

TEXT SPAM MESSAGES CLASSIFICATION USING ARTIFICIAL  
IMMUNE SYSTEM (AIS) ALGORITHMS

Nurul Fadhilah Binti Sulaiman  
(Matric No. 3130271)

Thesis submitted in fulfillment for the degree of  
MASTER OF SCIENCE

Faculty of Science and Technology  
UNIVERSITI SAINS ISLAM MALAYSIA  
Nilai

June 2016

## AUTHOR BIOGRAPHY

Nurul Fadhilah binti Sulaiman (3130271) was born on the August, 5 1990. She is currently residing at No. 50, Felcra Keruak, 22010 Jerteh, Terengganu. She previously was a student of Universiti Sains Islam Malaysia (USIM) and obtained Bachelor of Computer Science (Information Security and Assurance) from the Faculty of Science and Technology. Presently she studies Master of Science within the same institution. In this Masters journey, she managed to publish two papers related to this field titled “Integrated Mobile Spam Model Using Artificial Immune System Algorithms” in Knowledge Management International Conference 2014 (KMICe2014) and another paper, “A New SMS Spam Detection Method Using Both Content-Based and Non Content-Based Features”, in 2015 2nd International Conference on Communication and Computer Engineering (ICOCOE’2015). She can be contacted via fadhilahsulaiman90@gmail.com.



