tests the robustness of the classification

accuracy of ANN across all random samples for each size. (3) Once the samples are constructed, a **two layer** ANN (one input layer, one hidden layer and one output layer) is trained using the generated random samples for both attribute sets ATTR01 and ATTR02. (4) The power of both the attribute sets to differentiate winner and loser groups is subsequently tested using the standard statistical method - a two-group MANOVA. This is important because the results of the network training cannot be interpreted correctly unless it is established that the group means of stock winners and losers can be differentiated by MANOVA. (5) Since it is possible that out of all the attributes present in both attribute sets, not all are important for classifying winners and losers, a step-wise discriminant analysis is then performed to find out the attributes that are most important, from each of the two attribute sets, in classifying stocks into the two respective groups. (6) In order to investigate the affect of removal of redundant attributes on the performance of ANN, the network is re-trained using only those attributes that are found to be significant in the previous step. (7) As one of the objectives of the preliminary analysis is also to compare the performance of ANN with its statistical counterpart, the classical discriminant analysis is also performed. This demonstrates how the classical statistical methods perform compared to the ANN in the context of the present study. (8) The preliminary analysis also tests the performance of a network with a different architecture in the context of the present study by designing a network with **three layers** (one input layer, two hidden layers and one output layer) and subsequently training it using the attributes that are found to be significant in the step-wise discriminant analysis. This is done in order to find out that the results obtained

in Step 3 are not architecture specific.

This completes the preliminary analysis. The results of the preliminary analysis demonstrate the performance of ANN for different randomly generated samples for different combinations of attributes. In addition to this, the results also present a comparison of the performance of ANN with discriminant analysis which enables us to better evaluate the strengths and weaknesses of ANN in the context of the present study. The results also help and guide in designing the main analysis for the present study which is conducted as outlined in the next section.

4.3.2 Main Analysis (Step #2):

The objective of the main analysis is to further investigate the performance of ANN for each of the 16 quarters over the entire period of study from 1990 to 1993. The analysis compares the performance of ANN with that of discriminant analysis. It also aims at determining the influence of redundant attributes, if there are any, on the classification accuracy of ANN and if the influence persists across the period of the study. The main analysis investigates the performance of ANN on the hold-out samples for a few quarters as well.

In order to better understand the performance of ANN, the analysis uses two different strategies explained as follows.

- (1) **Strategy 1:** The strategy in which the analysis is done using all the attributes in both attribute sets ATTR01 and ATTR02.
- (2) **Strategy 2:** The strategy in which a step-wise discriminant analysis is first

performed in order to find the redundant attributes from each of the two attribute sets. Only those attributes that are found to be significant by the step-wise discriminant analysis are used.

Comparing the results of these two strategies would demonstrate the effect of the redundant attributes on the classification accuracy of ANN.

The steps that are followed in the main analysis are summarized in Table 9.

Table 8: Steps followed in the Main Analysis:

STEP #	WHAT	WHY
1.	Only bigger samples used.	Greater possibility of capitalizing on chance in smaller samples than in bigger ones.
2.	Multiple random samples for each size are constructed.	To test robustness of the classification accuracy of ANN.
3.	ANN is trained for both sets of attributes for both strategies.	To investigate the effect of the redundant attributes on the classification accuracy of ANN.
4.	Compare the results of discriminant analysis and ANN.	To show the relative classification accuracy of the two techniques.
5.	Change in the classification accuracy of ANN on moving from Strategy 1 to Strategy 2 is calculated.	To demonstrate the effect of both strategies on the performance of ANN.
6.	Conduct an out-of-sample test.	To investigate the performance of ANN on a hold-out sample.

The steps summarized in the above table are explained as follows. (1) As there is a greater likelihood of capitalizing on chance in smaller samples than in bigger ones, the main analysis is done only on samples with 80, 100 and all observations for both

attribute sets ATTR01 and ATTR02 for both strategies. (2) For each of the 16 quarters, five random samples are generated with 80 observations and similarly five random samples are generated with 100 observations for both attribute sets ATTR01 and ATTR02 for both strategies. Thus, in total for 16 quarters, for each of the two attribute sets and for each of the two strategies, a total of 640 samples are constructed. (3) Once samples are constructed, ANN is trained on each of the 640 samples and the in-sample classification accuracies are recorded. (4) As mentioned earlier, ANN is an alternative to the classical discriminant analysis therefore it is logical to compare performance of ANN with its statistical counterpart. This is accomplished by also performing discriminant analysis on all 640 samples. (5) Once the in-sample classification accuracies of ANN and discriminant analysis are obtained, the changes in the classification accuracies, as we move from the Strategy 1 to Strategy 2, are calculated. This demonstrates the increase/decrease in the classification accuracy on moving from one strategy to the other and in-turn show the effect of the presence/absence of certain attributes on the performance of ANN. (6) The last thing that the main analysis focuses on is to investigate the performance of ANN on a hold-out sample. Instead of doing it on all the quarters for the period of the study, the analysis focuses only on four quarters that are arbitrarily chosen from the 16 quarters for the period of the study. The motivation behind this is to get a representative performance of ANN on hold-out samples in the context of the present study. In order to proceed in the analysis, training samples and hold-out samples are constructed for the chosen quarters. The network is subsequently trained using the training samples and is tested on the hold-out samples.

The next chapter presents the results of the preliminary and main analyses.

5. EMPIRICAL RESULTS

5.1 RESULTS OF THE PRELIMINARY ANALYSIS:

The mean in-sample classification accuracy across the random samples of different sizes for both attribute sets ATTR01 and ATTR02 is presented in Table 10 for the first quarter of 1993.

Table 10: Mean Classification Accuracy of ANN Across 10 Random Samples, each for ATTR01 and ATTR02:

Sample Size	ATTR01	ATTR02
20 Observations	59.50 %	56.00 %
40 Observations	56.25 %	53.05 %
60 Observations	56.00 %	53.00 %
80 Observations	52.88 %	52.50 %
130 Observations	51.86 %	51.25 %

The results in Table 10 suggest that the classification accuracy of ANN is generally very poor and decreases as the sample size increases. These results may be due to the following possibilities. First, it is possible that the two groups (winners and losers) that the network is trying to differentiate are drawn from the same population. That is to say that there are no differences between the group means and hence, they can't be differentiated. Second, the higher classification accuracy in smaller sample sizes may have been achieved because of the greater likelihood of capitalizing on chance. Increasing the sample size reduces the possibility of capitalization on chance which may be the reason for the gradual drop in the classification accuracy. In order to find out if

the group means are different or not, a two-group MANOVA is done on the first quarter of 1993 for both attribute sets ATTR01 and ATTR02. The results are presented in Tables 11 and 12.

