

Universiti Teknologi MARA

**Comparative Revision of Artificial Intelligent
Methods for Discovering Gene Expression**

Mohd Dzulkarnain B. Zainal Ahbiddin

Thesis submitted in fulfillment of the requirements for
Bachelor of Science (Hons) in Information Technology
Faculty of Information Technology and
Quantitative Science

May 2007

DECLARATION

I certify that this thesis and the system to which it refers, are the product of my own individual work, and that any ideas from the work of other people, published or otherwise, are fully acknowledged in accordance with the standard referencing practices of the discipline. I also acknowledge the helpful guidance and support of my supervisor, Pn.Shuzlina bt. Abd Rahman.

I authorize the University of Technology MARA to lend this thesis to other institutions or individuals for the purpose of scholarly research. I further authorize the University of Technology MARA to reproduce the thesis by photocopying or other means, in total or in part, at the request of institutions or individuals for the purpose of scholarly research.

.....
MOHD DZULKARNAIN B. ZAINAL AHBIDDIN

2004618559

May 04 2007

ACKNOWLEDGEMENTS

In the name of ALLAH, who is the Most Gracious, Most Merciful and HIM alone is worthy of all praise. To HIM all the praise go and to HIM all the thankfulness of giving me the opportunity to live day in and day out.

It is with great honor to have the opportunity to express my highest gratitude towards my supervisor, Pn. Shuzlina bt Abd Rahman for giving me the encouragement and mostly support throughout the duration of this study. All the hours spent in discussions were highly remembered as valuable experiences to guide more for my future. Without the expertise of yours, this project would not be successful as I hoped it will be.

In inclusion to that list are Dr. Azlinah and Pn. Rozianawaty that throughout the years never seems tired giving valuable advice to help improve the project and wisely conducted the Information System Project. With no exception, a millions of thanks goes to both of my parents who never give up teaching me to become the person I am today.

To all my friends whom I shared and discuss expertise and experiences until today, many thanks. It has been a bumpy and tricky road. To whom I failed to mentioned, who indirectly or directly contributed to this project, I thank you so very much.

ABSTRACT

Gene expression analysis is one of the studies in bioinformatics. There are many methods and approaches use in gene expression analysis. Some methods that are currently being used, such as fuzzy ART, neural network, and Bayesian method were used in gene expression analysis. The problem that occurred in supervised learning is that the output and error rates that been provided were momentous. The reason in conducting this research is to determine the best methods, between two approaches for gene expression analysis. For this research, the approach used is supervised learning and the methods that were used are multi layer feedforward and k-Nearest Neighbour. The methodology that will be use for this research are knowledge acquisition, implementation that consists of three phase; experiment, result and analysis, experiments and observation, and evaluation and findings. After a series of experiments, the multi layer feedforward is the better method in determining gene expression especially protein genes rather than k-Nearest Neighbour. It is because the accuracy of the output is more precise and can be used for further analysis. The presentation of multilayer feedforward is clear and well-defined. The accuracy of the results is important for usage of others. This research can be a good reference for the advancement and development of gene expression analysis.

Table of Contents

CONTENT	PAGE
Declaration	
Approval	
Acknowledgement	
Abstract	
Chapter 1 - Introduction	
1.1 Description of Background Study	1
1.2 Problem Statement	2
1.3 Project objective	
1.4 Project Scope	
1.5 Significance of Project	3
Chapter 2 – Literature Reviews	
2.0 Literature Reviews	4
2.1 Introduction	
2.2 Supervised and Unsupervised Learning Approaches in Bioinformatics	
2.3 Current Methods of Gene Prediction, Their Strengths and Weakness	5
2.4 Challenges in Bioinformatics	
2.5 Gene Expression Analysis Using Fuzzy ART	6
2.6 Multilayer Feedforward Neural Networks	
2.7 Supervised Neural Networks	7
2.8 Predictive Gene Clusters with Supervised Algorithms	