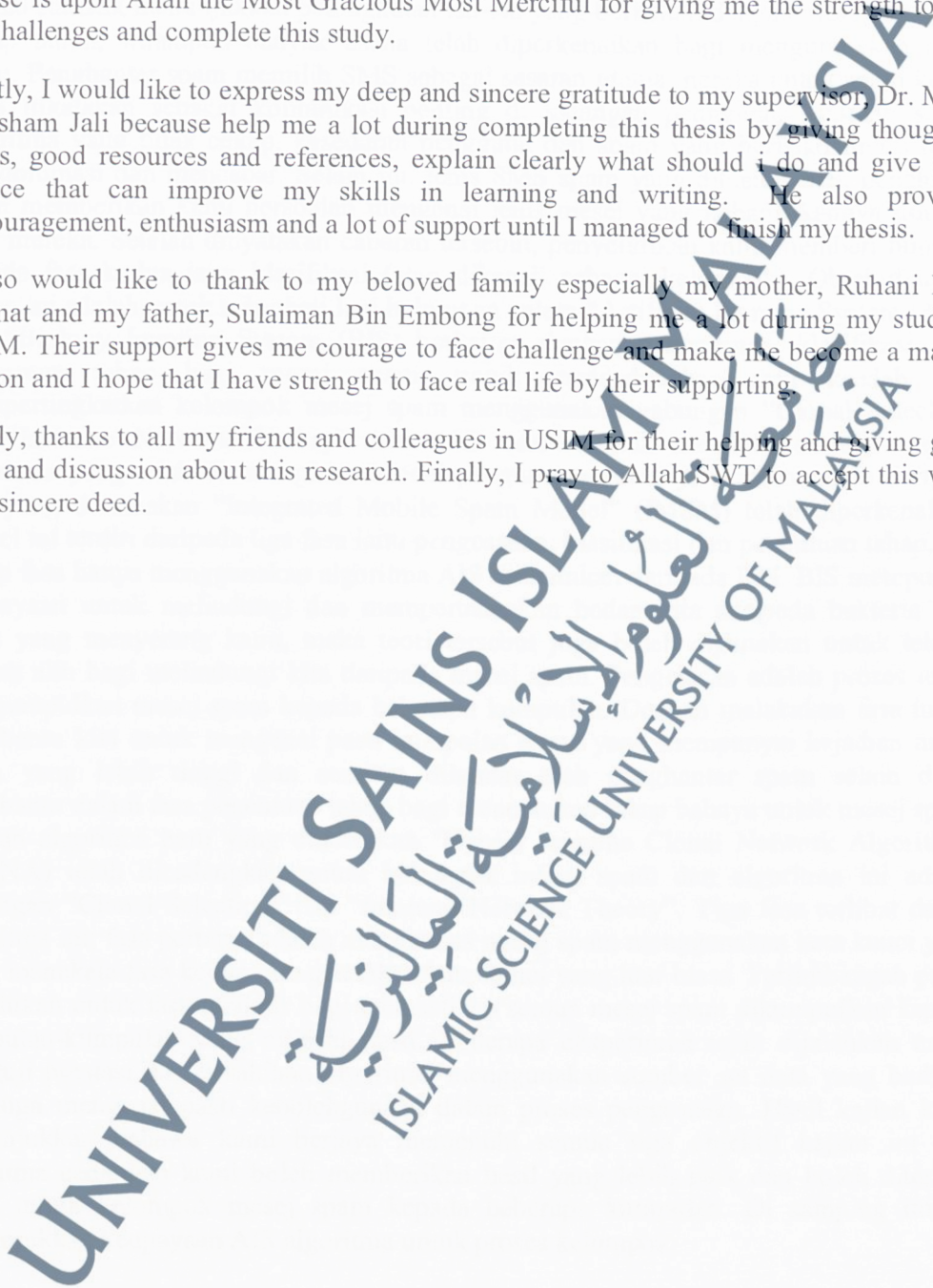
## ACKNOWLEDGEMENTS

Praise is upon Allah the Most Gracious Most Merciful for giving me the strength to face all challenges and complete this study.

Firstly, I would like to express my deep and sincere gratitude to my supervisor, Dr. Mohd Zalisham Jali because help me a lot during completing this thesis by giving thoughtful ideas, good resources and references, explain clearly what should i do and give good advice that can improve my skills in learning and writing. He also provided encouragement, enthusiasm and a lot of support until I managed to finish my thesis.

I also would like to thank to my beloved family especially my mother, Ruhani Binti Mamat and my father, Sulaiman Bin Embong for helping me a lot during my study in USIM. Their support gives me courage to face challenge and make me become a mature person and I hope that I have strength to face real life by their supporting.

Lastly, thanks to all my friends and colleagues in USIM for their helping and giving good idea and discussion about this research. Finally, I pray to Allah SWT to accept this work as a sincere deed.