Table 11: Results of Two-Group MANOVA on 1993 First Quarter for ATTR01:

Test Name	Value	Exact F	Hypoth DF	Error DF	p-Value	
Wilks	.87997	3.15992	6.00	139.00	.006'	
Univariate F-Tests with (1,144) D.F.						
Variable	Hypoth SS	Error SS	Hypoth MS	Error MS	F	p-Value
PMR93Q1	.19348	4.84831	.19348	.03367	5.746	.018*
QEM93Q1	486994807	6.175E+10	486994807	428837962	1.136	.288
RE8_93Q1	4157.1740	587729.32	4157.1740	4081.4536	1.018	.315
RE_93Q1	56.22803	14736.592	56.22803	102.33744	.549	.460
RP8_93Q1	.07631	1.73633	.07631	.01206	6.329	.013*
UE93Q1	155.88081	7964.0134	155.88081	55.30565	2.819	.095**

Table 12: Results of Two-Group MANOVA on 1993 First Quarter for ATTR02:

Test Name	Value	Exact F	Hypoth DF	Error DF	p-Value	
Wilks	.80604	2.86356	10.00	119.00	.003*	
Univariate F-Tests with (1,144) D.F.						
Variable	Hypoth SS	Error SS	Hypoth MS	Error MS	F	p-Value
MC93Q1	2.493E+10.1	1.518E+12	2.493E+10	1.186E+10	2.10	.149
PBR93Q1	1785.1165	260575.18	1785.1165	2035.743	.876	.351
PMR93Q1	.16243	3.95849	.16243	.03093	5.25	.024*
PSR93Q1	.00015	.03017	.00015	.00024	.652	.421
QEM93Q1	209923366.1	1.142E+10	209923366	89254870	2.35	.028*
QPRQ1_93	427.56239	5189.3199	427.56239	40.54156	10.5	.001*
RE8_93Q1	3.63932	74372.91	3.63932	581.0383	.006	.937
RE_93Q1	57.23582	10102.655	57.23582	78.92699	.725	.396
RP8_93Q1	.04482	1.32373	.04482	.01034	4.33	.039*
UE93Q1	149.90681	7891.9678	149.90681	61.65600	2.43	.121

The results in Tables 11 and 12 show that both attribute sets are able to strongly differentiate between winners and losers ($p=0.006$ & $p=0.003$, respectively). Thus, we reject the first possibility that the poor classification accuracy may have resulted from the absence of a group difference. The univariate tests of significance also show that, for the attribute set ATTR01, four attributes out of ten are significant at α level of 5%, and may contribute to the multivariate significance.

A step-wise discriminant analysis was performed next to determine the attributes that are redundant in the discrimination between winners and losers. The attributes that

are found to be redundant from each attribute set are then removed and the network is re-trained on larger samples to provide more accurate estimates of accuracy. In addition, classification using discriminant analysis is also performed. The comparative results for ANN and discriminant analysis are summarized in Tables 13.

Table 13: Mean Classification Accuracy of ANN and Discriminant Analysis for ATTR01 and ATTR02 After Redundant Attributes are Removed:

	80 Observations in Sample		130 Observations in Sample	
	ANN	Discriminant	ANN	Discriminant
Mean Classification Accuracy for ATTR01	82.25 %	68.00 %	75.57 %	63.57 %
Mean Classification Accuracy for ATTR02	81.85 %	70.54 %	75.25 %	64.26 %

The results presented in Tables 13 are remarkably different from those in Table 10. We can clearly observe a lot of increase in the classification accuracy of ANN. The increase is approximately 30 percent points for the 80 observation sample and about 25 percent points for the 130 observation sample. This increase in the classification accuracy of ANN suggests the following.

- (1) The network seems to be very sensitive to the presence of redundant attributes. As the redundant attributes are removed, an increase in the classification accuracy is observed. This contradicts the notion that ANNs are robust to the presence of noise in the input data.
- (2) The increase in the classification accuracy of ANN may be because of the following. The number of parameters²⁵ that ANN has to estimate goes down

²⁵ The number of parameters that the network has to estimate can be calculated by counting the number of connections in the network as the network estimates a weight for each connection. For example, for the attribute set ATTR01, there are six nodes in the input layer, four in the hidden layer and one in the output layer. For this network there are 24 connections between the input and the hidden layer (the network is fully connected) and 4 connections between the hidden and the output layer. Thus, there are

rapidly as the redundant attributes are removed. It may be because of this reduced number of estimates, that the classification accuracy of ANN increases.

Next, the robustness of the network architecture is investigated to ensure that these conclusions are not architecture dependent. This is done by repeating the analysis with the reduced set of attributes, as were used in the second iteration, on a **three layer network** (one input, two hidden, and one output layer) for both attribute sets ATTR01 and ATTR02. The results are presented in Table 14.

Table 14: Classification Accuracy of Four Layer ANN for ATTR01 and ATTR02:

	80 Observations in Sample	130 Observations in Sample
Mean In-Sample Classification Accuracy for ATTR01	82.00 %	75.14 %
Mean In-Sample Classification Accuracy for ATTR02	81.70 %	75.20 %

Comparing the results of Table 13 with Table 14 shows that there are hardly any differences between the results in individual tables. The most important thing is that the two statements that were made earlier regarding the performance of ANN hold true for the three layer architecture as well.

28 parameters (weights) that the network has to estimate. Similarly, the number of parameters that the network has to estimate for the ATTR02 can be calculated.

5.1.1 Implications of the Preliminary Analysis:

The results obtained in the preliminary analysis indicate the following.

- (1) As ANN is found to be very sensitive to the attributes in the input data, it suggests that this technique must be used after careful formulation of the attribute sets. In other words, the ANN is sensitive to the problem description.
- (2) It is found that the results of ANN are not architecture specific as there is hardly any change observed in the classification accuracy of ANN on changing the architecture.
- (3) It is also possible that above findings are specific to the data set or the specific quarter that is used. This will be further investigated in the main analysis.

It is essential that the results of the main analysis are interpreted with the above findings in mind. The next section presents the results of the main analysis.

5.2 RESULTS OF THE MAIN ANALYSIS:

Tables 15 through 17 present the mean in-sample classification accuracies of ANN and discriminant analysis for all the samples for each of the sixteen quarters for 80, 100 and all observations, respectively, using both strategies. Strategy 1 uses all attributes in each of the attribute sets whereas Strategy 2 uses only those attributes which are found to have discriminating power using the classical statistical analysis. The classification accuracy of discriminant analysis is also presented to provide a comparison of the performance of ANN with that statistical methods in the context of the present study.

Table 15: Mean Classification Accuracy of ANN and Discriminant Analysis for Samples with 80 Observations:

	Strategy 1				Strategy 2*			
	ATTR01 Disc (%)	ATTR01 ANN (%)	ATTR02 Disc (%)	ATTR02 ANN (%)	ATTR01 Disc (%)	ATTR01 ANN (%)	ATTR02 Disc (%)	ATTR02 ANN (%)
1990 Q1	67.50	66.75	68.00	56.63	66.88	81.25	68.13	76.25
1990 Q2	87.75	76.75	89.25	59.00	86.75	70.13	86.75	70.13
1990 Q3	88.13	76.13	90.25	64.88	85.50	85.50	76.50	76.50
1990 Q4	95.38	80.63	95.75	71.50	95.13	80.63	95.13	80.63
1991 Q1	65.00	57.25	63.25	52.75	63.75	66.75	63.75	66.75
1991 Q2	75.25	65.00	79.00	53.25	73.50	83.25	76.75	82.75
1991 Q3	82.00	70.50	83.75	57.00	82.00	88.00	83.75	87.00
1991 Q4	87.50	66.50	85.10	58.00	88.50	66.50	88.50	66.50
1992 Q1	71.75	59.25	73.75	52.80	67.50	87.50	71.50	87.75
1992 Q2	77.00	59.25	78.00	54.00	72.25	75.25	74.00	78.00
1992 Q3	85.25	59.50	87.00	64.75	86.40	87.50	87.00	86.75
1992 Q4	94.25	70.75	94.75	60.50	94.25	94.75	95.00	94.75
1993 Q1	68.50	57.50	68.50	51.75	65.25	61.25	63.50	63.25
1993 Q2	85.25	54.00	85.25	52.25	83.95	75.25	80.25	90.25
1993 Q3	87.75	55.31	88.25	50.75	83.25	96.25	86.00	50.00
1993 Q4	91.50	62.88	88.00	50.25	90.25	97.75	90.25	97.75
Mean	81.86	64.87	82.37	56.88	50.32	81.09	80.42	78.44
Std Dev	9.46	7.93	9.41	5.83	10.21	10.75	9.94	12.32
Min	65.00	54.00	63.25	50.25	63.75	61.25	63.50	50.00
Max	95.38	80.63	95.75	71.50	95.13	97.75	95.13	97.75