## ABSTRAK

Masalah mesej spam agak membimbangkan terutama bagi pengguna telefon mudah alih kerana statistik menunjukkan peningkatan isu-isu yang berkaitan dengan mesej teks spam setiap tahun, walaupun banyak usaha telah diperkenalkan bagi mengurangkan risiko spam. Penghantar spam memilih SMS sebagai sasaran utama mereka untuk spam kerana SMS dikatakan sebagai komunikasi penting di kalangan pengguna. Masalah seperti algoritma yang tidak cekap, kesedaran pengguna dan spam yang berisiko tinggi masih mendominasi dan mencabar. Selain itu, jenis SMS spam yang dihantar oleh penghantar spam memberikan kami persoalan mengenai jenis mesej yang kebanyakannya dihantar oleh mereka. Setelah dinyatakan cabaran tersebut, penyelidikan kami memberi tumpuan kepada fasa kedua iaitu klasifikasi (atau dikenali sebagai kelompok). Objektif utama kajian ini adalah untuk mengkaji lagi hubungan antara “Artificial Immune System (AIS)” dan “Biology Immune System (BIS) berkaitan dengan pengesanan, klasifikasi dan penentuan tahap bagi mesej spam, untuk mencadangkan satu kaedah bagi mempertingkatkan kelompok mesej spam menggunakan gabungan “Clonal Selection” dan “Immune Network Theory” dan akhir sekali untuk menjalankan dan menilai algoritma yang telah dicadangkan. Model pengurusan spam yang diilhamkan daripada BIS yang dinamakan “Integrated Mobile Spam Model” (IMSM) telah diperkenalkan. Model ini terdiri daripada tiga fasa iaitu pengesanan, klasifikasi dan penentuan tahap, dan setiap fasa hanya menggunakan algoritma AIS diilhamkan daripada BIS. BIS mempunyai keupayaan untuk melindungi dan mempertahankan badan kita daripada bakteria atau virus yang menyerang kami, maka teori tersebut juga boleh digunakan untuk telefon mudah alih bagi melindungi kita daripada mesej spam. Pengelasan adalah proses untuk mengumpulkan mesej spam kepada beberapa kumpulan. Dengan melakukan fasa ini, ia membantu kita untuk mengenal pasti kumpulan mana yang mempunyai kejadian mesej spam yang lebih tinggi dan sentiasa dihantar oleh penghantar spam selain dapat membantu dalam fasa penentuan tahap bagi menentukan tahap bahaya untuk mesej spam. Sebuah algoritma baru yang dinamakan “Hybrid Immune Clonal Network Algorithm” (HICNA) telah dicadangkan untuk kelompok mesej spam dan algoritma ini adalah gabungan “Clonal Selection” dan “Immune Network Theory”. Tiga fasa terlibat dalam algoritma ini; fasa pertama adalah mengimbas mesej spam menggunakan kata kunci yang biasa manakala fasa kedua menggunakan kata kunci yang luar biasa. Pertimbangan pakar diperlukan untuk fasa terakhir bagi memastikan semua mesej spam dikumpulkan kepada kumpulan-kumpulan yang dikenal pasti. Beberapa eksperimen telah dijalankan untuk menguji prestasi dan kesahihan algoritma menggunakan sumber set data yang berbeza dan juga mengenal pasti kebolehgunaan dalam proses pengesanan. Hasil kajian kami menunjukkan bahawa kami berjaya memenuhi semua tiga objektif kajian ini dan algoritma cadangan kami boleh memberikan hasil yang lebih baik dan boleh diterima pakai untuk kelompok mesej spam kepada beberapa kumpulan. Di samping itu, ia menunjukkan keupayaan AIS algoritma untuk proses kelompok.

**KATAKUNCI:** Artificial Immune System (AIS), Clonal Selection, Immune Network Theory, Klasifikasi, Kelompok, SPAM, SMS.

## ABSTRACT

The problem of spam messages is quite worrying especially for mobile users because statistics show increasing issues albeit many efforts have been introduced to reduce the risk of spam. Spammers chose SMS as their main target for spamming because SMS is considered as an important communication among them. Problems such as inefficient algorithm, users awareness and high risk of spam are still dominating and challenging. Besides, the varieties of SMS spam sending by spammers giving us a question on the types of messages that are mostly sent by them. Having stated the aforementioned challenges, this research focuses on the second phase which is the classification (or known as clustering). The main objectives of this research are to study the relationship between Artificial Immune System (AIS) and Biology Immune System (BIS) related to spam detection, classification and severity determination, to propose an enhance method for clustering spam messages using the combination of Clonal Selection and Immune Network Theory and lastly to conduct and evaluate the proposed algorithms. A spam management model inspired from the ideology of BIS named Integrated Mobile Spam Model (IMSM) is introduced. This model consists of three phases which are detection, classification and severity determination, and each phase uses only AIS algorithms inspired from BIS. BIS has the capability to protect and defend the body from bacteria or virus that attacks us, so this theory can be applied to the mobile phone to protect from spam messages as well. Classification is the process to cluster spam messages into several groups. By doing this phase, it helps us to identify which group of spam messages that has higher occurrence and is always sent by spammers besides can help in the severity determination phase to determine the level of danger for spam messages. A new algorithm named "Hybrid Immune Clonal Network Algorithm" (HICNA) is proposed for clustering spam messages and this algorithm is a combination of Clonal Selection and Immune Network Theory. Three phases involved in this algorithm; phase one is scanning the spam messages using common keywords while phase two is using uncommon keywords. Expert judgement is needed for the last phase to ensure all spam messages are clustered into identified groups. A number of experiments have been conducted to test the performance and validity of the algorithm using different source of datasets and also to identify its usability in the detection process. The research results show that three defined objectives were fulfilled and the proposed algorithm gives better results in clustering spam messages into several groups. In addition, it shows the capability of AIS algorithm for the clustering process.

**KEYWORDS:** Artificial Immune System (AIS), Clonal Selection, Immune Network Theory, Classification, Clustering, SPAM, SMS.

## MULAKHKHAS AL-BAHTH

مشكلة الرسائل غير المرغوب فيها مقلق خاصة لمستخدمي الهواتف المحمولة الإحصاءات تظهر على زيادة القضايا المتعلقة برسائل غير مرغوب فيها سنويا، وإن كان كثير من الجهود استحدثت للحد من خطر الرسائل غير المرغوب فيها. الاطر اختيار SMS كهدف رئيسي للسبام لأن يقال SMS كخطاب هاما فيما بينهم. مشاكل مثل عدم كفاءة الحوار زمية، وعي المستخدمين ومخاطر عالية من البريد المزج لا تزال المسيطرة وتكن الصعبة. بجانب ذلك، نوعا من البريد المزجج SMS المرسل من الاطر يعطينا سؤال حول أنواع الرسائل التي يتم إرسالها معظم منهم. وبعد أن ذكر التحديات المذكورة آنفا، وهذا نموذج إدارة البريد المزجج مستوحاة من فكر (BIS) Biology Immune System (BIS) يدعو Integrated Mobile Spam Model (IMSM) قد عرض. هذا الودم يكون على ثلاث مراحل وهي كشف وتصنيف وتحديد الخطورة، كل مرحلة ويستخدم حوار زميات (AIS) Artificial Immune System (AIS) الوحيدة مستوحاة من BIS. BIS لديها القدرة على حماية والدفاع جسمنا من البكتيريا أو الفيروسات التي تتاجم لنا، لذلك يمكن تطبيق نظريتها إلى الهاتف المحمول لحماية من الرسائل غير المرغوب فيها كذلك. ويتركز بحثنا على المرحلة الثانية وهو تصنيف (أو المعروف باسم التكتل). التصنيف هو عملية تجميع رسائل البريد المزجج الى عدة مجموعات . عن طريق القيام هذه المرحلة، أننا تساعدنا على تحديد أي مجموعة من الرسائل غير المرغوب فيها التي لديها أعلى حدوثها ويتم إرسالها دائما من الاطر وإلى جانب آخر تساعد في مرحلة تحديد شدة لتحديد مستوى الخطر لرسائل البريد المزجج. تقوم خوارزمية جديدة باسم "Hybrid Immune Clonal Network Algorithm (HICNA)" يقترح لتجميع رسائل البريد المزجج وهاته الخوارزمية هي مزيج من اختيار نسيلي ونظرية الشبكة المناعية. ثلاثة مراحل المشاركة في هذه الخوارزمية. المرحلة الأولى بالاتقاط الرسائل غير المرغوب فيها باستخدام كلمات رئيسية مشتركة مع المرحلة الثانية يستخدم كلمات غير مألوفة. هناك حاجة إلى حكم الإنسان للمرحلة الأخيرة لضمان تتجمع كل الرسائل غير المرغوب فيها إلى مجموعات محددة. وقد أجريت العديدة من التجارب لاختبار أداء وصحة الحوار زمية باستخدام مصادر مختلفة من قواعد البيانات وأيضا تحديد قابليته للاستخدام في عملية الكشف. تظهر نتائج بحثنا إلى أن تمكنا من تحقيق كل الأهداف الثلاثة في هذا البحث وأنظمتنا المقترحة تعطي أفضل نتائج وتكون قابلة للتطبيق لتجميع رسائل البريد المزجج الى عدة مجموعات. بالإضافة إلى ذلك، فإنه يدل على قدرة خوارزمية AIS لعملية التجميع.