* The number of attributes in the attribute sets ATTR01 and ATTR02 vary from quarter to quarter depending upon the number of attributes that were found significant in the step-wise discriminant analysis.

Table 16: Mean Classification Accuracy of ANN and Discriminant Analysis for Samples with 100 Observations:

	Strategy 1				Strategy 2			
	ATTR01 Disc (%)	ATTR01 ANN (%)	ATTR02 Disc (%)	ATTR02 ANN (%)	ATTR01 Disc (%)	ATTR01 ANN (%)	ATTR02 Disc (%)	ATTR02 ANN (%)
1990 Q1								
1990 Q2								
1990 Q3								
1990 Q4								
1991 Q1	69.80	56.04	65.40	55.42	65.60	68.54	65.60	68.54
1991 Q2	75.20	65.00	78.40	54.79	72.80	83.84	76.20	82.29
1991 Q3	82.60	64.79	84.20	57.92	83.00	83.55	83.40	84.17
1991 Q4	88.00	59.88	89.40	61.46	89.00	66.67	89.00	66.67
1992 Q1	69.40	58.20	70.20	53.00	65.60	81.40	68.20	81.60
1992 Q2	76.80	57.00	78.80	55.60	74.40	76.18	74.60	77.60
1992 Q3	85.60	65.60	87.20	52.80	85.60	88.80	85.80	88.40
1992 Q4	93.60	75.20	94.20	75.40	93.80	94.20	94.60	94.40
1993 Q1	66.80	54.00	70.20	53.00	65.40	67.20	64.60	66.00
1993 Q2	80.00	53.60	81.40	53.20	76.40	77.80	81.80	82.20
1993 Q3	86.60	50.85	87.20	54.40	82.80	83.80	87.60	89.00
1993 Q4	93.60	54.20	91.60	53.00	92.00	97.20	92.00	97.20
Mean	72.87	55.01	73.88	52.25	71.20	72.66	72.62	73.41
Std Dev	8.27	6.74	8.63	6.32	9.46	8.58	9.63	9.04
Min	66.80	50.85	65.40	52.80	65.40	66.67	64.60	66.00
Max	93.60	75.20	94.20	75.40	93.80	97.20	94.60	97.20

Table 17: Mean Classification Accuracy of ANN and Discriminant Analysis for Samples with all Observations:

	Strategy 1				Strategy 2			
	ATTR01 Disc (%)	ATTR01 ANN (%)	ATTR02 Disc (%)	ATTR02 ANN (%)	ATTR01 Disc (%)	ATTR01 ANN (%)	ATTR02 Disc (%)	ATTR02 ANN (%)
1990 Q1	67.77	67.78	70.00	53.33	67.77	88.89	71.11	74.44
1990 Q2	87.78	77.78	88.88	73.33	87.78	77.78	86.67	77.78
1990 Q3	85.56	75.56	86.67	55.56	85.56	75.56	85.56	75.56
1990 Q4	94.44	81.11	95.56	47.78	94.44	81.11	93.33	81.11
1991 Q1	68.63	56.86	63.73	55.88	65.69	69.61	65.69	67.65
1991 Q2	73.53	66.67	79.41	55.88	72.55	83.33	74.50	84.31
1991 Q3	81.37	69.61	85.29	58.82	80.39	84.31	82.35	90.20
1991 Q4	87.25	66.67	89.21	62.75	88.23	68.63	88.23	66.67
1992 Q1	70.09	56.07	67.29	57.01	65.42	96.16	67.29	97.85
1992 Q2	77.57	63.55	79.44	55.14	75.70	77.57	75.70	77.57
1992 Q3	85.98	66.36	87.85	55.14	85.98	86.92	85.98	86.92
1992 Q4	93.46	75.70	95.33	76.64	93.46	93.46	94.39	92.52
1993 Q1	66.15	53.08	70.00	53.08	65.38	53.08	63.85	65.38
1993 Q2	80.77	55.38	81.54	51.54	79.23	55.38	62.31	86.92
1993 Q3	83.84	56.15	85.38	52.31	81.54	92.31	86.92	51.54
1993 Q4	93.08	57.69	89.23	52.31	90.76	97.69	90.76	97.69
Mean	81.08	65.38	82.18	57.28	79.99	80.11	79.67	79.01
Std Dev	9.21	8.65	9.50	7.44	9.87	12.90	10.60	11.59
Min	66.15	53.08	63.73	47.78	65.38	53.08	62.31	51.54
Max	94.44	81.11	95.56	76.64	94.44	97.69	94.39	97.69

Following observations can be made by looking at the results presented in the previous three tables.

- (1) The classification accuracy of discriminant analysis is always considerably greater than that of ANN for Strategy 1.
- (2) For Strategy 2, the classification accuracy of ANN and discriminant analysis are on average very close.
- (3) As found in the preliminary analysis also, we can again observe an increase in the classification accuracy of ANN as we move from Strategy 1 to Strategy 2. The amount of increase/decrease in the classification accuracy of ANN and discriminant analysis on moving from Strategy 1 to Strategy 2 is presented in Tables 18.

Table 18: Difference in Classification Accuracy for Strategy 1 and Strategy 2 for Discriminant Analysis and ANN for Samples with 80, 100 and all Observations:

	80 Observations				100 Observations				All Observations			
	ATTR01 Disc (%)	ATTR01 ANN (%)	ATTR02 Disc (%)	ATTR02 ANN (%)	ATTR01 Disc (%)	ATTR01 ANN (%)	ATTR02 Disc (%)	ATTR02 ANN (%)	ATTR01 Disc (%)	ATTR01 ANN (%)	ATTR02 Disc (%)	ATTR02 ANN (%)
1990 Q1	-0.63	14.50	0.13	19.63					0.00	21.11	1.11	21.11
1990 Q2	-1.00	-6.63	-2.50	11.13					0.00	0.00	-2.21	4.45
1990 Q3	-2.63	9.38	-13.75	11.63					0.00	0.00	-1.11	20.00
1990 Q4	-0.25	0.00	-0.62	9.13					0.00	0.00	-2.23	33.33
1991 Q1	-1.25	9.50	0.50	14.00	-4.20	12.50	0.20	13.12	-2.94	12.75	1.96	11.77
1991 Q2	-1.75	18.25	-2.25	29.50	-2.40	18.84	-2.20	27.50	-0.98	16.66	-4.91	28.43
1991 Q3	0.00	17.50	0.00	30.00	0.40	18.76	-0.80	26.25	-0.98	14.70	-2.94	31.38
1991 Q4	1.00	0.00	3.40	8.50	1.00	6.79	-0.40	5.21	0.98	1.96	-0.98	3.92
1992 Q1	-4.25	28.25	-2.25	34.95	-3.80	23.20	-2.00	28.60	-4.67	40.09	0.00	30.84
1992 Q2	-4.75	16.00	-4.00	24.00	-2.40	19.18	-4.20	22.00	-1.87	14.02	-3.74	22.43
1992 Q3	1.15	28.00	0.00	22.00	0.00	23.20	-1.40	35.60	0.00	20.56	-1.87	31.78
1992 Q4	0.00	24.00	0.25	34.25	0.20	19.00	0.40	19.00	0.00	17.76	-0.94	15.88
1993 Q1	-3.25	3.75	-5.00	11.50	-1.40	13.20	-5.60	13.00	-0.77	0.00	-6.15	12.30
1993 Q2	-1.30	21.25	-5.00	38.00	-3.60	24.20	0.40	29.00	-1.54	0.00	-19.23	35.38
1993 Q3	-4.50	40.94	-2.25	-0.75	-3.80	32.95	0.40	34.60	-2.30	36.16	1.54	-0.77
1993 Q4	-1.25	34.88	2.25	47.50	-1.60	43.00	0.40	44.20	-2.32	40.00	1.53	45.38
Mean	-1.54	16.22	-1.94	21.56	-1.67	17.65	-1.27	21.16	-1.09	14.74	-2.51	21.73
Std Dev	1.80	12.83	3.82	12.79	1.85	6.61	1.91	9.16	1.40	13.88	4.87	12.63
Min	-4.75	-6.63	-13.75	-0.75	-4.20	6.79	-5.60	5.21	-4.67	0.00	-19.23	-0.77
Max	1.15	40.94	3.40	47.50	1.00	43.00	0.40	44.20	0.98	40.09	1.96	45.38

A further review of the tables 15, 16 and 17, indicates that the mean classification accuracy of ANN for the third quarter of 1993 for ATTR02 for Strategy 2, is highly volatile for samples with different number of observations. The mean accuracies for samples that have all the observations and those that have 80 observations are very poor for this cell, whereas for the sample with 100 observations the accuracy is quite high and not very different from the rest of the quarters. This, along with the fact that the individual samples are constructed by randomly selecting the observations, indicates that observations that might have been left out in the 100 observations sample have heavy influence on the classification accuracy of ANN.

This, in turn, suggests the possible presence of influential data points in the sample for that quarter and raises the following question - how robust is ANN to the presence of influential data points in the sample?, and do influential data points, if there are any, affect the classification accuracy of ANN? The first step, therefore, is to find influential data points in the sample for the third quarter of 1993. The methodology adopted for identifying this is as follows.

Before proceeding ahead in the analysis, five new random samples are constructed for the third quarter of 1993 and the network is re-trained. This is done in order to find that the results for the cell under investigation are not specific to the five random samples that were used. The results are presented in Table 19 which show that the classification accuracy of ANN is not very different from what was observed earlier and therefore, rejects the possibility that earlier results may have been sample specific.

Table 19: Mean Classification Accuracy of ANN and Discriminant Analysis for Samples with 80 Observations for Five Different Random Samples:

	Strategy 1				Strategy 2			
	ATTR01 Disc (%)	ATTR01 ANN (%)	ATTR02 Disc (%)	ATTR02 ANN (%)	ATTR01 Disc (%)	ATTR01 ANN (%)	ATTR02 Disc (%)	ATTR02 ANN (%)
1993 Q3	86.65	65.7	90.30	52.5	83.50	96.00	86.00	58.05

The methodology used for identifying influential data points in the input data set uses regression analysis approach to MANOVA²⁶. This is done in order to utilize the number of measures, such as Mahalonobis Distance, Cook's Distance etc., that regression analysis offers to detect the presence of influential data points in the input data set. Of these measures, Mahalonobis Distance is used to detect the presence of influential data points²⁷. If any influential data points are detected, then the identified cases are removed from the input data set and the residual statistics are re-calculated. The ANN is trained again on the new data set and discriminant analysis is also performed. This would demonstrate the effect of the removal of influential data points, if there are any, on the classification accuracy of ANN.

Accordingly, the regression analysis is performed and residual statistics are calculated which are presented in Table 20.

²⁶ This is accomplished by performing regression analysis by dummy coding the group membership. In a typical regression problem, the predictor(s) are usually continuous variables. Here, for MANOVA, the predictor is a categorical or nominal variable, and is used to determine how much of the variance in the variables is accounted for by the group membership. A group membership value of 0 and 1 is used for the Stock Loser and Winner groups respectively in this study. It should be noted that the values other than 0 and 1 could have been used as the dummy codes without affecting the results. All that is necessary is to distinguish between the subjects in the two groups by a different value. (Steven, 1992).

²⁷ The critical values of Mahalanobis Distances are obtained from Steven (1992, p115). For the quarter under investigation, the critical value is ≈ 20 . Therefore, any case having the Mahalanobis Distance over 20 would be considered as influential..

Table 20: Residuals Statistics for the Third Quarter of 1993:

	Min	Max	Mean	Std Dev
*MAHAL	0.02362	64.97091	2.98	9.04
*COOK D	0.00000	7.49987	0.06	0.66
*LEVER	0.00018	.050365	0.02	0.07

The maximum Mahalanobis' Distance value of 64.97 in the above table clearly indicates the presence of influential point(s). If the value of Mahalanobis' Distance for a data point is above the critical value, it suggests that the data point is influential, but it may not be an outlier. Cook's Distance should then be used to find out if the influential data point is an outlier as well. As a general rule, a Cook's Distance value of over 1 suggests that the data point is an outlier. In Table 20, the maximum Cook's Distance value of 7.49 suggests that the influential cases in the sample are outliers as well and therefore will have a lot of leverage.

In order to exactly pin-point the case(s) with influential data point(s), a sequence plot of Mahalanobis' Distance is plotted for all the cases as shown in Figure 5.1.

Figure 5.1: Sequence Plot of Mahalanobis' Distances With Influential Data Points

The Figure 5.1 shows that there are four influential data points present in the data set as Four cases have Mahalanobis' Distance values of well over 20. These four cases are identified as cases 22, 48, 63 and 102. In order to see the effect of removal of these cases on the performance of F^* , the four identified cases are removed and residual statistics are re-calculated which are presented in Table 21. The sequence plot of the Mahalanobis' Distance is shown in Figure 5.2.

Table 21: Residuals Statistics for the Third Quarter of 1993:

	Min	Max	Mean	Std Dev
*MAHAL	0.01717	15.61929	2.432	3.13
*COOK D	0.00000	0.13202	0.011	0.01
*LEVER	0.00014	0.03529	0.012	0.03

Figure 5.1: Sequence Plot of Mahalanobis' Distances With Influential Data Points

Looking at the Mahalanobis' Distance values in Table 21, we find that the maximum value is 15.92 which is less than 20 and therefore, there are no more influential data points now present in the new data set for the third quarter of 1993. The ANN is subsequently trained and discriminant analysis is performed for this new data set without influential data points. The results are presented in table 22.