الكلمات الرئيسية: (AIS) Artificial Immune System , Clonal Selection , Network Theory Immune

, المجموعات , تصنيف , SMS, SPAM

## TABLE OF CONTENTS

CONTENTS	PAGE
AUTHOR DECLARATION .....	i
AUTHOR BIOGRAPHY .....	ii
ACKNOWLEDGEMENTS .....	iii
ABSTRAK .....	iv
ABSTRACT .....	v
MULAKHKHAS AL-BAHTH .....	vi
TABLE OF CONTENTS .....	vii
LIST OF TABLES .....	ix
LIST OF FIGURES .....	xi
LIST OF APPENDICES .....	xiii
LIST OF ALGORITHMS .....	xiv
ABBREVIATION .....	xv
CHAPTER 1 .....	1
INTRODUCTION .....	1
1.1 BACKGROUND .....	1
1.2 PROBLEM STATEMENTS .....	7
1.3 RESEARCH QUESTIONS .....	9
1.4 RESEARCH OBJECTIVES .....	10
1.5 SCOPE .....	10
1.6 METHODOLOGICAL FRAMEWORK .....	12
1.7 THESIS STRUCTURE .....	15
CHAPTER 2 .....	16
LITERATURE REVIEW .....	16
2.1 BACKGROUND .....	16
2.2 DEFINITION OF TERMS .....	19
2.3 IMMUNE SYSTEM .....	20
2.3.1 BIOLOGICAL IMMUNE SYSTEMS .....	20
2.3.2 ARTIFICIAL IMMUNE SYSTEM .....	28
2.4 THE RELATIONSHIP BETWEEN AIS AND BIS .....	35
2.5 SPAM MESSAGES .....	39
2.5.1 RESEARCH IN SPAM .....	40
2.6 SUMMARY .....	55
CHAPTER 3 .....	56
SPAM MANAGEMENT MODEL .....	56
3.1 BACKGROUND .....	56
3.2 INTEGRATED MOBILE SPAM MODEL .....	56
3.3 CLUSTERING/ CLASSIFICATION USING AIS ALGORITHMS .....	63
3.4 SUMMARY .....	68
CHAPTER 4 .....	69
HYBRID IMMUNE CLONAL NETWORK ALGORITHM .....	69
4.1 BACKGROUND .....	69
4.2 EXPERIMENTS IN DETECTION PHASE .....	69

4.2.1	PROCEDURES AND METHODS .....	70
4.2.2	ENHANCING AIS DETECTION .....	72
4.3	EXPERIMENTS IN CLASSIFICATION (OR CLUSTERING) PHASE .....	89
4.3.1	HYBRID IMMUNE CLONAL NETWORK ALGORITHM (HICNA) .....	89
4.3.2	CLUSTERS OF SPAM.....	96
4.3.3	CLASSIFICATION USING HICNA .....	99
4.4	DETECTION AND CLUSTERING USING WEKA .....	102
4.5	SUMMARY .....	104
CHAPTER 5 .....		105
EVALUATIONS, RESULTS AND DISCUSSION .....		105
5.1	BACKGROUND .....	105
5.2	PROOF OF CONCEPT .....	105
5.2.1	EXPERIMENT 1 (FORMULA IDENTIFICATION AND UNDERSTANDING) .....	106
5.2.2	EXPERIMENT 2 (DETECTION USING LABEL AND UNLABELLED MESSAGES).....	121
5.2.3	EXPERIMENT 3 (CLASSIFICATION CONCEPT USING WEKA).....	128
5.2.4	EXPERIMENT 4 (PERFORMANCE OF ALGORITHM FOR DETECTION AND CLASSIFICATION).....	131
5.2.5	EXPERIMENT 5 (TRAINING AND TESTING DATASET FOR DETECTION AND CLASSIFICATION).....	138
5.3	DISCUSSION PROOF OF CONCEPT .....	151
5.4	PROOF OF PERFORMANCE .....	155
5.4.1	DETECTION PHASE.....	155
5.4.2	CLASSIFICATION PHASE .....	180
5.5	DISCUSSION PROOF OF PERFORMANCE .....	194
5.6	SUMMARY .....	198
CHAPTER 6 .....		199
CONCLUSION .....		199
6.1	BACKGROUND .....	199
6.2	ACHIEVEMENTS .....	199
6.3	LIMITATIONS .....	201
6.4	FUTURE WORKS .....	202
6.5	SUMMARY .....	203
REFERENCES .....		204
APPENDICES .....		213

## LIST OF TABLES

CONTENTS	PAGE
Table 1. 1: Research Activities in each phase .....	13
Table 1. 2: Summary on the research activities.....	14
Table 2. 1: Functions of Granulocytes White Blood Cells.....	23
Table 2. 2: Innate and Adaptive Immune System.....	26
Table 2. 3: The summary of AIS with BIS.....	39
Table 2. 4: Differences between Email and SMS .....	40
Table 2. 5: Existing researches related to the performance of classifiers.....	43
Table 2. 6: Existing studies related to the performance of clustering algorithms.....	51
Table 3. 1: The implementation of AIS in the proposed model.....	62
Table 4. 1: Number of messages in each dataset.....	71
Table 4. 2: The difference between spam and ham messages.....	72
Table 4. 3: Spam messages containing special characters using numbers.....	73
Table 4. 4: Different features in each algorithm .....	74
Table 4. 5: Characteristics of used datasets.....	76
Table 4. 6: Mapping component between AIS and HICNA.....	90
Table 4. 7: Explanation of the algorithm.....	91
Table 4. 8: The explanation of the process in HICNA.....	93
Table 4. 9: Hybrid Immune Clonal Network Algorithm with AIS understanding theory..	95
Table 4. 10: Comparison of clusters available between Delaney et al., (2012) with proposed clusters.....	97
Table 4. 11: Results comparing Delany et al., (2012) with the proposed method.....	98
Table 4. 12: Number of spam messages in each dataset.....	101
Table 5. 1: Classification using NB.....	109
Table 5. 2: Explanation results from Confusion Matrix for NB .....	111
Table 5. 3: Classification using SVM.....	112
Table 5. 4: Explanation results from Confusion Matrix for SVM .....	114
Table 5. 5: Classification using k-NN.....	115
Table 5. 6: Explanation results for Confusion Matrix for k-NN.....	117
Table 5. 7: Classification using DT.....	118
Table 5. 8: Explanation results for Confusion Matrix for DT.....	120
Table 5. 9: Use Training Set.....	123
Table 5. 10: Supplied Test Set.....	124
Table 5. 11: Percentage Split.....	125
Table 5. 12: Output of clustering using Use Training Set.....	129
Table 5. 13: Output of clustering using Classes to Clusters Evaluation.....	130
Table 5. 14: Performance of classification algorithms in WEKA using Use Training Set .....	132
Table 5. 15: Performance of classification algorithms in WEKA using Cross-Validation .....	133
Table 5. 16: Number of messages in each cluster manually tested.....	135