Table 22: Mean Classification Accuracy of ANN and Discriminant Analysis for Samples with 89 Observations without Influential Data Points:

	Strategy 1				Strategy 2			
	ATTR01 Disc (%)	ATTR01 ANN (%)	ATTR02 Disc (%)	ATTR02 ANN (%)	ATTR01 Disc (%)	ATTR01 ANN (%)	ATTR02 Disc (%)	ATTR02 ANN (%)
1993 Q3	89.25	63.13	85.75	54.70	86.40	96.25	83.75	80.15

The result for the cell under study in the Table 22 is very different from Tables 15, 17 and 19. This demonstrates that as with any other statistical method, the presence of influential data points in the input data set has a profound influence on the performance of ANN also.

The results of the main analysis confirm the findings of the preliminary analysis. For discriminant analysis, there is a slight decrease in the classification accuracy of ANN as we move from the Strategy 1 to the Strategy 2. This is because the decrease in the number of attributes, as we move from Strategy 1 to Strategy 2, reduces the possibility of capitalizing on chance and therefore, the accuracy drops down slightly. Thus, the results for discriminant analysis are consistent with what one would expect. On the other hand, the results for ANN are rather interesting because of the following.

- (1) We find that for the Strategy 1, as we move from the attribute set ATTR01 to ATTR02, that is increase the number of attributes, there is a drop in the classification accuracy. This may be because of an increase in the number of parameters, that the network has to estimate, from 28 for ATTR01 to 44 for ATTR02²⁸. Because of this increased number of estimates, we might be observing a drop in classification accuracy.

²⁸ See footnote 24.

- (2) We also find that the classification accuracy of ANN is always a lot lower than that of the discriminant analysis for the Strategy 1 for both attribute sets ATTR01 and ATTR02. But, as we move from the Strategy 1 to Strategy 2, we find that there is a substantial increase in the classification accuracy of ANN (about 15 percentage points for ATTR01 and over 20 percentage points for ATTR02). It is also found that for Strategy 2 the classification accuracy of ANN and discriminant analysis are generally very close; its only in a few cases that the accuracy of ANN is better than that of discriminant analysis.
- (3) The results also demonstrate that the network is sensitive to the number of attributes in the input data. These results show that, the greater the number of attributes that are present in the input data, the poorer is its classification accuracy.

5.3 RESULTS OF THE OUT-OF SAMPLE TEST:

To complete the main analysis, the performance of ANN and discriminant analysis are also investigated on a hold-out sample. The results obtained from this analysis are summarized in Tables 23 through 25. The results again present the same picture as outlined earlier. We can observe that in general there is a decrease in the classification accuracy as the sample size increases. This may be because of the greater likelihood of capitalizing on chance in smaller samples than in the larger ones. In that case, we should observe an increase in the out-of-sample classification accuracy as the sample size increases. The results show that there is an increasing trend in the out-of-sample

classification accuracy as the sample size increases, although the classification accuracy of the network is generally not very good. This confirms the following.

- (1) Some of the high in-sample and out-of-sample classification accuracies appear to be more of a chance rather than real prediction.
- (2) As we find that the out-of sample classification accuracy of ANN increases with the sample size, it suggests that bigger samples are better for training and is consistent with the theory.

Table 23: In-sample and Out-of-Sample Classification Accuracies of ANN and Discriminant Analysis for Sample with 20 Observations:

IN-SAMPLE RESULTS (%)								
	Strategy 1				Strategy 2			
	ATTR01 Disc	ATTR01 ANN	ATTR02 Disc	ATTR02 ANN	ATTR01 Disc	ATTR01 ANN	ATTR02 Disc	ATTR02 ANN
1991 Q1	70.00	65.00	63.00	55.00	69.00	85.00	67.00	90.00
1992 Q2	77.00	75.00	72.00	55.00	73.00	70.00	70.00	85.00
1993 Q3	61.00	60.00	65.00	50.00	65.00	90.00	66.00	55.00
1993 Q4	69.00	70.00	80.00	50.00	65.00	95.00	78.00	95.00
Mean	69.25	67.50	70.00	52.50	68.00	85.00	70.25	81.25
Std Dev	5.67	5.59	6.67	2.50	3.31	9.35	4.71	15.56
MIN	61.00	60.00	63.00	50.00	65.00	70.00	66.00	55.00
MAX	77.00	75.00	80.00	55.00	73.00	95.00	78.00	95.00
OUT-OF-SAMPLE RESULTS (%)								
	Strategy 1				Strategy 2			
	ATTR01 Disc	ATTR01 ANN	ATTR02 Disc	ATTR02 ANN	ATTR01 Disc	ATTR01 ANN	ATTR02 Disc	ATTR02 ANN
1991 Q1	60.50	54.88	55.29	54.88	60.27	53.20	57.27	54.88
1992 Q2	71.32	60.92	66.76	48.28	72.24	51.72	60.64	49.43
1993 Q3	55.27	53.64	61.27	51.82	59.67	80.91	59.00	53.64
1993 Q4	65.32	56.36	75.29	51.82	61.29	77.27	70.67	74.55
Mean	63.10	56.45	64.65	51.70	63.37	65.78	61.90	58.13
Std Dev	5.92	2.75	7.36	2.34	5.16	13.39	5.20	9.70
Min	55.27	53.64	55.29	48.28	59.67	51.72	57.27	49.43
Max	65.32	60.92	75.29	54.88	72.24	80.91	70.67	74.55

Table 24: In-sample and Out-of-Sample Classification Accuracies of ANN and Discriminant Analysis for Sample with 60 Observations:

IN SAMPLE RESULTS (%)								
	Strategy 1				Strategy 2			
	ATTR01 Disc	ATTR01 ANN	ATTR02 Disc	ATTR02 ANN	ATTR01 Disc	ATTR01 ANN	ATTR02 Disc	ATTR02 ANN
1991 Q1	71.25	56.67	72.75	55.00	67.33	73.33	69.75	75.00
1992 Q2	71.33	58.33	71.00	53.33	70.00	81.67	69.00	86.67
1993 Q3	59.00	58.33	68.25	50.00	67.67	93.33	67.33	50.00
1993 Q4	71.00	75.00	79.67	60.00	70.00	96.67	76.50	96.67
Mean	68.15	62.08	72.91	54.58	68.75	86.25	70.65	77.09
Std Dev	5.28	7.49	4.21	3.61	1.25	9.31	3.49	17.42
Min	59.00	56.67	68.25	50.00	67.33	73.33	67.33	50.00
Max	71.33	75.00	79.67	60.00	78.00	96.67	76.5	96.67
OUT-OF-SAMPLE RESULTS (%)								
	Strategy 1				Strategy 2			
	ATTR01 Disc	ATTR01 ANN	ATTR02 Disc	ATTR02 ANN	ATTR01 Disc	ATTR01 ANN	ATTR02 Disc	ATTR02 ANN
1991 Q1	69.33	57.14	69.33	42.86	61.29	45.24	63.75	40.48
1992 Q2	65.67	70.21	70.00	59.57	65.50	72.34	69.00	72.34
1993 Q3	60.00	56.67	61.29	61.67	62.67	81.67	63.33	38.33
1993 Q4	69.24	54.29	72.14	57.14	69.29	87.14	74.27	90.00
Mean	66.06	59.58	68.19	55.31	64.69	71.60	67.59	60.29
Std Dev	33.80	6.23	4.12	7.36	3.06	16.11	4.46	21.81
Min	60.00	54.29	61.29	42.86	61.29	45.24	63.75	38.33
Max	69.33	70.21	72.14	61.67	69.29	87.14	74.27	90.00