Table 5. 17: Dataset in training and testing .....	139
Table 5. 18: Detection method .....	140
Table 5. 19: Results for method 1 .....	141
Table 5. 20: Results for training phase .....	143
Table 5. 21: Results in training phase .....	144
Table 5. 22: Results in testing phase .....	144
Table 5. 23: Naïve Bayes .....	146
Table 5. 24: Support Vector Machine .....	147
Table 5. 25: k-Nearest Neighbour .....	148
Table 5. 26: Results in clustering .....	149
Table 5. 27: Comparison of Results between Negative Selection algorithm with Danger Theory algorithm .....	156
Table 5. 28: Results in WEKA using four classifiers .....	158
Table 5. 29: Results for Algorithm 1 using raw dataset .....	159
Table 5. 30: Process in each phase for Algorithm 1 .....	160
Table 5. 31: Results for Algorithm 2 using raw dataset .....	161
Table 5. 32: Process in each phase for Algorithm 2 .....	162
Table 5. 33: Results for Algorithm 3 using raw dataset .....	162
Table 5. 34: Process in each phase for Algorithm 3 .....	163
Table 5. 35: Results for Algorithm 4 using raw dataset .....	164
Table 5. 36: Process in each phase for Algorithm 4 .....	165
Table 5. 37: Results for Algorithm 5 using raw dataset .....	166
Table 5. 38: Process in each phase for Algorithm 5 .....	167
Table 5. 39: Summary result for raw dataset using five algorithms .....	168
Table 5. 40: Results for Algorithm 1 using clean dataset .....	170
Table 5. 41: Results for Algorithm 2 using clean dataset .....	171
Table 5. 42: Results for Algorithm 3 using clean dataset .....	172
Table 5. 43: Results for Algorithm 4 using clean dataset .....	173
Table 5. 44: Results for Algorithm 5 using clean dataset .....	174
Table 5. 45: Summary results of detection in clean dataset .....	175
Table 5. 46: Results of classification using different dataset .....	182
Table 5. 47: Results of classification using FadhilahSpam dataset .....	186
Table 5. 48: Results using HICNA for detection .....	191
Table 5. 49: Comparison results of classification using HICNA and k-Means in WEKA .....	193

## LIST OF FIGURES

CONTENTS	PAGE
Figure 1.1: CIA triad .....	1
Figure 1.2: Statistic of consumers who use mobile phones around the world .....	3
Figure 1.3: Model for managing 'spam' text messages inspired by AIS algorithms.....	10
Figure 1.4: Research Phases .....	12
Figure 2. 1: The methodological framework of the research .....	18
Figure 2. 2: Formation of blood cells.....	22
Figure 2. 3: White Blood Cells classification.....	23
Figure 2. 4: Process of phagocytosis by Macrophage.....	24
Figure 2. 5: Main classes of lymphocytes.....	25
Figure 2. 6: Theoretical summary of BIS .....	27
Figure 2. 7: AIS Theories .....	29
Figure 2. 8: The algorithm for Immune Network Theory.....	30
Figure 2. 9: The algorithm for Clonal Selection.....	31
Figure 2. 10: The algorithm for Negative Selection.....	32
Figure 2. 11: The algorithm for Danger Theory.....	34
Figure 2. 12: Negative Selections in Thymus .....	36
Figure 2. 13: The Immune Network Theory in an Antibody and Antigen.....	36
Figure 2. 14: Mapping the Danger Theory and Clonal Selection with BIS .....	38
Figure 2. 15: Clustering techniques .....	48
Figure 2. 16: Agglomerative and Divisive clustering.....	49
Figure 3. 1: Integrated Mobile Spam Model (IMSM).....	56
Figure 3. 2: Block diagram for spam model.....	59
Figure 3. 3: Epitope and paratope .....	64
Figure 3. 4: The basic of process cloning in Clonal Selection.....	65
Figure 3. 5: Flow chart of Clonal Selection and Immune Network Theory.....	67
Figure 4. 1: Generalized Pseudo-code for five algorithms.....	74
Figure 4. 2: Algorithm 1.....	77
Figure 4. 3: Algorithm 2.....	79
Figure 4. 4: Algorithm 3.....	81
Figure 4. 5: Algorithm 4.....	83
Figure 4. 6: Algorithm 5.....	86
Figure 4. 7: Algorithm for HICNA .....	90
Figure 5. 1: Confusion Matrix in WEKA .....	108
Figure 5. 2: Results for each class of messages in NB .....	110
Figure 5. 3: Confusion Matrix for NB .....	111
Figure 5. 4: Results for each class of messages in SVM.....	113
Figure 5. 5: Confusion Matrix for SVM.....	114
Figure 5. 6: Results for each class of messages in k-NN.....	116
Figure 5. 7: Confusion Matrix for k-NN.....	117

Figure 5. 8: Results for each class of messages in DT.....	119
Figure 5. 9: Confusion Matrix for DT.....	120
Figure 5. 10 : SMS messages with label spam and ham.....	122
Figure 5. 11: Unlabeled SMS messages .....	126
Figure 5. 12: Interface after entering dataset in WEKA .....	126
Figure 5. 13: Classifier output .....	127
Figure 5. 14: k-Means .....	127
Figure 5. 15: Hierarchical.....	127
Figure 5. 16: Cobweb.....	127
Figure 5. 17: Time taken to cluster dataset using Use Training Set.....	134
Figure 5. 18: Number of instances in each cluster using k-Means.....	136
Figure 5. 19: Problem in testing phase.....	142
Figure 5. 20: Train and test set are not compatible.....	142
Figure 5. 21: Comparison results for DIT dataset.....	177
Figure 5. 22: Comparison results for BEC dataset.....	178
Figure 5. 23: Comparison results for UCI dataset.....	179
Figure 5. 24: Number of messages in each cluster between experiment part 1 and part 2 .....	187

## LIST OF APPENDICES

CONTENTS	PAGE
Appendix A: List of dataset used for classification .....	213
Appendix B: Detection phase.....	216
Appendix C: Classification phase .....	231
Appendix D: Detection and classification using WEKA.....	235
Appendix E: Interface for classification application.....	242

UNIVERSITI SAINS ISLAM MALAYSIA  
 جَامِعَةُ الْعُلُومِ الْإِسْلَامِيَّةِ الْمَالِيزِيَّةِ  
 ISLAMIC SCIENCE UNIVERSITY OF MALAYSIA

## LIST OF ALGORITHMS

Clonal Selection  
Danger Theory  
Immune Network Theory  
Negative Selection  
Hybrid Immune Clonal Network Algorithm (HICNA)

UNIVERSITI SAINS ISLAM MALAYSIA  
جامعة العلوم الإسلامية الماليزية  
ISLAMIC SCIENCE UNIVERSITY OF MALAYSIA

## ABBREVIATION

aiNet	= Artificial Immune Network (an immune system based algorithm which combines the pre-processing and clustering procedures)
AIN	= Artificial Immune Network
AIS	= Artificial Immune System
AUC	= Area Under Curve
BIS	= Biology Immune System
CI	= Computational Intelligence
CIA	= Confidentiality, Integrity and Availability
CLONACH	= Clonal Selection Algorithm for Learning Concept Hierarchies
DCDD	= Dynamic Concept Drift Detection
DIT	= Dublin Institute Technology
DT	= Decision Tree
DTL	= Danger Theory based Learning
FFNN	= Feed Forward Neural Networks
FN	= False Negative
FP	= False Positive
GAHC	= Guided Agglomerative Hierarchical clustering
GCAIN	= Guided Clustering and Artificial Immune Network
GT	= Grumble Text
HAC	= Hierarchical Agglomerative Clustering
HICNA	= Hybrid Immune Clonal Network Algorithm
ICOCOE	= International Conference on Communication and Computer Engineering
IMSM	= Integrated Mobile Spam Mobile
KMICE	= Knowledge Management International Conference
k-NN	= k-Nearest Neighbour
LMT	= Logistic Model Tree
MHC	= Major Histocompatibility Complex
Msg	= Message
NB	= Naïve Bayes
PAM	= Partitioning Around Medoids
PCA	= Principle Component Analysis
ROC	= Receiver Operating Characteristics
SKMM	= Suruhanjaya Komunikasi dan Multimedia
SMS	= Short Message Service
SOM	= Self-Organizing Maps (SOM)
SVM	= Support Vector Machine
TF-IDF	= Term frequency-inverse document frequency
TN	= True Negative
TP	= True Positive
UCS	= Supervised Classifier System
URL	= Uniform Resource Locator
U.S	= United State
WEKA	= Waikato Environment for Knowledge Learning