Table 25: In-sample and Out-of-Sample Classification Accuracies of ANN and Discriminant Analysis for Sample with 80 Observations:

IN SAMPLE RESULTS (%)								
	Strategy 1				Strategy 2			
	ATTR01 Disc	ATTR01 ANN	ATTR02 Disc	ATTR02 ANN	ATTR01 Disc	ATTR01 ANN	ATTR02 Disc	ATTR02 ANN
1991 Q1	65.25	55.00	67.00	57.57	67.00	67.50	68.75	70.00
1992 Q2	60.00	50.00	61.50	55.00	60.25	77.50	59.00	81.25
1993 Q3	58.25	56.25	62.25	51.25	59.75	90.00	60.00	50.00
1993 Q4	65.25	50.00	63.25	51.25	60.00	97.50	58.50	97.50
Mean	62.19	52.81	63.50	53.77	61.75	83.13	61.56	74.69
Std Dev	3.12	2.85	2.11	2.68	3.04	11.51	4.18	17.28
Min	58.25	50.00	61.50	51.25	59.75	67.50	58.50	50.00
Max	65.25	56.25	67.00	57.57	67.00	97.50	68.75	97.50
OUT-OF-SAMPLE RESULTS (%)								
	Strategy 1				Strategy 2			
	ATTR01 Disc	ATTR01 ANN	ATTR02 Disc	ATTR02 ANN	ATTR01 Disc	ATTR01 ANN	ATTR02 Disc	ATTR02 ANN
1991 Q1	66.00	63.64	63.75	63.64	64.00	63.64	62.25	59.09
1992 Q2	58.50	49.63	58.50	40.74	58.25	66.67	55.00	70.37
1993 Q3	58.25	54.00	60.25	54.00	61.50	88.00	59.75	56.00
1993 Q4	60.25	46.00	61.00	46.00	57.50	92.00	55.50	92.00
Mean	60.75	53.32	60.88	51.10	60.31	77.58	58.12	69.37
Std Dev	3.13	6.60	1.89	8.65	2.61	12.55	3.01	14.12
Min	58.25	46.00	58.50	40.74	58.25	63.64	55.00	56.00
Max	66.00	63.64	63.75	63.64	64.00	92.00	62.25	92.00

The results presented in Sections 5.1, 5.2, and 5.3 suggest that prior to using ANN, a careful investigation of the sample size and the nature of attributes that are used for training the network must be performed. This is because ANN is found to be sensitive to the attributes that are used for training. The next section presents a comparison of the findings of the current study with some of the previous studies.

5.4 COMPARISON WITH PREVIOUS STUDIES:

Table 26 presents the comparison of the current study with previously done studies.

Table 26: Comparison of Current Study with Previous Studies:

Study	Learning Algorithm	Data	Network Performance
Halbert White (1988)	Backpropagation	IBM Daily Stock Returns	NOT GOOD
Kamijo & Tanigawa (1990)	Backpropagation	Weekly Index Returns	GOOD
Kimoto <i>et al.</i> (1990)	Backpropagation	Sharp Price Rises	GOOD
Swales & Yoon (1992)	Backpropagation	Qualitative Information	GOOD
Wong, Wang, Goh & Quek (1992)	Backpropagation	Accounting Attributes	SATISFACTORY
Kryzanowski, Galler & Wright (1993)	Boltzman Machine	Accounting Information	GOOD
Olson & Chou (1993)	Backpropagation	Index Returns	NOT GOOD
Current Study	Backpropagation	Attributes Modelled after Value Line used Attributes and from the Anomaly Literature	NOT GOOD

The table above shows that the current study is not the first one to have found problems with ANNs for security market applications indicating that the findings of the current study cannot be considered as a mere coincidence. Also as shown in previous sections, the results of the study are quite consistent for the entire analysis. Halbert White (1988) found that network has a poor out-of-sample prediction power in his study which illustrates the search for previously undetected regularities that might exist in the historical asset price data using an ANN. Later in 1993, Olson and Chou found similar results while examining the non-linear relationships that might exist in the cross correlation patterns in index returns for three major U.S. stock exchanges. Both of these studies used only one attribute, stock returns and index returns respectively, for training the network and found the same results. Coming to the current study, even though a variety of attributes were used, capturing a wide range of information along with network testing done over several randomly generated samples of varying sizes and varying number of attributes, same results were found. This suggests that usefulness of ANNs may be limited for the applications that do not involve a careful analysis of data and choice of the attributes. The conclusions that can be drawn from this study are presented in the next section.

5.5 CONCLUSIONS:

The results presented in Sections 5.1 through 5.3 clearly show that some thinking and careful investigation of attributes and data used for network training is required before ANN can be used. The conclusions that can be drawn from this study are

presented as follows.

The results from the analysis of the study suggest that a blind, black box approach to ANN may not be a very useful. Some pragmatism, knowledge of data sets and nature of problem is required before this technique could be effectively used. Some of the main findings of the study are as follows. (1) The network is found to be sensitive to the attributes that are present in the input data. As demonstrated in Section 5.2, as the number of attributes increase (from 6 to 10) the classification accuracy of ANN drops. However, when the attribute sets are constructed based on discriminating power of the individual attributes in each set, it reduces the number of attributes that are used and we observe that the classification accuracy of ANN increases. This happens because as the number of attributes decrease, it rapidly decreases the number of parameters that ANN has to estimate and leads to a better classification accuracy. (2) Discriminant analysis, as we know, is not heavily affected by the redundant attributes, at least for the in-sample case and therefore, performs much better than ANN in this case. (3) When redundancies in the attribute sets are removed, we find that ANN performs very similar to discriminant analysis; sometimes a bit better and sometimes a bit worse. The slight advantage that ANN may have over discriminant analysis may be explained by the difference in the principles on which the two techniques operate. Since the present study addresses a problem that deals with classification, Artificial Neural Networks would be best suited. Therefore, one would expect that the technique that optimizes classification would perform better. (4) Like any other classical statistical method, the performance ANN is also found to be heavily influenced by the presence of the influential data points in the

input data set.

The findings of the present study strongly suggest that ANN, especially for security market applications, is not a technique in which one could naively supply a data set to the network and expect good results to automatically flow out. In reality, one may need to actually massage the data a bit before it could be supplied to the network. Therefore, more caution should be exercised before accepting this tool for financial applications.

Despite some of the shortcomings of applying ANN for the applications such as the one in the present study, it can be said that ANN does compare favourably with multiple discriminant analysis but with some caveats. Thus, an application that takes into consideration, issues such as sample size requirements for good network training, choice of attributes, and the presence of influential data points in the input data set and uses classical statistical techniques such as discriminant analysis to enhance the performance of ANN, may result in a potentially good system that might be better capable of identifying stock winners from stock losers. In fact, the results of the present study give indications of how such a system may perform.

6. BENEFITS, LIMITATIONS AND SUGGESTIONS FOR FUTURE RESEARCH

6.1 BENEFITS:

The proposed study offers several benefits that are outlined as follows.

- (1) The first benefit is academic in nature. As this is the first study that tests the stability of ANN, it enhances the understanding of the suitability of ANNs for financial applications, stock selection in particular.
- (2) The study also provides an extensive comparison of the performances of Artificial Neural Network and Multiple Discriminant Analysis.
- (3) The present study is the first study in the context of stock selection to analyze the robustness of ANN to samples of varying sample sizes and the choice of attribute selection.
- (4) And finally, the present study is the first study to investigate the performance of ANN in the presence of influential data points.

6.2 LIMITATIONS:

As with any other research, this study also has some limitations. The first limitation arises due to lack of bigger sample size and ability to test over a long period.

limitation arises due to lack of bigger sample size and ability to test over a long period. For example, the data from Stock Guide database is available only from 1985 onwards. In addition to this, the study deals with a total sample of about 300 stocks per year. This may not be sufficient for an ANN application.

As mentioned earlier, the literature provides very little guidance for deciding the number of hidden layers and the number of neurons in the hidden layer. It is thus possible that none of the architectures that are used in the study are appropriate for the chosen application and further research may be warranted for deciding the appropriate architecture for such problems.

6.3 SUGGESTIONS FOR THE FUTURE RESEARCH:

In the view of the results obtained in the analysis, the present study can be extended in at least following four areas.

The results in Table 18 demonstrate that the change in accuracy of ANN from the Strategy 1 to Strategy 2 is significant - 15% increase for ATTR01 and over 25% increase for ATTR02. The direction of this change is exactly opposite to that of discriminant analysis. Future studies may be designed to address this issue to pre-select the choice of attributes based on their discriminating power.

As demonstrated in Swales and Yoon (1992), use of additional attributes constructed from qualitative data should be considered. In addition to this, attributes representing industry dummies may also be used. Future studies could also use a

different approach to sample creation, such as - the stock winner sample may be constructed by using stocks from quintiles 1, 2 & 3 instead of just quintile 1 and stock loser sample may be constructed from quintiles 3, 4 & 5 instead of just quintile 5.

The third suggestion stems from the limitation of the present study in having lack of data for longer periods. Future studies should try to use bigger sample size and longer periods for the network training. Also, the network should be tested for its performance by training the network across the period of the study and subsequently testing it.

As demonstrated by the present study, further research is warranted for deciding the network architecture. In addition to this, it would be interesting to see how a self organizing neural network would perform in the context of applications such as the one in the present study as self organizing neural networks can learn to detect regularities and **correlation** in their input and adapt their future responses to that input accordingly. It is there ability to detect the correlation in the input data that makes them interesting for such applications.

REFERENCES

- Bailey, D.L. and D.M. Thompson "Developing Neural Network Applications." *AI Expert*, sep 1990, pp.34-41.
- Bernhard, A. "Value Line methods of evaluating common stocks", Arnold Bernhard, New York.
- Black, F. "Yes, Virginia, there is hope: Tests of the Value Line ranking system", *Financial Analysts Journal*, 1973, vol. 29, pp.10-14.
- Casimir C. "Casey" Klimasauskas "Applying Neural Networks": in Trippi, R.R. and E. Turban *Neural Networks in Finance and Investing*, Chicago, Probus Publishing Co., 1993.
- Coats P.K. "Why Expert Systems Fail", *Financial Management*, Autumn 1988, pp.77-86.
- Coleman K.G., T.J. Graettinger, and W.F. Lawrence "Neural Networks for Bankruptcy Prediction: The Power to Solve Financial Problems", *AI Review*, July/Aug 1991, pp.48-50.
- Collins, E., S. Ghosh, and C. Scofield " An Application of a Multiple Neural Network Learning System to Emulation of Mortgage Underwriting Judgements", *Proceedings of the IEEE International Conference on Neural Networks*, July 1988, pp.II459-466.
- Conover, W.J. "Practical Nonparametric Statistics", 2ed., *John Wiley & Sons*, New York, 1980.
- Copeland, T.E. & D. Myers "The Value Line Enigma (1965-1978): A Case Study of Performance Evaluation Issues", *Journal of Financial Economics*, Nov 1982, vol. 10, no. 3, pp.289-322.
- Dutta S. and S. Shekhar " Bond Rating: A Non-Conservative Application of Neural Networks", *Proceedings of the IEEE International Conference on Neural Networks*, July

1988, pp.II443-II450.

Forester S., A. Prihar, and J. Schmitz "Back to the Future", *Canadian Investment Review*, Winter 1994/95, pp.9-13.

Grossberg, S. "Embedding fields: A Theory of Learning with Physiological Implications", *Journal of Mathematical Psychology* 6, 1969, pp.209-239.

Hawley, D.D., J.D. Johnson, & D. Raina "Artificial Neural Systems: A new tool for Financial Decision-Making", *Financial Analysts Journal*, Nov-Dec 1990, pp.63-72.

Huberman G. & S. Kandel "Value Line Rank and Firm Size", Oct 1987, vol. 60, no. 4, pp.577-590.

Kaplan, R.S. & R.L. Weil "Risk and The Value Line Contest", *Financial Analysts Journal*, July-Aug 1973, vol. 29, no.4, pp.56-62.

Kamijo, K., & T. Tanigawa, "Stock Price Pattern Recognition: A Recurrent Neural Network Approach", *Proceedings of the IEEE International Joint Conference on Neural Networks*, San Diego, 1990, pp.I215-221.

Kimoto ,T., K. Asakawa, M. Yoda, & M. Takeoka "Stock Market Prediction System with Modular Neural Networks", *Proceedings of the IEEE International Joint Conference on Neural Networks*, San Diego, 1990, pp.I1-I6.

Klein, M. and L.B. Methlie "Expert Systems: A Decision Support Approach With Applications in Management and Finance, *Addison Wesley Pub. Co. Inc.*, 1990.

Kryzanowski, L., M. Galler, & D.W. Wright "Using Artificial Neural Networks to Pick Stocks", *Financial Analysts Journal*, July-Aug 1993, pp.21-27.

Li, Bo "Price Related Common Stock Anomalies: Canadian Evidence", MMS Thesis, School of Business, Carleton University, 1994.

Odom, M., & R. Sharda "A Neural Network Model for Bankruptcy Prediction", *Proceedings of the IEEE International Conference on Neural Networks*, San Diego, 1990, pp.II163-II168.

Olson, D.O., & N.T. Chou "Cross Correlations and Neural Network forecasts of Daily Stock Returns", *Meetings of the Northern Finance Association*, Halifax, Sep 1993.

Reilly, D.L. Collins E., C. Scofield, and S. Ghosh "Risk Assessment of Mortgage Applications with a Neural Network System: An Update as the Test Portfolio Ages", *Proceedings of the IEEE International Conference on Neural Networks*, July 1991, pp.II479-II482.

Ritter J.R., "The Long Term Performance of Initial Public Offerings", *Journal of Finance*, 1991, pp.3-27.

Raghupati, W., L.L. Schkade, & B.S. Raju "A Neural Network Model for Bankruptcy Prediction", *Proceedings of the IEEE 24th Annual Hawaii International Conference of Systems Sciences*, 1991.

Salchenberger, L.M., E.M. Cinar, and N.A. Lash "Neural Networks: A New Tool for Predicting Thrift Failures", *Decision Sciences*, July/Aug 1992, vol. 23, no. 4, pp.899-916.

Shelton J.P. "The Value Line Contest: A Test of the Predictability of Stock Price Changes", *Journal of Business*, July 1967, vol. 40, pp.251-269.

Silverman, B. "Expert Systems for Business", *Addison Wesley Pub. Co. Inc.*, 1987.

Stickel, S.E. "The Effect of Value Line Rank Changes on Common Stock Prices", *Journal of Financial Economics*, March 1985, vol. 14, no. 1, pp.121-144.

Steven, James "Applied Multivariate Statistics for the Social Sciences", 2nd ed, *Lawrence Erlbaum Associates Pub*, New Jersey, 1992.

Surkan A.J. and J.C. Singleton "Neural Networks for Bond Rating Improved by Multiple Hidden Layers" *Proceedings of the IEEE International Conference on Neural Networks*, 1990, pp.II163-II168.

Swales, G.S. Jr., & Y. Yoon "Applying Artificial Neural Networks to Investments Analysis", *Financial Analysts Journal*, Sep-Oct 1992, vol. 48, no. 1, pp.78-80.

Tam, K.Y. and M.Y. Kiang "Managerial Applications of Neural Networks: The Case of Bank Failure Predictions", *Management Science*, July 1992, vol. 38, no. 7, pp.926-947.

White, Halbert "Economic Prediction Using Neural networks: The Case of IBM Daily Stock Returns", *Proceedings of the IEEE International Conference on Neural Networks*, July 1988, pp.II451-II458.

Winograd, T. and F. Flores "Understanding Computers and Cognition", *Addison Wesley Pub. Co. Inc.*, 1986.

Wong, F.S., P.Z. Wang, T.H. Goh, & B.K. Quek "Fuzzy Neural Systems for Stock Selection", *Financial Analysts Journal*, Jan-Feb 1992, vol. 48, no. 1, pp.47-52.

APPENDIX A

STOCK GROUPINGS ACCORDING TO THE STOCKGUIDE DATABASE Ver 4.9

10	Integrated Mines	55	Machinery & Structural Steel
11	Mining	56	Fabricating & Engineering
13	Gold and Precious metals	57	Speciality Industries
14	Mining Exploration - No Production	60	Steel Products
16	Integrated oils	61	Steel Distributors & Servicing
18	Oils & Gas Producers	62	Transportation Equipment
21	Oil & Gas, Mining, or Forest Services	63	Industrial Contractors
23	Pipelines	70	Technology - Hardware
26	Paper & Forest Products	71	Technology - Software
27	Building Materials	75	Transportation & environment
31	Autos & Parts	77	Gas & Electrical Utilities
32	Breweries & Beverages	79	Telephone Utilities
34	Distilleries	80	Broadcasting
36	Food Processing	81	Cable & Entertainment
37	Household Goods	83	Publishing & Printing
39	Tobacco	85	Real Estate & Construction
41	Department Stores	90	Banks & Trusts
42	Food Stores	92	Leasing, Financing, Mortgages
43	Biotechnology	95	Investment Companies & funds
44	Hospitality	96	Insurance
45	Restaurants	97	Financial Management Companies
46	Speciality Stores	98	Conglomerates
47	Wholesale Distributors & Importers		
50	Business Services		
52	Chemicals & Fertilizers		
54	Electrical & Electronic Products		

APPENDIX B

OVERVIEW OF MATLAB NEURAL NETWORK TOOLBOX:

MATLAB stands for *matrix laboratory*. It is a technical computing environment developed by MathWorks Inc. for high-performance numeric computation and visualization. The basic data element for MATLAB is a matrix that does not require dimensioning. In addition to this, MATLAB also features a family of application-specific solutions that they call *toolboxes*. These toolboxes are comprehensive collections of MATLAB functions that extend the MATLAB environment in order to solve particular classes of problems. Neural Network Toolbox is one such toolbox that contains functions for designing and simulating neural networks. The functions in the toolbox fall into the categories of (1) transfer function specification for network simulation, (2) learning rules for updating network parameters, and (3) training functions for training networks on data. This software was chosen because of its easy-to-use environment along with its comprehensiveness and flexibility in creating customized applications.

APPENDIX C

LEARNING IN NEURAL NETWORKS:

Learning in neural networks refers to the process by which the network adjusts its weights and biases in order to achieve some desired network behaviour. Learning method is the most important distinguishing factor in various neural networks. There are two distinct types of learning techniques

- Supervised
- Unsupervised

Supervised learning is the process in which changes in a network's weights and biases are due to the intervention of any external teacher. The teacher typically provides output targets. The main idea in supervised learning is that the system is repeatedly given facts about various cases, along with the expected outputs. The system uses the learning method to adjust the weights in order to produce outputs close to what are expected. Supervised learning has an analogy of a child learning new lessons with the help of a teacher. The most popular learning algorithm used for supervised learning which has found wide application in many areas including finance and accounting, is backpropagation. This is a learning rule in which weights and biases are adjusted by error derivative (delta) vectors backpropagated through the network. Backpropagation is commonly applied to feedforward multilayered networks. Sometimes this is also called as the generalized delta rule.

On the other hand, unsupervised learning is a learning process in which changes

in a network's weights and biases are not due to the intervention of any external teacher. Commonly changes are a function of the current network input vectors, output vectors, and previous weights and biases. This type of learning conjures up the image of a child who learns by repeated exposure to its environment.

Most of the neural networks employ the supervised learning method because it is of more immediate use for problem solving. On the other hand, unsupervised learning has the potential for creating intelligent systems more powerful than what other methods could accomplish.

GLOSSARY

Annealing or Simulated Annealing :

This is the discrete optimization method that uses a gradually decreasing amount of random noise introduced into the weights and/or input signals of a network as training progresses while searching some problem space for globally optimal solution.

Artificial Neural System :

It is another term that is used for the neural networks which distinguishes between biological and non-biological neural systems.

Backpropagation :

The error backpropagation rule is a two-pass weight-learning algorithm for neural networks. Error is first estimated during a forward pass through the network, then the measured error is propagated backward through the network while connection weights are adjusted to reduce the error.

Boltzmann Machine :

It is a distributed parallel processing algorithm developed by Geoffrey Hinton and Terrence Sejnowski based on statistical mechanics that uses simulated annealing to find stable solution states to problems.

Input Vector :

It is the pattern that is presented to the input neuronodes of a network.

Network Paradigm :

It includes the features of a neural network solution that makes it different from

other neural networks, typically including the connection scheme, the transfer function, and the learning law of the network. The specific sizes of the network layers are not considered as distinguishing features

Neural Network :

It is a type of data processing system whose physical architecture is inspired by the structure of biological neural systems. It is analogous to one of many neurons, dendrites, axons, synapses, and other biological structures making up the neural systems of living beings.

Neuron :

It is the basic processing element of a neural network. Artificial neurons are abstractions of biological neurons.

Output Vector:

The output vector is the output of a neural network. Each element of the output vector is the output of a neuron.

Threshold :

It is the activation level that the neuronode must attain before its output rises above zero.

Supervised Learning:

A learning process in which changes in a network's weights and biases are due to the intervention of any external teacher. The teacher typically provides output targets.

Transfer Function:

It is the function which maps a neuron's (or layer's) inputs to its outputs.